Foregut development: an act of balance

Next generation sequencing and Copy Number Variation profiling in EA/TEF

Erwin Brosens

ISBN/EAN: 978-90-5335-879-5

Cover illustrations: Crestock corporation
Lay-out cover: Ridderprint BV
Printed by: Ridderprint BV

The research presented in this dissertation was performed at the Departments of Pediatric Surgery and Clinical Genetics, Erasmus University Medical Center, Rotterdan, the Netherlands. The study was financially supported by the 'Sophia Stichting Wetenschappelijk Onderzoek' (project 493 and S13-09) Rotterdam, the Netherlands and Erasmus MC Grant DRP/RT/MP/054.

The printing of this thesis was financially supported by:

The Erasmus University Rotterdam

Foregut Development: an Act of Balance

Next generation sequencing and Copy Number Variation profiling in EA/TEF

Voordarmontwikkeling: een kwestie van evenwicht

DNA sequentie en kopie nummer variatie in oesophagusatresie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 17 juni 2014 om 15.30 uur
door
Erwin Brosens
geboren te Breda

Promotiecommissie:

Promotor:	Prof.dr. D. Tibboel
Overige leden:	Prof.dr. R.C.M. Hennekam
	Prof.dr. R.M.W Hofstra
	Prof.dr.ir. J.A. Veltman
Copromotor:	Dr. J.E.M.M. de Klein

Table of contents

Part 3 Genetic studies in discordant monozygotic twins

Part 4 Next generation sequencing in familial and consanguineous patients

Part 5 General discussion and Summary

Appendices

PART 1

General Introduction & Aim and outline

1.1

General Introduction

Adapted from:

Clinical and etiological heterogeneity in patients with tracheo-esophageal malformations and associated anomalies. Eur.J. Med. Genet. *submitted*

Erwin Brosens, Mirjam Ploeg, Yolande van Bever, Anna E. Koopmans, Hanneke IJsselstijn, Robbert J. Rottier, Rene Wijnen, Dick. Tibboel, Annelies de Klein

Abstract

Esophageal Atresia (EA) is a severe developmental defect of the foregut that presents with or without a Tracheo-Esophageal Fistula (TEF). The prevalence of EA/TEF over time and around the world has been relatively stable. EA/TEF is manifested in a broad spectrum of anomalies: in some patients it manifests as an isolated atresia or fistula, but in over half it affects several organ systems. While the associated malformations are often those of the VACTERL spectrum (Vertebral, Anorectal, Cardiac, Tracheo-Esophageal, Renal and Limb), many patients are affected by other malformations, such as microcephaly, micrognathia, pyloric stenosis, duodenal atresia, a single umbilical artery, and anomalies of the genitourinary, respiratory and gastrointestinal systems. Though EA/TEF is a genetically heterogeneous condition, recurrent genes and loci are sometimes affected. Tracheo-Esophageal (TE) defects are in fact a variable feature in several known single gene disorders and in patients with specific recurrent Copy Number Variations and structural chromosomal aberrations.

At present, a causal genetic aberration can be identified in 11-12% of patients. In most, EA/TEF is a sporadic finding; the familial recurrence rate is low (1%). As this suggests that epigenetic and environmental factors also contribute to the disease, non-syndromic EA/TEF is generally believed to be a multifactorial condition. Several population-based studies and case reports describe a wide range of associated risks, including age, diabetes, drug use, herbicides, smoking and fetal alcohol exposure. The phenotypical and genetic heterogeneity seen in EA/TEF patients indicates not one underlying cause, but several.

Unraveling the complex multifactorial and heterogeneous etiology of EA/TEF and associated features will require large cohorts of patients. Combined statistical analysis of component findings, genome sequencing, and genome wide association studies will elucidate new causal genetic defects and predisposing loci in the etiology within specific sub-populations. Improved knowledge of environmental risk factors, genetic predisposition and causal genetic syndromes may improve prediction and parental counseling, and prevent co-morbidity.

Introduction

Esophageal Atresia (EA) with or without Tracheo-Esophageal Fistula (TEF), is a developmental defect of the foregut characterized by the absence of continuity of the esophagus. EA/TEF (MIM 189960) can be classified in three ways: 1.) the Gross anatomical classification based on the presence and location of atresia and fistula; 2) a classification based on the association with other congenital anomalies (isolated or non-isolated); and 3.) a classification based on the presence of Tracheo-Esophageal anomalies (TE) in a known genetic syndrome (syndromal or non-syndromal).[1, 2] In the vast majority of patients (78.0-91.8%), the atresia is associated with a TEF, i.e., a distal connection of the esophagus to the trachea. A minority of patients have only an atresia (5.0-13.0%), a fistula (2.4-6.5%), an atresia with a proximal connection to the trachea (0.4-5.7%), or an atresia combined with both a proximal and distal fistula (0.1-2.6%).[3, 4] In approximately half of patients, TE anomalies are associated with other congenital defects.[5, 6] Non-syndromal EA/TEF is considered to be a multifactorial disease resulting from a variety of genetic and environmental influences.[5, 7, 8]

Prevalence, diagnosis and treatment

At ~ 2.5 per 10,000 births, the average prevalence of EA/TEF has been stable over time; regional prevalence ranges roughly between 1 and 4 per 10,000 births, including stillbirths and terminations of pregnancies.[9, 10] Contributing to these regional fluctuations are the quality of registries. While boys are affected more often than girls, with a sex ratio of 3:2[5, 11], this gender disparity can be confounded by genetic and environmental factors.[12-14] Although it is preferable to detect EA before birth—in order to schedule delivery at a pediatric surgery and perinatal center, and also to improve parental counseling—EA is usually not detected by ultrasound or MRI [15]. Prenatal clinical manifestations of EA are polyhydramnios combined with non-visualization of the fetal stomach during ultrasound, at least in the case of an absent TEF. However, these signs are not exclusive to EA; if a TEF is present, amniotic fluid may flow into the fetal stomach resulting in normal fetal stomach filling.

Prenatal detection rates differ substantially between reference centers (10-50%) [10], and were recently enhanced by combining ultrasound and the so-called amniotic fluid EA index, a promising biochemical approach that measures amniotic fluid alpha-fetoprotein and gamma-glutamyl transpeptidase.[16]

After birth, EA can be suspected if newborns have excessive saliva and/or are in respiratory distress. The diagnosis is confirmed if a nasogastric tube cannot be passed to the stomach. In most cases, tracheo-esophageal defects are repaired using a right sided thoracotomy within 48 hours of birth. Ten percent of European centers nowadays use a thoracoscopic procedure [17].

Early tracheo-esophageal development

Esophagus and respiratory structures develop from one common structure, the foregut. At the end of the third week of development of a human embryo, the endodermal layer folds to form a primitive gut tube. The primitive gut is regionalized and eventually differentiates into specific organs and derivatives of the gut tube, by a time dependent and localized expression and signaling actions of several growth factors (NODAL, FGF4), transcription factors (HEX, SOX2, FOXA2 and CDX2) and molecular pathways.[18] For example, high NODAL levels prime the endodermal layer to an anterior fate, and low levels of NODAL and high levels of FGF4 to a posterior fate.[19] The midgut is eventually formed—and the gut tube completed—by the inward-growing foregut, now expressing HEX, SOX2 and FOXA2, and the hindgut, expressing CDX2.[20] Homeobox transcription factors are important for regionalizing the gut tube, and eventually define regional gut identity and specification.[21]

In the fourth week of development, the foregut is arranged in a ventral respiratory field marked by high NKX2.1, the absence of SOX2 expression, and a dorsal gastrointestinal tube marked by the reverse NKX2.1/SOX2 expression pattern. In mice, dorsal-ventral patterning of Sox2 and Nkx2.1 is essential for proper foregut morphogenesis.[22, 23] This developmental process is excellently reviewed by Jacobs et al and Morrisey et al.[23, 24] The separation site of the dorsal and ventral foregut is marked by Barx-1, which is expressed in the mesenchyme.[25] After this specification, the foregut separates into two sections: a ventral respiratory part with two lung buds, and a dorsal gastrointestinal structure. Key biological processes involved in this separation process are regulated by signals to the epithelium and from the surrounding mesenchyme (Wnt2, Wnt2b, Fgf10 and Bmp4) and notochord (Nog, Shh).[24] Abnormal foregut morphogenesis can lead to disturbances in dorso-ventral patterning, expression pattern, and the timing of signaling factors in key regulatory networks such as those in Bmp-signaling [26], Wnt-signaling, RA-signaling[27] and Sonic Hedgehog signaling.[28]

Animal models and affected genes in patients with foregut abnormalities provide clues about a number of important biological processes during foregut separation and morphogenesis, including cell differentiation, proliferation, apoptosis, polarity and cytoskeletal rearrangements and cell-to-cell adhesion.[24]

Tracheo-esophageal defects and the VACTERL association

A broad phenotypical spectrum of anomalies is associated with EA. In some cases there is an isolated atresia, but in many more cases, several organ systems are affected.[5, 29] Certain malformations are associated with TE anomalies, more often than one would expect by chance; many are of the VACTERL kind (Vertebral, Anorectal, Cardiac, Tracheo-esophageal, Renal or urinary tract and Limbs malformations). VACTERL association is diagnosed if three or more of the VACTERL component features are present and known genetic syndromes are excluded.[30] As the inclusion criteria for VACTERL-associated malformations differ between institutions, VACTERL incidence ranges between 1 in 10,000 to 40,000 live born infants.[30, 31]

From a developmental perspective, it is difficult to explain this co-occurrence of multiple congenital anomalies. For instance, the VACTERL-associated structures are formed at different points during development, with vertebral organogenesis starting at around day 23, and anorectal development around day 45.[32] Many theories attempt to explain why the development of various organ systems is impaired: they include environmental exposure before or during organogenesis, epigenetic factors, hemodynamic instability in a monochorionic conception, a malformation sequence after abnormal notochord development that is followed by subsequent vertebral malsegmentation; and disturbances in developmental processes or key regulatory genes and pathways such as Sonic Hedgehog signaling (*SHH, GLI2, GLI3*).[32, 33]

Tracheo-esophageal defects and other associated malformations

TE anomalies are often associated with other, non-VACTERL, malformations such as microcephaly, duodenal atresia [29], single umbilical artery, pyloric stenosis[34], malformations of the genitourinary, respiratory, gastrointestinal and central nervous system; and diaphragmatic hernia, micrognathia and other craniofacial anomalies.[5, 29] These malformations are often associated with TE anomalies and one or more of the VACTERL components.[5] Specific combinations of associated features may indicate a common etiology.

The Erasmus University MC-Sophia TE cohort

From 1988 to 2014, 582 patients with TE have been included in our Erasmus University MC-Sophia TE cohort (EMC cohort). At present, 186 patients (32%) have isolated TE (no other major congenital anomalies); 67 (11%) have TE with other, non-VACTERL component features; 149 (26%) have TE plus one other VACTERL component feature; and 131 (23%) patients have two or more VACTERL component features. Fifty-three patients (9%) have a confirmed genetic syndrome, 12 of whom with three or more VACTERL component features. These fifty-three patients and their genetic diagnosis are depicted in figure 1.

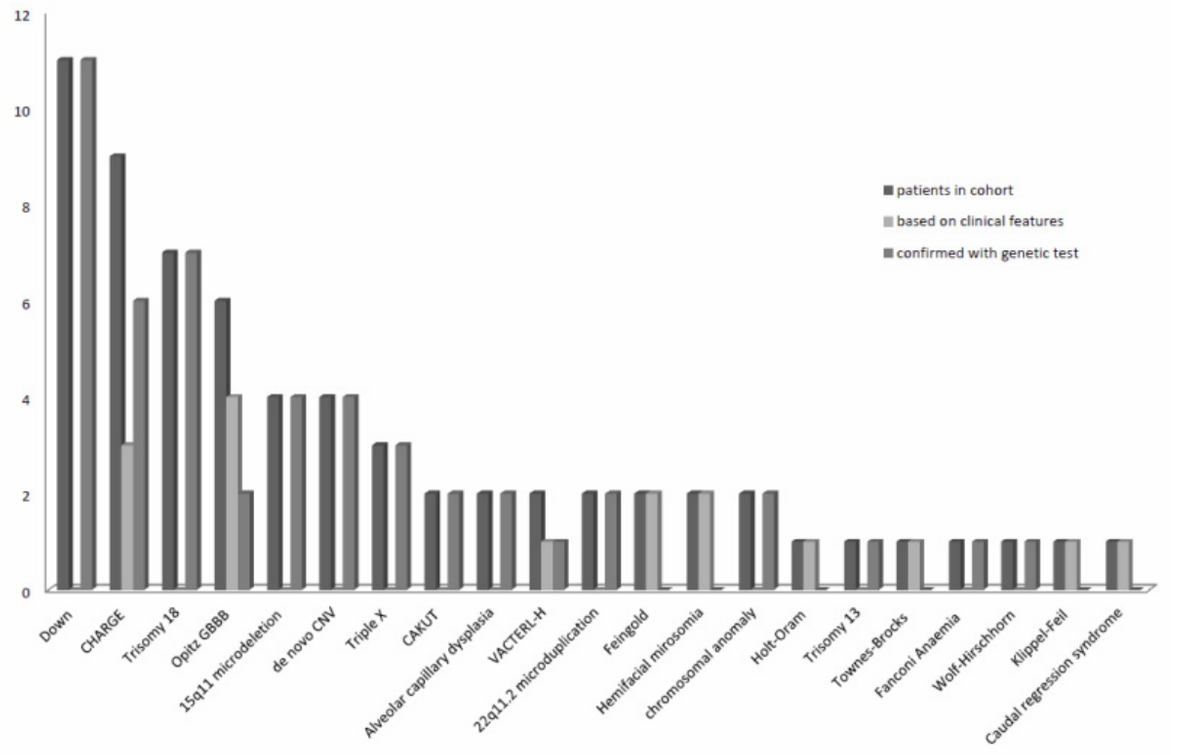

Figure 1. Genetic syndromes in Erasmus University MC-Sophia-Children's Hospital cohort of TE anomalies. The first dark gray bar for each syndrome represents the total number of patients with this syndrome in the Erasmus MC-Sophia TE cohort. The second light gray bar indicates the number of patients whose medical records do not describe a genetic test; these patients' diagnoses were based on phenotypical characteristics alone. The last dark gray bar represents the number of patients whose syndrome was diagnosed or confirmed on the basis of a genetic test. For instance, as karyotyping identified three patients with a triple X karyotype, the first bar for this syndrome represents the three patients in the cohort, and the second bar has a value of zero because the diagnosis was confirmed by a genetic test (karyotyping and micro-array, the value of three in the third bar). Abbreviations: CAKUT= congenital anomalies of the kidney and urinary tract, VACTERL-H: Vertebral, Anorectal, Cardiac, Tracheo-Esophageal, Renal and Limb anomalies with Hydrocephalus.

Four (0.7%) patients have a confirmed *de novo* non-syndromal Copy Number Variation, two (0.3%) have a chromosomal anomaly, four (0.6%) have a 15q11 micro deletion, and two (0.3%) a 22q11 micro duplication. The remaining 121 patients (21%) with TE-anomalies and two or more additional VACTERL component features do not have a genetic diagnosis and are classified as VACTERL associated. An overview of the number of affected organ systems and genetic diagnosis is given in figure 2.

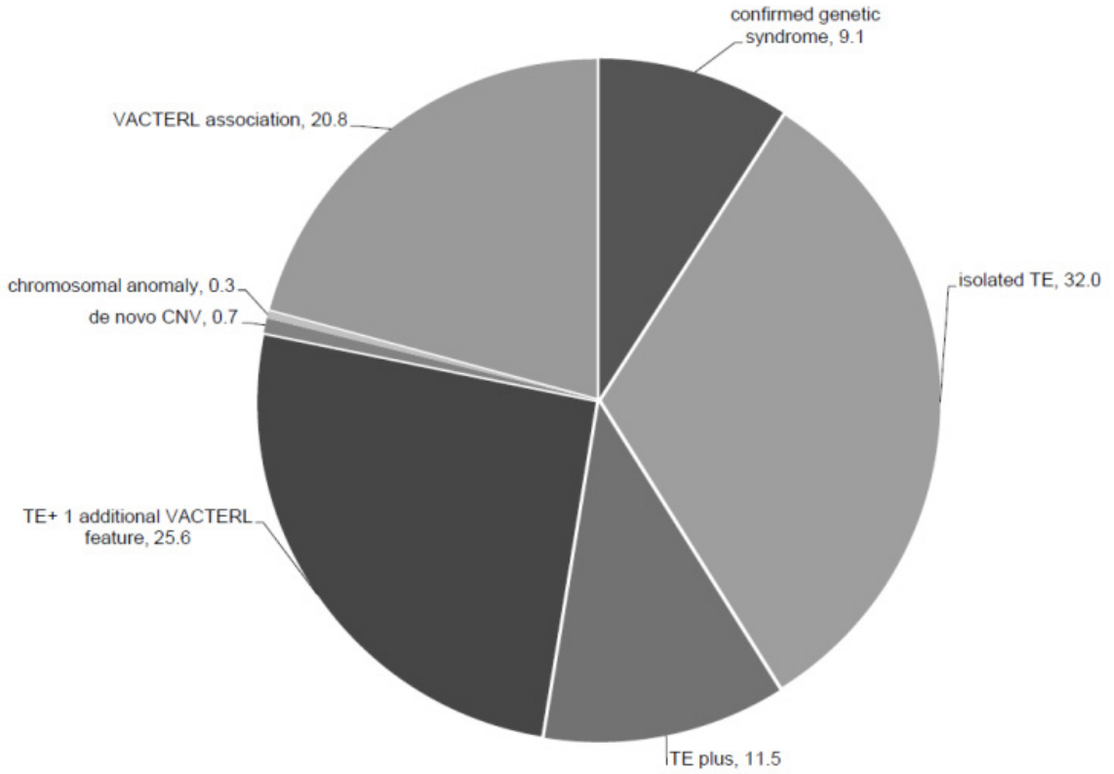

Figure 2. Distribution of anomalies in the Erasmus University MC-Sophia-Children's Hospital cohort of TE anomalies. TE; Tracheo-esophageal anomalies, TE plus; Tracheo-esophageal anomalies and one or more non-VACTERL associated major anomalies, VACTERL; Vertebral, Anorectal, Cardiac, Tracheo-Esophageal, Renal or Limb anomalies.

Cluster analysis of VACTERL-associated features has produced several proposals for subdividing VACTERL patients. Kallen and co-workers suggest a subdivision into a lower and upper group of VACTERL anomalies with cardiac anomalies clustering largely in the upper group, and renal anomalies clustering largely in the lower group.[35] Comparison of the Erasmus MC-Sophia (VACTERL) TE cohort with the figures published by Jenetzky and co-workers for the (VACTERL)-ARM cohort [36] does indeed suggest that renal anomalies are more prevalent in VACTERL patients with anorectal malformations than in patients with TE anomalies. Solomon et al describe five major subgroups, the largest two groups of which contain either anorectal malformations and no TE anomalies, or vice versa.[37]

However, subdivisions depend greatly on inclusion criteria and study design, i.e., the influence of sampling bias and possible differences in the classification of a "major" anomaly.[36]

	Esophageal atresia [1, 4, 10] (n>1200), %	**EMC-Esophageal atresia (n=582), %**	**VACTERL / VATER [37] (n=79), %**	**EMC- VACTERL (n=139), %**
Vertebral	6-21	21	59-82	60
Anorectal	10-16	14	55-90	52
Cardiac	13-34	27	38-80	60
Tracheo-esophageal	100	100	52-82	100
Renal	5-14	16	52-81	53
Limb	5-19	12	39-52	38

Table 1: Distribution of VACTERL features in EA/TEF patients and in VACTERL patients. EMC cohort; Erasmus University Medical Centre-Sophia Children's hospital cohort of TE anomalies n=582, 139 patients with three or more VACTERL components, genetic syndrome patients not excluded.

For instance, while the Rotterdam-Sophia cohort is based on patients with TE anomalies, anorectal malformations are the main inclusion criterion in the cohort described by Jenetzky and co-workers. Use of the χ^2 and phi association tests in the Erasmus MC-Sophia TE cohort only showed a significant ($P<0.0001$) but moderate (phi. 0.421) association between the C-TE-R components; no other combinations were significant. An overview of the Erasmus MC-Sophia TE cohort VACTERL component feature distribution is given in table 1.

Table 2a and b Genetic syndromes and abnormalities affecting tracheo-esophageal (TE) development. EMC-cohort; Erasmus University Medical Centre-Sophia Children's hospital cohort of tracheo-esophageal anomalies, TE: tracheo-esophageal anomaly, Rare = not a common feature but at least 2 case reports known. Table based on information obtained from a literature search using Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/), the OMIM database (http://www.ncbi.nlm.nih.gov/omim) [38], MGD database (http://www.informatics.jax.org/)[39] and London Dysmorphology Database.[40] Incidental reports, e.g. 1 or 2 cases described in literature and syndromes in which TE is a variable feature without a known gene or locus are depicted in the complete table, available in the supplementary information (suppl. table 1) In some of the patients with syndrome TE the described inheritance pattern are not clear, these are marked with "Uncertain" in the inheritance column.

Syndrome	Clinical features	Gene	Locus	Inheritance	TE frequency	Cases in EMC-cohort	Reference
Microphthalmia and esophageal atresia	Anophthalmia/microphthalmia, esophageal atresia with or without tracheo-esophageal fistula, and urogenital anomalies—most commonly cryptorchidism, hypospadias and micropenis	*SOX2*	3q26.3-q27	Uncertain; Autosomal dominant	100%	0	[1, 41]
VACTERL association	Vertebral anomalies, anorectal malformations, cardiac malformations, esophageal atresia, tracheoesophageal fistula, renal anomalies and limb anomalies.	unknown	unknown	Uncertain; Sporadic	50-80%	121	[1, 31, 37]
Feingold	Microcephaly, limb malformations, esophageal and duodenal atresias, learning disability/mental retardation, hand and foot abnormalities, cardiac and renal malformations, vertebral anomalies and deafness	*MYCN*	2p24.1	Autosomal dominant	25-40%	2	[1, 2, 6, 41, 42]
Trisomy 18	Congenital heart defects, kidney malformations, mental retardation, omphalocele, esophageal atresia, growth deficiency.	multiple	18	Autosomal dominant	25%	7	[43]
CHARGE	Coloboma, heart anomaly, choanal atresia, retardation, genital and ear anomalies/deafness, facial palsy, cleft palate, and dysphagia	*CHD7, SEMA3E*	8q12; 7q21	Autosomal Dominant; Microdeletion	10-20%	9	[1, 2, 6, 41, 44]
Fanconi Anaemia	Developmental abnormalities in major organ systems, early onset bone-marrow failure, and a high predisposition towards cancer. VACTERL-associated defects and hydrocephalus.	*FANC A,B,C,D1,G*	16q, 9q22, 13q12,3p25, 9p13,Xp22	-	1-14%	1	[1, 2, 6, 45]
Hemifacial microsomia	Craniofacial, cardiac, vertebral, and central nervous system defects. Incomplete development of the ear, nose, soft palate, lip, and mandible.	*BAPX1?; TCOF1?*	14q32; 5pter?, 22q11.2	Autosomal dominant	5%	2	[1, 46]
Down	Mental retardation, delayed physical growth, facial characteristics, congenital heart diseases, thyroid, gastrointestinal, eye and hearing disorders.	multiple	21	Autosomal dominant	0.5-1.0%	11	[43, 47]
Opitz GBBB	Hypertelorism, hypospadias, cleft lip/palate, laryngotracheoesophageal abnormalities, imperforate anus, developmental delay, and cardiac defects	*MID1*	Xp22	X-linked recessive	rare	6	[1, 2, 8, 41]

Table 2a Genetic syndromes and abnormalities affecting tracheo-esophageal (TE) development

Syndrome	Clinical features	Gene	Locus	Inheritance	TE frequency	Cases in EMC-cohort	Reference
Triple X	Variable, ranging from absence of symptoms to mental retardation and developmental defects such as limb and gastro-intestinal anomalies	multiple	X	Uncertain	rare	3	[48, 49]
15q11 deletion	Mental retardation, movement and behavior disorders, facial dysmorphisms, genital anomalies and developmental delay	*UBE3A*, *NDN*, *SNRPN*	15q11	Uncertain; Autosomal dominant	rare	5	[50]
22q11.2 del/ dup.	Parathyroid hypoplasia, thymic hypoplasia, outflow-tract defects of the heart, cleft palate, facial dysmorphism, hypocalcaemia, hypertelorism, and midline defects.	*TBX1*	22q11.2	Uncertain; Autosomal dominant	rare	2	[51]
Renal adysplasia; Potter; CAKUT	Kidney anomalies (renal dysplasia, duplex kidney, and hydronephrosis) and ureter anomalies (vesicoureteral reflux, megaureter, and ureterovesical junction (UVJ) vesicoureteral reflux obstruction)	*PAX2*, *HNF1B*, *DSTYK*, *UPK3A*, *ROBO2*, *TRAP1*	10q24.31, 17q12, 1q32, 3p12.3, 8q11.23, 16p13.3	-	rare	2	[52, 53]
Alveolar capillary dysplasia	Alveolar Capillary Dysplasia, VACTERL-associated defects, urinary tract obstruction	*FOXF1*	16q24.1	-	rare	2	[54]
Klippel-Feil	Fused cervical vertebrae, short neck, low posterior hairline, limited neck movement, Cardiac defects, craniofacial anomalies, skeletal and ocular anomalies, malformation of the larynx	*GDF6*, *GDF3*, *MEOX2*	8q22, 12p13.3, 17q21	-	rare	1	[55, 56]
Trisomy 13	Mental retardation, microcephaly, structural eye defects, meningo-myelocele, polydactyly, cleft palate, genital, kidney and heart defects.	multiple	13	Autosomal dominant	rare	1	[43]
Pallister-Hall	Hypothalamic hamartoma, pituitary dysfunction, central polydactyly and visceral malformations, anal atresia and occasionally laryngotracheo-esophageal cleft.	*GLI3*	7p13	-	rare	0	[41, 57]
Mandibulofacial dysostosis with microcephaly	Microcephaly, midface and malar hypoplasia, micrognathia, microtia, dysplastic ears, preauricular skin tags, significant developmental delay, and speech delay.	*EFTUD2*	17q21	-	rare	0	[58]
Thrombocytopenia-absent radius	Platelet number reduction, limb malformations, cardiac and renal abnormalities	*RBM8A*	1q21	Autosomal recessive; Microdeletion	rare	0	[59-61]

Table 2b Genetic syndromes and abnormalities affecting tracheo-esophageal (TE) development.

Heritability of tracheo-esophageal anomalies and associated malformations

Several lines of evidence suggest that TE has a genetic background: 1.) in 72 twins, the concordance rate is higher in monozygotic twins with isolated EA (67%) than in dizygotic twins (42%)[62]; 2.) murine knockout models indicate candidate genes in humans; 3.) TE defects are a variable feature in several known single-gene disorders, and 4.) TE can be present in patients with specific mutations, Copy Number Variations, structural chromosomal aberrations or whole chromosome duplications.[2, 43, 63, 64] Finally, although TE and VACTERL association are usually sporadic findings, familial TE does exist [65, 66]: in the EMC-Sophia TE cohort we have ten familial TE patients in 5 families. (1.7%) (*data not shown*), and other studies indicate a comparable number of familial cases.[67]

VACTERL-association heritability

In 69 twins with VACTERL association there does not seem to be a higher concordance rate in monozygotic twins (27%) than in dizygotic twins (31%).[68] Familial recurrence of VACTERL association component features has also been reported in patients with EA .[69] McMullen *et al* describe that 1.4% of sibs of patients in their TE cohort (n=140) have one VACTERL component feature or more.[70] Two of the children (n=41) of these patients also were affected, one of them with TE anomalies, indicative of a recurrence risk of 2-3%. Similarly, Solomon found one or more component feature in 5% of first-degree relatives of VACTERL patients, with 9% of the probands having affected relatives.[71] However, Bartels and co-workers did not find a higher prevelance of component features in first-degree relatives of patients with VACTERL association.[72]

TE and genetic syndromes

TE can be present as a variable feature in several known genetic syndromes. EA/TEF is a fairly common finding in syndromes such as Anophthalmia-esophageal-genital (AEG) syndrome, Feingold syndrome, CHARGE syndrome (i.e., Coloboma, Heart anomaly, choanal Atresia, Retardation, Genital and Ear anomalies) and trisomy 13. TE defects in other syndromes or conditions are more incidental (table 2). AEG syndrome also known as Microphthalmia and esophageal atresia syndrome (OMIM #206900) is a rare autosomal dominant disorder caused by mutations and deletions in the *SOX2* gene.[73] Clinical characteristics are specific eye anomalies (anophthalmia and microphthalmia), esophageal atresia (100%), genital abnormalities, and brain and neurodevelopmental anomalies.[74, 75] Feingold syndrome(OMIM #164280) is an autosomal dominant disorder whose clinical

manifestations include microcephaly, brachy-mesophalanangy, learning disabilities, and gastrointestinal atresia. The frequency of TE-anomalies in Feingold syndrome range around 30-40%. Feingold syndrome is caused by heterozygous mutations and deletions in the *MYCN* gene, a SHH signaling target gene.[76] Although heterozygous *mycn* mutated mice are healthy, homozygous knockout mice die around embryonic day 11; defects include those in the esophageal epithelium.[77] CHARGE syndrome (OMIM #214800)is caused by heterozygous mutations in the chromodomain *CHD7* gene, and is characterized not only by anomalies of the eye (coloboma) and ear (semicircular canal anomalies), but also by heart and genital anomalies, choanal atresia, and cranial nerve defects. Cleft palate, esophageal atresia (10-20%) and dysphagia are also commonly associated with this syndrome.[78, 79]

In the extended version of table 2 in the online supplementary information several other syndromes/conditions that have been reported as single case-reports; these still have to be confirmed in other EA/TEF patients. For instance, EA/TEF, duodenal atresia and hearing loss were observed in one of two monozygotic male siblings with a *de novo* deletion of the *WHSC1* gene, one of the genes suspected in Wolf-Hirschorn syndrome (*Erasmus MC-Sophia cohort unpublished data*). A second example is a female EA patient with a 15q11 deletion, who had mild cardiac anomalies and pectus excavatum.[72] Wong and co-workers suggest that the phenotypical spectrum of 15q11 deletions should be expanded to contain these more severe congenital defects. In our database we also found four additional 15q11 deletion patients, all with EA and cardiac anomalies (table 2 and figure 1A).

As well as Klippel-Feil in a male patient with TE, macrocephaly and infantile hypertrophic pyloric stenosis, other syndromal diagnoses in our cohort include Townes Brocks syndrome in a male patient with TE, cardiac anomalies, anorectal malformation, macrocephaly, micrognathia and ear anomalies; a ZIC3 mutation in a patient with X-linked heterotaxy[80]; and, more recently, a compound heterozygous *TRAP1* mutation in a patient with CAKUT and esophageal atresia.[52] Incidental reports by others of TE in syndromal patients include Thrombocytopenia- absent radius syndrome. [59-61] and a missense mutation in *PTEN* in a patient with a TEF, macrocephaly and hypoplasia of the thumbs.[81] The literature search resulted in over 70 TE-associated genetic and environmental syndromes. Supplementary table 1 describes syndromes in which TE are an incidental finding and whose locus or cause are currently unknown.

VACTERL association and genetic syndromes

Other VACTERL spectrum malformations are also a variable or defining feature of many of the genetic syndromes of which TE anomalies are a variable component. As well as these syndromes, there are of course syndromes with VACTERL components in which the esophagus is not affected, or only rarely. For instance, many patients with Alagille syndrome (*JAG1*, *NOTCH2*) have vertebral, cardiac and renal anomalies, while patients with Cornelia de Lange syndrome (*NIPBL*, *SMC1A*, *SMC3*) have renal and limb malformations in addition to their defining features.

Since the malformations described in many genetic syndromes overlap with those seen in patients with TE and VACTERL association, the patients in question may have a much higher frequency of known syndromic cases than previously described, and would benefit from genetic screening based on current knowledge. As differential diagnosis can be challenging if it is based on phenotype alone, reverse phenotyping might provide diagnoses in many of these previously "unresolved" cases. For instance, a search through these "unresolved cases" in the our database identified one male patient with cardiac, genitourinary and TE anomalies as well as choanal atresia and cleft palate—who may thus have CHARGE syndrome. Several patients had anorectal malformations, TE, and dysplastic ears or microtia that would warrant *SALL1* (Townes-Brocks syndrome) mutational screening. Interestingly, SALL1 is an interaction partner of SOX2 and Nanog[82], the genes affected in AEG-syndrome and involved in a recurrent TE-associated chromosomal anomaly in the 17q21.3-q24.2 region (*NOG*).[43] Our database contains 11 patients registered with VACTERL-associated malformations combined with hydrocephaly (VACTERL-H or hemifacial microsomia); similarly, a number of patients with hypospadias or other genitourinary malformations and two or more VACTERL component features could have Opitz G/ BBB syndrome. In EA/TEF patients with consanguineous parents or microcephaly, Fanconi anemia should be considered.

Chromosomal anomalies, Copy Number Variations

Tracheo-esophageal anomalies can be present in patients with whole-chromosome duplication syndromes such as Down syndrome (0.5-1.0%), trisomy 13 and trisomy 18, and Triple X syndrome.[48, 49] Since many genes are involved in these triplications, it is difficult to determine the causal gene. Chromosomal rearrangements and aneuploidies have also been described in patients with TE.[43, 83] Chromosomal studies have described hotspots such as deletions on chromosome 2q37, 4q35, 6q13-q15 and duplications on 3p25-

pter and 5q34-qter.[43, 83] Chromosomal syndromes such as chromosome 10 - paternal disomy, chromosome 17q22-q23.2 - submicroscopic deletion, chromosome 6q27 - submicroscopic deletion and mosaic trisomy 16 are also described in the London Dysmorphology Database.[40]

These hotspots are all *de novo* in origin and could harbor disease-causing genes. For instance, we know from animal knockout studies that loss of function of *Nog* and *Tbx4* can cause TE abnormalities, and that these genes are located within the frequently deleted human 17q21.3-q24.2 region.[84-86] Although patients with chromosomal deletions that include the *SHH* locus have been described, only two of them have TE anomalies.[87-91] As new technologies such as micro-array have largely replaced GTG-band karyotyping, and allow us to detect small Copy Number Variations (CNV), *de novo* and rare recurrent CNV have been detected in VACTERL and TE patients.[50, 91-99] However, sequencing of candidate genes involved in TE anomalies through CNV Profiling or Next Generation Sequencing of sporadic TE cases [52, 95, 100] did not identify other patients with similar mutations or CNVs. These incidental findings demonstrate the wide heterogeneity of the underlying gene defects that cause TE and associated anomalies.

Animal knockout models

Animal models can be used to study the effect of gene mutation, deletion or knockout. They have successfully identified candidate genes involved in human disease, and help to characterize the genes, signaling pathways and biological processes involved in normal tracheo-esophageal development. In 2001, for instance, Mahlapuu et al. described the importance of murine *Foxf1* in foregut development[101]; and in 2009 Stankiewicz et al. showed that *FOXF1* mutations and deletions are responsible for Alveolar Capillary Dysplasia (ACD).[54, 102] TE phenotypes are found in animal studies in which other forkhead box transcription factors *are* targeted (*Foxp1*, *Foxp2*, *Foxp4)* [103], as they are in various other transcription factors such as homeobox (e.g. *Hoxc4*, *MEOX2*, *NKX2.1*).[104-108] These genes are important in foregut development and candidate genes in human patients with TE-anomalies. Dorsal-ventral patterning and localization of Sox2 and Nkx2.1 are crucial for proper foregut morphogenesis and tracheo-esophageal separation.[22] Sox2 is involved in esophageal and eye malformations. In mice, Sox2 and Chd7 interact[78] and regulate genes of the Notch and SHH signaling pathway, including genes involved in VACTERL-like syndromes, such as Alagille syndrome (JAG1), Feingold syndrome (MYCN) and Pallister-Hall syndrome (GLI3).[78]

Other factors, such as Bone Morphogenic Proteins(*Bmp7*, *Bmp4*) and Noggin (*Nog*), are also important for dorsal-ventral patterning in the proximal foregut [24, 26]. *Bmp7* and *Nog* are expressed in the dorsal endoderm and *Bmp4* in the ventral mesenchyme; knockout of *Bmp4* or *Nog* can disturb dorsal-ventral patterning. Noggin is an inhibitor of the Bone Morphogenic Proteins and over 70 percent of *Nog* knockout mice have EA/TEF [24, 26]. Although *NOG* mutations in humans have been described, no associations with EA/TEF has been reported thus far [26]; and no loss of function mutations was found after sequencing of *NOG* in EA patients [109]. Sonic Hedgehog (*shh*) knockout mice have EA/TEF, and two patients with TE anomalies and deletion of this gene have been described. However, mutation analysis of *SHH* in an EA cohort did not reveal any pathogenic mutations.[110]

Another link between disturbed SHH signaling and esophageal developmental defects comes from the incidental finding of esophageal atresia in patients with ACD. (OMIM# 265380). These patients have reduced or absent expression of *FOXF1* due to mutation/deletion of this gene [102] or a deletion of the upstream lncRNA, the intragenic GLI2 binding domain[54] or CTCF and CEBPB binding domains.[111] Although *Foxf1* heterozygous knockout mice often develop EA, manifestation of TE anomalies is rare in humans.[101] (table2 and figure 1). Sometimes animal knockout models share the TE anomalies seen in their human disease counterparts, although in most models there is little overlap (table 3). Remarkably, the most prevalent syndromic EA/TEF conditions—such as ACD, CHARGE syndrome and Feingold syndrome—have an association with the SHH signaling pathway (see figure 3)

Shh signaling induces Bmp4 [112] and *Bmp4* knockouts do not develop a trachea.[113] SHH, GlI2 and GLI3 are dependent on proper cilia functioning.[114] While mice with loss of function or knockout mutations in genes involved in cilia function, formation and Hedgehog signal transduction (*IFT172*, *Dync2h1*, *Fuz*, *Wdr35)*[115-119] have TE in animal models, they do not (yet) have a TE defect counterpart in human disease (table 3). The importance of cilia formation and function, which are crucial for hedgehog signaling[114], is perhaps underestimated in foregut development, and new initiatives to establish a role for cilliopathy in VACTERL and other TE anomalies should be encouraged.[114]

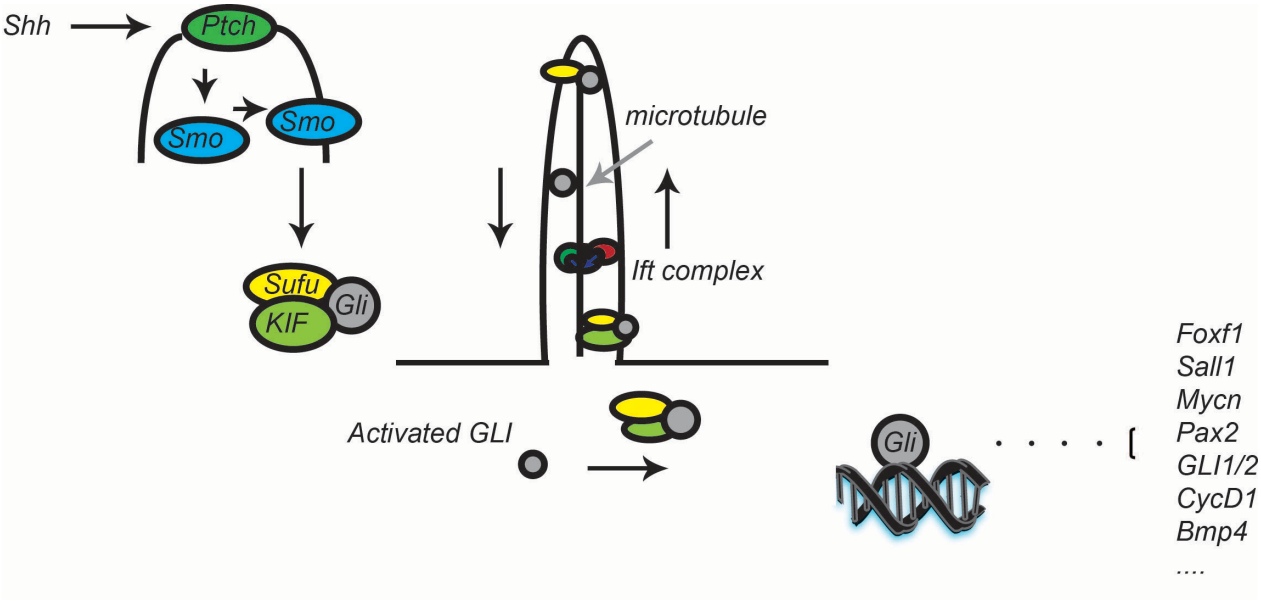

Figure 3. Schematic presentation of the Shh signaling pathway. Hedgehog signaling enables the release of Smo by Ptch and its sub sequent accumulation in the primary cilium. Active Smo interacts with the Sufu complex, which translocates to the tip of the cilium. GLI is activated and transported back in the direction of the basel body of the cilium and sub sequent to the nucleus. There GLI can bind to the DNA and activate transcription of target genes. Figure based on [114] and [120]

TE anomaly	Genes
In animal models only	*IFT172, Rarα/ Rarβ , Nkx2.1, Rab25, Hoxc4 , Chrd, Ctnnb1, Dync2h1, Efnb2, Foxp4, Fuz, Lec, Sox4, Wdr35, Foxp2, Foxp1*
In animal models and involved in human TEF	*NOG, TBX4, PCSK5, SHH, GLI2, GLI3, SOX2, SOX17, FOXF1, TBX1, MEOX2, RIPK4*
Involved exclusively in human syndromes of which TE is a variable feature	*HOXD13, CHD7, MYCN, SALL1, MID1, TRAP1, FANC-genes, ZIC3, PTEN, EFTUD2, NKX3.2, FBN2, RBM8A, GDF6, GDF3, TBX5, WHSC1, UBE3A, NDN, SNRPN, PAX2, HNF1B, DSTYK, UPK3A, ROBO2, POR, FGFR2, FGFR£, AXIN1, VANGL1, FGFR3, TERC, NOP10, TINF2, DKC1, ITGA6, FRAS1, FREM2, FLNA, AAAS, MKKS, WNT7a, WNT3, TCOF1*

Table 3. Animal and human genes involved in TE malformations.
Table based on information obtained from a literature search using Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/), the OMIM database (http://www.ncbi.nlm.nih.gov/omim) [121], MGD database (http://www.informatics.jax.org/)[122] and London Dysmorphology Database.[82] Database search was done in January 2014

In summary, there is a genetic contribution to EA and other TE- and VACTERL-associated defects. As the underlying genetic defect is still unknown in a subset of patients, it is yet to be discovered. The phenotypical and genetic heterogeneity seen in EA/TEF patients is indicative of not one, but several underlying causes. New causal genetic defects and predisposing loci in the etiology in specific sub-populations will be identified by combining statistical analysis of component findings in large patient cohorts with genome sequencing and genome-wide association studies.

Seemingly pathogenic chromosomal rearrangements, CNV or mutations in human genes that are involved in TE anomalies, identified in animal studies, and detected *de novo* in patients are also detected as events inherited from unaffected parents. For instance, the only person to be affected in a family with an inherited t(1;13)(p8;q12) was the patient herself.[121] Similarly, a chromosome 22q11 micro-duplication in another patient was inherited from a healthy parent.[92] But this duplication has also been described in patients as a *de novo* CNV[96] and in translocations.[122] Mutations in *PCSK5* in VACTERL patients were also inherited from unaffected parents.[106] It is also indicated by the variable penetrance of the TE–anomalies in several known genetic syndromes that there is more to these defects than Mendelian genetics alone: low heritability and monozygotic twin concordance rate indicate an environmental contribution, possibly a large one. As discordance has been described in several syndromes in which TE- and VACTERL-associated anomalies are variable features [123], monozygotic twin discordance is not unique to patients with TE-anomalies.

Environmental and other risk factors

Several risk and environmental factors have been associated with an increased risk for EA and VACTERL association (table 4). Environmental exposures can be estimated with questionnaires, either retrospectively during follow-up counseling visits in hospitals, or prospectively in a cohort of pregnant women. Several population-based studies and case-reports have highlighted a wide variety of associated risks, ranging from excessive hot-tub use during pregnancy to fetal alcohol exposure. Higher paternal and maternal age and low parity have also been associated with an increased risk. Some case reports have linked maternal diabetes to congenital anomalies, including those of the VACTERL spectrum.[124] At least seven case reports have described children with EA who were exposed to Methimiazole/ Carbimazole during pregnancy .[125] Affected patients can have features such as facial dysmorphisms, scalp vertex cutis aplasia, choanal atresia, EA/TEF,

developmental delay, and growth retardation. Methimiazole and Carbimazole are anti-thyroid drugs used to treat Graves' disease, a disorder characterized by excessive thyroid hormone secretion and thyromegaly caused by activation of the thyroid gland by immunoglobulin G autoantibodies. It has been debated whether or not these abnormalities are caused by the use of the anti-thyroid drugs or by the hyperthyroidism itself .[126, 127]

Another example of a teratogen is Diethylstilbestrol (DES), a synthetic estrogen used to prevent miscarriages. This endocrine-disrupting chemical has a trans-generational effect and associated congenital anomalies include genitourinary anomalies such as hypospadias[128, 129] and vaginal and cervical adenocarcinomas.[130, 131] Felix and coworkers describe nine women exposed to DES in utero who have children with EA/TEF.[132] Three women (2.4%) reported in utero DES exposure in an epidemiological study in which questionnaires were sent to parents who were contacted through a patient organization. Their children had trachea-esophageal anomalies as well as other congenital anomalies. One girl with isolated EA/TEF, a girl with meningocele and agenesis of the corpus callosum and a boy with EA/TEF and renal, anorectal and vertebral anomalies. Furthermore, Felix reported six EA/TEF children of the Erasmus MC-Sophia cohort with mothers exposed to DES (3.1%). These children had EA and or TEF, and four of them had one or more additional major anomaly, mainly of the VACTERL spectrum.[132] Although DES is significantly associated with EA/TEF, the exact mechanism whereby DES causes this anomaly in the offspring of women exposed *in utero* is unclear.

DES disturbs estrogen signaling via the *esr1* receptor [133], thereby inhibiting *trp63* expression.[134, 135] Interestingly, *trp63* is important in esophageal development and epithelial morphogenesis.[24, 136] However, the EA/TEF children have never been exposed to DES. Conceivably, epigenetic modifications are involved in the trans-generational effect of DES. While a pilot study in the Erasmus MC-Sophia children's cohort (n=48) showed a trend towards a significant association between gardening during pregnancy and EA/TEF, herbicide use was not significantly associated, making it hard to draw conclusions from the gardening trend.[5]Although numerous targeted and induced animal models have examined the development of the foregut [24, 115-119, 137-139], their importance to our understanding of human EA is not always clear. Critical foregut developmental pathways can be experimentally manipulated. For instance, we know from animal models that maternal cadmium exposure and smoking disturb Wnt- and SHH-signaling in neonatal mice and chickens.[140, 141]

Environmental or risk factor	Association	reference(s)
Low maternal parity	Associated	[65, 142, 143]
Maternal age	Associated	[5, 142, 144, 145]
Paternal age	Associated	[146, 147]
Ethnicity	Associated	[142]
Obesity	Absent	[148, 149]
Socio-economic status	Uncertain	[148]
Seasonal effects	Absent	[150, 151]
Maternal diabetes	Associated	[152-155]
Frequent maternal use of hot tub	Associated	[156]
Maternal dyspepsia	Absent	[157]
Exogeneous sex hormones	conflicting evidence	[158-162]
Maternal diethylstilbestrol exposure	Associated	[132]
Methimazole	Associated	[126, 163, 164]
Carbimazole	Associated	[163, 165]
Mycophenolate	Associated	[166]
Propylthiouracil	conflicting evidence	[163, 164]
Herbicides or insecticides	conflicting evidence	[5, 167]
Tabacco smoke	Absent	[5, 148, 168]
Alcohol consumption	Absent	[5, 148, 168]
Caffeine intake	Associated	[169]
Twinning	Associated	[6, 170, 171]
Maternal phenylketonuria	Associated	[172, 173]
Infectious disease	Absent	[151]
Vitamin A deprivation	associated in animal model	[174]
Adriamycin antibiotic	associated in animal model; absent in human	[175-177]
Ethylnitrosourea	associated in animal model	[106]

Table 4. Associated and non-associated environmental components and risk factors described in tracheo-esophageal anomalies

Another example is Adriamycin, which interferes with DNA replication and inhibits DNA and RNA synthesis.[173], and therefore affects many tissues and organ systems. While rodents develop esophageal atresia after prenatal exposure to Adriamycin, this has not been described in humans exposed to even higher dosages.[176] Neither does

adriamycin completely cross the placenta in humans, since only very low concentrations have been detected in the fetus.[176] VACTERL and non-VACTERL types of defects have been reported in the offspring of Adriamycin-treated rats. [178] While Adriamycin is known to cause apoptosis of monocytes and macrophages [179], nephropathy [180] and cardiomyopathy[181], Gillick and co-workers found no evidence of generalized cell death in the Adriamycin-induced VACTERL-association animal model.[182] Although the rodent Adriamycin model has been widely studied, it is unclear whether it is relevant to our understanding of EA/TEF.

Even though there are many associations, it is difficult to move beyond them and prove causality of these environmental risk factors. Two findings may help establish a link between genetics (null genotype, *de novo* CNV) and a pathway involved in biotransformation of teratogens (Gluthatione transferases).[94, 183] The first is the association of a null genotype of Gluthatione transferase (*GSTM1*) in mothers of EA patients and in the patients themselves[183]; the second is the presence of a large *de novo* deletion encompassing another Gluthatione transferase gene (*GSTP1*) in a patient with TE.[94] Overall, however, progress has been made in gene-environment studies in general: it has become increasingly evident that environmental burdens can leave epigenetic marks. These marks in the fetal epigenome— for instance after maternal diabetes[184] or after prenatal exposure to cigarette smoke[185]—are detected or suspected and can lead to disturbances in key developmental pathways. Studying the TE patient's epigenome may facilitate the unbiased measurement of prenatal environmental exposure.

Future developments

The wide genetic and phenotypical heterogeneity of TE defects indicate a multifactorial etiology. Next-Generation Sequencing and high resolution SNP arrays now enable us to detect smaller and smaller *de novo* CNV and mutations. Their use in the few familial cases and phenotypically clustered groups of sporadic patients may identify new genetic syndromes. Such new syndromes will explain an increasing part of the etiology in patients currently diagnosed with EA/TEF and VACTERL-associated malformations. Describing patient phenotypes uniformly will help to categorize this heterogeneous patient population prior to investigating genetic or environmental associations.

To date, the majority of EA/TEF and VACTERL-association patients have been sporadic, many of them caused by private *de novo* CNV or mutations. Improved surgical treatment has increased patient survival. Now that they are at an age when they may wish to have children, these patients may be at risk of transmitting their *de novo* mutations to the next generation in an autosomal dominant manner. Genetic counseling and genome profiling of families with recurrence using techniques that were unavailable at their time of their birth may be an important instrument not only for identifying genes involved in disease etiology, but also for predicting future high-risk pregnancies.

However, such developments will be very unlikely to explain the entire EA/TEF and VACTERL-association etiology. If association studies are to detect rare and low-effect loci, large patient and control cohorts are necessary, which will only be possible by combining patient cohorts. The effect of multiple rare or private inherited variants detected with NGS and studied with a burden test, also need large sample sizes. Uniform parental questionnaires should be used in multicenter studies, possibly in combination with one specific type of high-density SNP array or sequencing pipeline, thus enabling epidemiological and Genome Wide Association Studies in large cohorts.

The greatest challenge in elucidating EA/TEF and VACTERL association etiology will be to categorize the non-genetic contribution to their etiology. While the use of uniform questionnaires will lead to new associations, proving the causality of these associations will be much more difficult. It may be best first to focus on the effect of known mutations and teratogenic risk factors, possibly by using functional tests to study suspected biological pathways or processes. Patient specific genetic and environmental causality presents challenges with regard to counseling and informing parents on recurrence risk and future co-morbidity, and also the risk of familial recurrence. Greater knowledge of environmental risk factors, genetic predisposition and causal genetic syndromes may even produce preventive strategies and the ability to predict co-morbidity.

Acknowledgments

The authors thank D. Alexander for editorial support. Erwin Brosens is funded by the Sophia Foundations for Scientific Research, projects SSWO- 493 and SWOO13-09

References

1. Genevieve, D., et al., *An overview of isolated and syndromic oesophageal atresia.* Clin Genet, 2007. **71**(5): p. 392-9.
2. de Jong, E.M., et al., *Etiology of esophageal atresia and tracheoesophageal fistula: "mind the gap".* Curr Gastroenterol Rep, 2010. **12**(3): p. 215-22.
3. Bax, K.N., A.M. Roskott, and D.C. van der Zee, *Esophageal atresia without distal tracheoesophageal fistula: high incidence of proximal fistula.* J Pediatr Surg, 2008. **43**(3): p. 522-5.
4. Spitz, L., *Oesophageal atresia.* Orphanet J Rare Dis, 2007. **2**: p. 24.
5. de Jong, E.M., et al., *Non-VACTERL-type anomalies are frequent in patients with esophageal atresia/tracheo-esophageal fistula and full or partial VACTERL association.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 92-7.
6. Shaw-Smith, C., *Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology.* J Med Genet, 2006. **43**(7): p. 545-54.
7. Oddsberg, J., *Environmental factors in the etiology of esophageal atresia.* J Pediatr Gastroenterol Nutr, 2011. **52 Suppl 1**: p. S4-5.
8. Shaw-Smith, C., *Genetic factors in esophageal atresia, tracheo-esophageal fistula and the VACTERL association: roles for FOXF1 and the 16q24.1 FOX transcription factor gene cluster, and review of the literature.* Eur J Med Genet, 2010. **53**(1): p. 6-13.
9. Nassar, N., et al., *Prevalence of esophageal atresia among 18 international birth defects surveillance programs.* Birth Defects Res A Clin Mol Teratol, 2012. **94**(11): p. 893-9.
10. Pedersen, R.N., et al., *Oesophageal atresia: prevalence, prenatal diagnosis and associated anomalies in 23 European regions.* Arch Dis Child, 2012. **97**(3): p. 227-32.
11. Depaepe, A., H. Dolk, and M.F. Lechat, *The epidemiology of tracheo-oesophageal fistula and oesophageal atresia in Europe. EUROCAT Working Group.* Arch Dis Child, 1993. **68**(6): p. 743-8.
12. Cui, W., et al., *Sex differences in birth defects: a study of opposite-sex twins.* Birth Defects Res A Clin Mol Teratol, 2005. **73**(11): p. 876-80.
13. Lisi, A., et al., *Sex and congenital malformations: an international perspective.* Am J Med Genet A, 2005. **134a**(1): p. 49-57.
14. Lubinsky, M.S., *Classifying sex biased congenital anomalies.* Am J Med Genet, 1997. **69**(3): p. 225-8.
15. de Jong, E.M., et al., *Pre- and postnatal diagnosis and outcome of fetuses and neonates with esophageal atresia and tracheoesophageal fistula.* Prenat Diagn, 2010. **30**(3): p. 274-9.
16. Czerkiewicz, I., et al., *Biochemical amniotic fluid pattern for prenatal diagnosis of esophageal atresia.* Pediatr Res, 2011. **70**(2): p. 199-202.
17. Zani, A., et al., *International Survey on the Management of Esophageal Atresia.* Eur J Pediatr Surg, 2013.
18. Faure, S. and P. de Santa Barbara, *Molecular embryology of the foregut.* J Pediatr Gastroenterol Nutr, 2011. **52 Suppl 1**: p. S2-3.
19. Zorn, A.M. and J.M. Wells, *Vertebrate endoderm development and organ formation.* Annu Rev Cell Dev Biol, 2009. **25**: p. 221-51.
20. Carlson, B.M., *Human Embryology and Developmental Biology.* 2013: Elsevier.
21. Roberts, D.J., et al., *Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut.* Development, 1995. **121**(10): p. 3163-74.
22. Que, J., et al., *Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm.* Development, 2007. **134**(13): p. 2521-31.
23. Morrisey, E.E. and B.L. Hogan, *Preparing for the first breath: genetic and cellular mechanisms in lung development.* Dev Cell, 2010. **18**(1): p. 8-23.
24. Jacobs, I.J., W.Y. Ku, and J. Que, *Genetic and cellular mechanisms regulating anterior foregut and esophageal development.* Dev Biol, 2012. **369**(1): p. 54-64.
25. Woo, J., et al., *Barx1-mediated inhibition of Wnt signaling in the mouse thoracic foregut controls tracheo-esophageal septation and epithelial differentiation.* PLoS One, 2011. **6**(7): p. e22493.
26. Que, J., et al., *Morphogenesis of the trachea and esophagus: current players and new roles for noggin and Bmps.* Differentiation, 2006. **74**(7): p. 422-37.
27. Bayha, E., et al., *Retinoic acid signaling organizes endodermal organ specification along the entire antero-posterior axis.* PLoS One, 2009. **4**(6): p. e5845.
28. Litingtung, Y., et al., *Sonic hedgehog is essential to foregut development.* Nat Genet, 1998. **20**(1): p. 58-61.
29. Stoll, C., et al., *Associated malformations in patients with esophageal atresia.* Eur J Med Genet, 2009. **52**(5): p. 287-90.
30. Solomon, B.D., et al., *Clinical geneticists' views of VACTERL/VATER association.* Am J Med Genet A, 2012. **158a**(12): p. 3087-100.
31. Solomon, B.D., *VACTERL/VATER Association.* Orphanet J Rare Dis, 2011. **6**: p. 56.

32. Stevenson, R.E. and A.G. Hunter, *Considering the Embryopathogenesis of VACTERL Association.* Mol Syndromol, 2013. **4**(1-2): p. 7-15.
33. Pharoah, P.O., *Causal hypothesis for some congenital anomalies.* Twin Res Hum Genet, 2005. **8**(6): p. 543-50.
34. van Beelen, N.W., et al., *Increased incidence of hypertrophic pyloric stenosis in esophageal atresia patients.* Eur J Pediatr Surg, 2014. **24**(1): p. 20-4.
35. Kallen, K., et al., *VATER non-random association of congenital malformations: study based on data from four malformation registers.* Am J Med Genet, 2001. **101**(1): p. 26-32.
36. Jenetzky, E., et al., *Bias in patient series with VACTERL association.* Am J Med Genet A, 2011. **155a**(8): p. 2039-41; author reply 2042-3.
37. Solomon, B.D., et al., *Analysis of component findings in 79 patients diagnosed with VACTERL association.* Am J Med Genet A, 2010. **152a**(9): p. 2236-44.
38. Hamosh, A., et al., *Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.* Nucleic Acids Res, 2005. **33**(Database issue): p. D514-7.
39. Eppig, J.T., et al., *The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D881-6.
40. Guest, S.S., C.D. Evans, and R.M. Winter, *The Online London Dysmorphology Database.* Genet Med, 1999. **1**(5): p. 207-12.
41. Brunner, H.G. and H. van Bokhoven, *Genetic players in esophageal atresia and tracheoesophageal fistula.* Curr Opin Genet Dev, 2005. **15**(3): p. 341-7.
42. Courtens, W., et al., *Feingold syndrome: report of a new family and review.* Am J Med Genet, 1997. **73**(1): p. 55-60.
43. Felix, J.F., D. Tibboel, and A. de Klein, *Chromosomal anomalies in the aetiology of oesophageal atresia and tracheo-oesophageal fistula.* Eur J Med Genet, 2007. **50**(3): p. 163-75.
44. Lee, K.D., et al., *Esophageal atresia and tracheo-esophageal fistula associated with coarctation of the aorta, CHARGE association, and DiGeorge syndrome: a case report and literature review.* Pediatr Surg Int, 2008. **24**(10): p. 1153-6.
45. Umana, L.A., et al., *A male newborn with VACTERL association and Fanconi anemia with a FANCB deletion detected by array comparative genomic hybridization (aCGH).* Am J Med Genet A, 2011. **155a**(12): p. 3071-4.
46. Fischer, S., et al., *Histone acetylation dependent allelic expression imbalance of BAPX1 in patients with the oculo-auriculo-vertebral spectrum.* Hum Mol Genet, 2006. **15**(4): p. 581-7.
47. Solomon, B.D., et al., *Consideration of VACTERL association in patients with trisomy 21.* Clin Dysmorphol, 2010. **19**(4): p. 209-11.
48. Brosens, E., et al., *Structural and numerical changes of chromosome X in patients with esophageal atresia.* Eur J Hum Genet, 2014.
49. Zizka, J., et al., *[Trisomy X in a neonate with esophageal atresia].* Cesk Pediatr, 1977. **32**(12): p. 723-4.
50. Wong, D., et al., *Expanding the BP1-BP2 15q11.2 Microdeletion Phenotype: Tracheoesophageal Fistula and Congenital Cataracts.* Case Rep Genet, 2013. **2013**: p. 801094.
51. Digilio, M.C., et al., *Microdeletion 22q11 and oesophageal atresia.* J Med Genet, 1999. **36**(2): p. 137-9.
52. Saisawat, P., et al., *Whole-exome resequencing reveals recessive mutations in TRAP1 in individuals with CAKUT and VACTERL association.* Kidney Int, 2013.
53. Eghbalian, F., A. Monsef, and S.H. Mousavi-Bahar, *Urinary tract and other associated anomalies in newborns with esophageal atresia.* Urol J, 2009. **6**(2): p. 123-6.
54. Szafranski, P., et al., *Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder.* Genome Res, 2013. **23**(1): p. 23-33.
55. Clarke, R.A., et al., *Heterogeneity in Klippel-Feil syndrome: a new classification.* Pediatr Radiol, 1998. **28**(12): p. 967-74.
56. Tracy, M.R., J.P. Dormans, and K. Kusumi, *Klippel-Feil syndrome: clinical features and current understanding of etiology.* Clin Orthop Relat Res, 2004(424): p. 183-90.
57. Leboulanger, N. and E.N. Garabedian, *Laryngo-tracheo-oesophageal clefts.* Orphanet J Rare Dis, 2011. **6**: p. 81.
58. Gordon, C.T., et al., *EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia.* J Med Genet, 2012. **49**(12): p. 737-46.
59. Eren, E., et al., *An unusual association of TAR syndrome with esophageal atresia: a variant?* Pediatr Hematol Oncol, 2005. **22**(6): p. 499-505.
60. Bajaj, R., et al., *Tar syndrome with unilateral absent radius and associated esophageal atresia: a variant?* Indian J Pediatr, 1999. **66**(3): p. 460-3.
61. Peker, E., et al., *TAR syndrome and esophagial atresia: a concomitant or variant condition?* J Matern Fetal Neonatal Med, 2011. **24**(2): p. 226-8.

62. Schulz, A.C., et al., *Nine new twin pairs with esophageal atresia: a review of the literature and performance of a twin study of the disorder.* Birth Defects Res A Clin Mol Teratol, 2012. **94**(3): p. 182-6.
63. Felix, J.F., et al., *Genetics and developmental biology of oesophageal atresia and tracheo-oesophageal fistula: lessons from mice relevant for paediatric surgeons.* Pediatr Surg Int, 2004. **20**(10): p. 731-6.
64. Bednarczyk, D., M.M. Sasiadek, and R. Smigiel, *Chromosome Aberrations And Gene Mutations In Patients With Esophageal Atresia.* J Pediatr Gastroenterol Nutr, 2013.
65. Robert, E., et al., *An international collaborative study of the epidemiology of esophageal atresia or stenosis.* Reprod Toxicol, 1993. **7**(5): p. 405-21.
66. Veenma, D., et al., *Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts.* Eur J Hum Genet, 2012. **20**(3): p. 298-304.
67. Choinitzki, V., et al., *Second study on the recurrence risk of isolated esophageal atresia with or without trachea-esophageal fistula among first-degree relatives: no evidence for increased risk of recurrence of EA/TEF or for malformations of the VATER/VACTERL association spectrum.* Birth Defects Res A Clin Mol Teratol, 2013. **97**(12): p. 786-91.
68. Bartels, E., et al., *VATER/VACTERL association: identification of seven new twin pairs, a systematic review of the literature, and a classical twin analysis.* Clin Dysmorphol, 2012. **21**(4): p. 191-5.
69. Brown, A.K., et al., *Oesophageal atresia, related malformations, and medical problems: a family study.* Am J Med Genet, 1999. **85**(1): p. 31-7.
70. McMullen, K.P., et al., *Familial recurrence of tracheoesophageal fistula and associated malformations.* Am J Med Genet, 1996. **63**(4): p. 525-8.
71. Solomon, B.D., et al., *Evidence for inheritance in patients with VACTERL association.* Hum Genet, 2010. **127**(6): p. 731-3.
72. Bartels, E., et al., *Inheritance of the VATER/VACTERL association.* Pediatr Surg Int, 2012. **28**(7): p. 681-5.
73. Fantes, J., et al., *Mutations in SOX2 cause anophthalmia.* Nat Genet, 2003. **33**(4): p. 461-3.
74. Bardakjian, T.M. and A. Schneider, *Association of anophthalmia and esophageal atresia: four new cases identified by the anophthalmia/microphthalmia clinical registry.* Am J Med Genet A, 2005. **132a**(1): p. 54-6.
75. Shah, D., et al., *Bilateral microphthalmia, esophageal atresia, and cryptorchidism: the anophthalmia-esophageal-genital syndrome.* Am J Med Genet, 1997. **70**(2): p. 171-3.
76. Kenney, A.M., M.D. Cole, and D.H. Rowitch, *Nmyc upregulation by sonic hedgehog signaling promotes proliferation in developing cerebellar granule neuron precursors.* Development, 2003. **130**(1): p. 15-28.
77. Stanton, B.R., et al., *Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop.* Genes Dev, 1992. **6**(12a): p. 2235-47.
78. Engelen, E., et al., *Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes.* Nat Genet, 2011. **43**(6): p. 607-11.
79. Verloes, A., *Updated diagnostic criteria for CHARGE syndrome: a proposal.* Am J Med Genet A, 2005. **133a**(3): p. 306-8.
80. Wessels, M.W., et al., *Polyalanine expansion in the ZIC3 gene leading to X-linked heterotaxy with VACTERL association: a new polyalanine disorder?* J Med Genet, 2010. **47**(5): p. 351-5.
81. Reardon, W., X.P. Zhou, and C. Eng, *A novel germline mutation of the PTEN gene in a patient with macrocephaly, ventricular dilatation, and features of VATER association.* J Med Genet, 2001. **38**(12): p. 820-3.
82. Karantzali, E., et al., *Sall1 regulates embryonic stem cell differentiation in association with nanog.* J Biol Chem, 2011. **286**(2): p. 1037-45.
83. Lurie, I.W., *Structural autosomal imbalance and oesophageal defects: addendum to the article by Felix et al. (2007).* Eur J Med Genet, 2007. **50**(4): p. 322-5.
84. Dallapiccola, B., et al., *Interstitial deletion del(17) (q21.3q23 or 24.2) syndrome.* Clin Genet, 1993. **43**(1): p. 54-5.
85. Marsh, A.J., et al., *Interstitial deletion of chromosome 17 (del(17)(q22q23.3)) confirms a link with oesophageal atresia.* J Med Genet, 2000. **37**(9): p. 701-4.
86. Park, J.P., et al., *A unique de novo interstitial deletion del(17)(q21.3q23) in a phenotypically abnormal infant.* Clin Genet, 1992. **41**(1): p. 54-6.
87. Speleman, F., et al., *Detection of subtle reciprocal translocations by fluorescence in situ hybridization.* Clin Genet, 1992. **41**(4): p. 169-74.
88. Schinzel, A., *Cyclopia and cebocephaly in two newborn infants with unbalanced segregation of a familial translocation rcp (1;7)(q32;q34).* Am J Med Genet, 1984. **18**(1): p. 153-61.
89. Schinzel, A., *A further case of cyclopia due to unbalanced segregation of a previously reported rcp(1;7)(q32;q34) familial translocation.* Am J Med Genet, 1986. **24**(1): p. 205-6.
90. Schinzel, A., *Catalogue of unbalanced chromosome aberrations in man.* 2 ed. 2001, New York: De Gruyter.

91. Zen, P.R., et al., *Esophageal stenosis in a child presenting a de novo 7q terminal deletion.* Eur J Med Genet, 2010. **53**(5): p. 333-6.
92. Brosens, E., et al., *VACTERL Association Etiology: The Impact of de novo and Rare Copy Number Variations.* Mol Syndromol, 2013. **4**(1-2): p. 20-6.
93. Faguer, S., et al., *A 17q12 chromosomal duplication associated with renal disease and esophageal atresia.* Eur J Med Genet, 2011. **54**(4): p. e437-40.
94. Ferreira de Almeida, T. and D.R. Bertola, *Microdeletion 11q13.1.q13.2 in a patient presenting with developmental delay, facial dysmorphism, and esophageal atresia: possible role of the GSTP1 gene in esophagus malformation.* Birth Defects Res A Clin Mol Teratol, 2013. **97**(7): p. 463-6.
95. Hilger, A., et al., *De novo microduplications at 1q41, 2q37.3, and 8q24.3 in patients with VATER/VACTERL association.* Eur J Hum Genet, 2013. **21**(12): p. 1377-82.
96. Schramm, C., et al., *De novo microduplication at 22q11.21 in a patient with VACTERL association.* Eur J Med Genet, 2011. **54**(1): p. 9-13.
97. Solomon, B.D., et al., *De novo deletion of chromosome 20q13.33 in a patient with tracheo-esophageal fistula, cardiac defects and genitourinary anomalies implicates GTPBP5 as a candidate gene.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(9): p. 862-5.
98. van Binsbergen, E., et al., *A fetus with de novo 2q33.2q35 deletion including MAP2 with brain anomalies, esophageal atresia, and laryngeal stenosis.* Am J Med Genet A, 2014. **164a**(1): p. 194-8.
99. Smigiel, R., et al., *Oesophageal atresia with tracheoesophageal fistula and anal atresia in a patient with a de novo microduplication in 17q12.* Eur J Med Genet, 2014. **57**(1): p. 40-3.
100. Hernandez-Garcia, A., et al., *Contribution of LPP copy number and sequence changes to esophageal atresia, tracheoesophageal fistula, and VACTERL association.* Am J Med Genet A, 2012. **158a**(7): p. 1785-7.
101. Mahlapuu, M., S. Enerback, and P. Carlsson, *Haploinsufficiency of the forkhead gene Foxf1, a target for sonic hedgehog signaling, causes lung and foregut malformations.* Development, 2001. **128**(12): p. 2397-406.
102. Stankiewicz, P., et al., *Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations.* Am J Hum Genet, 2009. **84**(6): p. 780-91.
103. Shu, W., et al., *Foxp2 and Foxp1 cooperatively regulate lung and esophagus development.* Development, 2007. **134**(10): p. 1991-2000.
104. Garber, E.D., *"Bent-Tail," A Dominant, Sex-Linked Mutation in the Mouse.* Proc Natl Acad Sci U S A, 1952. **38**(10): p. 876-9.
105. Boulet, A.M. and M.R. Capecchi, *Targeted disruption of hoxc-4 causes esophageal defects and vertebral transformations.* Dev Biol, 1996. **177**(1): p. 232-49.
106. Szumska, D., et al., *VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5.* Genes Dev, 2008. **22**(11): p. 1465-77.
107. Mankoo, B.S., et al., *The concerted action of Meox homeobox genes is required upstream of genetic pathways essential for the formation, patterning and differentiation of somites.* Development, 2003. **130**(19): p. 4655-64.
108. Minoo, P., et al., *Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos.* Dev Biol, 1999. **209**(1): p. 60-71.
109. Murphy, A.J., et al., *Mutational analysis of NOG in esophageal atresia and tracheoesophageal fistula patients.* Pediatr Surg Int, 2012. **28**(4): p. 335-40.
110. Aguinaga, M., et al., *Sonic hedgehog mutation analysis in patients with VACTERL association.* Am J Med Genet A, 2010. **152a**(3): p. 781-3.
111. Szafranski, P., et al., *Novel FOXF1 deep intronic deletion causes lethal lung developmental disorder, alveolar capillary dysplasia with misalignment of pulmonary veins.* Hum Mutat, 2013. **34**(11): p. 1467-71.
112. Astorga, J. and P. Carlsson, *Hedgehog induction of murine vasculogenesis is mediated by Foxf1 and Bmp4.* Development, 2007. **134**(20): p. 3753-61.
113. Li, Y., et al., *Bmp4 is required for tracheal formation: a novel mouse model for tracheal agenesis.* Dev Biol, 2008. **322**(1): p. 145-55.
114. Ngan, E.S., K.H. Kim, and C.C. Hui, *Sonic Hedgehog Signaling and VACTERL Association.* Mol Syndromol, 2013. **4**(1-2): p. 32-45.
115. Information submitted by the NHLBI Cardiovascular Development Consortium (CvDC), B.t.B.P., *Modeling the Genetic Basis for Human Congenital Heart Disease in Mice* M.D.D.S. (B2B/CvDC), Editor. 2011.
116. Mill, P., et al., *Human and mouse mutations in WDR35 cause short-rib polydactyly syndromes due to abnormal ciliogenesis.* Am J Hum Genet, 2011. **88**(4): p. 508-15.
117. Gilissen, C., et al., *Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome.* Am J Hum Genet, 2010. **87**(3): p. 418-23.

118. Friedland-Little, J.M., et al., *A novel murine allele of Intraflagellar Transport Protein 172 causes a syndrome including VACTERL-like features with hydrocephalus.* Hum Mol Genet, 2011. **20**(19): p. 3725-37.
119. Halbritter, J., et al., *Defects in the IFT-B component IFT172 cause Jeune and Mainzer-Saldino syndromes in humans.* Am J Hum Genet, 2013. **93**(5): p. 915-25.
120. Gill, P.S. and N.D. Rosenblum, *Control of murine kidney development by sonic hedgehog and its GLI effectors.* Cell Cycle, 2006. **5**(13): p. 1426-30.
121. Cetinkaya, M., et al., *Oesophageal atresia in a neonate with a familial translocation t(1;13)(p8;q12).* Genet Couns, 2012. **23**(1): p. 73-5.
122. Prieto, J.C., et al., *Phenotypic expansion of the supernumerary derivative (22) chromosome syndrome: VACTERL and Hirschsprung's disease.* J Pediatr Surg, 2007. **42**(11): p. 1928-32.
123. Zwijnenburg, P.J., H. Meijers-Heijboer, and D.I. Boomsma, *Identical but not the same: the value of discordant monozygotic twins in genetic research.* Am J Med Genet B Neuropsychiatr Genet, 2010. **153b**(6): p. 1134-49.
124. Castori, M., et al., *VACTERL association and maternal diabetes: a possible causal relationship?* Birth Defects Res A Clin Mol Teratol, 2008. **82**(3): p. 169-72.
125. Cassina, M., et al., *Pharmacologic treatment of hyperthyroidism during pregnancy.* Birth Defects Res A Clin Mol Teratol, 2012. **94**(8): p. 612-9.
126. Di Gianantonio, E., et al., *Adverse effects of prenatal methimazole exposure.* Teratology, 2001. **64**(5): p. 262-6.
127. Momotani, N., et al., *Maternal hyperthyroidism and congenital malformation in the offspring.* Clin Endocrinol (Oxf), 1984. **20**(6): p. 695-700.
128. Klip, H., et al., *Hypospadias in sons of women exposed to diethylstilbestrol in utero: a cohort study.* Lancet, 2002. **359**(9312): p. 1102-7.
129. Brouwers, M.M., et al., *Hypospadias: a transgenerational effect of diethylstilbestrol?* Hum Reprod, 2006. **21**(3): p. 666-9.
130. Greenwald, P., et al., *Vaginal cancer after maternal treatment with synthetic estrogens.* N Engl J Med, 1971. **285**(7): p. 390-2.
131. Herbst, A.L., H. Ulfelder, and D.C. Poskanzer, *Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women.* N Engl J Med, 1971. **284**(15): p. 878-81.
132. Felix, J.F., et al., *Esophageal atresia and tracheoesophageal fistula in children of women exposed to diethylstilbestrol in utero.* Am J Obstet Gynecol, 2007. **197**(1): p. 38.e1-5.
133. Couse, J.F., et al., *Estrogen receptor-alpha knockout mice exhibit resistance to the developmental effects of neonatal diethylstilbestrol exposure on the female reproductive tract.* Dev Biol, 2001. **238**(2): p. 224-38.
134. Kurita, T. and G.R. Cunha, *Roles of p63 in differentiation of Mullerian duct epithelial cells.* Ann N Y Acad Sci, 2001. **948**: p. 9-12.
135. Kurita, T., A.A. Mills, and G.R. Cunha, *Roles of p63 in the diethylstilbestrol-induced cervicovaginal adenosis.* Development, 2004. **131**(7): p. 1639-49.
136. Daniely, Y., et al., *Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium.* Am J Physiol Cell Physiol, 2004. **287**(1): p. C171-81.
137. Nam, K.T., et al., *Loss of Rab25 promotes the development of intestinal neoplasia in mice and is associated with human colorectal adenocarcinomas.* J Clin Invest, 2010. **120**(3): p. 840-9.
138. Holland, P., et al., *RIP4 is an ankyrin repeat-containing kinase essential for keratinocyte differentiation.* Curr Biol, 2002. **12**(16): p. 1424-8.
139. Dravis, C. and M. Henkemeyer, *Ephrin-B reverse signaling controls septation events at the embryonic midline through separate tyrosine phosphorylation-independent signaling avenues.* Dev Biol, 2011. **355**(1): p. 138-51.
140. Blacquiere, M.J., et al., *Maternal smoking during pregnancy decreases Wnt signalling in neonatal mice.* Thorax, 2010. **65**(6): p. 553-4.
141. Fujiwara, N., et al., *Disturbance of SHH signalling pathway during early embryogenesis in the cadmium-induced omphalocele chick model.* Pediatr Surg Int, 2013. **29**(2): p. 165-70.
142. Oddsberg, J., et al., *Influence of maternal parity, age, and ethnicity on risk of esophageal atresia in the infant in a population-based study.* J Pediatr Surg, 2008. **43**(9): p. 1660-5.
143. Duong, H.T., et al., *Is maternal parity an independent risk factor for birth defects?* Birth Defects Res A Clin Mol Teratol, 2012. **94**(4): p. 230-6.
144. Gill, S.K., et al., *Association between maternal age and birth defects of unknown etiology: United States, 1997-2007.* Birth Defects Res A Clin Mol Teratol, 2012. **94**(12): p. 1010-8.
145. Loane, M., H. Dolk, and J.K. Morris, *Maternal age-specific risk of non-chromosomal anomalies.* Bjog, 2009. **116**(8): p. 1111-9.
146. Green, R.F., et al., *Association of paternal age and risk for major congenital anomalies from the National Birth Defects Prevention Study, 1997 to 2004.* Ann Epidemiol, 2010. **20**(3): p. 241-9.

147. Yang, Q., et al., *Paternal age and birth defects: how strong is the association?* Hum Reprod, 2007. **22**(3): p. 696-701.
148. Oddsberg, J., et al., *Maternal tobacco smoking, obesity, and low socioeconomic status during early pregnancy in the etiology of esophageal atresia.* J Pediatr Surg, 2008. **43**(10): p. 1791-5.
149. Stothard, K.J., et al., *Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis.* Jama, 2009. **301**(6): p. 636-50.
150. Castilla, E.E., et al., *Monthly and seasonal variations in the frequency of congenital anomalies.* Int J Epidemiol, 1990. **19**(2): p. 399-404.
151. Fraser, C., P.A. Baird, and A.D. Sadovnick, *A comparison of incidence trends for esophageal atresia and tracheoesophageal fistula, and infectious disease.* Teratology, 1987. **36**(3): p. 363-9.
152. Frias, J.L., et al., *Infrequently studied congenital anomalies as clues to the diagnosis of maternal diabetes mellitus.* Am J Med Genet A, 2007. **143a**(24): p. 2904-9.
153. Correa, A., et al., *Diabetes mellitus and birth defects.* Am J Obstet Gynecol, 2008. **199**(3): p. 237.e1-9.
154. Correa, A., et al., *Lack of periconceptional vitamins or supplements that contain folic acid and diabetes mellitus-associated birth defects.* Am J Obstet Gynecol, 2012. **206**(3): p. 218.e1-13.
155. Oddsberg, J., Y. Lu, and J. Lagergren, *Maternal diabetes and risk of esophageal atresia.* J Pediatr Surg, 2010. **45**(10): p. 2004-8.
156. Duong, H.T., et al., *Maternal use of hot tub and major structural birth defects.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(9): p. 836-41.
157. Acs, N., et al., *A possible association between maternal dyspepsia and congenital rectal/anal atresia/stenosis in their children: a population-based case-control study.* Acta Obstet Gynecol Scand, 2009. **88**(9): p. 1017-23.
158. Lammer, E.J. and J.F. Cordero, *Exogenous sex hormone exposure and the risk for major malformations.* Jama, 1986. **255**(22): p. 3128-32.
159. Lammer, E.J., J.F. Cordero, and M.J. Khoury, *Exogenous sex hormone exposure and the risk for VACTERL association.* Teratology, 1986. **34**(2): p. 165-9.
160. Bracken, M.B., et al., *Role of oral contraception in congenital malformations of offspring.* Int J Epidemiol, 1978. **7**(4): p. 309-17.
161. Rothman, K.J. and C. Louik, *Oral contraceptives and birth defects.* N Engl J Med, 1978. **299**(10): p. 522-4.
162. Wilson, J.G. and R.L. Brent, *Are female sex hormones teratogenic?* Am J Obstet Gynecol, 1981. **141**(5): p. 567-80.
163. Andersen, S.L., et al., *Birth Defects After Early Pregnancy Use of Antithyroid Drugs: A Danish Nationwide Study.* J Clin Endocrinol Metab, 2013.
164. Yoshihara, A., et al., *Treatment of graves' disease with antithyroid drugs in the first trimester of pregnancy and the prevalence of congenital malformation.* J Clin Endocrinol Metab, 2012. **97**(7): p. 2396-403.
165. Douchement, D., et al., *Choanal atresia associated with tracheoesophageal fistula: the spectrum of carbimazole embryopathy.* Pediatrics, 2011. **128**(3): p. e703-6.
166. Hoeltzenbein, M., et al., *Teratogenicity of mycophenolate confirmed in a prospective study of the European Network of Teratology Information Services.* Am J Med Genet A, 2012. **158a**(3): p. 588-96.
167. McDonald, A.D., et al., *Congenital defects and work in pregnancy.* Br J Ind Med, 1988. **45**(9): p. 581-8.
168. Wong-Gibbons, D.L., et al., *Maternal periconceptional exposure to cigarette smoking and alcohol and esophageal atresia +/- tracheo-esophageal fistula.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(11): p. 776-84.
169. Browne, M.L., et al., *Maternal caffeine intake and risk of selected birth defects in the National Birth Defects Prevention Study.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(2): p. 93-101.
170. Mastroiacovo, P., et al., *Congenital malformations in twins: an international study.* Am J Med Genet, 1999. **83**(2): p. 117-24.
171. Chambers, C.D., et al., *Intrauterine growth restriction in like-sex twins discordant for structural defects.* Birth Defects Res A Clin Mol Teratol, 2006. **76**(4): p. 246-8.
172. Rogowiecka, E., et al., *[Maternal PKU syndrome as an obstetric problem: literature review and own clinical experience].* Ginekol Pol, 1998. **69**(12): p. 1007-15.
173. Felix, J.F., et al., *Genetic and environmental factors in the etiology of esophageal atresia and/or tracheoesophageal fistula: an overview of the current concepts.* Birth Defects Res A Clin Mol Teratol, 2009. **85**(9): p. 747-54.
174. Wilson, J.G., C.B. Roth, and J. Warkany, *An analysis of the syndrome of malformations induced by maternal vitamin A deficiency. Effects of restoration of vitamin A at various times during gestation.* Am J Anat, 1953. **92**: p. 189-217.
175. Diez-Pardo, J.A., et al., *A new rodent experimental model of esophageal atresia and tracheoesophageal fistula: preliminary report.* J Pediatr Surg, 1996. **31**(4): p. 498-502.
176. Germann, N., F. Goffinet, and F. Goldwasser, *Anthracyclines during pregnancy: embryo-fetal outcome in 160 patients.* Ann Oncol, 2004. **15**(1): p. 146-50.

177. Matalon, S.T., A. Ornoy, and M. Lishner, *Review of the potential effects of three commonly used antineoplastic and immunosuppressive drugs (cyclophosphamide, azathioprine, doxorubicin on the embryo and placenta).* Reprod Toxicol, 2004. **18**(2): p. 219-30.
178. Mc Laughlin, D., et al., *Adriamycin-Induced Models of VACTERL Association.* Mol Syndromol, 2013. **4**(1-2): p. 46-62.
179. Krysko, D.V., et al., *TLR-2 and TLR-9 are sensors of apoptosis in a mouse model of doxorubicin-induced acute inflammation.* Cell Death Differ, 2011. **18**(8): p. 1316-25.
180. Okuda, S., et al., *Adriamycin-induced nephropathy as a model of chronic progressive glomerular disease.* Kidney Int, 1986. **29**(2): p. 502-10.
181. Umlauf, J. and M. Horky, *Molecular biology of doxorubicin-induced cardiomyopathy.* Exp Clin Cardiol, 2002. **7**(1): p. 35-9.
182. Gillick, J., et al., *Cell death in the early adriamycin rat model.* Pediatr Surg Int, 2002. **18**(7): p. 576-80.
183. Filonzi, L., et al., *Evidence that polymorphic deletion of the glutathione S-transferase gene, GSTM1, is associated with esophageal atresia.* Birth Defects Res A Clin Mol Teratol, 2010. **88**(9): p. 743-7.
184. Vrachnis, N., et al., *Impact of maternal diabetes on epigenetic modifications leading to diseases in the offspring.* Exp Diabetes Res, 2012. **2012**: p. 538474.
185. Knopik, V.S., et al., *The epigenetics of maternal cigarette smoking during pregnancy and effects on child development.* Dev Psychopathol, 2012. **24**(4): p. 1377-90.

1.2

Aim and outline of this thesis

Aim and outline of this thesis

In chapter 1 of this thesis, a *general introduction*, the current knowledge on the etiology of esophageal atresia and tracheo-esophageal fistula is reviewed. EA/TEF is a clinical heterogeneous condition and has a heterogeneous etiology. Tracheo-esophageal anomalies are a variable feature in over 70 genetic syndromes. Moreover, chromosomal anomalies, single base pair mutations and Copy Number Variations have been described in sporadic patients, a causal genetic aberration can be identified in 11-12% of patients. EA/TEF is mostly a sporadic finding. As this suggests that epigenetic and environmental factors also contribute to the disease, non-syndromic EA/TEF is generally believed to be a multifactorial condition. Indeed several population-based studies and case reports describe a wide range of associated risks, including age, diabetes, drug use, herbicides, smoking and fetal alcohol exposure.

In chapter 2, *Copy Number Variations in patients with EA or VACTERL spectrum malformations*, we address chromosomal anomalies and CNV in these patients. We have reviewed the literature for *de novo* CNV and describe the recurrent inherited and de novo CNV in the Erasmus University MC-Sophia Children's hospital TE- cohort. We show that although described in literature and present in our cohort, de novo CNV is rare in these patients. Furthermore, we hypothesize about the role of rare an private recurrent CNV in these patients which, although inherited, could point to mechanisms or biological processes contributing to this constellation of developmental defects.

In chapter 3, *Genetic studies in discordant monozygotic twins*, we describe why twin studies can be a valuable asset in the geneticists' toolbox. Also, we discuss the pitfalls e.g. caution has to be taken by the interpretation of these twin-study results since environmental and genetic components are not always exactly identical in monozygous twins. We compare the DNA of discordant monozygotic twins using SNP-array and Whole-Exome sequencing in order to identify somatic changes, either Copy Number Variations, small insertions/deletions or single nucleotide changes, that could explain the twin discordance.

Chapter 4, *Next generation sequencing in familial and consanguineous patients*, describes two other methods we used to determine underlying genetic factors in EA/TEF and renal agenesis/hypoplasia, one of the VACTERL spectrum associated malformations. With genetic studies in families with multiple affected members and with studies in consanguineous patients we aim to identify the causal genetic variants in these patients. TE

and VACTERL association familial recurrence rate is low (1-3%). We describe two genetic studies, one in 5 familial cases and the other study in 3 consanguineous patients with TE anomalies and a family with two siblings with renal anomalies.

In chapter five, the *General discussion*, we discuss the results of our experiments and address the prospects and challenges of our and future research. We highlight the phenotypical and genetic heterogeneity in EA/TEF patients, indicating not one underlying cause, but several causal and contributing factors. This wide genetic and phenotypical heterogeneity of TE defects indicate a multifactorial etiology. We give recommendations for further research, stressing the importance of international collaboration and joining of large cohorts of patients. With joint efforts we will be able to speed up the process of identifying new causal and predisposing, genetic and environmental, factors in EA/TEF etiology. Next-Generation Sequencing and high resolution SNP arrays now enable us to detect smaller and smaller *de novo* CNV and mutations: their use in the few familial cases and phenotypically clustered groups of sporadic patients will reveal causal genetic variation and may identify new genetic syndromes. Such new syndromes will explain an increasing part of the etiology in patients currently diagnosed with EA/TEF and VACTERL-associated malformations. Describing patient phenotypes uniformly will help to categorize the heterogeneous patient population prior to genetic or environmental associations.

To date, the majority of EA/TEF and VACTERL-association patients have been sporadic, many of them caused by private *de novo* CNV or mutations. Improved surgical treatment has increased patient survival. Now they are at an age when they may wish get pregnant, these patients are at risk of transmitting their *de novo* mutations to the next generations in an autosomal dominant manner. Genetic counseling and genome profiling of young adults using techniques that were unavailable at their time of their birth may be an important instrument not only for identifying genes involved in disease etiology, but also for predicting high-risk pregnancies. However, such developments will be very unlikely to explain the entire EA/TEF and VACTERL-association etiology. If association studies are to detect rare and low-effect loci, large patient and control cohorts are necessary, which will only be possible by combining patient cohorts. Uniform parental questionnaires should be used in multicenter studies, possibly in combination with one specific type of high-density SNP array, thus enabling epidemiological and Genome Wide Association Studies in large cohorts. After all, our main goal in these type of patient based studies is to gain more knowledge of risk factors in order to improve parental counseling, and perhaps predict and prevent co-morbidity.

PART 2

Copy Number Variations in patients with EA or VACTERL

2.1

VACTERL association etiology: the impact of *de novo* and rare copy number variations

Adapted from:

VACTERL association etiology: the impact of de novo and rare Copy Number Variations.
E. Brosens, H. Eussen, Y. van Bever, R.M. van der Helm, H. IJsselstijn, H.P. Zaveri, R.M.H. Wijnen, D.A. Scott, D. Tibboel, A. de Klein. *Mol Syndromol. 2013 Feb;4(1-2):20-6*

Abstract

Copy number variations (CNVs), either DNA gains or losses, have been found at common regions throughout the human genome. Most CNVs neither have a pathogenic significance nor result in disease related phenotypes but, instead, reflect the normal population variance. However, larger CNVs, which often arise de novo, are frequently associated with human disease. A genetic contribution has long been suspected in VACTERL (Vertebral, Anal, Cardiac, TracheoEsophageal fistula, Renal and Limb anomalies) association. The anomalies observed in this association overlap with several monogenetic conditions associated with mutations in specific genes, e.g. Townes Brocks (SALL1),Feingold syndrome (MYCN) or Fanconi anemia. So far VACTERL association is typically considered a diagnosis of exclusion. Identifying recurrent or de novo genomic variations in individuals with VACTERL association could make it easier to distinguish VACTERL association from other syndromes and could provide insight into disease mechanisms.

Sporadically, de novo CNVs associated with VACTERL are described in literature. In addition to this literature review of genomic variation in published VACTERL association patients we describe CNVs present in 68 VACTERL association patients collected in our institution. De novo variations (>30kb) are absent in our VACTERL association cohort. However, we identified recurrent rare CNVs which, although inherited, could point to mechanisms or biological processes contributing to this constellation of developmental defects.

Introduction

Copy number analysis has proven to be a powerful tool for identifying genes and genomic regions that contribute to the occurrence of congenital malformations. Common copy number variations (CNVs), regions of variable DNA gains or losses, account for a significant proportion of the healthy human genome. [1, 2] Most CNVs are inherited polymorphisms that have no appreciable effect on health. However, there are many examples of de novo or rare CNVs that have clearly been associated with human diseases, e.g. Wolf-Hirschhorn syndrome (OMIM #194190) and 22q11.2 deletion syndrome (OMIM #192430).[3] Pathologic CNVs are often larger (>500kb) in size and are usually not inherited from an unaffected parent.[4]

Implementation of new molecular cytogenetic techniques, such as microarray-based comparative genomic hybridization and single nucleotide polymorphism arrays, has revealed previously unidentified genotypic aberrations which can now be correlated with phenotypic anomalies. As a result, numerous publications have implicated specific pathogenic CNVs in intellectual disability, congenital anomalies like cleft lip, microcephaly , renal malformations [5], and neurological conditions including autism and schizophrenia[3]It may very well be that, like in other congenital anomalies, there is role for pathogenic CNVs in VACTERL association (OMIM #192350) etiology. VACTERL association is a heterogeneous condition defined by six core structural defects (Vertebral, Anal, Cardiac, Trachea-Esophageal fistula, Renal and Limb anomalies) which occur together more commonly than would be expected by chance alone.

These defects are also observed in several other monogenetic conditions caused by intragenic mutations, e.g. Townes-Brocks syndrome (OMIM #107480; SALL1)—whose features include imperforate anus, cardiac defects, renal anomalies and hand defects, most often affecting the thumb—and Feingold syndrome (OMIM #164280; MYCN)—whse features can include esophageal atresia, cardiac anomalies, renal anomalies and abnormalities of the hand and fingers. [6]) In addition to gene mutations, CNVs have been described as causal factor in several VACTERL-like syndromes. These include Goldenhar/OAVS (OMIM #141400[7], Townes-Brocks syndrome [8], X-linked VACTERL-H (OMIM #314390) [9], MURCS association (OMIM #601076) [10], OEIS complex (OMIM #258040) [11, 12], TAR syndrome (OMIM #274000) [13], 13q32 deletion syndrome [14]and 22q11.2 deletion syndrome.[15] Due to the abundance of overlapping defects in various organs, the scarcity of known causal factors and its heterogeneous phenotype, VACTERL association is typically considered a diagnosis of exclusion. In general the

diagnosis is made when at least three of the six associated core defects are present and all other phenotypical overlapping syndromes have been excluded. [16]

The role of genetics in VACTERL association has long been suspected. VACTERL is usually a sporadic finding, but familial cases do exist. [17] Moreover, in about 9% of VACTERL patients one of the relatives has one of the six core VACTERL features.[18]In some rare cases a genetic defects have been described such as a polyalanine expansion[19], nuclear [20-22] or mtDNA [23, 24] mutations and numerical or structural chromosome aberrations. [25-27] The resolution to detect these chromosomal anomalies has increased significantly with the introduction of micro-array technology; as current technologies allow detection of genomic imbalances down to only a few kb in size. Although the role of CNV and chromosomal aberrations in congenital anomalies is well established, little is known of their role in VACTERL association etiology. It is possible that recurrent or de novo genomic variations contribute to the development of some cases of VACTERL association. Identifying such changes could make it easier to distinguish VACTERL association from other syndromes and other potentially-related conditions, and could provide insight into disease mechanisms

Materials and Methods

Literature Review

We reviewed the literature to identify both numerical or structural chromosomal anomalies and copy number variations described in individuals with VACTERL association. We followed the inclusion criteria for VACTERL association (three or more of the core VACTERL elements, Figure 1) and excluded the patients with a confirmed genetic syndrome.

Study Population

Since 1988, the Erasmus MC - Sophia Children's Hospital Department of Pediatric Surgery unit has been collecting clinical data and, when possible, DNA, from VACTERL patients. This cohort is part of a larger EA/TEF cohort (n=567) in which patient sampling and registration are based on the existence of either Esophageal Atresia and/or TracheoEsophageal fistula regardless of additional anomalies. Patients were selected using the same criteria as for the literature review. In total, DNA from 68 out of 121 VACTERL patients was analyzed for copy number variations.

Analysis of copy number variation

Copy number variation analysis was performed using Illumine 12-HumanCytoSNP, Human 610-Quad or Omni Express Bead Chips (Illumina, Inc., San Diego, USA) according to manufacturer's instructions. Arrays are processed according to their manufacturer's standard protocol. Normalized output was generated with Illumina's Genome Studio program (Illumina, San Diego, CA, USA) and copy number variation was visualized in Nexus CN6.1. (Biodiscovery Inc, El Segundo, CA, USA) Inheritance of CNVs was determined only if they were larger than 30kb, contained genes and were either unique or had a low frequency in the general population.

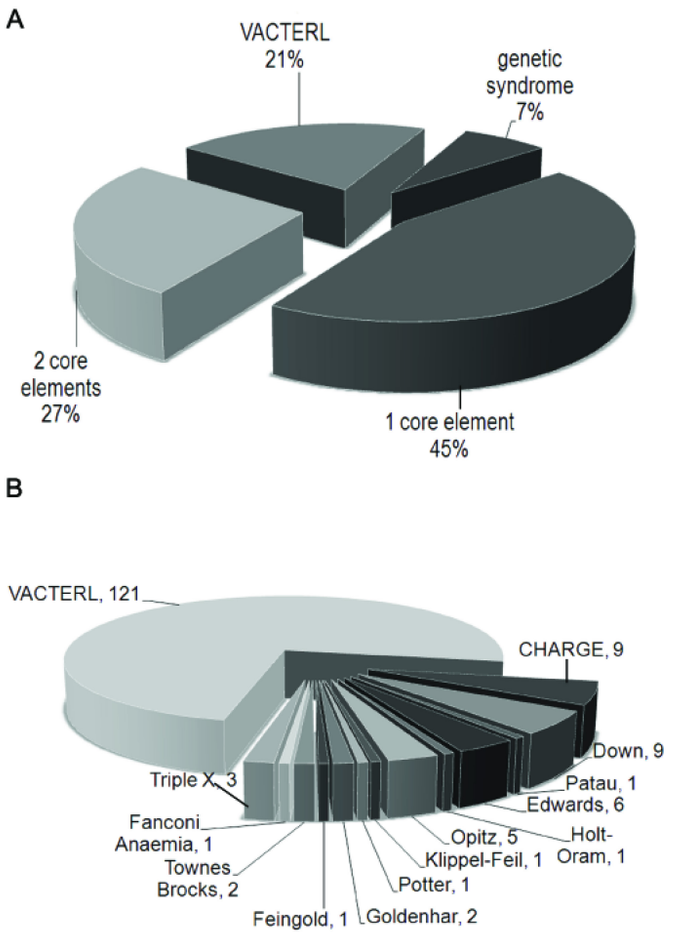

Figure 1. Inclusion criteria. (A) In total 45% of the EA/TEF cohort had 1 core component (TEF) and 27% had one additional core element. 163 patients out of 567 (21%) entries in the Rotterdam EA/TEF cohort met the criteria for VACTERL association. (B) 42 Patients had a confirmed genetic syndrome (7%) and were excluded from the VACTERL cohort.

Results and Discussion

Published CNVs identified in patients with VACTERL association

No large studies looking for CNVs in individuals with VACTERL association have been published. However, several case reports have been published that describe chromosomal anomalies and recurrent and de novo CNVs in VACTERL patients (Table 1). Most of the published de novo genetic anomalies that have been identified in individuals with VACTERL association are unique. However, some changes are recurrent and have been identified in more than one VACTERL patient.

The first is on chromosome 17 where two overlapping deletions have been reported affecting chromosome band 17q23 in both patients. This region contains many genes, but includes 2 candidate genes, TBX2and TBX4, which encode T-box transcription factors. Heterozygous loss of function mutations in TBX4 have been shown to cause small patella syndrome (OMIM #147891) an autosomal dominant skeletal dysplasia characterized by patellar aplasia or hypoplasia and by anomalies of the pelvis and feet [28]. TBX2 has not been implicated in human disease but homozygous Tbx2knockout mice are embryonic lethal and have cardiac anomalies and polydactyly.[29] The second shared locus is chromosomal band 8q24.3 which is duplicated in two individuals with VACTERL. This locus harbours many genes including GLI4 that encodes a member of the krueppel C2H2-type zinc-finger protein family. Although the exact function of GLI4 has not been determined, we note that Gli2$^{-/-}$ and Gli3$^{-/-}$ mice have a VACTERL phenotype [30]. Therefore we consider GLI4 to be an excellent candidate gene.

CNVs identified in the Rotterdam VACTERL cohort

In the Rotterdam VACTERL cohort, we did not observe any clearly de novo CNVs. However, one VACTERL patient with trachea agenesis had a maternally inherited 488 kb 16p12.1 deletion and a 3.7MB deletion on chromosome 5q11.2[31]The 5q11.2 deletion was not inherited from the mother. There was no DNA available from the unknown sperm donor. Several genes are located in this large deletion. Among the top ranked genes by the Endeavour gene prioritization tool are ITGA1, which regulates mesenchymal stem cell proliferation [32]and FST, an activin binding protein. Although *de novo* CNVs were not identified, all of the patients in this cohort had one or more large (>100kb) CNVs (Figure 2). Most of these CNVs were known polymorphisms whose frequencies in normal controls make them unlikely to contribute to VACTERL association.

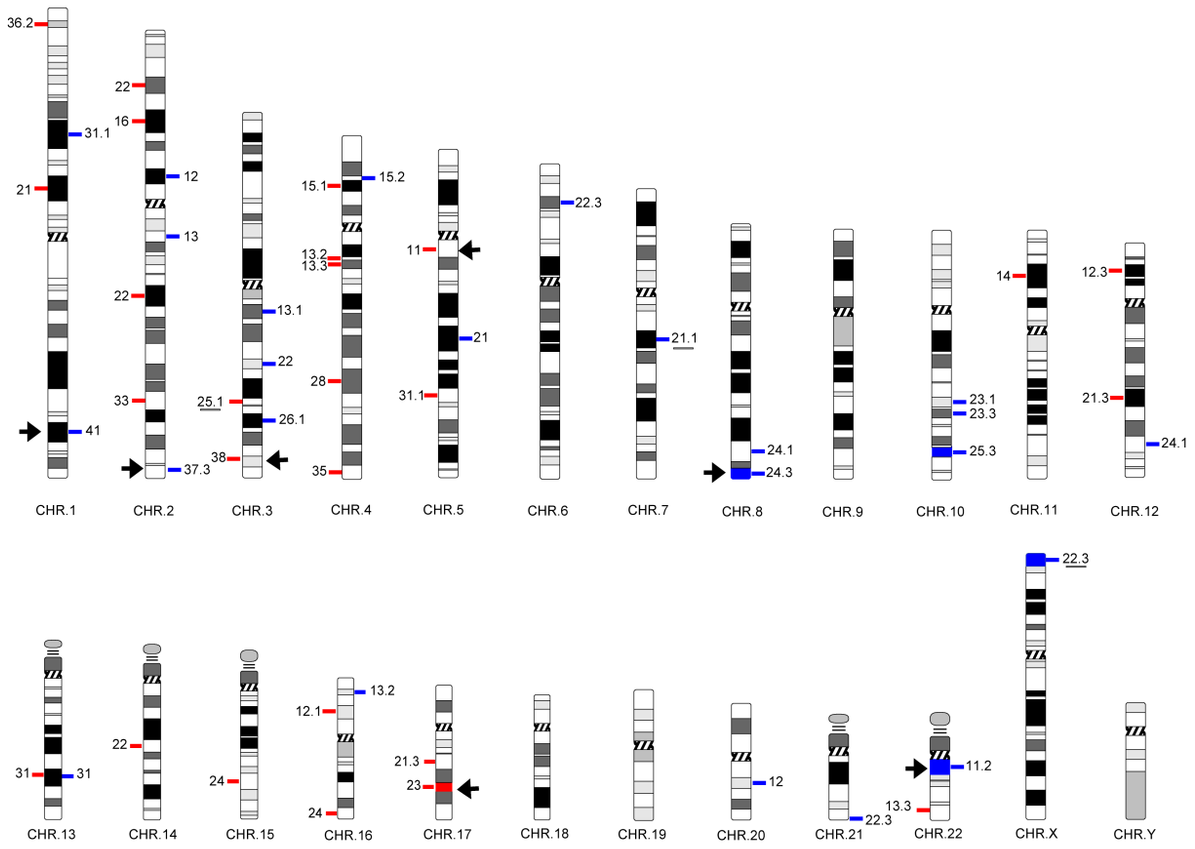

Figure 2. CNVs in the Rotterdam Cohort and recurring published CNV and structural chromosomal anomalies. In this karyogram our institution's unique and rare (underlined) gain (blue) and loss (red) are depicted alongside the chromosomal bands in which they are located. Common polymorphisms are not visualised. At three loci, recurrent gains either from literature (8q24.3), in our cohort (10q25.3) or both (22q11) (blue regions on ideogram). Two published recurrent chromosomal anomalies lead to a deletion of band 17q23 (red region on ideogram). With arrows, *de novo* CNV are depicted.

However, we observed 3 regions with CNVs which are rarely seen in the general population but were shared by more than one of our VACTERL patients (Table 1). These recurring variations can point to pathways or mechanisms involved in disease etiology, especially when they have an extremely low frequency in the general population.

The first of these rare recurrent CNV in our cohort consisted of maternally inherited 300kb duplications in band 10q25.3. This region contains the actin binding LIM protein family member 1 gene (*ABLIM1*) which encodes a protein that may play a role in binding cytoplasmatic proteins to the actin cytoskeleton.[33] Interestingly, Arrington et al. found a 451 kb interstitial deletion on chromosome 3q28 involving only the LIM domain containing preferred translocation partner in lipoma (*LPP*) gene in an individual with esophageal atresia with tracheoesophageal fistula, hypospadia, cardiac-, renal- and rib anomalies.[34]This change was not found in the individual's mother but a paternal sample

was not available, making it impossible to determine if this was a *de novo* change or was inherited from an unaffected family member. In our cohort, no CNVs affecting *LPP* were identified [35] but the presence of deletions affecting both *ABLIM1* and *LPP* in some VACTERL association patients (e.g. esophageal atresia with tracheoesophageal fistula, hypospadia, horseshoe kidneys, hemivertebrae and urinary reflux) suggests that disturbances of the cytoskeleton may contribute to VACTERL phenotypes.

The second recurring CNV was a duplication affecting chromosome 22q11.2. The patient, with all of the 6 core VACTERL features affected and her mother had a 22q11.2 micro duplication overlapping 1.4 Mb of the *de novo* duplication in a VACTERL patient described by Schramm and colleagues.[36]This patient had vertebral fusion, anal atresia, right sided duplicated kidney and additional non-VACTERL deformations.

The third recurring CNV involves a gain of the Short stature HOmeoboX-containing gene (*SHOX*), which plays an important role in limb development. [37] Duplications involving *SHOX* were identified in two VACTERL patients, both of whom had limb anomalies and esophageal atresia with tracheoesophageal fistula. [38]) Moreover, one of the two patients had horseshoe kidneys, hypospadia and dysmorphic features and the second patient an atrial septum defect. The duplications were inherited, one from the patient's mother and the other from the patient's father.

CNVs inherited from an unaffected parent are often considered to be non-pathogenic. [39] However, the absence of a phenotype in the parent does not exclude a causal relationship. Differences in phenotypes seen in individuals carrying the same CNV can be due to several mechanisms, i.e. differences in environmental exposures; the combination of two recessive alleles in the affected individual, variable expressivity, incomplete penetrance, skewed X-inactivation or a two-hit CNV model. In this last model, it is postulated that second hits, or alterations in another genomic region, can affect the same biological process and that the additive effects of these changes result in a particular phenotype.[40, 41]Large CNVs with low population frequencies are excellent candidates for a two hit model.

Due to the scarcity of published *de novo* CNVs and absence of such changes in our cohort, we believe that relatively large *de novo* copy number variations are not a common cause of VACTERL association. Furthermore, we cannot exclude the possibility that smaller, inherited CNVs can contribute to the development of VACTERL phenotypes. We did observe several small unique and rare inherited CNVs in our cohort, which on their own are likely benign but could contribute to VACTERL association in a two hit model. The

latest high resolution arrays contain millions of probes and can detect CNVs as small as 2kb. If we want to progress our knowledge about the role of small CNV in VACTERL association etiology, we must know more about frequency and distribution of these small variants in large normal control populations.

To summarize, we believe that copy number variations can play a role in VACTERL association by shifting the balance from normal to abnormal development in combination with other genetic, environmental, and/or stochastic factors. These changes may also focus attention on genes, pathways, or processes that are frequently affected, by mutations or CNVs, in individuals with VACTERL association. By studying larger cohorts it may be possible to identify additional recurrent CNVs that contribute to VACTERL phenotypes which could, in turn, provide insight into the etiology of VACTERL association.

Chromosome	Type	Remarks (hg18)	Inheritance	Ref
1q41	Gain	chr1: 215 945 774–216 077 064	de novo	[42]
2q37.3	Gain	chr2: 241 202 666–241 227 781	de novo	[42]
2q22-q24.2	Deletion	del(2) (q22q24.2)	de novo	[43]
3q28	Loss	chr3:189395885-189951376	de novo	[34]
5q11	Loss	chr5:51,185,650-55,001,348	ICSI; father NA	[31]
6q25.3-q27	Loss	#1	NA	-
6q13-q15	Deletion	(6)(q13-q15)	de novo	[44]
7	Duplication	trisomy 7	de novo	[45]
8q24.22-q24.3	Gain	#1	NA	-
8q24.3	Gain	chr8: 145 012 210–145 132 100	de novo	[42]
10q22-qter	Duplication	dup(10)(q22-qter)	de novo	[46]
10q25.3	Gain	chr10:116,250,268-116,546,953	Inherited-Mat	
10q25.3	Gain	chr10:116,261,258-116,515,586	Inherited-Mat	
11q23-qter	Duplication	47,XY+der(22)t(11;22)(q23;q11.2)	de novo	[47]
12	Duplication	r(12)	de novo	[48]
13,r(13)	Duplication	trisomy 13	de novo	[26, 49]
16p12.1	Loss	chr16: 21,854,140-22,342,140	Inherited-Mat	[31]
16q24.1	Loss	chr16:82908199–86405076	de novo	[50]
17q23-2q24.3	Deletion	del (17) (q23.2q24.3)	de novo	[51]
17q22-q23.3	Deletion	del (17)(q22q23.3)	de novo	[52]
18q12.1	Loss	#1	NA	-
18q22.2-qter	Deletion	18q22.2-->qter	de novo	[53]
18,r(18)	Duplication	trisomy 18	de novo	[26, 54]
21	Duplication	trisomy 21	de novo	[26]
22q11.2	Gain	chr22:17,281,004-19,792,353	de novo	[36]
22q11.2	Gain	chr22:17,017,139-18,665,833	Inherited-Mat	
22pter-q11.2	Duplication	47,XY+der(22)t(11;22)(q23;q11.2)	de novo	[47]
X	Duplication	Triple X	de novo	[38]
Xp22.3	Gain	chrX:242,432-1,318,727	Inherited-Pat	[38]
Xp22.3	Gain	chrX:327,015-1,889,115	Inherited-Mat	[38]

Table 1 Chromosomal anomalies, recurring and *de novo* CNVs seen in VACTERL patients . #1 abstract Shin et al American Society of Human Genetics 2011, NA data not available; Mat (maternal) and Pat (paternal); recurrent CNV in bold; ICSI (Intracytoplasmic sperm injection)

References

1. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.
2. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome.* Science, 2004. **305**(5683): p. 525-8.
3. Girirajan, S., C.D. Campbell, and E.E. Eichler, *Human copy number variation and complex genetic disease.* Annu Rev Genet, 2011. **45**: p. 203-26.
4. Miller, D.T., et al., *Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies.* Am J Hum Genet, 2010. **86**(5): p. 749-64.
5. Southard, A.E., L.J. Edelmann, and B.D. Gelb, *Role of copy number variants in structural birth defects.* Pediatrics, 2012. **129**(4): p. 755-63.
6. Solomon, B.D., *VACTERL/VATER Association.* Orphanet J Rare Dis, 2011. **6**: p. 56.
7. Huang, X.S., et al., *Two neighboring microdeletions of 5q13.2 in a child with oculo-auriculo-vertebral spectrum.* Eur J Med Genet, 2010. **53**(3): p. 153-8.
8. Bardakjian, T.M., et al., *Association of a de novo 16q copy number variant with a phenotype that overlaps with Lenz microphthalmia and Townes-Brocks syndromes.* BMC Med Genet, 2009. **10**: p. 137.
9. Chung, B., et al., *From VACTERL-H to heterotaxy: variable expressivity of ZIC3-related disorders.* Am J Med Genet A, 2011. **155a**(5): p. 1123-8.
10. Uliana, V., et al., *Expanding the phenotype of 22q11 deletion syndrome: the MURCS association.* Clin Dysmorphol, 2008. **17**(1): p. 13-7.
11. El-Hattab, A.W., et al., *OEIS complex associated with chromosome 1p36 deletion: a case report and review.* Am J Med Genet A, 2010. **152A**(2): p. 504-11.
12. Kosaki, R., et al., *OEIS complex with del(3)(q12.2q13.2).* Am J Med Genet A, 2005. **135**(2): p. 224-6.
13. Klopocki, E., et al., *Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome.* Am J Hum Genet, 2007. **80**(2): p. 232-40.
14. Walsh, L.E., G.H. Vance, and D.D. Weaver, *Distal 13q Deletion Syndrome and the VACTERL association: case report, literature review, and possible implications.* Am J Med Genet, 2001. **98**(2): p. 137-44.
15. Ryan, A.K., et al., *Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study.* J Med Genet, 1997. **34**(10): p. 798-804.
16. Solomon, B., et al., *Clinical Geneticists' Views of VACTERL/VATER Association.* 2012.
17. Hilger, A., et al., *Familial occurrence of the VATER/VACTERL association.* Pediatr Surg Int, 2012. **28**(7): p. 725-9.
18. Solomon, B.D., et al., *Evidence for inheritance in patients with VACTERL association.* Hum Genet, 2010. **127**(6): p. 731-3.
19. Wessels, M.W., et al., *Polyalanine expansion in the ZIC3 gene leading to X-linked heterotaxy with VACTERL association: a new polyalanine disorder?* J Med Genet, 2010. **47**(5): p. 351-5.
20. Damian, M.S., et al., *VACTERL with the mitochondrial np 3243 point mutation.* Am J Med Genet, 1996. **62**(4): p. 398-403.
21. Garcia-Barcelo, M.M., et al., *Identification of a HOXD13 mutation in a VACTERL patient.* Am J Med Genet A, 2008. **146a**(24): p. 3181-5.
22. Szumska, D., et al., *VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5.* Genes Dev, 2008. **22**(11): p. 1465-77.
23. Solomon, B.D., et al., *VACTERL association and mitochondrial dysfunction.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(3): p. 192-4.
24. Thauvin-Robinet, C., et al., *Another observation with VATER association and a complex IV respiratory chain deficiency.* Eur J Med Genet, 2006. **49**(1): p. 71-7.
25. Solomon, B.D., et al., *Consideration of VACTERL association in patients with trisomy 21.* Clin Dysmorphol, 2010. **19**(4): p. 209-11.
26. Felix, J.F., D. Tibboel, and A. de Klein, *Chromosomal anomalies in the aetiology of oesophageal atresia and tracheo-oesophageal fistula.* Eur J Med Genet, 2007. **50**(3): p. 163-75.
27. Shaw-Smith, C., *Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology.* J Med Genet, 2006. **43**(7): p. 545-54.
28. Bongers, E.M., et al., *Mutations in the human TBX4 gene cause small patella syndrome.* Am J Hum Genet, 2004. **74**(6): p. 1239-48.
29. Harrelson, Z., et al., *Tbx2 is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development.* Development, 2004. **131**(20): p. 5041-52.
30. Kim, P.C., R. Mo, and C. Hui Cc, *Murine models of VACTERL syndrome: Role of sonic hedgehog signaling pathway.* J Pediatr Surg, 2001. **36**(2): p. 381-4.

31. de Jong, E.M., et al., *5q11.2 deletion in a patient with tracheal agenesis.* Eur J Hum Genet, 2010. **18**(11): p. 1265-8.
32. Ekholm, E., et al., *Diminished callus size and cartilage synthesis in alpha 1 beta 1 integrin-deficient mice during bone fracture healing.* Am J Pathol, 2002. **160**(5): p. 1779-85.
33. Roof, D.J., et al., *Molecular characterization of abLIM, a novel actin-binding and double zinc finger protein.* J Cell Biol, 1997. **138**(3): p. 575-88.
34. Arrington, C.B., et al., *Haploinsufficiency of the LIM domain containing preferred translocation partner in lipoma (LPP) gene in patients with tetralogy of Fallot and VACTERL association.* Am J Med Genet A, 2010. **152a**(11): p. 2919-23.
35. Hernandez-Garcia, A., et al., *Contribution of LPP copy number and sequence changes to esophageal atresia, tracheoesophageal fistula, and VACTERL association.* Am J Med Genet A, 2012. **158a**(7): p. 1785-7.
36. Schramm, C., et al., *De novo microduplication at 22q11.21 in a patient with VACTERL association.* Eur J Med Genet, 2011. **54**(1): p. 9-13.
37. Vickerman, L., S. Neufeld, and J. Cobb, *Shox2 function couples neural, muscular and skeletal development in the proximal forelimb.* Dev Biol, 2011. **350**(2): p. 323-36.
38. Brosens, E., et al., *Structural and numerical changes of chromosome X in patients with esophageal atresia.* Eur J Hum Genet, 2014.
39. Shaw-Smith, C., et al., *Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features.* J Med Genet, 2004. **41**(4): p. 241-8.
40. Girirajan, S., et al., *A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay.* Nat Genet, 2010. **42**(3): p. 203-9.
41. Veltman, J.A. and H.G. Brunner, *Understanding variable expressivity in microdeletion syndromes.* Nat Genet, 2010. **42**(3): p. 192-3.
42. Hilger, A., et al., *De novo microduplications at 1q41, 2q37.3, and 8q24.3 in patients with VATER/VACTERL association.* Eur J Hum Genet, 2013. **21**(12): p. 1377-82.
43. Woods, C.G., et al., *A case report of a child with an interstitial deletion of chromosome 2 (q22;q24.2) and the VATER association phenotype.* Proc. Greenwood Genet. Center, 1993(12): p. 119-120.
44. McNeal, R.M., R.R. Skoglund, and U. Francke, *Congenital anomalies including the VATER association in a patient with del(6)q deletion.* J Pediatr, 1977. **91**(6): p. 957-60.
45. Schinzel, A., *Catalogue of Unbalanced Chromosone Aberrations in Man.* 2 ed. 2001, Berlin: Walter de Gruyter GmbH & co.
46. Field, B., A. Smith, and D. Sillence, *Malformation syndrome of chromosome 10q duplication and the Vater association.* Ann. Genet (Paris), 1983. **26**: p. 31-33.
47. Prieto, J.C., et al., *Phenotypic expansion of the supernumerary derivative (22) chromosome syndrome: VACTERL and Hirschsprung's disease.* J Pediatr Surg, 2007. **42**(11): p. 1928-32.
48. Cinti, R., et al., *Molecular characterisation of a supernumerary ring chromosome in a patient with VATER association.* J Med Genet, 2001. **38**(2): p. E6.
49. Lorentz, C.P., et al., *Mosaic r(13) resulting in large deletion of chromosome 13q in a newborn female with multiple congenital anomalies.* Am J Med Genet, 2002. **111**(1): p. 61-7.
50. Stankiewicz, P., et al., *Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations.* Am J Hum Genet, 2009. **84**(6): p. 780-91.
51. Levin, M.L., et al., *Unique de novo interstitial deletion of chromosome 17, del(17) (q23.2q24.3) in a female newborn with multiple congenital anomalies.* Am J Med Genet, 1995. **55**(1): p. 30-2.
52. Marsh, A.J., et al., *Interstitial deletion of chromosome 17 (del(17)(q22q23.3)) confirms a link with oesophageal atresia.* J Med Genet, 2000. **37**(9): p. 701-4.
53. Dowton, S.B., et al., *Chromosome 18q22.2-->qter deletion and a congenital anomaly syndrome with multiple vertebral segmentation defects.* J Med Genet, 1997. **34**(5): p. 414-7.
54. van der Veken, L.T., et al., *Low grade mosaic for a complex supernumerary ring chromosome 18 in an adult patient with multiple congenital anomalies.* Mol Cytogenet, 2010. **3**: p. 13.

2.2

Recurrent Copy Number Variation in patients with Esophageal Atresia and VACTERL associated anomalies

Adapted from:

Recurrent Copy number variation in patients with Esophageal Atresia and VACTERL associated anomalies.

E. Brosens, E.M. de Jong , H. Eussen[1], J. Douben, Y. van Bever, H. IJsselstijn, H.P. Zaveri, R. Wijnen, D.A. Scott, D. Tibboel, A. de Klein. *Manuscript in preparation*

Abstract

Esophageal Atresia (EA) with or without TracheoEsophageal Fistula (TEF) are common congenital anomalies whose cause is unknown in over 90% of affected patients. EA/TEF can either be present as an isolated defect or in association with other developmental defects: e.g. as one of the core features of the VACTERL (Vertebral, Anal, Cardiac, Tracheo-esophageal fistula, Renal and Limb anomalies) association. The hypothesis that genetic defects contribute to both EA/TEF and VACTERL etiology is supported by the fact that EA/TEF is a variable feature of several known monogenetic syndromes. Among these possible defects are Copy Number Variations (CNVs). As de novo CNVs can help to identify causal genes or affected biological pathways, the recurrence of unique and rare inherited CNVs may, in combination with other factors, predispose for the development of EA/TEF and other features of the VACTERL association.

We therefore profiled 268 affected individuals with micro-array. All had one or more relatively large (≥ 30kb) CNV, most of them known polymorphisms. However, sixteen loci contained putative de novo CNVs, most of which were seen only once in our patient cohort. These (putative) de novo variants were observed at two previously published (5q11 and 16q24.1), four loci affecting genes of known syndromes (4p16.3, 4q21, 7p14.1, 7q36) and twelve new loci: i.e. 4q35, 6q23.2, 7p22, 8p22, 8q13.1, 11p14, 15q21, 13q12.11, 16p13.11, 21q22, Xp21. We also identified 295 inherited CNVs which were either absent or rare in published cohorts of control individuals and our in-house control database. Interestingly, 26 of these inherited variants were recurrent in our patient cohort. Using our genome-wide multiplatform approach we identified several loci that may interfere with biological pathways disturbed in EA/TEF or VACTERL association patients.

Introduction

Foregut-related anomalies are complex diseases believed to be caused by multiple genetic and environmental factors.[1, 2] With a prevalence of 2-3 in 10,000 live births [3] Esophageal Atresia (EA) with or without Tracheo-Esophageal fistula (TEF) is one of the more common foregut-related anomalies. Although over half of these patients have EA/TEF as an isolated defect, many also have additional congenital malformations; 23% of them have at least two other of the six core defects of VACTERL (Vertebral, Anal, Cardiac, Tracheo-Esophageal fistula, Renal, Limb anomalies) association.[4]

Chromosomal anomalies, including aneuploidies as trisomy 13, 18 and 21, can be identified in 6-10% of patients.[2, 5, 6] These aneuploidies and other genetic disorders represent the 9-10% of EA/TEF patients with a confirmed genetic syndrome[7, 8], and there is a strong suspicion that genetic factors are involved in the remainder. One of the reasons for this is that EA/TEF is a variable feature of many monogenetic syndromes associated with single base-pair mutations. Other VACTERL-associated anomalies have a large overlap with these syndromes, e.g. Feingold syndrome (*MYCN*), CHARGE syndrome (*CHD7*), Anophthalmia-Esophageal-Genital (*AEG*) syndrome (*SOX2*) and Fanconi anemia. [2, 9] Non-syndromic TE is usually a sporadic findings, although there are reports describing familial recurrence.[10-12]

Few studies describe genetic aberrations in EA/TEF and VACTERL association. [13, 14] EA/TEF is reported in patients with aneuploidy, structural chromosomal anomalies, single-nucleotide mutations and *de novo* Copy Number Variations (CNV).[5, 15] These CNV, which are detected with microarray technology, have a well-established role in congenital anomalies: this technique has identified many genes or loci involved in developmental defects.[16-18] CNV can contribute to disease etiology in several genetic syndromes which have TE anomalies as a variable feature such as Feingold syndrome[19], 22q11 deletion syndrome, Fanconi anemia[20], CHARGE syndrome[21], alveolar capillary dysplasia[22] and mandibulofacial dysostosis.[23] Despite sporadic evidence that *de novo* CNV can play a role in non-syndromic EA/TEF and the VACTERL association[24] no frequently recurring risk loci have yet been identified.

We hypothesize that *de novo* genomic and inherited rare recurrent CNV could predispose to EA/TEF and VACTERL association. These CNV harbor one or more disease-related genes or phenotype-modifying factors. We describe the variation detected in our cohort (n=268) enabling us to identify several rare recurring and *de novo* CNV. Our multiplatform array screening provides insight into the biological pathways and disease mechanisms involved.

Methods

Study design

Candidate genes for rare genetic conditions are usually identified using homozygosity mapping or linkage analysis (and subsequent sequencing of these loci) in large families with several affected and unaffected relatives. Unfortunately, because familial recurrence is extremely low in EA/TEF [25, 26] we had to use a different strategy to test our hypothesis. Our strategy was based on three assumptions: (1) CNV are usually considered to be causal for the abnormal phenotype in congenital anomalies if this CNV is absent in the parents of the affected individual and if it targets relevant genes. (2) Copy number variation recurrent in single cases could indicate loci harboring genes mutated or otherwise affected in larger disease cohorts. (3) Although a CNV is inherited, it could act as a modifier or in a multiple hit model if it has an extremely low population frequency and is also seen more than once in patients with a rare disease.

Patient cohort

This research was approved by the Medical Ethical Review Boards of Erasmus MC - Sophia Children's Hospital (Rotterdam, NL) and Baylor College of Medicine (Houston, USA). Patients with EA/TEF and VACTERL associated anomalies were identified from the medical records and/or referral to clinical geneticists. We reviewed these records; patients with a confirmed genetic syndrome or chromosomal aberration, were excluded from further analysis. After retrieval of parental informed consent, blood was drawn from 268 patient and their parents. Two hundred and thirty nine patients were treated in the Erasmus MC-Sophia, twenty eight in the Baylor College of Medicine. The major abnormalities of each patient are available on request, the phenotypes of patients with de novo CNV (table 2) and the patients with overlapping CNV and phenotypes are described in supplementary table 2. VACTERL patients were defined as those with three or more anomalies of the VACTERL core components with the diagnostic criteria described earlier [7], and absence of a confirmed genetic syndrome.[27]

DNA isolation

DNA for genomic analysis was extracted from peripheral blood and fibroblast cells with the Puregene DNA purification kit (Gentra Systems, USA) or QIAamp DNA Blood Midi Kit. (Qiagen, Inc., Hilden, Germany) DNA was extracted from Formalin Fixed Paraffin Embedded (FFPE) samples (thymus tissue) as instructed by the supplier of the microarrays (Agilent Inc., San Diego, CA, USA). DNA quality was evaluated with the Thermo Scientific Nano Drop 2000 (Thermo Fisher Scientific Inc., Waltham, USA), and dsDNA quantity with the Quant-iT™ PicoGreen ® dsDNA kit. (Invitrogen, Carlsbad, CA, USA). A total of 200 ng dsDNA was used as input for microarray analysis.

Detection of genomic variation

Karyotyping was done according to standard analytical methods. All DNA was tested for subtelomeric aberrations with Multiplex ligation-dependent probe amplification analysis, using the P036E1 and P070A2 Salsa telomere kits (MRC Holland, Amsterdam, the Netherlands) as published previously. [28]

High-resolution analyses were performed using single-nucleotide polymorphism (SNP) microarrays (Illumina Inc., San Diego, CA, USA and Affymetrix Inc. Santa Clara, CA, USA) and CGH (Agilent Inc., San Diego, CA, USA) oligonucleotide-based arrays using standard protocols. Micro-array analysis was initially performed using three types of array chips (GeneChip Human Mapping 250K Nsp, 12-HumanCytoSNP DNA Analysis BeadChips v1-v2.1, and Agilent Human Genome CGH 105K and 244K), later on these were replaced by chips with a higher resolution (Illumina Human 610-Quad Beadchip and Illumina HumanOmniExpress BeadChip, Illumina Infinium CytoSNP-850k BeadChip and Agilent SurePrint G3 Human CGH 1M Oligo Microarray Kits G4411B/G4447). Most of the patients were screened with more than one array chip type.

We generated normalized output with Feature Extraction software (version 9.1), with CGH analytics software (version 3.3.28), with Affymetrix GTYPE v4.1 (Affymetrix, Santa Clara, CA) or with Illumina Genomestudio version 2011.1, depending on chip type. (Illumina, San Diego, CA, USA) CNVs in patient samples were visualized as log2-Ratios (Log2R) detected through comparison of patient probe intensity data with those of a virtual reference set of 400 female CEU samples (SNP-array) or by comparison with sex matched controls of unaffected, unrelated individuals or reference DNA. (Promega Corporation, Madison, WI)

Analysis settings

We visualized SNP-array Copy Number State (Log2R) and allelic ratio (B-Allele Frequency, BAF) in Biodiscovery Nexus CN7.5 (Biodiscovery inc, Hawthorne, CA, USA) This program estimates SNP array copy number state using the SNP-FASST2 Segmentation algorithm, a Hidden-Markov-Model -based segmentation algorithm that uses a combination of the BAF and LogR states to determine copy number and allelic segments. Segmentation significance threshold was set at 5.0E-7 with a minimum of 5 probes per segment and a maximum probe spacing of 1000kb. The log2R-ratio thresholds were set at +0.18 (single copy gain), -0.18 (single copy loss), 0.4 (two or more copy gain) and -1.1 (homozygous loss). The Homozygous Frequency/ Homozygous Value/ Heterozygous Imbalance Threshold were set at 0.95/0.8/ 0.4. The minimum LOH length was set at 100kb and minimum SNP probe density, at 10 probes/Mb. Gender correction was used with a 3:1 sex chromosome gain threshold of 1.2 and a 4:1 sex chromosome gain threshold of 1.7. Log2R ratios of CGH-array results were determined with the ADM2 algorithm with filtering options of a minimum of 3 probes and abs. (log2Ratio) >0.3. Each segment deviating from the normal situation was reviewed by visual inspection in Nexus CN 7.5. During the course of this study, the genome build switched from build hg18 to hg19. Not all arrays-chips could be re-processed or re-analyzed in the new genome build. Detected regions were transformed using the UCSC lift-over tool. Re-analysis of SNP-array in the new genome build, improved segmentation and waving correction algorithms resulted in the loss of several low confidence CNV calls.

Validation of microarray results

The frequency of each individual CNV in our cohort was compared with CNV frequency at that locus in the Toronto Database of Genomic Variants [29], the Copy Number Variation project at the Children's Hospital of Philadelphia [30] and an in-house control database of common copy number variations found in unaffected individuals. Genomic changes were considered to be a *private CNV* if similar changes were absent, and to be a *rare CNV* if their frequency was less than 0,25% in these databases. All CNV above this last threshold were considered to be common polymorphisms.

We used different array chips in this experiment, each with different marker spacing, distribution and content. Therefore, we set a size CNV restriction threshold of minimum 30kb for single events and 10-30kb, depending on array type and probe content and spacing, if there were multiple CNVs affecting the same locus in our patient cohort. We only confirmed CN state and inheritance pattern of unique and rare CNV if their size exceeded

the restrictions thresholds and contained genes. Similar to the size settings, if a CNV did not contained genes it was only further examined when there were multiple CNVs affecting the same locus.

To confirm the *de novo* results, patient and parental CN quantification was done by either additional SNP array, Real Time Quantitative PCR, Fluorescence In Situ Hybridization (FISH) and/or Multiplex Amplicon Quantification (Multiplicon N.V., Gent, Belgium). Two Primer pairs for qPCR were designed within the possible CNVs using Primer Express Software v2.0 (Applied Biosystems, Foster City, CA, USA), absence of SNPs in primer sequences confirmed in dbSNP (build 135) and specificity of amplified region determined with the University of Santa Cruz (UCSC) *in sillco* PCR and melting-curve analysis. Primer pairs with repeats in their resulting amplicon, according to CENSOR repeat masker[31] were excluded. QPCR experiments were performed using a Lightcycler 1.5 instrument and LightCycler FastStart DNA Master SYBER Green I kit with C14ORF145 as a control locus. (Roche Molecular Diagnostics, Indianapolis, IN, USA) Each sample, including the no-template control (NTC) and control DNA, was run in triplicate. Locus exon 4 of the KIAA1279-gene was used as a control.[11] Cut-off values of <0.7 were used for deletions and values of >1.3 for duplications. For FISH confirmation, BAC-clones were selected from the UCSC genome browser and ordered from BACPAC Resources. After isolation of the BAC-DNA, the probes were amplified, labeled and used for FISH, according to standard protocols and described earlier.[32]

The MAQ assay is a PCR-based-amplification method which uses 6 primer pairs on different loci for sample-internal copy number normalization, maximal 5 CNV specific primer pairs developed with the manufacturers' software package, and one type of FAM labeled primer specific to the sequence-tagged forward primers to amplify 20ng of dsDNA input. The DNA of 4 unaffected, unrelated individuals is used as sample-external copy number normalization. We amplified the DNA according to the manufacturer's protocol in a thermocycler with a heated lid, and analysed the resulting Fam-labelled amplicons of patient, parental and CEU controls on an automated sequencer (ABI 3730XL, Applied Biosystems, Foster City, CA, USA). This capillary electrophoresis step separates fragments on the basis of their length, these differences in amplicon length make multiplexing possible. The differences in fluorescence intensity reflect copy number state and are visualized in the MAQ-S analysis tool (Multiplicom Inc., Niel, Belgium) which compares amplicon size to the Genescan LIZ 500 size standard and normalizes copy number state to the internal amplicons and 4 external references.

Results

Patient cohort

In this study we determined the copy number profile of 268 patients of the Erasmus MC Rotterdam and Houston Baylor College of Medicine cohorts of Esophageal Atresia patients. Supplementary table 2 describes the patients' phenotypes, which have recurrent CNV. We performed CNV profiling on 268 out of 583 patients of the combined cohorts. A total of 60 had EA/TEF as an isolated defect, 86 had one other additional major VACTERL associated defect, and 122 patients fulfilled all of the criteria or VACTERL association. This means that 77.6 % of patients had additional, predominantly VACTERL associated, major congenital anomalies.

Screening our large cohort with high-resolution oligonucleotide and SNP microarrays led to the identification of many copy number polymorphisms already described in unaffected individuals (*data not shown*).[29, 30, 33] Comparison of the recorded *de novo* CNV data with known EA/TEF loci from animal models or syndromic EA/TEF yielded five loci of interest. One *de novo* CNV covering *FRAS1*, a small intronic *GLI3* deletion, a *WHSC1* deletion, and two previously described *FOXF1* regulatory site deletions.[22]

We observed 295 rare and private inherited CNV (~1 per patient), widely distributed over the genome. Interestingly, our analysis identified 26 rare CNV observed in multiple patients, 14 patients had *de novo* CNV of which three patients had more than one *de novo* CNV. Several, not all, of these loci affect genes with a possible role in the abnormal phenotype. Table 1 depicts the *de novo* CNV seen in the Erasmus MC-Sophia and Baylor College of Medicine TE-cohorts. Three examples of de novo CNV are depicted in figure 1, 2 and 3. Recurrent CNV are visualized in figure 4 and described in online supplementary table 1, single rare and private CNV regions are available on request. In addition to the current described genomic aberration, two patients have a chromosomal anomaly, twenty-two patients have a whole chromosome duplications: three have Triple X syndrome, 11 have Down syndrome, seven have trisomy 13 and one has trisomy 18.[34]

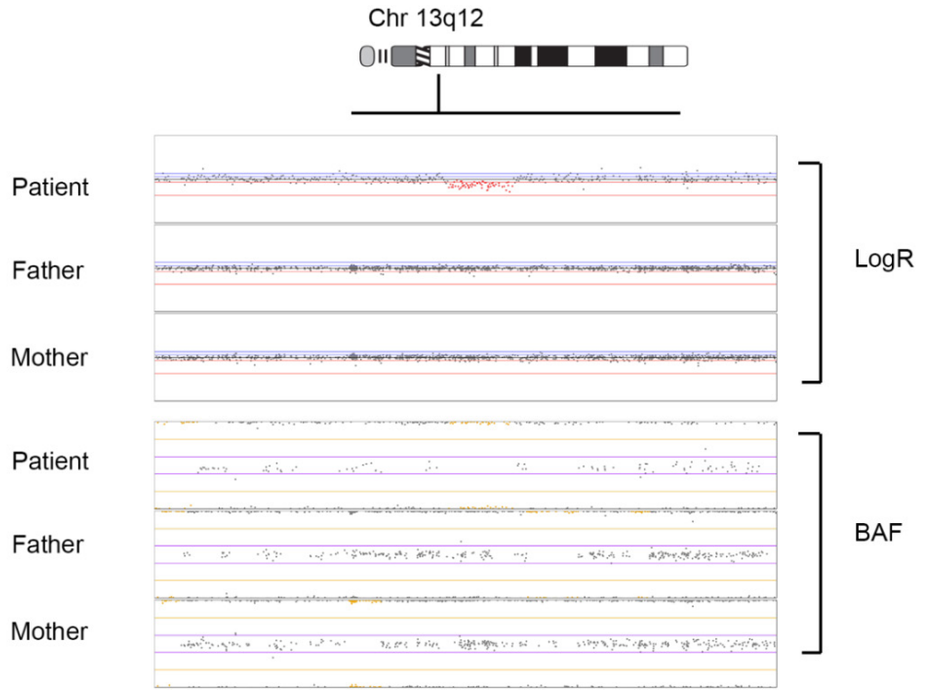

Figure 1. De novo deletion on chromosome 13q21. Note the loss (red dots) in the patients logR track and the loss of Heterozygosity (yellow dots) in the patients B-allele frequency (BAF) plot.

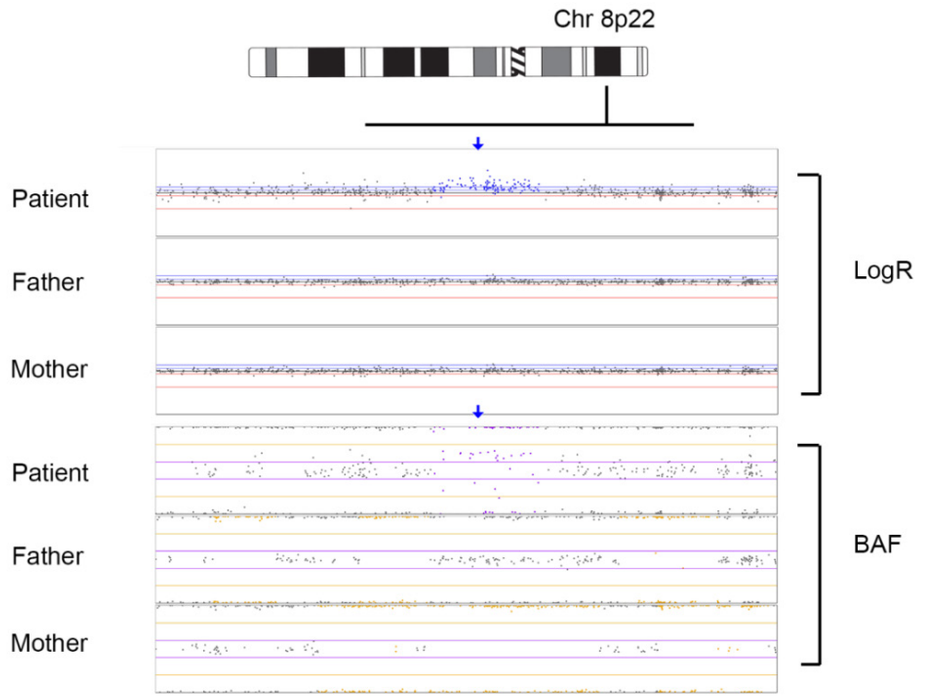

Figure 2. De novo duplication on chromosome 8p22. Note the gain (blue dots/arrow) in the patients logR track and allelic imbalance (purple dots/arrow) in the patients B-allele frequency (BAF) plot.

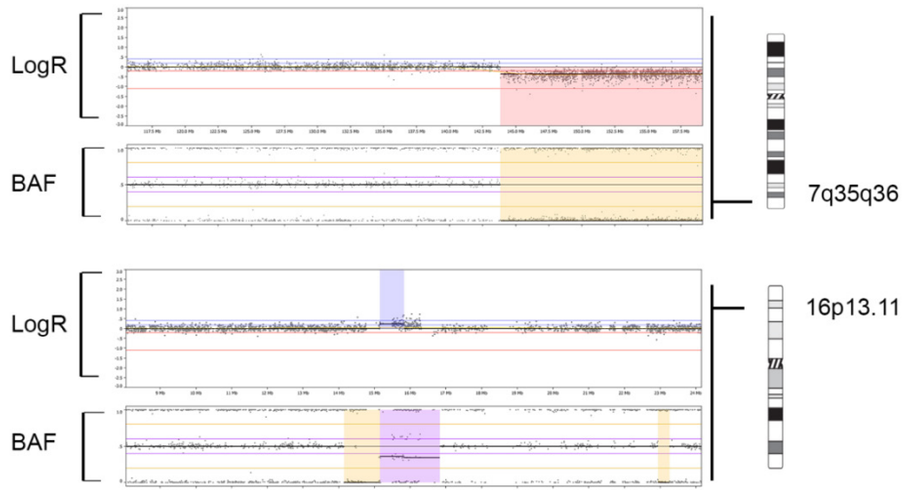

Figure 3. De novo deletion ranging from chromosomal band 7q35 to 7q36.3. Note the loss (red) in the patients logR track and the loss of Heterozygosity (yellow) in the patients B-allele frequency (BAF) plot. Lower panel depicts the de novo 16p13.11 duplication in the same patient. Note the gain (blue) in the patients logR track and allelic imbalance (purple) in the patients B-allele frequency (BAF)

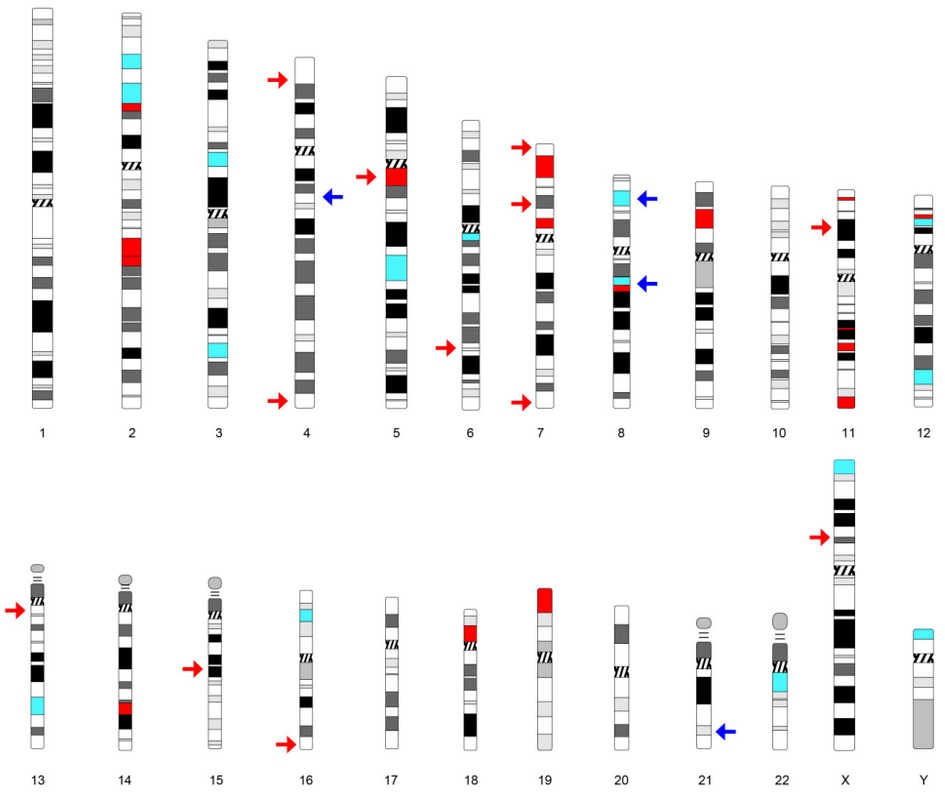

Figure 4 Overview of de novo deletions (red arrows), de novo duplications (blue arrows) and recurrent inherited losses (red colored chromosomal bands) and gains (blue) in the Erasmus MC-Sophia and Baylor College of Medicine TE cohorts.

Type	Cytoband	Chromosome Region	Length	(Candidate) genes	Reference
Gain	1q41	chr1:217879151-218010441*	131290	*SPATA17*	[35]
Gain	8q24.3	chr8:144940222-145060112*	119890	*PLEC-1, PARP10*	
Loss	2q33.2q35	chr2:204,394,564–219,189,331	14794767	*MAP2*	[36]
Loss	3q28	chr3:187913191-188468682*	555491	*LPP*	[37]
Loss	4p16.3	chr4:1889113-1966468	77355	*WHSC1*	This manuscript
Gain	4q21.1q21.21	chr4:78968955-78988693	19738	*FRAS1*	This manuscript
Gain	21q22.2	chr21:40100880-40154748	53868	*LINC00114*	
Loss	4q35.2	chr4:189025846-189042743	16897	*TRIML2*	This manuscript
Loss	5q11	chr5:51149893-54965591*	3815698	*Multiple*	[38]
Loss	6q23.2	chr6:134730991-135020544*	289553	*AJ606330, LOC154092, LINC01010, AJ606331*	This manuscript
Loss	7p22	chr7:6123758-6139997*	16239	*Distal to USP42*	This manuscript
Loss	7p14.1	chr7:42125411-42131780	6369	*GLI3-intronic*	This manuscript
Loss	7q35q36.3	chr7:143,839,360-159,138,663	15299304	*Multiple incl. SHH, MNX1*	This manuscript
Gain	16p13.11				
Loss	7q35q36	chr7:147683847-159088636*	11404789	*SHH, MNX1*	[39, 40]
Gain	8p22	chr8:17625479-17813225	187746	*FGL1, LFIRE1, MTUS1, PCM1*	This manuscript
Loss	11p14.3	chr11:21853276-22016176	162900	.	
Loss	15q21.3	chr15:53994879-54022499	27620	*WDR72*	
Gain	8q13.1	chr8:66955527-66980813*	25286	*DNAJC5B*	This manuscript
Loss	11q13.1q13.2	chr11:65508902-67473140*	1964238	*GSTP1*	[41]
Loss	13q12.11	chr13:22688792-22981935	293143	*AK054845, LINC00540*	This manuscript
Loss	16q24.1	chr16:85979509-86361236*	381727	*LINC01082, LINC01081*	[22]
Loss	16q24.1	chr16:86202409-86286888*	84479	*LINC01082, LINC01081*	[22]
Gain	17q12	chr17:34727386-36297053**	1569667	*AATF, TADA2L, HNF1B*	[42]
Loss	20q13.33	chr20:60238426-60895697*	657271	*GTPBP5*	[43]
Null#	Xp21.11	chrX:22961251-23106879	145628	*DDX53, LOC100873065*	This manuscript

Table 1. De novo CNV in the Erasmus MC-Sophia and Baylor College of Medicine TE cohorts and de novo CNV described in literature. *Region from hg18-hg19 liftover, ** Region from hg17-hg19 liftover, #nullizygous

Discussion

We hypothesized that both *de novo* genomic and inherited rare recurrent CNV could predispose to, or modify the phenotype of, EA/TEF and VACTERL association patients. We indeed identified several *de novo*, unique and rare recurring CNV with this multiplatform micro-array cohort screening. Genes affected by these CNV could help to give insight in disease mechanisms and biological pathways involved in EA/TEF and VACTERL association disease etiology. CNVs larger than 500 kb are not frequent in our genome, in the general population ~ 9% of individuals have a CNV larger than 500kb of which only 3% is larger than 1 Mb. The *de novo* rate of these large CNV is estimated at 1.2 x 10-2 per generation.[44, 45] We identified 295 rare and private inherited CNV larger than 30kb (~1 per patient), widely distributed over the genome. Interestingly, our analysis identified 26 of those rare CNV in multiple patients. Fifteen patients had *de novo* CNV, two patients had multiple putative *de novo* CNV. All *de novo* CNV was non-recurrent, except two previously published deletions on chromosome 16q24.1[22] However, de novo CNV had overlap with structural chromosomal anomalies previously described in EA/TEF.[5, 15]

Aneuploidies and structural chromosomal anomalies

These chromosomal anomalies and aneuploidies are however, not frequent observed phenomena, except for those seen in genetic syndromes such as trisomy 13, 18 or 21. Some of the recurrent structural chromosomal anomalies have overlap with the de novo CNV seen in our study. For instance, 46,XY,der(4)t(3;4)(p25;q35)mat[46] has overlap with the 4q35 deletion and 46,XX,-13,+der(18)t(13;18)(q12;p11.2)[47] has overlap with the 13q12 deletion. Genes in these overlapping regions could have impact on EA/TEF aetiology. Unfortunately, little is known about the genes within the smallest region of overlap For instance, the tripartite motif family-like 2 gene (*TRIML2*), the only gene affected by the 4q35 *de novo* deletion is expressed in whole brain and in lymphoblast B-cells. These type of B-box and SPRY domain containing proteins might play a role in early embryonic development[48], but the exact role or mechanism is unknown.

Balanced structural chromosomal anomalies are not detected by micro-array and, since this technology has replaced high resolution GTG-banded karyotyping as a first tier diagnostic procedure[49], these will not be detected as much as in the past. Inherited, balanced translocations are seen in EA/TEF patients. For example, Cetinkaya and co-workers [50] describe a familial translocation (1;13)(p8;q12) in a patient with esophageal

atresia and a translocation 46,XY,t(5;10)(q13;q23)pat, inherited from an unaffected father was described by Felix and co-workers.[5]

Rare recurrent inherited CNV

Inheritance of a single CNV from a healthy parent is generally regarded as a benign variant. These rare and private are proposed to arise after replication errors.[16] These CNV have such a low population frequency that they have either arisen recently and have no biological meaning or are somehow detrimental and are virtually extinct from the population. Interpretation of these CNV is difficult, for instance they can be ancestry specific[51] or have only modifying role, like the 1q21 deletion in TAR syndrome [52] or the 16p21.1 micro-deletion in patients with developmental delay.[53] Other factors complicating interpretation of rare and private inherited CNV are variable gene expressivity, incomplete penetrance, skewed X-inactivation and/or mutations elsewhere in the genome.[16]

Recurrent inherited rare or private CNV and an overlapping patient phenotype extending trachea-esophageal anomalies are 2p16 gain (patients have EA/TEF and anorectal malformations), 3q21.1 gain (patients have EA/TEF and cardiac anomalies), 7p21 loss, (patients have EA/TEF and cardiac anomalies), 8p22 gain (patients have EA/TEF and vertebral anomalies), 11q21 loss (patients have EA/TEF, anorectal and genitourinary anomalies), a 22q11 duplication (patients have EA/TEF, renal anomalies) and PAR1 duplication (patients have EA/TEF, thumb anomalies). Previously we have described these last two patients with inherited duplications of Short stature HOmeoboX-containing gene (*SHOX*) within the X/Yp PAR1 region.[34] We proposed that these CNV contributed to the patients'phenotypes and might be a determinant in the secondary phenotypical anomalies.

Reduced penetrance of these CNV is also described in patients with EA. Since the introduction of micro-array technology reports describing chromosomal rearrangements, of which detection is often dependent on karyotyping, are less frequent. Cetinkaya describes an inherited translocation involving chromosome 1 and 13, t(1;13)(p8;q12), in a patient with esophageal atresia. [50] His translocation was inherited from a healthy parent. We also observed incomplete penetrance in our cohort: one patient with an inherited 22q11 microduplication syndrome had a severe phenotype; all 6 VACTERL features were present. In contrast her mother, with similar 22q11 duplication, had only mild dysmorphic features. An overlapping duplication has also been described as a *de novo* micro duplication [54] and

in a translocation.[55] This regions overlaps the *TBX1* gene, , a dosage sensitive gene, responsible for the phenotype in DiGeorge syndrome [MIM 188400][56] the associated gene in the 22q11 deletion syndrome. Deletions of this region have been observed in patients with trachea-esophageal anomalies.[57, 58]

Yet another, explanations for an absence of phenotypical characteristics is variable expressivity. For instance, while a part of their phenotype overlaps, Faguer describes differences in expression of a micro-duplication. [59] Faguer reports a father with bilateral vesico-ureteric reflux and renal hypodysplasia and his child with left multicystic hypodysplastic kidney with megaureter, vesico-ureteric reflux and bladder diverticulae and esophageal atresia, both having a duplication harboring the *HNF1B* gene mutated in one fifth of patients with hypodysplastic kidneys.[59]

Rare and private CNV could have a subclinical phenotype in the patients parents. Or on their own they are not enough to disturb normal development, but act in a so-called two-'hit' model. Two factors, genetic or non-genetic, tilt the balance from normal to abnormal development. This two hit hypothesis, first proposed by Knudson in tumor genesis[60], states that multiple factors disturb a similar biological process. The co-occurrence of these rare CNV in a cohort of a rare, heterogeneous, developmental defects as EA/TEF could be a chance effect, representing the increased screening of patients compared to unaffected individuals. Although their virtual absence in large cohorts such as the database of genomic variants suggests otherwise. Proving polygenetic inheritance is difficult. Recently Chan and co-workers describe this phenomenon in schizophrenia and diabetes[61], but they also point out false positive results in GWAS association studies.

If these rare CNV are modifiers and second, genetic, hits are present than these will most likely be gene mutations. Tracheo-esophageal anomalies are frequent in single gene disorders and perhaps diagnosis of these disorders is difficult due to the changed phenotypical characteristics of patients carrying both the modyfing factor and the gene mutation. Duplications might be rescue mechanisms in which a normal copy is duplicated to balance out a copy affected by a mutation, resulting in increased gene expression[62] or deletions might worsen an otherwise less severe condition.

De novo CNV in this study cohort

A CNV is most likely to be pathogenic if it is absent in both unaffected parents and affects meaningful genes. These (putative) de novo variants were observed at two

previously published (5q11[38] and 16q24.1[22]), four loci affecting genes of known syndromes (4p16.3, 4q21, 7p14.1, 7q36) and twelve new loci: i.e. 4q35, 6q23.2, 7p22, 8p22, 8q13.1, 11p14, 15q21, 13q12.11, 16p13.11, 21q22, Xp21. Excluding aneuploidies and confirmed genetic syndromes this would total the frequency of patients with *de novo* CNV at 5.2% and the *de novo* rate of rare and private CNV at 6.1%. Although this *de novo* rate is increased compared to the general population, we describe much smaller CNV. Assuming variation increases with decreasing variation size, these numbers are not as different from the population background as seen in Congenital Diaphragmatic Hernia[63] or intellectual disability.[64] Moreover, the *de novo* CNV seen in EA/TEF is in general non-recurrent and does not always affect genes with clear association to abnormalities seen in patients. Therefore, these *de novo* CNV could very well be non-causal.

Two de novo deletions involve twin pregnancies. The first, male patient SKZ_2038, born 2nd from a monochorionic diamniotic pregnancy, with reduced hearing, duodenal and esophageal atresia and trachea-esophageal fistula. *SALL1* mutation screening was negative. Wolf-Hirschhorn syndrome patients have seizures, typical craniofacial malformations and facial dysmorphisms. It is a contiguous gene syndrome, with variable deletion sizes. The characteristic facial features are not always present in patients with smaller deletions.[65] In patient SKZ_2038 *only* the *WHSC1* gene was deleted. Many syndromes have TE anomalies as an incidental finding[4] and now this syndrome can be added to the list. His brother did not have trachea-esophageal anomalies, but will be screened for Wolf-Hirschorn features later in life. The twin of patient SKZ_1662 patient, with a 13q12.11 *de novo* deletion died in utero. No information was available regarding observed congenital anomalies. The female index patient has EA/TEF, tracheal stenosis, a sacral abnormality and her left kidney was abnormaly positioned in the midline. Within this region one transcribed mRNA (*AK054845*) and one lncRNA (*LINC00540*) are located. No information was available on their biological role. At approximately 500kb distance the *FGF9* gene is located. Mice fgf9 knockouts can have a wide variety of abnormalities, including developmental problems of the skeletal, respiratory and the gastrointestinal system.[66] Perhaps a *FGF9* regulatory region is affected by the *de novo* deletion.

The DNA of female patient SKZ_1307 was previously analyzed on a low-resolution beadchip. reanalysis, transformation to genome build hg19 and comparison to high resolution SNP-array data of his parents hinted at the presence of a putative *de novo FRAS1* deletion. Patients with Fraser syndrome can have VACTERL associated anomalies [27]

and homozygous recessive or compound heterozygous *FRAS1* mutations are characterized by cardiac, renal and genitourinary malformations as well as crypophthalmos and syndactyly. This patient has long fingers and toes, anal atresia and EA/TEF, none of the typical Fraser syndrome associated anomalies. However, this patient has a de novo duplication, which could be a contributing factor in this uncharacteristic phenotype. Moreover, an additional putative *de novo* gain in copy number is present in this patient which affected a long non-coding RNA, LINC00114, on chromosome 21.

The importance and biological impact of most detected *de novo* deletions is uncertain. For instance, the *de novo* 7p22 deletion does not affect genes. It is located distal to the USP42 gene, which might involve in p53 stability.[67] A 6q23.2 deletion in a patient with dysmorphic features and a ventricular septal defect also did not involve any genes. However several transcribed noncoding RNAs are affected and these might influence the expression of important genes. For instance, at 300 kb distance the *TCF21* gene is located, a gene that interacts with *TBX1*.[68] Yet another, *de novo DNAJC5B* duplication, also observed three additional times as an inherited duplication in this cohort, has no evident relationship with the patients' phenotype. Female patient SKZ_416 has EA/TEF, hydronephrosis, large ears, palpebral fissures slant down, thin fingers and long phalanges. The patients with an inherited deletion had EA/TEF and cranial malformations (SKZ_641) or isolated EA/TEF (SKZ_1510 and 1560). In general, DNAJ proteins are involved in protein folding. There is no information about the specific role of *DNAJC5B* in embryonic development.

TE and de novo CNV described in literature

Not many de novo CNV, either micro deletions or micro duplications, are described in patients with nonsyndromic EA/TEF. Van Binsbergen et al describes a patient with brain malformations, micrognathia, esophageal atresia, laryngeal stenosis and intrauterine growth retardation [36]This patient has an interstitial deletion in chromosomal bands 2q33q35 and propose that the microtubule associated protein 2 (*MAP2*) as a candidate gene for the brain malformations. Other de novo micro deletions described in literature are a 20q13.33 micro deletion in a boy with EA/TEF, cardiac and genital anomalies[43], a 11q13.1q13.2 micro deletion in a boy with EA/TEF, developmental delay and minor facial dysmorphisms [41] and a deletion of the long arm of chromosome 7q, including SHH, in a patient with multiple congenital anomalies including esophageal stenosis.[39]

LIM domain containing preferred translocation partner in lipoma (*LPP*) is deleted in an patient with esophageal atresia with tracheoesophageal fistula, hypospadia, cardiac-, renal- and rib anomalies.[37] Screening our cohort for abnormalities in this gene did not result in additional cases.[69]

Micro duplications have also been described. Recently Smigiel et al describe a 17q12 duplication in a boy with EA/TEF, anal atresia, sacral bone defect and cryptorchidism. Hilger and coworkers describe two micro duplications; one at 1q41 in a male with vertebral, anorectal, cardiac, trachea-esophageal and renal malformations and the other at 8q24.3 in a boy with vertebral, anorectal, trachea-esophageal, renal and genitourinary malformations.[35]

Concluding remarks

The micro-arrays used in these experiments increased in resolution over the years. Not all samples are analysed at the same resolution. We can conclude that relatively large ≥ 30- 50kb *de novo* CNV do not play a major role in EA/TEF disease etiology. New platforms with higher resolutions, such as the Affymetrix HD GeneChip have resolutions up to 2kb. Whole genome sequencing will, eventually, bring down the resolution to the base pair level. Re-evaluating our cohort with these techniques could increase the level of detected *de novo* variation. Large cohorts of unaffected individuals will need to be screened with these techniques in order to have a clear picture about the pathogenic nature of small CNV. Also a better understanding of the impact of (small) CNV on regulatory elements, insulators and other untranscribed DNA segments is necessary. As previously shown with the lncRNA deletions in patients with EA/TEF alveolar capillary dysplasia[22] these CNV can have a severe impact. A better understanding of where and how much variation is tolerated in our genomes (without an immediate pathogenic impact on developmental) is necessary because else an increase in resolution to detect CNV may lead to data interpretation problems, such high false positive rates and the discovery of smaller CNVs of unknown clinical significance.[70] Perhaps we should even first focus on those parts of healthy genomes that are *never* affected by CNV.[71]

Although we screened a large number of patients of our cohorts, the number represented approximately half. Isolated EA, although present in half of patients, was severly underrepresented in this screening. Also, as time progressed (we have collected data since 1988) more and more genetic syndromes were discovered. Screening our and other "historical" cohorts may be necessary to determine the exact impact of genetic factors. If such an endeavor is undertaken, it may be wise to use (the same) high density SNP-arrays or perform whole genome sequencing. Using genotyping information, genome wide association studies (GWAS) and homozygosity mapping in outbred individuals has proven to be successful in other congenital anomalies.[72-74] GWAS and homozygosity mapping require large sample sizes to obtain sufficient statistical power especially in heterogeneous conditions as VACTERL association.

We hypothesized that *de novo* genomic and inherited rare recurrent CNV could cause or predispose to EA/TEF. The data presented in this study provides limited evidence for a moderate impact of *de novo* CNVs in EA/TEF and VACTERL disease etiology, ~5%. Although we found several *de novo* and rare recurring changes, they do not point to one specific pathway or biological process. In addition comparison of the recorded overlapping CNV data with known EA/TEF loci from animal models or recorded chromosomal anomalies yielded no new candidates. Although predisposing loci are present (e.g. *SHOX*, 22q11dup) and probably influence or modify patient phenotype in sporadic occasions, their individual impact on patient population phenotype is moderate.

To conclude, there is a contribution of genetic factors to EA/TEF and the VACTERL association etiology. With aneuploidy and structural chromosomal anomalies (~4%), single base pair mutations (~6 %) now Copy Number Variation (~5 %) totals the genetic contribution to about ~15%. Whole genome and whole Exome DNA sequencing large patient cohorts will reveal new DNA variations increasing the contribution of genetics and our knowledge of EA/TEF and VACTERL disease etiology.

Acknowledgments

The authors thank David Alexander for editorial support. Erwin Brosens is funded by the Sophia Foundations for Scientific Research, projects SSWO- 493 and SWOO13-09

References

1. Felix, J.F., et al., *Environmental factors in the etiology of esophageal atresia and congenital diaphragmatic hernia: results of a case-control study.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 98-105.
2. Genevieve, D., et al., *Genetic factors in isolated and syndromic esophageal atresia.* J Pediatr Gastroenterol Nutr, 2011. ***52* Suppl 1**: p. S6-8.
3. Pedersen, R.N., et al., *Oesophageal atresia: prevalence, prenatal diagnosis and associated anomalies in 23 European regions.* Arch Dis Child, 2012. **97**(3): p. 227-32.
4. Brosens, E., et al., *Locus and phenotype heterogeneity in patients with trachea-esophageal malformations and associated anomalies.* Journal of Medical Genetics, 2014.
5. Felix, J.F., D. Tibboel, and A. de Klein, *Chromosomal anomalies in the aetiology of oesophageal atresia and tracheo-oesophageal fistula.* Eur J Med Genet, 2007. **50**(3): p. 163-75.
6. Shaw-Smith, C., *Genetic factors in esophageal atresia, tracheo-esophageal fistula and the VACTERL association: roles for FOXF1 and the 16q24.1 FOX transcription factor gene cluster, and review of the literature.* Eur J Med Genet, 2010. ***53***(1): p. 6-13.
7. de Jong, E.M., et al., *Non-VACTERL-type anomalies are frequent in patients with esophageal atresia/tracheo-esophageal fistula and full or partial VACTERL association.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 92-7.
8. Genevieve, D., et al., *An overview of isolated and syndromic oesophageal atresia.* Clin Genet, 2007. **71**(5): p. 392-9.
9. Solomon, B.D., *VACTERL/VATER Association.* Orphanet J Rare Dis, 2011. **6**: p. 56.
10. Robert, E., et al., *An international collaborative study of the epidemiology of esophageal atresia or stenosis.* Reprod Toxicol, 1993. **7**(5): p. 405-21.
11. Veenma, D., et al., *Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts.* Eur J Hum Genet, 2012. **20**(3): p. 298-304.
12. Brosens, E., et al., *Whole-exome mutation and CNV analysis in familial esophageal atresia* manuscript in preparation, 2014.
13. Reutter, H. and M. Ludwig, *VATER/VACTERL Association: Evidence for the Role of Genetic Factors.* Mol Syndromol, 2013. **4**(1-2): p. 16-9.
14. Bednarczyk, D., M.M. Sasiadek, and R. Smigiel, *Chromosome Aberrations And Gene Mutations In Patients With Esophageal Atresia.* J Pediatr Gastroenterol Nutr, 2013.
15. Lurie, I.W., *Structural autosomal imbalance and oesophageal defects: addendum to the article by Felix et al. (2007).* Eur J Med Genet, 2007. **50**(4): p. 322-5.
16. Stankiewicz, P. and J.R. Lupski, *Structural variation in the human genome and its role in disease.* Annu Rev Med, 2010. **61**: p. 437-55.
17. Zhang, F., et al., *Copy number variation in human health, disease, and evolution.* Annu Rev Genomics Hum Genet, 2009. **10**: p. 451-81.
18. Southard, A.E., L.J. Edelmann, and B.D. Gelb, *Role of copy number variants in structural birth defects.* Pediatrics, 2012. **129**(4): p. 755-63.
19. Cognet, M., et al., *Dissection of the MYCN locus in Feingold syndrome and isolated oesophageal atresia.* Eur J Hum Genet, 2011. **19**(5): p. 602-6.
20. Umana, L.A., et al., *A male newborn with VACTERL association and Fanconi anemia with a FANCB deletion detected by array comparative genomic hybridization (aCGH).* Am J Med Genet A, 2011. **155a**(12): p. 3071-4.
21. Corsten-Janssen, N., et al., *More Clinical Overlap between 22q11.2 Deletion Syndrome and CHARGE Syndrome than Often Anticipated.* Mol Syndromol, 2013. **4**(5): p. 235-45.
22. Szafranski, P., et al., *Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder.* Genome Res, 2013. **23**(1): p. 23-33.
23. Gordon, C.T., et al., *EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia.* J Med Genet, 2012. **49**(12): p. 737-46.
24. Brosens, E., et al., *VACTERL Association Etiology: The Impact of de novo and Rare Copy Number Variations.* Mol Syndromol, 2013. **4**(1-2): p. 20-6.
25. Hilger, A., et al., *Familial occurrence of the VATER/VACTERL association.* Pediatr Surg Int, 2012. **28**(7): p. 725-9.
26. Solomon, B.D., et al., *Evidence for inheritance in patients with VACTERL association.* Hum Genet, 2010. **127**(6): p. 731-3.
27. Solomon, B.D., et al., *Clinical geneticists' views of VACTERL/VATER association.* Am J Med Genet A, 2012. **158a**(12): p. 3087-100.
28. Van Opstal, D., et al., *Rapid aneuploidy detection with multiplex ligation-dependent probe amplification: a prospective study of 4000 amniotic fluid samples.* Eur J Hum Genet, 2009. **17**(1): p. 112-21.

29. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.
30. Shaikh, T.H., et al., *High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications.* Genome Res, 2009. **19**(9): p. 1682-90.
31. Kohany, O., et al., *Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor.* BMC Bioinformatics, 2006. **7**: p. 474.
32. Eussen, B.H., et al., *A familial inverted duplication 2q33-q34 identified and delineated by multiple cytogenetic techniques.* Eur J Med Genet, 2007. **50**(2): p. 112-9.
33. Cooper, G.M., et al., *A copy number variation morbidity map of developmental delay.* Nat Genet, 2011. **43**(9): p. 838-46.
34. Brosens, E., et al., *Structural and numerical changes of chromosome X in patients with esophageal atresia.* Eur J Hum Genet, 2014.
35. Hilger, A., et al., *De novo microduplications at 1q41, 2q37.3, and 8q24.3 in patients with VATER/VACTERL association.* Eur J Hum Genet, 2013. **21**(12): p. 1377-82.
36. van Binsbergen, E., et al., *A fetus with de novo 2q33.2q35 deletion including MAP2 with brain anomalies, esophageal atresia, and laryngeal stenosis.* Am J Med Genet A, 2014. **164a**(1): p. 194-8.
37. Arrington, C.B., et al., *Haploinsufficiency of the LIM domain containing preferred translocation partner in lipoma (LPP) gene in patients with tetralogy of Fallot and VACTERL association.* Am J Med Genet A, 2010. **152a**(11): p. 2919-23.
38. de Jong, E.M., et al., *5q11.2 deletion in a patient with tracheal agenesis.* Eur J Hum Genet, 2010. **18**(11): p. 1265-8.
39. Zen, P.R., et al., *Esophageal stenosis in a child presenting a de novo 7q terminal deletion.* Eur J Med Genet, 2010. **53**(5): p. 333-6.
40. Speleman, F., et al., *Detection of subtle reciprocal translocations by fluorescence in situ hybridization.* Clin Genet, 1992. **41**(4): p. 169-74.
41. Ferreira de Almeida, T. and D.R. Bertola, *Microdeletion 11q13.1.q13.2 in a patient presenting with developmental delay, facial dysmorphism, and esophageal atresia: possible role of the GSTP1 gene in esophagus malformation.* Birth Defects Res A Clin Mol Teratol, 2013. **97**(7): p. 463-6.
42. Smigiel, R., et al., *Oesophageal atresia with tracheoesophageal fistula and anal atresia in a patient with a de novo microduplication in 17q12.* Eur J Med Genet, 2014. **57**(1): p. 40-3.
43. Solomon, B.D., et al., *De novo deletion of chromosome 20q13.33 in a patient with tracheo-esophageal fistula, cardiac defects and genitourinary anomalies implicates GTPBP5 as a candidate gene.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(9): p. 862-5.
44. Itsara, A., et al., *De novo rates and selection of large copy number variation.* Genome Res, 2010. **20**(11): p. 1469-81.
45. Itsara, A., et al., *Population analysis of large copy number variants and hotspots of human genetic disease.* Am J Hum Genet, 2009. **84**(2): p. 148-61.
46. Martin, N.J. and B.G. Steinberg, *The dup(3)(p25 leads to pter) syndrome: a case with holoprosencephaly.* Am J Med Genet, 1983. **14**(4): p. 767-72.
47. Jackson, J., et al., *How accurate does rapid fetal karyotyping need to be? Case of unbalanced t(13;18).* Prenat Diagn, 1993. **13**(8): p. 767-70.
48. Ikeda, M., et al., *B-box and SPRY domain containing protein (BSPRY) is associated with the maintenance of mouse embryonic stem cell pluripotency and early embryonic development.* J Reprod Dev, 2012. **58**(6): p. 691-9.
49. Miller, D.T., et al., *Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies.* Am J Hum Genet, 2010. **86**(5): p. 749-64.
50. Cetinkaya, M., et al., *Oesophageal atresia in a neonate with a familial translocation t(1;13)(p8;q12).* Genet Couns, 2012. **23**(1): p. 73-5.
51. Oei, L., et al., *A genome-wide copy number association study of osteoporotic fractures points to the 6p25.1 locus.* J Med Genet, 2014. **51**(2): p. 122-31.
52. Klopocki, E., et al., *Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome.* Am J Hum Genet, 2007. **80**(2): p. 232-40.
53. Girirajan, S., et al., *A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay.* Nat Genet, 2010. **42**(3): p. 203-9.
54. Schramm, C., et al., *De novo microduplication at 22q11.21 in a patient with VACTERL association.* Eur J Med Genet, 2011. **54**(1): p. 9-13.
55. Prieto, J.C., et al., *Phenotypic expansion of the supernumerary derivative (22) chromosome syndrome: VACTERL and Hirschsprung's disease.* J Pediatr Surg, 2007. **42**(11): p. 1928-32.
56. Shaikh, T.H., et al., *Low copy repeats mediate distal chromosome 22q11.2 deletions: sequence analysis predicts breakpoint mechanisms.* Genome Res, 2007. **17**(4): p. 482-91.

57. Digilio, M.C., et al., *Three patients with oculo-auriculo-vertebral spectrum and microdeletion 22q11.2.* Am J Med Genet A, 2009. **149A**(12): p. 2860-4.
58. Ensenauer, R.E., et al., *Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients.* Am J Hum Genet, 2003. **73**(5): p. 1027-40.
59. Faguer, S., et al., *A 17q12 chromosomal duplication associated with renal disease and esophageal atresia.* Eur J Med Genet, 2011. **54**(4): p. e437-40.
60. Knudson, A.G., *Two genetic hits (more or less) to cancer.* Nat Rev Cancer, 2001. **1**(2): p. 157-62.
61. Chan, Y., et al., *An excess of risk-increasing low-frequency variants can be a signal of polygenic inheritance in complex diseases.* Am J Hum Genet, 2014. **94**(3): p. 437-52.
62. Beckmann, J.S., X. Estivill, and S.E. Antonarakis, *Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability.* Nat Rev Genet, 2007. **8**(8): p. 639-46.
63. Wat, M.J., et al., *Genomic alterations that contribute to the development of isolated and non-isolated congenital diaphragmatic hernia.* J Med Genet, 2011. **48**(5): p. 299-307.
64. Hehir-Kwa, J.Y., et al., *De novo copy number variants associated with intellectual disability have a paternal origin and age bias.* J Med Genet, 2011. **48**(11): p. 776-8.
65. Battaglia, A., T. Filippi, and J.C. Carey, *Update on the clinical features and natural history of Wolf-Hirschhorn (4p-) syndrome: experience with 87 patients and recommendations for routine health supervision.* Am J Med Genet C Semin Med Genet, 2008. **148c**(4): p. 246-51.
66. Harada, M., et al., *FGF9 monomer-dimer equilibrium regulates extracellular matrix affinity and tissue diffusion.* Nat Genet, 2009. **41**(3): p. 289-98.
67. Hock, A.K., et al., *Regulation of p53 stability and function by the deubiquitinating enzyme USP42.* Embo j, 2011. **30**(24): p. 4921-30.
68. Harel, I., et al., *Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis.* Proc Natl Acad Sci U S A, 2012. **109**(46): p. 18839-44.
69. Hernandez-Garcia, A., et al., *Contribution of LPP copy number and sequence changes to esophageal atresia, tracheoesophageal fistula, and VACTERL association.* Am J Med Genet A, 2012. **158a**(7): p. 1785-7.
70. Bernardini, L., et al., *High-resolution SNP arrays in mental retardation diagnostics: how much do we gain?* Eur J Hum Genet, 2010. **18**(2): p. 178-85.
71. Macosko, E.Z. and S.A. McCarroll, *Genetics. Our fallen genomes.* Science, 2013. **342**(6158): p. 564-5.
72. Birnbaum, S., et al., *Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24.* Nat Genet, 2009. **41**(4): p. 473-7.
73. Mangold, E., et al., *Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate.* Nat Genet, 2010. **42**(1): p. 24-6.
74. Hildebrandt, F., et al., *A systematic approach to mapping recessive disease genes in individuals from outbred populations.* PLoS Genet, 2009. **5**(1): p. e1000353.

Casenr	Gender	Major anomalies
148-04-01	NA	NA
148-09-01	NA	NA
148-10-01	NA	NA
SKZ_0060	M	EA/TEF, upper limb anomalies, dysmorphisms
SKZ_0282	M	EA/TEF, renal anomalies
SKZ_0374	M	EA/TEF, cardiac anomalies, renal anomalies, upper limb anomalies, genital anomalies
SKZ_0399	M	EA/TEF, anal anomalies, upper limb anomalies, genital anomalies, dysmorphisms
SKZ_0416	F	EA/TEF, renal anomalies, dysmorphisms
SKZ_0641	M	EA/TEF, cranial abnormalities, deafness
SKZ_0671	F	EA/TEF, vertebral anomalies, cranial anomalies, dysmorphisms
SKZ_0708	F	EA/TEF, vertebral anomalies, renal anomalies, ureteral atresia, genital anomalies
SKZ_0744	F	EA/TEF, vertebral anomalies
SKZ_0773	F	EA/TEF, vertebral anomalies, anal , genital and ear anomalies, renal anomalies, upper limb anomalies, cleft lip+jaw+palate, duodenal atresia,
SKZ_0796		EA/TEF, pyloric stenosis
SKZ_0832	M	EA/TEF, vertebral anomalies, anal anomalies, renal anomalies, genital anomalies, lower limb anomalies
SKZ_0856	M	EA/TEF, anal anomalies, cardiac anomalies,vertebral anomalies, dysmorphisms
SKZ_1032	M	EA/TEF, cardiac anomalies
SKZ_1035	F	EA/TEF, cardiac anomalies, brain anomalies, cranial anomalies, dysmorphisms
SKZ_1150	F	EA/TEF + cardiac anomalies
SKZ_1307	F	EA/TEF, anal anomalies, upper limb anomalies
SKZ_1415	M	EA/TEF, cardiac anomalies, upper limb anomalies, duodenal atresia, dysmorphisms, spleen anomalies, pancreas andgenital anomalies
SKZ_1432	M	EA/TEF, anal anomalies, renal anomalies, urethral fistula and atresia, genital anomalies
SKZ_1497	M	EA/TEF, cardiac anomalies
SKZ_1508	M	EA/TEF, upper limb anomalies, cardiac anomalies, dysmorphisms
SKZ_1516	F	EA/TEF, cardiac anomalies
SKZ_1550	M	EA/TEF, anal anomalies, renal anomalies, urethral anomalies
SKZ_1730	M	EA/TEF, anal anomalies, cardiac anomalies, upper limb anomalies, cranial anomalies, genital anomalies
SKZ_1780	M	EA/TEF, vertebral anomalies, anal anomalies, cardiac anomalies, renal anomalies, upper+lower limb anomalies
SKZ_1790	F	EA/TEF, cardiac anomalies, anal atresia,
SKZ_1810	M	EA/TEF, vertebral anomalies, cardiac anomalies, renal anomalies, upper limb anomalies, brain anomalies
SKZ_1821	M	EA/TEF, cardiac anomalies, genital anomalies, ear anomalies
SKZ_1825	M	EA/TEF, cranial anomalies
SKZ_1900	M	TEF, lung hypoplasia/agenesis
SKZ_1910	M	EA/TEF, cardiac anomalies
SKZ_1988	F	EA/TEF, anal anomalies
SKZ_2010	M	EA, ACD
SKZ_2035	F	EA/TEF, anal anomalies, renal anomalies
SKZ_2041	M	EA/TEF, cardiac anomalies, renal anomalies
SKZ_2064	M	TEF, cardiac anomalies

Supplementary table 2 Major affected organ systems in patients with recurring CNV. Not depicted patients have isolated EA and/ or TEF

2.3

Structural and numerical changes of chromosome X in patients with esophageal atresia

Adapted from:

Structural and numerical changes of chromosome X in patients with esophageal atresia.

Erwin Brosens, Elisabeth M. de Jong, Tahsin Stefan Barakat, Bert H Eussen, Barbara D'haene[4], Elfride de Baere, Hannah Verdin, P, Pino P Poddighe, Robert-Jan Galjaard, Joost Gribnau, Alice S Brooks, Dick Tibboel, and Annelies de Klein. *Eur J Hum Genet. 2014 Jan 8; Epub ahead of print*

Abstract

Esophageal atresia with or without tracheo-esophageal fistula (EA/TEF) is a relatively common birth defect often associated with additional congenital anomalies as vertebral, anal, cardiovascular, renal and limb defects: the so-called VACTERL association. Little is known about causal genetic factors. Rare case- reports of gastro-intestinal anomalies in children with triple X syndrome prompted us to survey the incidence of structural and numerical changes of chromosome X in patients with EA/TEF.

All available (n= 269) karyotypes of our large (321) EA/TEF patient cohort were evaluated for X-chromosome anomalies. If sufficient DNA material was available, we determined genome-wide copy number profiles with SNP-array and identified subtelomeric aberrations on the difficult to profile PAR1 region using telomere-MLPA. Additionally we investigated X-chromosome inactivation (XCI) patterns and mode of inheritance of detected aberrations in selected patients.

Three EA/TEF patients had an additional, maternally inherited X chromosome. These three females had normal random XCI patterns. Two male EA/TEF patients had small inherited duplications of the XY-linked SHOX locus. Patients were small for gestational age at birth (<P5) and had additional, mostly VACTERL associated, anomalies.

Triple X syndrome is rarely described in patients with EA/TEF and no duplications of the SHOX gene were reported so far in these patients. Since normal patterns of XCI were seen, over-expression of X-linked genes that escape XCI, such as the *SHOX* gene, could be pathogenic by disturbing developmental pathways.

Introduction

Esophageal atresia (EA) with or without tracheo-esophageal fistula (TEF) is a relatively common birth defect affecting approximately 1:3500 newborns. These newborns can have a heterogeneous phenotype, some have EA and/or TEF as an isolated defect and others have more anomalies, predominantly VACTERL (Vertebral, Anal, Cardiovascular, Tracheo-Esophageal, Renal and Limb) associated. [1]

EA/TEF is a variable feature in several genetic syndromes e.g. Feingold (*MYCN*), CHARGE (*CHD7*), Anophthalmia-Esophageal-Genital (AEG) syndrome (*SOX2*) and Fanconi anemia. [2] In addition the genetic defects in these syndromes, other putative causal genetic aberrations are described in EA/TEF patients. Although there are chromosomal hotspots, these aberrations are mostly scattered across the genome.[2]

Structural and numerical chromosome abnormalities affecting sex chromosomes have been described in patients with congenital malformations. [3-5] These defects are rare in patients with EA/TEF, although EA/TEF is a variable feature in patients with Opitz G syndrome (*MID1)* and VACTERL association with hydrocephalus (*FANCB)*[1, 2] There are several reports describing X-chromosome duplication in association with gastro-intestinal anomalies.[6] This prompted us to retrospectively evaluate the cytogenetic results in our EA/TEF cohort.

We identified three patients with a triple X karyotype, strengthening the relationship of gastrointestinal anomalies and X-chromosome triplication. In addition to classical karyotyping, we examined patient DNA with telomere-multiplex ligation-dependent probe amplification (MLPA) and micro arrays for structural X-chromosome abnormalities. These molecular-genetic studies revealed Short stature HOmeoboX-containing gene (*SHOX*) duplications in two additional patients with esophageal atresia, tracheo-esophageal fistula and limb anomalies.

We hypothesize that genes on X and/or genes that escape X chromosome inactivation could influence essential developmental pathways in limb and foregut development.

Patients and Methods

Patient population

This study was approved by the Medical Ethical Review Board of Erasmus MC - Sophia Children's Hospital. After retrieving (parental) informed consent, 321 patients with EA/TEF, admitted to the department of Paediatric Surgery were included. Pregnancy, clinical and follow-up data were extracted from medical charts. Available DNA and cell lines of patients (n=180) and parents were collected and used for genetic analysis. Patients with a previous confirmed genetic syndrome, known chromosomal anomaly and/or pathogenic point mutation were excluded from further molecular-genetic evaluation. There is weak evidence for an association of EA/TEF with certain environmental components [7], however we did not exclude any of the patients in our cohort based on these risk factors.

The database (>100.000 patients) of prenatal and postnatal diagnostics of our department of Clinical Genetics was searched for triple X karyotypes and confirmed *SHOX* duplications.

Cytogenetic evaluation

Karyotyping was performed according to standard protocols on either lymphocytes from peripheral blood cultures or after amniocentesis. Karyotyping had been performed for 269 of the 321 patients of our cohort, as systematic cytogenetic follow-up of patients with congenital anomalies was not done before 1998.

Multiplex ligation-dependent probe amplification (MLPA) and quantitative PCR

Not all array chips used had sufficient marker density in the PAR1 region. Therefore, we additionally screened for copy number variations in this region with MLPA, using the P036E1 and P070A2 Salsa telomere kit (MRC Holland, Amsterdam, the Netherlands) as previously described.[8] Genemarker 1.6 (SoftGenetics, LLC, State College, PA, USA) was used for data analysis. If duplications of the PAR1 region were detected, copy number profiling of the *SHOX* region was confirmed with a newly developed qPCR assay, with 11 amplicons within *SHOX* and the PAR1 region.[9]

Fluorescent in situ hybridization (FISH)

The target BAC clone for Xp22 (RP11-800K15) and control probes on Xq25 (RP11-49N19) and 8p12 (RP11-489E7) were selected from the University of Santa Cruz (UCSC) genome browser (UC Santa Cruz, CA, USA, assembly march 2006) and ordered from BACPAC Resources. (Oakland, CA, USA) After isolation of the BAC DNA, the probes were labelled and used for FISH on chromosome preparations from patients and parents, according to standard protocols.[10]

RNA-FISH analysis and immunocytochemistry of human cell lines

RNA-FISH and immunocytochemistry were performed on fibroblast cell lines (>90% confluence) and EBV-transformed lymphocytes, as previously described by Jonkers et al [11, 12] using a 16.4 kb plasmid covering the complete XIST RNA sequence as described previously. [13, 14]

HUMARA analysis

To determine the parental origin and the methylation status of the additional X-chromosome, we used the HUMARA assay (Human Methylation of the Androgen Receptor Assay) using 40ng of genomic DNA input for the digestion reaction and gel electrophoresis to separate PCR products.[15]

Microarrays

The genome wide copy number profile of all patients (n=180) was determined using either Affymetrix GeneChip Human Mapping 250K NSP1 (Affymetrix, Santa Clara, CA), HumanQ610, HumanCytoSNP-12v1 to 2.1 or HumanOmniExpress (Illumina, San Diego, CA). We generated Affymetrix CEL files with the Affymetrix genotype command console v3.2 software. The HumanCytoSNP-12v2.1 chip (Illumina®, San Diego, CA, USA) was used in the cases with *SHOX* duplications and their parents, for better coverage of the PAR1 region. All procedures were done according to the manufacturer's protocol as previously described.[16]

SALL1 mutation analysis

"Sequencing of the coding region of the *SALL1* gene, including the splice sites, was done as described previously. [17] Primer sequences are available on request. We did not have sufficient DNA of Triple X patient 3 to perform additional Sanger sequencing"

Statistical analysis

Differences in two proportions were tested with the Pearson's chi-square (χ^2) test, reported with a 95% confidence interval (CI), performed in SPSS 15.0.

Results

Patient characteristics

Patients included in the Erasmus MC - Sophia Children's Hospital EA/TEF cohort can be subdivided into four categories: isolated EA and/or TEF (45%) , patients with one additional core VACTERL component (27%), VACTERL association e.g. three or more of the VACTERL core components(21%) and patients with a diagnosed genetic syndrome (7%) . [18, 19]

Triple X syndrome

A triple X karyotype (see supplementary figure 1) was identified in three EA/TEF patients resulting in an odds ratio for triple X syndrome of 11.3, (95% CI= 3.6 to 35.2). All three were small for gestational age at birth (<5th percentile) with maternal ages ranging from almost 26 to 28. The first patient had all VACTERL anomalies and additional genito-urinary anomalies. The second patient had TEF and mild dysmorphic features and the third patient had EA/TEF, ventricular septal defect and thin fingers (patients 1-3, Table 1). SNP array analysis performed to exclude copy number variations elsewhere in the genome, did not reveal additional possible pathogenic copy number variations in any of the three triple X patients.

Searching the database of our department of Clinical Genetics, yielded 59 non-mosaic 47,XXX karyotypes since 1988; 29 had been detected prenatally and 30 postnatal (apart from the above three patients). Indications (not mutually exclusive) for prenatal karyotyping were: maternal age >35 years (n=25), congenital malformations on ultrasound (n=4), increased risk for Down syndrome on first trimester screening ultrasound (n=2), increased echogenicity of the foetal bowel on ultrasound (n=1), and congenital anomalies in an earlier pregnancy (n=3). One screening, with an increased maternal age indication, concerned a twin pregnancy of which one sib had a 47,XXX karyotype. Postnatal patients were karyotyped based on the following indications: suspicion of Fragile X syndrome (n=5), mental retardation (n=5), multiple congenital anomalies (n=4), combined mental retardation and multiple congenital anomalies (n=3), repeating spontaneous abortions in the index (n=4),

failure to thrive (n=4), a chromosomal abnormality in the family (n=3), suspicion of trisomy 21 (n=1), and a possible chromosomal aberration in juvenile systemic lupus erythematosus (n=1).

Postnatal follow-up of thirteen pregnancies was not documented. Seven pregnancies were terminated, in one case because the foetus showed anencephaly. Congenital malformations had been documented for three of the nine births; a neural tube defect, hygroma colli with generalized edema, and osteogenesis imperfecta, respectively. A review of the medical charts of the postnatal diagnosed patients identified at least four patients with congenital heart defects.

X inactivation studies

Previous studies have indicated that only one X chromosome remains active in somatic cells of 47,XXX patients. RNA-FISH and immunocytochemistry were performed to assess the X inactivation status in cell lines derived from the 47,XXX patients of our cohort. RNA-FISH analysis revealed two *XIST* clouds in almost every fibroblast and lymphocyte cell (>95% of the nuclei), co-localizing with an area of low level COT-1 expression (figure 2A). Immunocytochemistry of 47,XXX fibroblasts detected enrichment of the facultative heterochromatin markers H3K27me3 and MacroH2A1 on two X-chromosomes, co-localizing with the DAPI-dense Barr-bodies (Figure 2B). HUMARA analysis confirmed a maternal origin of the additional X-chromosome in all three patients. In all these patients and their mothers random XCI was observed, with no skewed preference of inactivation of a particular X-chromosome (supplementary figure 3)

SHOX duplications

Enrichment of the triple X karyotype in our EA/TEF cohort prompted us to look further in our cohort for sex- chromosome aberrations with micro-array and telomere-MLPA. These results indicated inherited *SHOX* duplications in two boys: patient 4 with an paternal inherited partial PAR1 duplication and patient 5 with an partial PAR1 duplication inherited from his mother. Duplication of the *SHOX* gene is a rare event, it is present only 2 times in the database of genomic variation. Both *SHOX* duplications were confirmed with a *SHOX*-specific qPCR [9] (figure 1A, supplementary figure 2).

To exclude additional potential pathogenic copy number variations in other regions in the genome of both patients and their parents, we analysed their genome-wide copy number profile. These results revealed multiple copy number variations in both the patients and their parents. However, upon closer examination of these regions in the database of genomic variation, all of them were common polymorphisms. Given their relative high population frequency they were not considered as potential pathogenic copy number variations in a relatively rare condition as EA/TEF or VACTERL association.

Patient 4's twin sib was spontaneously aborted in the third month of gestation. His mother was diagnosed with Goldenhar syndrome. She has right-sided hemifacial microsomia, anotia, deafness, paresis of the pallatum molle and facial dysmorphisms. Two relatives on the mother's side have thumb anomalies. EA/TEF or other major anomalies were absent in the father. The boy was small for gestational age at birth (<2SD) and had associated cardiovascular, renal, limb and genital anomalies (Table 1).

SHOX-specific qPCR confirmed the duplication of the SHOX gene (amplicon 1-13) in the PAR1 region in patient 4 and his father. Cultured lymphocytes of proband or parents for FISH validation or to localize the duplication were not available. The 981Kb duplicated segment, corresponds to the PAR1 region on chromosome X and both SHOX variants, SHOXA g1-292dup and SHOXB g1-225dup are completely duplicated in patient 4, including all of its regulatory sequences.[20, 24, 25]. (arr [hg18] Xp22 or Yp11(248,968 _1,229,976)x3) (Figure 1A, 1B, supplementary fig. 4).

Patient 5, with apparently healthy parents, was also small for gestational age at birth and had cardiac and limb anomalies. He inherited a PAR1 duplication/suspected rearrangement from his mother overlapping the SHOX gene (arr [hg18] Xp22 (325941_593267)x3) and separated by a region with normal copy numbers a second Xp22 duplicated segment containing 7 other genes. (IL3RA, SLC25A6, ASMTL, PP1164, P2RY8, SFRS17A and ASMT) (arr [hg18] Xp22(1428051_1891174)x3, supplementary figure 5). SHOX-specific qPCR (duplication of amplicon 2-8) confirmed the SHOX gene duplication in patient 5 and his mother. Moreover, FISH (BAC clone RP11-800K15) confirmed the location of the duplication at chromosome band Xp22. The direction of the inserted duplicated segment was not determined

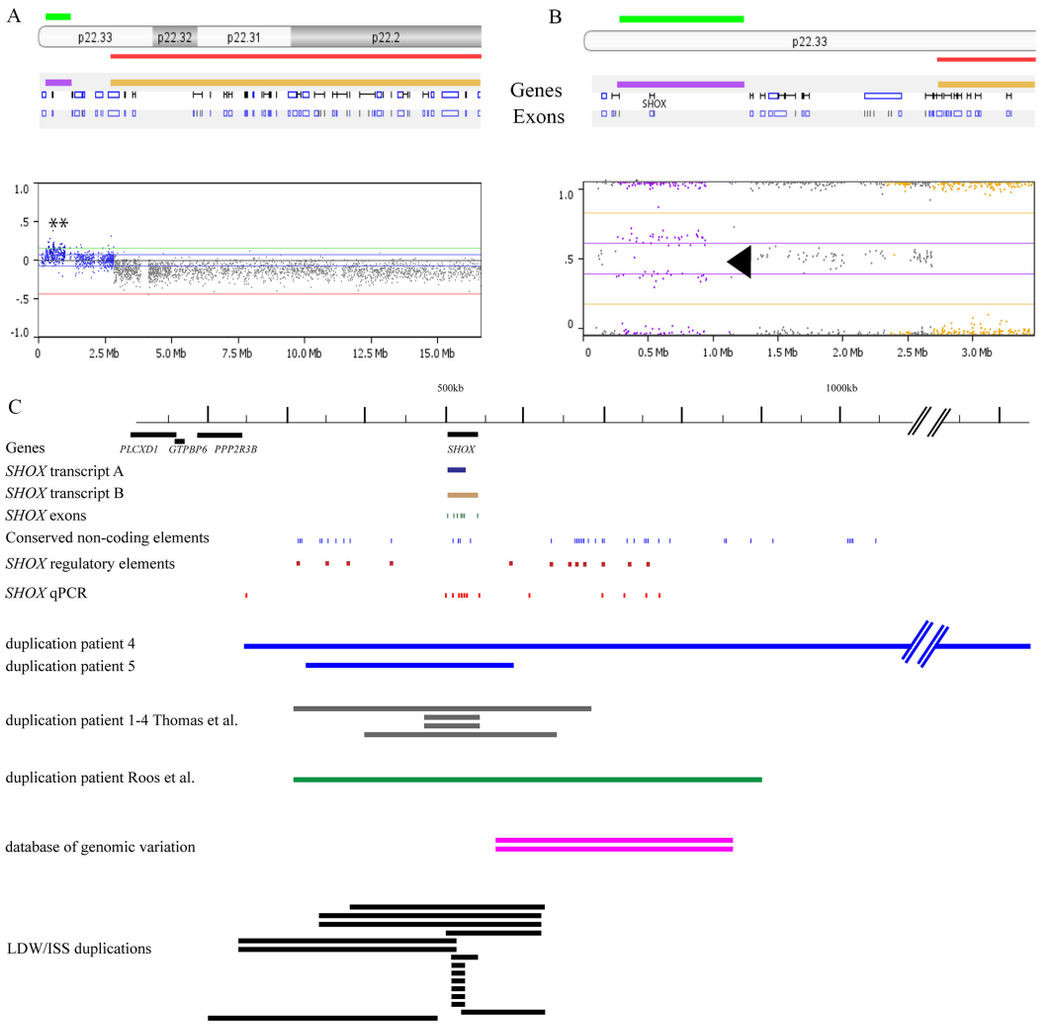

Figure 1 Triple X syndrome and SHOX aberrations in patients with OA/TOF. 1A. The HumanCytoSNP-12 chip (Illumina®, San Diego, CA, USA) showing an interstitial duplication of the PAR1 region (base pair position chrX:248,968-1,229,976), including the SHOX gene in patient 4. The elevated log2 ratio indicates the duplicated segment (green bar) 1B.The HumanCytoSNP-12 chip result visualized in Biodiscovery Nexus CN6.1. The B-allele frequency of patient 4 is indicated in the enlarged right panel (purple bar, arrow). The shift from a heterozygous state (gray dots, 0.5) to a 0.33/0.66 frequency (purple dots) is indicative for a copy number change. Together with the raise in the logR this state is indicative for a copy number gain. 1C.Schematic presentation of the SHOX A and B transcripts in dark blue and brown, the conserved non-coding elements (blue) [20], SHOX gene exons (green), SHOX qPCR (red) [9]and SHOX regulatory elements described in literature (dark-red), the SHOX duplications in patient 4 and 5 and their position compared the duplications observed in patients 1-4 in Thomas et al. [21], Roos et al. [22], the database of genomic variation and the duplication in LWD/ISS patients described by Benito-Sanz et al. [23]. The red lines indicate the length of the duplications, with a larger duplication in patient 4 (amplicon 1-13) and a SHOX duplication in patient 5 (amplicons 2-8). (supplementary fig. 2) Both MLPA probes (orange) were duplicated in patient 4 and 5.

Pt no	Karyotype	MLPA results	Maternal age	Clinical features
1	47,XXX	X/Yp SHOX enh	25.9	Absent sacrum Anal atresia Pulmonary stenosis EA + TEF Vesico-urethral reflux; urethral atresia Absent thumbs Cloacal malformation: abnormal labia, hydrometrocolpos
2	47,XXX	X/Yp SHOX enh	28.2	TEF Dysmorphic features
3	47,XXX	X/Yp SHOX enh	26.1	EA + TEF Ventricular septum defect Thin fingers
4	46,XY	X/Yp SHOX enh paternal inheritance	31.5	Aberrant subclavian artery EA + TEF Horseshoe kidneys Adducted thumbs; left thumb smaller than right Hypospadias Frontal bossing Dysmorphic features
5	46,XY	X/Yp SHOX enh maternal inheritance	34.3	Atrial septum defect (type II) EA + TEF Proximal placement of thumbs

Table 1 Congenital anomalies of EA/TEF patients with X-chromosome anomaly, EA; esophageal atresia; TEF: tracheo-esophageal fistula; enh, enhanced signal with MLPA kits P036E1 and P070A2

N	genital	urinary	gastrointestinal anomalies	V	A	C	TE	R	L	Other anomalies	Reference
3	+	+	cloacal extrophy incl. imperforate anus, esophageal atresia + TEF	+	+	+	+		+	dysmorphic features	present study
1			jejunal atresia								[26]
1			duodenal atresia								[27]
1			duodenal atresia *			+					[6]
1			omphalocele							Beckwith-Wiedemann syndrome	[28]
1			omphalocele					+			[29]
1	+	+						+		pulmonary hypoplasia, laryngeal atresia, craniofacial anomalies	[30]
1			ectopic anus		+				+	clinodactyly, inferior coloboma, clinodactyly, dysmorphic features	[31]
1	+	+	cloacal extrophy incl. imperforate anus		+						[32]
1	+	+	cloacal extrophy incl. imperforate anus and rectoperineal fistula, colonic atresia, omphalocele		+			+			[33]
1	+	+	imperforate anus, esophageal atresia + TEF	+	+		+	+		pulmonary hypoplasia, agenesis of gallbladder	[34]

Table 2 Congenital malformations of the gastro-intestinal tract and/or foregut-related structures in triple X syndrome N; number of patients, TEF; tracheo-oesophageal fistula, GI; gastro-intestinal malformations; V, vertebral defects; A; anorectal malformations; C, cardiovascular anomalies; TE, esophageal atresia and/or tracheo-esophageal fistula; R, renal anomalies; L, limb malformations; *, duodenal atresia due to annular pancreas

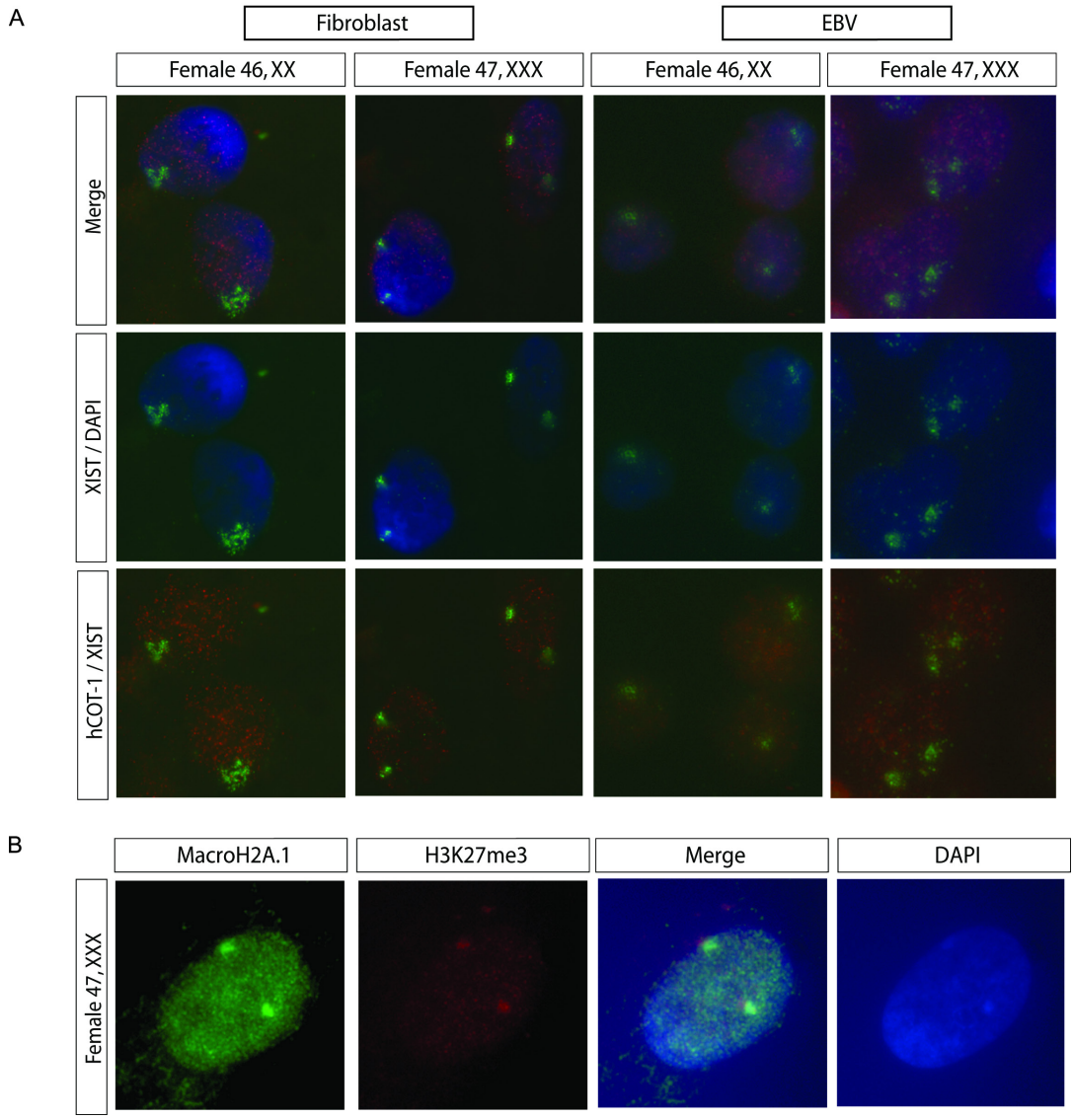

Figure 2. X-Chromosome inactivation studies in triple X patients. 2A. RNA-FISH analysis detecting *XIST* and COT-1 RNA. *XIST* RNA was detected using a digeoxin-labelled cDNA probe [12, 14] (FITC, green), and a biotin-labelled COT-1 DNA probe detected expression of repetitive sequences (Rhodamine, red). Characteristic for an inactive X-chromosome is the presence of *XIST* RNA accumulation and absence of COT-1 RNA. In 47, XXX patient cells two *XIST* clouds and COT-1 holes were detected in the majority of the cells (> 95% of the nuclei, n >100). Nuclei were counterstained with DAPI. 2B. Immunocytochemistry detecting H3K27me3 (Rhodamine, red) and MacroH2A1 (FITC, green). Enrichment of H3K27me3 and MacroH2A1 was found on two X-chromosomes in 47, XXX fibroblast cell lines, co-localizing with DAPI-dense Barr bodies, indicative of two inactive X-chromosomes. Nuclei are counterstained with DAPI.

SALL1 mutation analysis

Mutation analysis of the SALL1 coding region only identified common variants e.g. two missense variants, one in all patients and a control (rs4614723; minor allele frequency is 1.514%) and one in patient 4 (rs13336129; minor allele frequency is 6.651%). We also identified one intronic variant in patient 4 (rs13336129; minor allele frequency is 45.545%) and two synonymous variants in triple X patient 1 and a control (rs11645288; minor allele frequency is 17,816% and rs1965024; minor allele frequency is 49,216%)"

Discussion

Triple X syndrome and Gastro-intestinal anomalies

Congenital malformations and mental retardation syndromes have been linked to the X-chromosome. Our search for genes or loci involved in EA/TEF or other foregut-related anomalies identified five patients in our cohort with chromosome X/Yp aberrations; three with triple X syndrome and two with inherited *PAR1* duplications.

A large proportion of triple X females have a subclinical phenotype, therefore a triple X karyotype is usually a random finding in prenatal screening or cytogenetic follow-up of pregnancies. Most 47,XXX females remain undiagnosed, therefore triple X syndrome has an estimated incidence rate of 0.10% with an average maternal age of 33.[4] Affected girls have a lower birth weight, more accelerated growth until puberty, long legs and an increase in behavioural problems and psychiatric disorder prevalence.[4] In 90% of patients, the additional X-chromosome is the result of a maternal meiotic I error and the incidence of non-disjunction errors increases with maternal age.[35]

Sex chromosome triploidies (47,XXX/ 47,XXY/ 47,XYY) are a rare (0.42%) finding in a large (n=4282) prenatal cohort analysed with karyotyping and micro-array.[36] Haverty et al. calculate the incidence rate in females and to be 0.17%.[32] In a recent European study of the EUROCAT working group 0.054 triple X patients/1000 births are observed. [37] In our EA/TEF cohort the incidence rate of triple X syndrome is 1.12 % with an average maternal age of 30.8, 11 times higher than that in the estimated general population and 6.5 times higher than in the

calculated incidence rate by Haverty et. al. Guichet *et al.* reviewed prenatally and postnatally diagnosed 47,XXX karyotypes from 18 laboratories.[38]

Mental retardation or congenital malformations were described in over one third of the 190 patients reported. In all cases, weight-for-gestational-age at birth was under the 25th percentile. Congenital anomalies associated with the triple X syndrome described in case reports include anomalies of the urinary tract, genital anomalies and craniofacial anomalies, specifically a reduced head circumference and/or decreased brain volume.[4, 39, 40] Genito-urinary malformations are well-described in triple X patients. Since genito-urinary malformations often are associated with lower gastro-intestinal tractanomalies, perhaps a cloacal septation problem gives rise to the higher incidence of these types of malformations. Gastro-intestinal anomalies, including atresia of the esophagus, duodenum and jejunum, as well as omphalocele and anorectal malformations, have been reported sporadically.[6, 26-29, 31-34] All thirteen reported patients with gastro-intestinal and/or foregut-related anomalies are reviewed in Table 2. The patient described by Hoang *et al.* shows a similar phenotype compared to our patients; higher mesodermal defects (esophageal atresia) and lower mesodermal defects (anal atresia, genito-urinary defects).[34]

X inactivation patterns

The HUMARA assay demonstrated the maternal origin of the supernumerary X-chromosome and the absence of skewed preference for a particular X-chromosome in our triple X patients. RNA-FISH analysis and immunocytochemistry demonstrated inactivation of two out of three X-chromosomes

Overexpression of genes escaping X-inactivation could be responsible for the phenotypical abnormalities observed in our three EA/TEF patients and the gastro-intestinal patients described in literature. Ten percent of genes have a variable pattern of gene X-inactivation and expression.[41] The extent of this escape is tissue specific, and often results in variable or lower levels of expression from the inactive X chromosome compared to the active X chromosome.[42, 43] Why and how certain genes escape XCI, especially in humans, is still unknown.[44] Female 'escapees' may have a dosage-sensitive function, which would explain the phenotype in patients with sex chromosome anomalies, such as Turner and Klinefelter syndrome. The

observation that 47,XXX females also have decreased brain volume in the presence of normal pubertal maturation, suggests a possible direct dosage effect of X-chromosomal genes.[45] Two esophageal atresia candidate genes, *MID1 and FANCB* do not escape X-inactivation, although there is a X-inactivation preference for the X-chromosome that contains the mutated allele in Fanconi anemia.[36] Other genes that escape XCI could perhaps cause the gastro-intestinal anomalies found in our cohort.

SHOX duplications

One of those escapees, located in the PAR1 region, is *SHOX*. This gene has 2 isoforms; *SHOXA* and *SHOXB* which are surrounded by several conserved non coding regulators. [20] *SHOX* encodes a cell-specific homeodomain protein and isoform a plays an important role during human embryonic bone and limb development. [46] Two patients in our EA/TEF cohort have a duplication in the PAR1 region, the only overlapping duplicated gene is *SHOX*. The co-occurrence of *SHOX* duplication in a small cohort of a rare disease as esophageal atresia is intriguing, but we cannot exclude a chance finding. Large PAR1 duplications are relatively rare in the database of genomic variation (http://dgv.tcag.ca/), a database of "healthy" individuals and only 2 duplications of *SHOX* are described. In the ISCA consortium patient database (www.iscaconsortium.org) duplications of *SHOX* are more prevalent (92 in total) and generally classified as uncertain.

However, we observed a *SHOX* duplication in two EA/TEF patients and in both these patients limb development is disturbed. The limb anomalies, growth anomalies of the thump, are different from the wrist deformity usually associated with *SHOX* deletions and duplications. [23, 47] Furthermore, SHOX duplications are also associated with limb anomalies e.g. in Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome type 1[48], Idiopathic Short Stature [49] or Leri-Weill Dyschondrosteosis. [23]

The healthy carriers of the transmitted *SHOX* duplications in patient 4 and 5 seem to have the exact same duplication/rearrangement as their offspring. Absence of a phenotype in healthy parents could be caused by incomplete penetrance or variable expressivity as described for other microdeletion/ duplication syndromes.[50] Thomas *et al.* and Roos *et al.* describe five families with inherited *SHOX*

duplications.[21, 22] The duplication was associated with cleft palate in two cases and one patient had a Madelung deformity. These, and the *SHOX* duplications described by Benito-Sanz [21-23] are often inherited from an unaffected parent. However, it is important to know the exact location of the *SHOX* duplication since insertion of the duplicated segment could result in haploinsufficiency of *SHOX* by affecting the normal copy of the gene or its regulatory elements. It would certainly be beneficial to determine RNA and protein expression in e.g. esophageal or bone tissue, however no biopsies of affected and corresponding normal tissues are available at this moment

The mouse homologue of *SHOX*, Shox2 is involved in limb development [51] and in human *SHOX* enhancers are active in developing limbs.[52] Sall1 and hoxd mutant mice with limb anomalies quit similar to those observed in our patients have overexpression of shox2. [53, 54] *SALL1* is mutated in Townes-Brocks syndrome, patients suffering from this syndrome often have anal, renal and thump anomalies and esophageal atresia is a variable feature in this syndrome. Sequencing revealed no pathogenic *SALL1* mutations in the *SHOX* duplication patients.

We could speculate that *SHOX* duplication is the second hit in a two hit model modulating, not causing, the abnormal development in these patients.

SHOX duplication and other rare inherited Copy Number Variations could be modifying factors explaining responsible for the broad phenotypical spectrum characteristic of the EA/TEF patient population. [19]

Genetic aberration e.g. pathogenic mutations, aneuploidies and structural chromosomal changes like translocations, inversions or Copy Number Variations have previously been detected in ~12.5% of patients in our cohort. This number will steadily increase, since it is expected that screening previously unresolved cases with whole exome sequencing or improved high-resolution micro-array will identify both known and new causal genetic defects. Screening large patient cohorts for genetic defects can delineate new genetic syndromes when genotypes and phenotypes overlap, like recently published for the *EFTUD2* gene. [55]

In conclusion, we describe five patients with sex chromosomal aberrations and EA/TEF. All five patients had duplicated loci of pseudoautosomal genes, including *SHOX*, that escape X-inactivation and are candidates for a gene dosage effect. As a consequence of the additional X-chromosome, triple X females express more transcripts from genes that escape XCI. The expression of one or several of these genes could contribute to the phenotype. Overexpression of XCI escapees could shift the balance from normal to abnormal development in a small percentage of triple X patients. The expression levels of escaping genes on the inactive X-chromosomes may vary between individuals and different tissues.[56] As previously described the phenotypic variability of triple X syndrome ranges widely, from subclinical phenotypes to mental retardation and congenital malformations. [4]

The incidence of triple X syndrome in our EA/TEF cohort is 6.5-11 times higher than expected. Overexpression of XCI-escapees, *SHOX* or other X-linked genes could be responsible for or modulate the phenotype of EA/TEF patients.

Acknowledgments

The authors thank J. Hagoort for editorial support. Elisabeth de Jong and Erwin Brosens were funded by the Sophia Foundations for Scientific Research, projects SSWO- 493 and SWOO13-09

Conflict of Interest Statement

The authors declare no conflict of interest.

References

1. de Jong, E.M., et al., *Etiology of esophageal atresia and tracheoesophageal fistula: "mind the gap".* Curr Gastroenterol Rep, 2010. **12**(3): p. 215-22.
2. Shaw-Smith, C., *Genetic factors in esophageal atresia, tracheo-esophageal fistula and the VACTERL association: roles for FOXF1 and the 16q24.1 FOX transcription factor gene cluster, and review of the literature.* Eur J Med Genet, 2010. **53**(1): p. 6-13.
3. Iskender, C., et al., *Fetal axillary cystic hygroma: a novel association with triple X syndrome.* Birth Defects Res A Clin Mol Teratol, 2012. **94**(11): p. 955-7.
4. Otter, M., C.T. Schrander-Stumpel, and L.M. Curfs, *Triple X syndrome: a review of the literature.* Eur J Hum Genet, 2010. **18**(3): p. 265-71.
5. Visootsak, J. and J.M. Graham, Jr., *Klinefelter syndrome and other sex chromosomal aneuploidies.* Orphanet J Rare Dis, 2006. **1**: p. 42.
6. Bagci, S., et al., *Intestinal atresia, encephalocele, and cardiac malformations in infants with 47,XXX: Expansion of the phenotypic spectrum and a review of the literature.* Fetal Diagn Ther, 2010. **27**(2): p. 113-7.
7. Oddsberg, J., *Environmental factors in the etiology of esophageal atresia.* J Pediatr Gastroenterol Nutr, 2011. **52 Suppl 1**: p. S4-5.
8. Van Opstal, D., et al., *Rapid aneuploidy detection with multiplex ligation-dependent probe amplification: a prospective study of 4000 amniotic fluid samples.* Eur J Hum Genet, 2009. **17**(1): p. 112-21.
9. D'Haene, B., et al., *Improved molecular diagnostics of idiopathic short stature and allied disorders: quantitative polymerase chain reaction-based copy number profiling of SHOX and pseudoautosomal region 1.* J Clin Endocrinol Metab, 2010. **95**(6): p. 3010-8.
10. Eussen, B.H., et al., *A familial inverted duplication 2q33-q34 identified and delineated by multiple cytogenetic techniques.* Eur J Med Genet, 2007. **50**(2): p. 112-9.
11. Jonkers, I., et al., *RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation.* Cell, 2009. **139**(5): p. 999-1011.
12. Jonkers, I., et al., *Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle.* Mol Cell Biol, 2008. **28**(18): p. 5583-94.
13. Plath, K., et al., *Role of histone H3 lysine 27 methylation in X inactivation.* Science, 2003. **300**(5616): p. 131-5.
14. van den Berg, I.M., et al., *X chromosome inactivation is initiated in human preimplantation embryos.* Am J Hum Genet, 2009. **84**(6): p. 771-9.
15. Allen, R.C., et al., *Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation.* Am J Hum Genet, 1992. **51**(6): p. 1229-39.
16. Gribnau, J., et al., *X chromosome choice occurs independently of asynchronous replication timing.* J Cell Biol, 2005. **168**(3): p. 365-73.
17. Koopmans, A.E., et al., *Patient survival in uveal melanoma is not affected by oncogenic mutations in GNAQ and GNA11.* Br J Cancer, 2013. **109**(2): p. 493-6.
18. de Jong, E.M., et al., *Non-VACTERL-type anomalies are frequent in patients with esophageal atresia/tracheo-esophageal fistula and full or partial VACTERL association.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 92-7.
19. Brosens, E., et al., *VACTERL Association Etiology: The Impact of de novo and Rare Copy Number Variations.* Mol Syndromol, 2013. **4**(1-2): p. 20-6.
20. Kenyon, E.J., et al., *Functional analysis of conserved non-coding regions around the short stature hox gene (shox) in whole zebrafish embryos.* PLoS One, 2011. **6**(6): p. e21498.
21. Thomas, N.S., et al., *Clinical and molecular characterization of duplications encompassing the human SHOX gene reveal a variable effect on stature.* Am J Med Genet A, 2009. **149A**(7): p. 1407-14.
22. Roos, L., K. Brondum Nielsen, and Z. Tumer, *A duplication encompassing the SHOX gene and the downstream evolutionarily conserved sequences.* Am J Med Genet A, 2009. **149A**(12): p. 2900-1.
23. Benito-Sanz, S., et al., *Clinical and molecular evaluation of SHOX/PAR1 duplications in Leri-Weill dyschondrosteosis (LWD) and idiopathic short stature (ISS).* J Clin Endocrinol Metab, 2011. **96**(2): p. E404-12.
24. Sabherwal, N., et al., *Long-range conserved non-coding SHOX sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients.* Hum Mol Genet, 2007. **16**(2): p. 210-22.

25. Benito-Sanz, S., et al., *Identification of the first PAR1 deletion encompassing upstream SHOX enhancers in a family with idiopathic short stature.* Eur J Hum Genet, 2012. **20**(1): p. 125-7.
26. Trautner, M.C., et al., *Jejunal atresia in an infant with triple-X syndrome.* J Matern Fetal Neonatal Med, 2004. **16**(3): p. 198-200.
27. Rolle, U., et al., *Duodenal atresia in an infant with triple-X syndrome: a new associated malformation in 47,XXX.* Birth Defects Res A Clin Mol Teratol, 2007. **79**(8): p. 612-3.
28. Wilkins-Haug, L., et al., *Isolated fetal omphalocele, Beckwith-Wiedemann syndrome, and assisted reproductive technologies.* Birth Defects Res A Clin Mol Teratol, 2009. **85**(1): p. 58-62.
29. De Veciana, M., C.A. Major, and M. Porto, *Prediction of an abnormal karyotype in fetuses with omphalocele.* Prenat Diagn, 1994. **14**(6): p. 487-92.
30. Hood, O.J., et al., *Multiple congenital anomalies associated with a 47,XXX chromosome constitution.* Am J Med Genet, 1990. **36**(1): p. 73-5.
31. Johnston, K.M., N.C. Nevin, and J.M. Park, *Cloacal defect in a 23-year-old with 47,XXX karyotype and clinical features of Cat Eye syndrome.* J Obstet Gynaecol, 2002. **22**(6): p. 696.
32. Haverty, C.E., et al., *47,XXX associated with malformations.* Am J Med Genet A, 2004. **125A**(1): p. 108-11; author reply 112.
33. Lin, H.J., F. Ndiforchu, and S. Patell, *Exstrophy of the cloaca in a 47,XXX child: review of genitourinary malformations in triple-X patients.* Am J Med Genet, 1993. **45**(6): p. 761-3.
34. Hoang, M.P., et al., *Case report of a 22-week fetus with 47,XXX karyotype and multiple lower mesodermal defects.* Pediatr Dev Pathol, 1999. **2**(1): p. 58-61.
35. Spencer, K., N. Tul, and K.H. Nicolaides, *Maternal serum free beta-hCG and PAPP-A in fetal sex chromosome defects in the first trimester.* Prenat Diagn, 2000. **20**(5): p. 390-4.
36. Meetei, A.R., et al., *X-linked inheritance of Fanconi anemia complementation group B.* Nat Genet, 2004. **36**(11): p. 1219-24.
37. Boyd, P.A., et al., *Sex chromosome trisomies in Europe: prevalence, prenatal detection and outcome of pregnancy.* Eur J Hum Genet, 2011. **19**(2): p. 231-4.
38. Guichet, A., et al., *Trisomy X: ACLF (Association des Cytogeneticiens de Langue Francaise) retrospective study.* Ann Genet, 1996. **39**(3): p. 117-22.
39. Stewart, D.A., C.T. Netley, and E. Park, *Summary of clinical findings of children with 47,XXY, 47,XYY, and 47,XXX karyotypes.* Birth Defects Orig Artic Ser, 1982. **18**(4): p. 1-5.
40. Krusinskiene, V., L. Alvesalo, and A. Sidlauskas, *The craniofacial complex in 47, XXX females.* Eur J Orthod, 2005. **27**(4): p. 396-401.
41. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females.* Nature, 2005. **434**(7031): p. 400-4.
42. Barakat, T.S., et al., *X-changing information on X inactivation.* Exp Cell Res, 2010. **316**(5): p. 679-87.
43. Heard, E. and C.M. Disteche, *Dosage compensation in mammals: fine-tuning the expression of the X chromosome.* Genes Dev, 2006. **20**(14): p. 1848-67.
44. Brown, C.J. and J.M. Greally, *A stain upon the silence: genes escaping X inactivation.* Trends Genet, 2003. **19**(8): p. 432-8.
45. Lenroot, R.K., N.R. Lee, and J.N. Giedd, *Effects of sex chromosome aneuploidies on brain development: evidence from neuroimaging studies.* Dev Disabil Res Rev, 2009. **15**(4): p. 318-27.
46. Clement-Jones, M., et al., *The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome.* Hum Mol Genet, 2000. **9**(5): p. 695-702.
47. Binder, G., *Short stature due to SHOX deficiency: genotype, phenotype, and therapy.* Horm Res Paediatr, 2011. **75**(2): p. 81-9.
48. Gervasini, C., et al., *SHOX duplications found in some cases with type I Mayer-Rokitansky-Kuster-Hauser syndrome.* Genet Med, 2010. **12**(10): p. 634-40.
49. Iughetti, L., et al., *Unexpected phenotype in a boy with trisomy of the SHOX gene.* J Pediatr Endocrinol Metab, 2010. **23**(1-2): p. 159-69.
50. Ou, Z., et al., *Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes.* Genet Med, 2008. **10**(4): p. 267-77.
51. Vickerman, L., S. Neufeld, and J. Cobb, *Shox2 function couples neural, muscular and skeletal development in the proximal forelimb.* Dev Biol, 2011. **350**(2): p. 323-36.
52. Durand, C., et al., *Enhancer elements upstream of the SHOX gene are active in the developing limb.* Eur J Hum Genet, 2010. **18**(5): p. 527-32.
53. Cobb, J. and D. Duboule, *Comparative analysis of genes downstream of the Hoxd cluster in developing digits and external genitalia.* Development, 2005. **132**(13): p. 3055-67.

54. Kiefer, S.M., et al., *SALL1 truncated protein expression in Townes-Brocks syndrome leads to ectopic expression of downstream genes.* Hum Mutat, 2008. **29**(9): p. 1133-40.
55. Gordon, C.T., et al., *EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia.* J Med Genet, 2012. **49**(12): p. 737-46.
56. Talebizadeh, Z., S.D. Simon, and M.G. Butler, *X chromosome gene expression in human tissues: male and female comparisons.* Genomics, 2006. **88**(6): p. 675-81.

2.4

Contribution of *LPP* Copy Number and Sequence Changes to Esophageal Atresia, Tracheo-esophageal Fistula, and VACTERL Association

Adapted from:

Contribution of LPP Copy Number and Sequence Changes to Esophageal Atresia, Tracheoesophageal Fistula, and VACTERL Association

Andrés Hernández-García[1], Erwin Brosens, Hitisha P. Zaveri, Elisabeth M. de Jong, Zhiyin Yu, Maria Namwanje, Allison Mayle, Caraciolo J. Fernandes, Brendan Lee, Maria Blazo, Seema R. Lalani, Dick Tibboel, Annelies de Klein, Daryl A. Scott. *Am J Med Genet A. 2012 Jul;158A(7):1785-7*

To the editor:

Esophageal atresia (EA) and tracheoesophageal fistula (TEF) are common life-threatening birth defects that affect approximately 1 in 3,500 births [1]. Approximately 10% of individuals with EA/TEF also meet criteria for VACTERL association which requires the presence of three or more of the following non-randomly associated birth defects: Vertebral anomalies, Anal anomalies, Cardiac defects, TracheoEsophageal fistula and/or esophageal atresia, Renal anomalies and Limb.[2] The genetic factors that contribute to the development of most cases of TEF and VACTERL association have yet to be identified.

The LIM domain containing preferred translocation partner in lipoma (LPP) gene has recently been proposed as a candidate gene that may contribute to the development of EA/TEF, or phenotypes associated with VACTERL association [3]. LPP localizes in focal adhesions as well as in cell-to-cell contacts and binds to VASP, a protein implicated in the control of actin organization. This suggests that LPP has a structural role at sites of cell adhesion in maintaining cell shape and motility. [4, 5] LPP is also transiently located in the nucleus where it can activate transcription.[5]

The bulk of evidence for *LPP*'s role in EA/TEF and VACTERL association comes from a report by Arrington et al. in which they described a 451 kb interstitial deletion involving only the *LPP* gene in an individual diagnosed with EA/TEF and VACTERL association [3]. They also described an intronic deletion in *LPP* associated with reduced *LPP* expression levels that segregated with TOF in three individuals in a small nuclear family—a father and two children. However, in both cases, a lack of DNA from family members made it impossible to determine if the changes arose *de novo* or were inherited from an unaffected family member. Deletions affecting multiple exons of *LPP* have also been reported in 5 out of 2906 unselected adults (0.17%) from the island of Kosrae, Micronesia but the phenotype of these individuals was not reported, making it impossible to draw any conclusions about their effects.[6]

In an effort to clarify *LPP*'s role in the development of EA/TEF and VACTERL association, we looked for genomic alterations and deleterious sequence changes in the *LPP* gene in individuals with these diagnoses. First, we screened for genomic alterations in a convenience cohort of 195 patients with these diagnoses based on a review of their medical records. This cohort consisted of 48 individuals with isolated EA/TEF, 77 with EA/TEF and additional findings, 64 with non-isolated EA/TEF who met criteria for VACTERL

association, and 6 with VACTERL association in whom EA/TEF was not a feature. After obtaining informed consent, DNA samples were screened for genomic alterations using an array comparative genomic hybridization or SNP-based copy number detection platform. Although each of these platforms contained a minimum of 40 probes within the *LPP* gene, no copy number variations or regions of absence of heterozygosity (AOH) >1 Mb—which could be associated with consanguinity and the possible inheritance of a recessive inherited deleterious allele—affecting the *LPP* gene were detected in this cohort.

If we assume that these array-based copy number detection techniques can detect 90% of deleterious copy number changes, then screening 195 individuals would have a greater than 97% chance of identifying at least one copy number change in the *LPP* gene if such changes were present in at least 2% of a similarly affected cohort. The absence of *LPP* copy number changes in our cohort suggests that such changes are not a common cause of EA/TEF or VACTERL association.

To determine if deleterious sequence changes in *LPP* are a common cause of EA/TEF or VACTERL association, we screened for such changes in a partially overlapping cohort of 75 patients—21 with isolated EA/TEF, 18 with EA/TEF and additional findings, 27 with non-isolated EA/TEF who met criteria for VACTERL association, and 9 patients with VACTERL association in whom EA/TEF was not a feature. PCR amplification and sequencing of the *LPP* coding region and intervening introns (Supplemental Online Material) revealed four non-synonymous changes—pSer144Thr (S144T), p.Ser246Leu (S246L), p.Tyr346His (Y346H), and p.Arg388Cys (R388C).

These changes were identified in a total of seven patients whose clinical findings are summarized in Table I. The S246L, Y346H and R388C changes have been previously reported in the SNP database (dbSNP) and the Y346H, and R388C changes have also been seen in individuals sequenced as part of the 1000 Genomes project. Three publicly available on-line programs — the HumVar model of PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org/) and SNPs3D (http://www.snps3d.org/) —were used to predict the effects of non-synonymous changes in the *LPP* gene on protein function. These results are shown in Table II.

Patient ID	Clinical Synopsis [a]	Non-Synonymous Change (Inheritance [b])
TX-11	VACTERL association with TEF, vertebral, and cardiac anomalies	Het Y346H (Parental samples not available)
TX-22	VACTERL association with vertebral, renal, and limb anomalies	Homo Y346H (Inherited from heterozygous parents) Het R388C (Maternal)
N-A12	EA/TEF and cardiac anomalies	Het Y346H (Both parents are heterozygous)
N-A13	EA/TEF, cardiac	Het S246L (Maternal)
N-A15	EA/TEF, limb anomalies	Het S246L (Maternal)
N-A24	EA/TEF	Het S144T (Paternal)
N-A42	EA/TEF, upper limb, and genital anomalies	Homo Y346H (Paternal, maternal sample not available)

Table I. Clinical and molecular synopsis of individuals with non-synonymous changes in LPP.
Het = individuals were heterozygous for the change; Homo = individuals were homozygous for the change. a Phenotype based on an analysis of the medical record. b Parents were reported to be unaffected but no radiological evaluations were obtained to rule out subtle changes that may have eluded detection.

Although the S144T change is rare, it is unlikely to contribute to the phenotype of the patient in which it was found since it was inherited from a self-reportedly unaffected father and was predicted to be benign by all three of the prediction programs. The Y346H allele was predicted to be possibly damaging by one prediction program but benign by the other two— calling into question its effect on *LPP* function. The c.1036T>C change associated with the Y346H allele is also a relatively common variant with a reported minor allele frequency of 0.2099 in dbSNP. Based on Hardy-Weinberg equilibrium, the Y346H allele is expected to be found in the homozygous state in 4.4% of the population, a level that would make it unlikely to strongly contribute to the development of EA/TEF or VACTERL association even in the homozygous state.

In contrast to the Y346H change, both the S246L change that was seen in two patients and the R388C change seen in one patient are relatively rare. The S246L change was predicted to be deleterious by one of the three prediction programs and the R388C change was predicted to be deleterious by all three prediction programs. However, in all cases, these changes were inherited from self-reportedly unaffected parents. The R388C change was also seen in 1 out of 78 African American controls from the Baylor Polymorphism Resource—a collection of anonymous control samples from major ethnic and racial backgrounds. Although the anonymous nature of this resource makes it impossible to determine the phenotype of the individual carrying the R388C change, the presence of this change in these control individuals suggests that this allele may be seen more commonly in this ethnic group.

Taken together, these results suggest that genomic alterations and clearly deleterious sequence changes in the *LPP* gene are not a common cause of EA/TEF or VACTERL association. This makes it unlikely that clinical screening of the *LPP* gene would prove to be cost effective in individuals with EA/TEF or VACTERL association. Our data also suggests that if the non-synonymous *LPP* sequence changes identified in this study contribute to the development of these phenotypes, they are likely to have a relatively mild effect and to do so in combination with other genetic, environmental and/or stochastic factors.

Acknowledgments

The authors would like to thank the families that participated in this study and Alyssa Tran for her help with patient recruitment. Funding for this study was provided by Doris Duke Charitable Foundation grant 2007053 (DAS) and the Sophia Stichting Wetenschappelijk Onderzoek (SWOO, project 493), Rotterdam, The Netherlands. The authors have no conflicts of interest to declare.

Observed Change	Number of Patients	Present in dbSNP?	Present in 1000 Genomes?	Predicted Effect (PolyPhen 2)	Predicted Effect (SIFT)	Predicted Effect (SNPs3D)
Nonsynonymous Changes						
c.431G>C p.Ser144Thr S144T	1 Het	No	No	Benign	Tolerated	Non-Deleterious
c.737C>T p.Ser246Leu S246L	2 Het	Yes	No	Benign	Deleterious	Non-Deleterious
c.1036T>C p.Tyr346His Y346H	3 Het 1 Homo	Yes	Yes	Probably Damaging	Tolerated	Non-Deleterious
c.1162C>T p.Arg388Cys R388C	1 Het	Yes	Yes	Possibly Damaging	Deleterious	Deleterious
Synonymous Changes						
c.942C>T p.Asp314= D314D	24 Homo 34 Het	Yes	Yes	N/A	N/A	N/A
c.1605A>G p.Cys536= C536C	22 Homo 33 Het	No	No	N/A	N/A	N/A
c.1143G>A p.Ser382= S382S	1 Het	No	No	N/A	N/A	N/A
c.1395C>T p.Glu466= E466E	2 Het	No	No	N/A	N/A	N/A

Table II. Evaluation of changes in the *LPP* gene identified in patients with EA/TEF and VACTERL association. Het = individuals were heterozygous for the change; Homo = individuals were homozygous for the change; N/A = not applicable

Materials and Methods

Patient accrual and preparation of DNA

Informed consent was obtained from a convenience sample of patients with EA/TEF or a diagnosis of VACTERL association based on the presence of three or more of the following birth defects: Vertebral anomalies, Anal anomalies, Cardiac defects, TracheoEsophageal fistula or esophageal atresia, Renal anomalies and Limb defects. When possible, the parents of these individuals were similarly accrued. DNA was extracted from whole blood or immortalized lymphoblastic cultures. dsDNA extracted from whole blood or immortalized lymphoblastic cultures using a QIAamp DNA Blood Midi Kit (Qiagen Inc., Hilden, Germany) or a Qiagen Puregene Blood Kit (Qiagen Sciences, MD, USA) was quantified using a Quant-iT™ PicoGreen ® dsDNA kit (Invitrogen, CA, USA) or a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., DE, USA) .

In cases where the amount of genomic DNA available was limiting, ~10 ng of DNA was amplified using a GenomiPhi DNA (GE Healthcare Life Sciences, NJ, USA) amplification kit according to manufacturer's instructions. This amplified DNA was then used as a template for PCR amplification and sequencing but not for array studies.

African American control samples were obtained from the Baylor Polymorphism Resource, a collection of anonymous control samples from major ethnic and racial backgrounds.

Array comparative genomic hybridization and SNP-based copy number analyses

A screen for copy number changes involving LPP was carried out by array comparative genomic hybridization (aCGH) or SNP-based copy number analysis enabling both copy number and absence of heterozygosity (AOH) detection.

aCGH experiments were performed using Agilent Human Genome CGH 244K or SurePrint G3 Human CGH 1M Oligo Microarray Kits (G4411B, G4447, Agilent Technologies, CA, USA) as previously described [7]These arrays have a minimum of 82 probes that span the length of the LPP gene. Controls consisted of DNA from sex matched individuals with no personal or family history of EA/TEF or VACTERL association.

SNP-based analyses of copy number and AOH detection, were performed using either Affymetrix GeneChip Human Mapping 250K NSP1 (Affymetrix, CA, USA), HumanQ610, HumanCytoSNP-12v2.1 or HumanOmniExpress (Illumina, CA, USA)

platforms. Depending on array type, a minimum of 40 markers within the LPP gene allow for copy number and/or AOH detection. Information within the array allow for self-normalization of image intensities. We generated Affymetrix CEL files with the Affymetrix genotype command console v3.2 software or Illumina Final Reports with Illumina's Genome Studio. We analyzed the SNP arrays in Biodiscovery Nexus CN5.1 using Illumina's SNPFASST2 algorithm to call segments consisting of a minimum of 3 markers. SNP array data were compared to a virtual reference array. This virtual array was composed of a large pool of array specific CEU HapMap control samples. The copy number was considered abnormal if both logR and B-allele frequency were either above (gain) or below (loss) normal values (logR values between -0.18 and +0.18 and B-allele frequencies ranging from 0.45 to 0.55).

Sequencing of LPP and analysis of variants

The coding region and intervening exon/intron boundaries of LPP were amplified by PCR using previously published primers [Arrington et al., 2010]. Sequence changes in PCR amplified products were identified by comparison with control DNA sequences using Sequencher 4.7 software (Gene Codes Corporation).

Three publicly available on-line programs—the HumVar model of PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org/) and SNPs3D (http://www.snps3d.org/)—were used to predict the effects of non-synonymous changes in the LPP gene on protein function.

References

1. Shaw-Smith, C., *Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology.* J Med Genet, 2006. **43**(7): p. 545-54.
2. Chittmittrapap, S., et al., *Oesophageal atresia and associated anomalies.* Arch Dis Child, 1989. **64**(3): p. 364-8.
3. Arrington, C.B., et al., *Haploinsufficiency of the LIM domain containing preferred translocation partner in lipoma (LPP) gene in patients with tetralogy of Fallot and VACTERL association.* Am J Med Genet A, 2010. **152a**(11): p. 2919-23.
4. Gorenne, I., et al., *LPP expression during in vitro smooth muscle differentiation and stent-induced vascular injury.* Circ Res, 2006. **98**(3): p. 378-85.
5. Petit, M.M., et al., *LPP, an actin cytoskeleton protein related to zyxin, harbors a nuclear export signal and transcriptional activation capacity.* Mol Biol Cell, 2000. **11**(1): p. 117-29.
6. Gusev, A., et al., *Whole population, genome-wide mapping of hidden relatedness.* Genome Res, 2009. **19**(2): p. 318-26.
7. Scott, D.A., et al., *Genome-wide oligonucleotide-based array comparative genome hybridization analysis of non-isolated congenital diaphragmatic hernia.* Hum Mol Genet, 2007. **16**(4): p. 424-30.

2.5

Small non-coding differentially-methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder

Adapted from:

Small non-coding differentially-methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder

Przemyslaw Szafranski, Avinash V. Dharmadhikari, Erwin Brosens, Priyatansh Gurha, Katarzyna E. Kolodziejska, Zhishuo, Ou, Piotr Dittwald, Tadeusz Majewski, K. Naga Mohan, Bo Chen, Richard E. Person, Dick Tibboel, Annelies de Klein, Jason Pinner, Maya Chopra, Girvan Malcolm, Gregory Peters, Susan Arbuckle, Sixto F. Guiang III, Virginia A. Hustead, Jose Jessurun, Russel Hirsch, David P. Witte, Isabelle Maystadt, Neil Sebire, Richard Fisher, Claire Langston, Partha Sen, Paweł Stankiewicz. *Genome Res. 2013 Jan;23(1):23-33*

Abstract

An unanticipated and tremendous amount of the non-coding sequences of the human genome are transcribed. Long non-coding RNAs (lncRNAs) constitute a significant fraction of non-protein coding transcripts; however, their functions remain enigmatic.

We demonstrate that deletions of a small non-coding differentially-methylated region at 16q24.1, including lncRNA genes, cause a lethal lung developmental disorder, Alveolar Capillary Dysplasia with Misalignment of Pulmonary Veins (ACD/MPV), with parent of origin effects. We identify overlapping deletions 250 kb upstream to FOXF1 in nine patients with ACD/MPV that arose de novo specifically on the maternally inherited chromosome and delete two lung-specific lncRNA genes. These deletions define a distant cis-regulatory region that harbors, besides lncRNA genes, also a differentially methylated CpG island, binds GLI2 depending on the methylation status of this CpG island, and physically interacts with and up-regulates the FOXF1 promoter.

We suggest that lung-transcribed 16q24.1 lncRNAs may contribute to long-range regulation of FOXF1 by GLI and other transcription factors. Perturbation of lncRNA-mediated chromatin interactions may, in general, be responsible for position effect phenomenon and potentially cause many disorders of human development.

Introduction

Mammalian regulatory elements, including enhancers, repressors, and insulators are usually contained in regions that lack protein-coding genes, and can be located hundreds of kb from genes that they regulate[1]. High-throughput analysis of the human transcriptome revealed that the majority of noncoding portion of the genome is transcribed as regulatory RNAs [2, 3] Long non-coding RNAs (lncRNAs) are non-protein coding transcripts longer than 200 nucleotides, distinguished from small regulatory RNAs such as microRNAs, siRNAs, piwiRNAs, and snoRNAs. Little is known about the function of lncRNAs in humans with *XIST* being the earliest identified and best characterized to date human lncRNA. Recently, lncRNAs have been implicated in imprinting and long-range gene regulation with functional properties of enhancers [2-4]. Here, we demonstrate that loss of a distant enhancer region in a coding-gene desert at 16q24.1, that includes lung-expressed lncRNA genes, leads to Alveolar Capillary Dysplasia with Misalignment of Pulmonary Veins (ACD/MPV) (MIM 265380).

ACD/MPV is a rare neonatally lethal developmental disorder of the lungs defined by malposition of pulmonary veins adjacent to small pulmonary arteries, medial thickening of small pulmonary arteries, deficient lobular development, a paucity of alveolar wall capillaries, and occasional lymphangiectasis[5, 6]. A few familial ACD/MPV cases have been described ([7, 8]. During the last 10 years, we ascertained ACD/MPV samples (mainly formalin-fixed paraffin-embedded lung tissues) from over 90 families. In 2009, we reported genomic deletions and inactivating point mutations in the *FOXF1* gene in chromosome 16q24.1 in unrelated patients with sporadic histopathologically-verified ACD/MPV [9].

FOXF1 (Forkhead box protein F1) is a member of the FOX transcription factor family sharing a winged helix/forkhead DNA-binding domain [10-12]. *FOXF1* and its mouse orthologue are predominantly expressed in sub-epithelial mesenchymal tissues of developing lung and foregut[4, 13-15] (. Homozygous mice deficient for *Foxf1* die *in utero* by E10 due to defects in mesodermal differentiation and cell adhesion [16]. Interestingly, approximately 50% of *Foxf1*$^{+/-}$ mice die from pulmonary hemorrhage and show severe defects in alveolarization and vasculogenesis[16, 17], partially recapitulating histopathological pulmonary defects in infants with ACD/MPV, whereas the remainder appears normal.

Both *in vitro* and *in vivo* studies in mice have shown that none of the *in silico*-predicted transcription factor binding sites located within the *Foxf1* promoter confer its tissue specificity [4, 18], suggesting that the promoter is regulated by tissue specific distant regulatory elements. Also, deletions upstream to *FOXF1*, leaving *FOXF1* intact, have been described in two patients with ACD/MPV, suggesting the presence of distant regulatory elements for *FOXF1* [9]. In addition, *FOXF1* has been bioinformatically predicted to be paternally imprinted [19] and other studies have supported this prediction[9, 20].

We now report novel overlaping *de novo* non-coding deletion copy-number variants (CNVs) located 96-257 kb upstream to *FOXF1* in seven patients with ACD/MPV, who all died in the first month of life with severe respiratory distress and pulmonary hypertension. Based on extensive experimental characterization of the microdeletion-defined upstream regulatory region, SDR, we propose that paternally incompletely imprinted *FOXF1* is regulated *in cis* by an interplay between chromatin looping, possibly with contribution of lncRNAs, and methylation-controlled GLI2-binding.

Results

Non-coding deletion CNVs upstream to FOXF1

DNA samples from nine patients with ACD/MPV, negative for point mutations within the *FOXF1* coding exons, were analyzed by array comparative genomic hybridization (aCGH) for the presence of CNVs in 16q24.1. In sample 59.4, aCGH revealed a 15 kb deletion of the entire *FOXF1* and its promoter. Studies of the remaining eight cases identified CNVs that left the *FOXF1* coding region intact. In patient 41.4, a ~ 11 kb deletion, mapping 2.6 kb upstream to *FOXF1*, removed one copy of the distal portion of the *FOXF1* promoter, corresponding to *Foxf1* promoter region II in mice (Kim et al. 2005) and overlapping with a portion of the lncRNA gene *LOC400550* (*FOXF1-AS1*, Ota et al. 2004; *lncFOXF1*, Khalil et al. 2009). We show by RT-PCR and DNA sequencing that *lncFOXF1* is expressed in normal newborn lungs (Supplemental Fig. S1). In the remaining seven patients, overlapping deletions, ranging in size from 140 kb to 2,625 kb, mapped 96-257 kb upstream to *FOXF1* (Fig. 1; Supplemental Table S1).

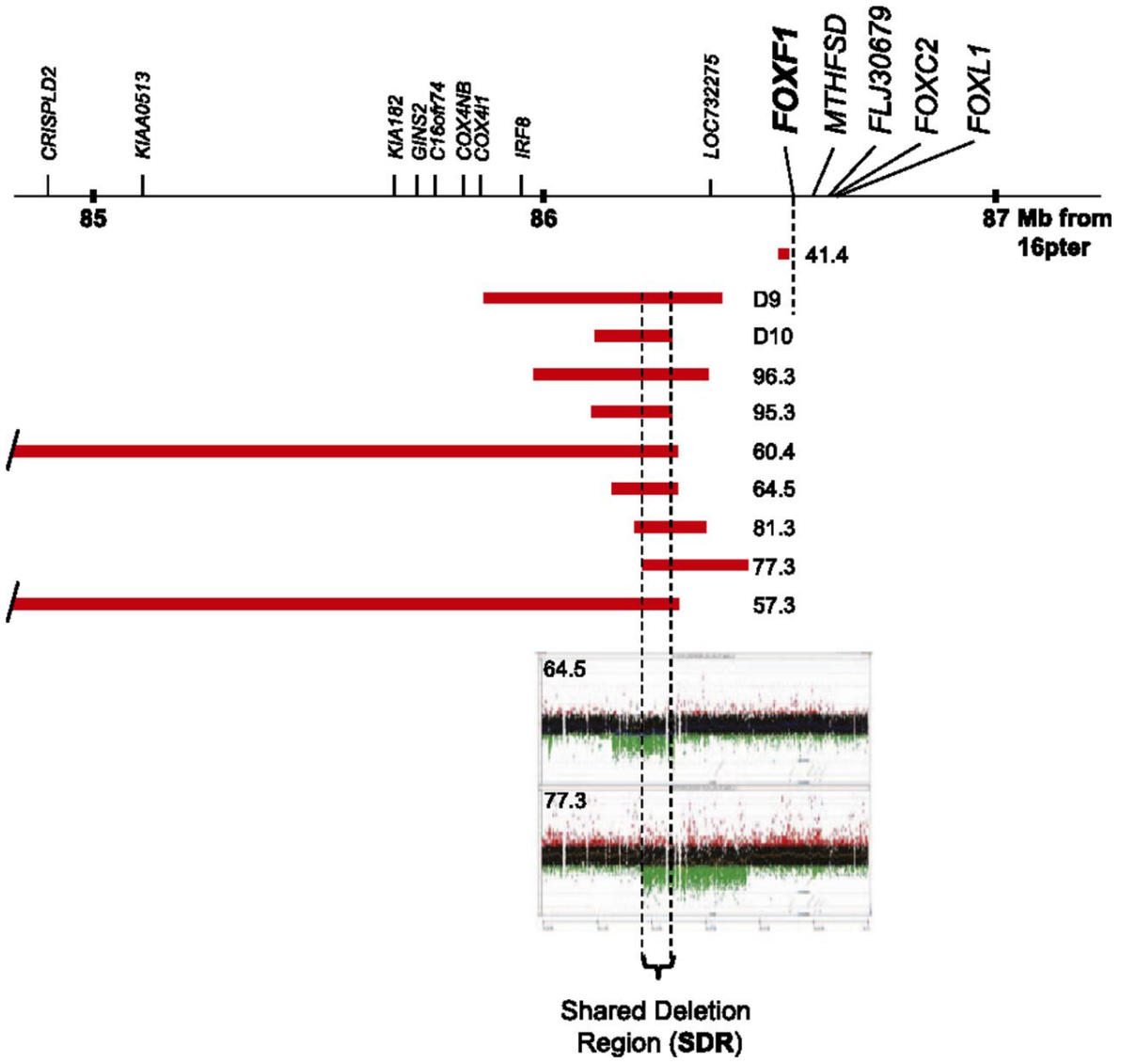

Figure. 1. Nine (two published D9, D10, and seven novel) ACD/MPV-causing microdeletions share a 75 kb putative regulatory region (SDR), mapping 257 kb upstream to FOXF1 (16q24.1). Array CGH plots of two deletions defining SDR are shown.

To better estimate the recurrence risk for these deletions, we mapped and sequenced the deletion breakpoints of six of the upstream microdeletions (patients 57.3, 60.4, 64.5, 77.3, 95.3, and 96.3) (Supplemental Fig. S2; Supplemental Table S1) and the microdeletion that included the *FOXF1* gene (patient 59.4) (Supplemental Fig. S2). Microhomology was identified in five of seven breakpoint junctions (patients 57.3, 60.4, 64.5, 77.3, and 95.3) consistent with the deletions arising by a template switching replicative mechanism such as FoSTeS/MMBIR [21, 22]. In three of these five cases, both deletion breakpoints mapped within *Alu* elements (Supplemental Table S1). The probability of finding both breakpoints in *Alu* elements by chance was estimated at less than 0.0066 (Supplemental Fig. S3), thus being much lower than the observed frequency of finding breakpoints of ACD/MPV

microdeletions in *Alu* repetitive elements. This finding suggests a mechanistic link between *Alu* sequences and the location of microdeletion breakpoints.

Interestingly, in patients 64.5 and 95.3, the distal breakpoints were located within the same copy of an *Alu*Sx element (about 100 bp apart), thus defining a potential breakage-prone hotspot. In patient 59.4, the breakpoints occurred within highly homologous regions of the low complexity/GC rich repeats and SINE/MIR repeats. We did not find any evidence for low-level somatic mosaicism in parental blood samples using PCR for patient-specific junction fragments; all microdeletions appear to have occurred *de novo* (Supplemental Fig. S4).

Regulatory region upstream to FOXF1

Sequence alignments of the seven distant microdeletions showed that they share a 75 kb region 257 kb upstream to *FOXF1* (chr16:86,212,040-86,287,054) (Shared Deletion Region, SDR) (Fig. 1; Supplemental Table S1). We found that this, deletion-defined SDR represents a protein-gene desert (http://genome.ucsc.edu). Nevertheless, this genomic region harbors DNA segments that are evolutionarily conserved among land vertebrates and have high 7x regulatory potential (Fig. 2A). Hence, we hypothesized that the SDR may contain regulatory site(s) controlling the *FOXF1* promoter.

Interestingly, we also found that the SDR encompasses one locus (chr16:86,223,827-86,234,547) and part of another locus (chr16:86,254,429-86,338,058), encoding differentially-spliced putative lncRNAs exhibiting increased expression in the lungs (http://genome.ucsc.edu). The 43.5 kb EST CR737045 (chr16:86,259,186-86,302,689), part of the 79-kb lncRNA TCONS_00024764, is specifically expressed in fetal lungs (SOURCE at http://source.stanford.edu). We show, by RT-PCR and sequencing, that *TCONS_00024764* lncRNAs is also expressed in normal newborn lung at much higher level than in HEK293 kidney cells and lymphoblasts (Supplemental Fig. S1).We further narrowed SDR to ~ 10 kb Segment 1 of the highest 7x regulatory potential and sequence conservation (Fig. 2A), containing a cluster of GLI-binding sites overlapping with CpG island (Fragment 1a) (Fig. 2B; Supplemental Fig. S5), and ~ 8 kb Segment 2 with multiple binding sites for other transcription regulators, and highest potential for histone methylation and acetylation. Both segments encode lncRNAs expressed in the lungs (Fig. 2A).

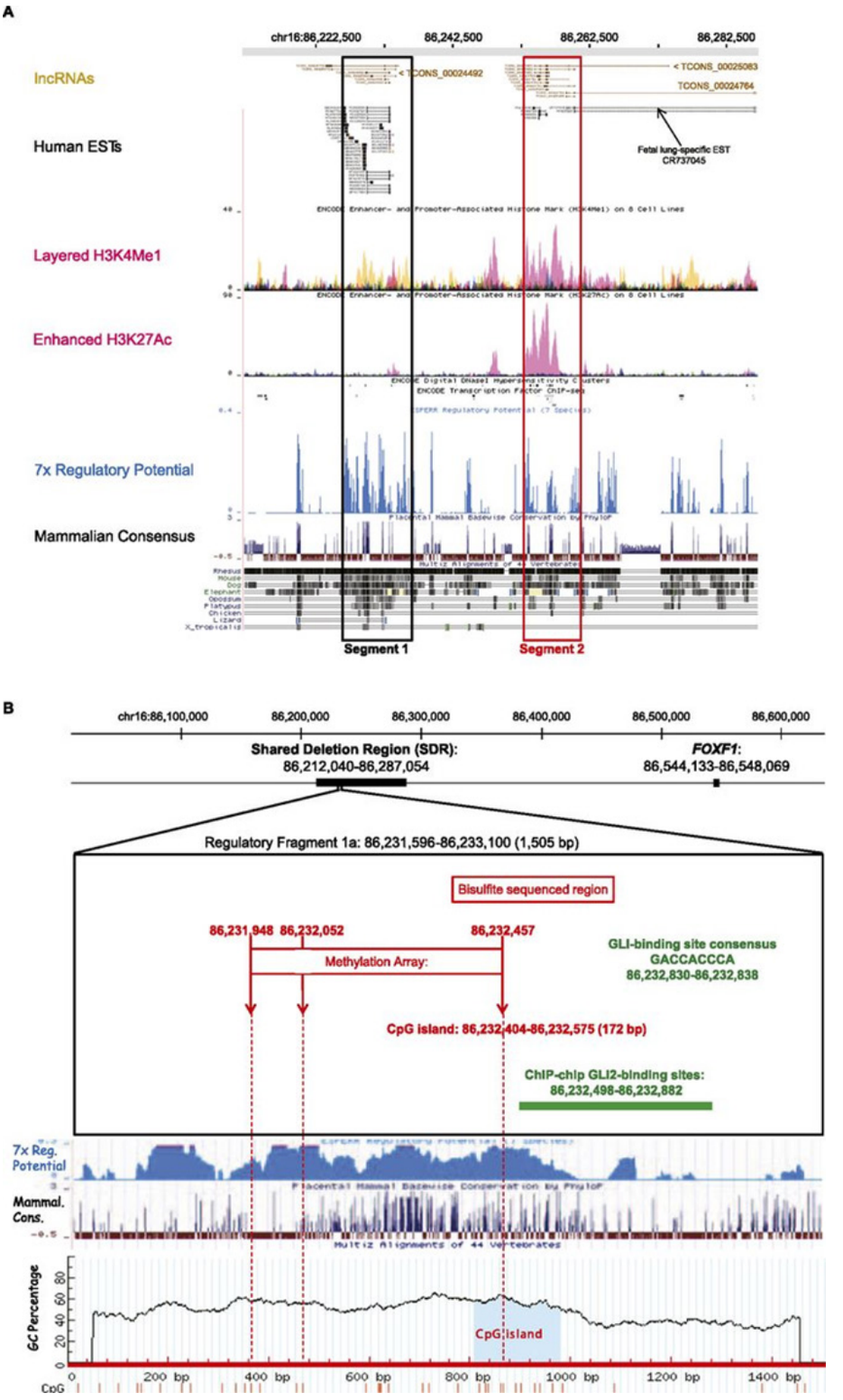

Figure. 2. Characterization of the SDR. (*A*) Segment 1 (~ 10 kb, black rectangle), and Segment 2 (~ 8 kb, red rectangle), harboring, among others, transcription factor binding sites and a part of the fetal lung-expressed lncRNA, are shown. The figure is drawn according to the 2009 human reference sequence (GRCh37/hg19). (*B*) Insight into the ~ 1.5 kb Fragment 1a showing (from top to bottom): bisulfite sequenced region, methylation array oligo probes that indicated differential methylation, differentially-methylated CpG island, ChIP-chip GLI2-binding region, 7x regulatory potential, mammalian evolutionary conservation, and CG content.

GLI2 binds within the upstream regulatory region

Sonic hedgehog (Shh) signaling targets a GLI2 transcription factor and is crucial for early lung development in mice [23]) and humans [24]. GLI1-3 have been shown to be expressed in human lung mesenchymal tissues, including endothelial cells [14]. Moreover, studies in mice have shown that Gli2 regulates *Foxf1* expression in the developing stomach and intestine [25], which, together with airway and alveolar tissues, are of endodermal origin. To determine whether *in silico* identified GLI-binding sites in the SDR bind GLI2 *in vivo*, we screened the SDR for the GLI2-binding using a ChIP-chip assay. We detected strong binding of GLI2 to SDR in the cultured human pulmonary microvascular endothelial cells (HPMEC) (p=0.007, n=2) at a region including several *in silico*-predicted GLI-binding sites (chr16:86,232,498-86,232,882) (Supplemental Fig. S5,S6). Although only one of the GLI-binding sites of SDR perfectly matches the consensus sequence, the clustering of nine of these sites within a 0.2 kb segment of the Fragment 1a may have synergistic effect on GLI2 binding. Given that HPMEC express both *FOXF1* and GLI2 (Supplemental Fig. S7), we hypothesized that GLI2 binding at SDR may regulate *FOXF1* transcription.

GLI2-binding to the distant enhancer regulates FOXF1 promoter

To this end, we performed a secreted alkaline phosphatase (SEAP) reporter assay in HPMEC (Fig. 3A,B). We prepared plasmid constructs that allow for transient expression of the SEAP reporter gene under the control of the *FOXF1* promoter and its putative regulatory sites. The upstream putative regulatory sequences were then tested both *in cis* (inserted into a vector upstream to the *FOXF1* promoter) and *in trans* (placed on another plasmid used in the co-transfection of HPMEC with the *FOXF1* promoter-containing reporter vector). The -5.5 kb *FOXF1* promoter region includes two intervals that remain highly conserved between mice and humans (Kim et al. 2005), and contains the RNA *Pol*II binding site (chr16:86,543,198-86,543,374) and multiple transcription factor binding sites (httm://genome.ucsc.edu). When placed upstream to the promoterless *SEAP*, the -5.5 kb promoter region activated *SEAP* transcription 11 to 20 fold, this range depended on whether the cells were additionally co-transfected with a pCS2Gli2 vector, constitutively expressing GLI2 (P<0.0001) (Fig. 3A).

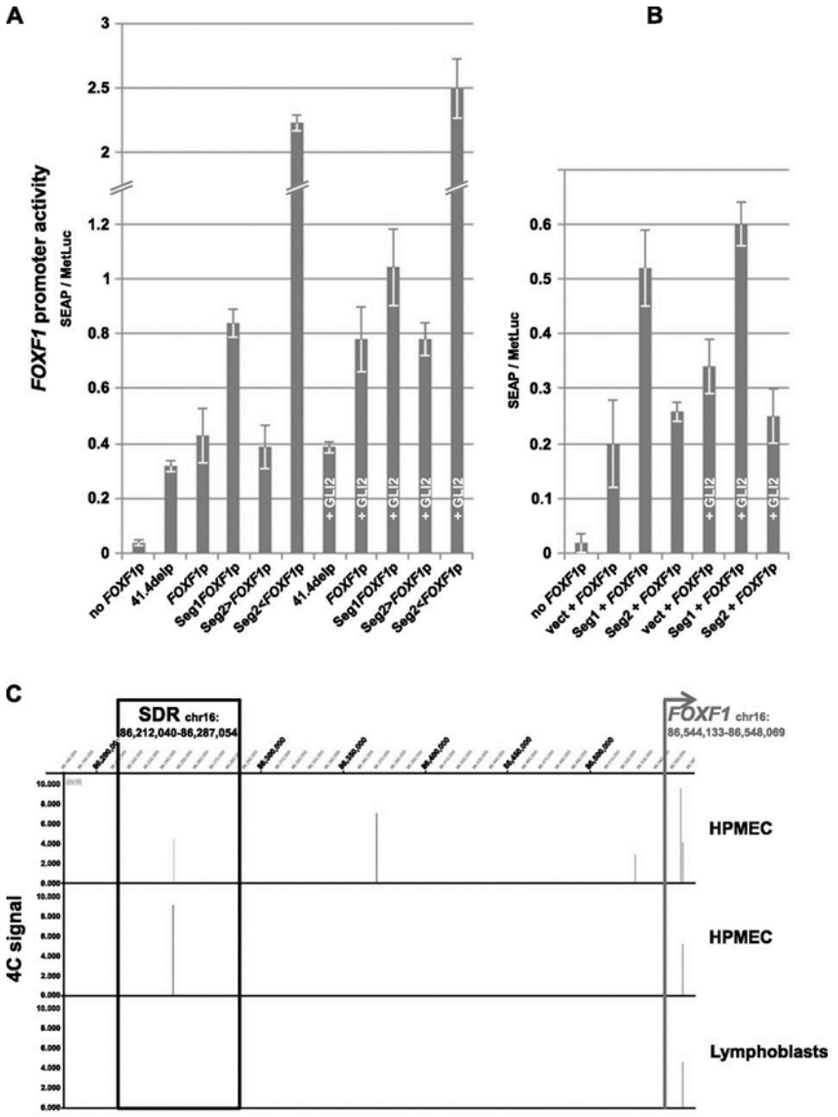

Figure. 3. Distant regulatory region, SDR, controls the activity of *FOXF1* promoter. (*A, B*) Results of the reporter assay experiments showing regulation of the *FOXF1* promoter in HPMEC by fragments of Segments 1 and 2 of the SDR. Abbreviations: *FOXF1*p, *FOXF1* promoter (0 to 5.5 kb upstream to ATG codon) cloned in pSEAP2Basic; 41.4delp, truncated *FOXF1* promoter (0 to -3.0 kb, ACD case 41.4) cloned in pSEAP2Basic; Seg1*FOXF1*p, distant upstream GLI-binding region (part of Fragment 1a) and *FOXF1* promoter cloned in pSEAP2Basic; Seg2>*FOXF1*p, Segment 2 putative enhancer and *FOXF1* promoter cloned in pSEAP2Basic; Seg2<*FOXF1*p, same enhancer in reverse orientation and *FOXF1* promoter cloned in pSEAP2Basic; vect, pGEM-T Easy vector; Seg1, GLI-binding region of Fragment 1a cloned in pGEM-T Easy; Seg2, putative enhancer of Segment 2 cloned in pGEM-T Easy. The presence of additional GLI2 constitutively expressed from pCS2Gli2 vector is denoted by + GLI2. The GLI-binding region of Fragment 1a stimulates activity of *FOXF1* promoter *in cis* (*A*) and *in trans* (*B*) settings. (*C*) Results of the 4C experiment revealing long-range physical interaction between the *FOXF1* promoter and SDR. Coordinates of the SDR region interacting with the *FOXF1* promoter (SDR-promoter interaction peaks) in HPMEC are (from top): 86,246,725-86,246,982 (0.05<FDR≤0.1) and 86,246,125-86,246,982 (0.1<FDR≤0.2). The two HPMEC panels represent results of independent experiments. No SDR-*FOXF1* promoter interaction peak was detected in lymphoblasts, which do not express *FOXF1*.

Madison et al. [25] demonstrated that the *FOXF1* promoter harbors the evolutionarily conserved (identical among mammalian, bird, fish, and amphibian species) consensus GLI binding site (chr16:86,539,377-86,539,385) that is deleted in patient 41.4 (Fig. 1). As expected, the truncated version of the *FOXF1* promoter, missing this GLI-binding site, showed weaker SEAP expression in comparison with the intact promoter, especially in the excess of GLI2 (P=0.038) (Fig. 3A). Of note, several variant GLI-binding sites are still present in the residual truncated promoter region from case 41.4, potentially explaining why it responded, although weakly, to excess of GLI2 (Fig. 3A). Placement of the 0.6 kb part of the Fragment 1a of the SDR (containing GLI-binding sites) upstream to the *FOXF1* promoter increased SEAP expression from the *FOXF1* promoter two-fold (P<0.0016). Activation of the promoter by the SDR GLI-binding sites *in trans* also increased reporter expression (P=0.027) (Fig. 3B), further suggesting that these GLI-binding sites can regulate *FOXF1* transcription from a distant location, likely through chromatin looping that juxtaposes *FOXF1* promoter and SDR. Moreover, co-transfection of HPMEC with the plasmid pCS2Gli2, constitutively expressing GLI2 protein, increased expression of the SEAP reporter from the *FOXF1* promoter in all construct combinations that we tested (Fig. 3A,B), further indicating that *FOXF1* can be regulated in lung endothelial cells by a GLI2 transcription factor. The increase of *FOXF1* transcription from -5.5 kb *FOXF1* promoter alone, in response to the extra GLI2 synthesis, most likely results from binding of GLI2 to the mentioned consensus GLI-binding site or any of several GLI-binding site variants located within 5.5 kb upstream to *FOXF1*.

Using the SEAP reporter assay, we also tested an *in silico* identified putative regulatory Segment 2, located ~ 16 kb distally to the Segment 1 (Fig. 2A). This region was indicated by ChIP-seq assay to bind CTCF insulators, enhancer-binding protein CEBPB, and several transcription factors other than GLI, including STAT3, c-Jun, p300 (N-15), JunD, KAP1, TBP, and AP-2 gamma (http://genome.ucsc.edu). In the presence or absence of the excess of GLI2, a 1.4 kb fragment (chr16:86,256,619-86,258,038) of the Segment 2 appeared to have no effect on the *SEAP* expression from the *FOXF1* promoter when subcloned next to the *FOXF1* promoter in the same orientation as the promoter, but increased the promoter activity five-fold when placed in the opposite orientation (p<0.0001) (Fig. 3A). When tested in *trans*, the 1.4 kb fragment of Segment 2 did not significantly affect the *FOXF1* promoter activity (Fig. 3B). Thus, in contrast to the GLI-binding sites of the Fragment 1a, the tested fragment of Segment 2 seems less likely to contribute to the long-range control of the *FOXF1* expression.

SDR-FOXF1 promoter chromatin looping

To determine whether there is a physical interaction between the *FOXF1* promoter and the SDR that would juxtapose distant GLI-binding sites and the *FOXF1* promoter, we performed a chromosome conformation capture-on-chip (4C) analysis. We detected interactions between the *FOXF1* promoter region and regions located upstream to it, including the 75 kb SDR (Fig. 3C). Using two independently prepared 4C libraries from HPMEC, we identified overlapping fragments (chr16:86,246,725-86,246,982 and 86,246,125-86,246,982) interacting with the *FOXF1* promoter around the *Pol*II-binding site that were not detected in the 4C library prepared from control lymphoblasts (Fig. 3C). These fragments map 12 kb distal to the Segment 1. Chromatine looping between SDR and *FOXF1* promoter could occure with a contribution of lncRNAs[2, 26-28] including any of those encoded within SDR. Moreover, CTCF binding within the SDR (Supplemental Fig. S8) also might contribute to the formation of looping architecture[29]. Since we did not detect interactions between SDR and the *FOXF1* promoter in lymphoblasts, we propose that chromatin looping between SDR and the *FOXF1* promoter (Supplemental Fig. S9) allows GLI2 to increase *FOXF1* activity specifically in lung endothelium.

Maternal origin of ACD/MPV microdeletions

We have determined parental origin for seven out of eight upstream microdeletions (patients 57.3, 60.4, 64.5, 77.3, 81.3, 95.3, and 96.3) and for one microdeletion (59.4) that included the entire *FOXF1*. Interestingly, all these microdeletions arose *de novo* on the maternal chromosome (Supplemental Table S1,S2). Six published *de novo* microdeletions that included *FOXF1* (D1, D3, D4, and D8) or mapped upstream to *FOXF1* (D9 and D10) also arose *de novo* on the maternal chromosome [9]. Moreover, in recently published unique familial case of ACD/MPV, with missense mutations in *FOXF1* affecting five subjects, the mutation was inherited from heterozygous healthy mother, in whom the mutation arose *de novo* on her paternal chromosome [20]. In aggregate, these findings strongly suggest a parental origin bias of *FOXF1*-associated CNVs ($P<0.001$, n=14).

Differential expression of FOXF1 parental alleles

To further test whether *FOXF1* is imprinted, we compared *FOXF1* expression levels from both parental chromosomes using two sets of semi-quantitative RT-PCRs with primers designed to differentiate between the two parental chromosomes. We found that whereas both primer sets amplified equal amounts of *FOXF1* from genomic DNA, their amplification from cDNA differed significantly (P=0.011, n=3) (Supplemental Fig. S10). Thus, our data indicate that *FOXF1* expression from the maternal chromosome is not equal to its expression from the paternal chromosome, further indicating that *FOXF1* is imprinted in humans albeit incompletely.

Differentially methylated CpG site within SDR

Genomic imprinting can be caused by differential methylation of CpG islands, histone modification, or interactions with lncRNAs [30]. Here, we have investigated in detail CpG methylation as one of the potential mechanisms of *FOXF1* imprinting. The *FOXF1* promoter resides within a large CpG island [31]. We have bisulfite sequenced this region of DNA from normal lung and have found that the *FOXF1* promoter is not methylated. Previous studies using immortalized normal cells and mammary organoids have also shown that the *FOXF1* promoter is not methylated [31]. Interestingly, the part of the *FOXF1* CpG island that is included in the *FOXF1* first exon is extensively methylated (Fig. 4A); however, the functional significance of this methylation remains unknown. We also analyzed the results of the methylation studies using Infinium 450K methylation arrays on 39 brain DNA samples from normal individuals (courtesy of Dr. A.L. Beaudet). One of the 13 potentially differentially methylated regions upstream to *FOXF1* encompassed three probes located within the SDR Fragment 1a in the vicinity of a cluster of GLI2-binding sites (Fig. 2B, Supplemental Fig. S5, Supplemental Table S3) identified in the ChIP-chip experiment, and functionally verified in the reporter assay. Bisulfite sequencing of the SDR 0.2 kb CpG island closest to the consensus GLI2-binding site and overlaping with several GLI-binding site variants confirmed that it is differentially methylated in DNA of normal fetal lung (Fig. 4A).

Methylation of CpG island regulates GLI-binding sites

We were interested to know whether methylation of the CpG island, including and neighboring the GLI-binding sites, interferes with the ability of these sites to regulate *FOXF1* expression. To this end, we have methylated (*in vitro*) half of the C residues of this CpG island in a 0.6 kb portion (chr16:86,232,261-86,232,908) of the Fragment 1a. We then

co-transfected the methylated construct with a *FOXF1* promoter-harboring reporter plasmid. We observed a reduction in the ability of the Fragment 1a GLI-binding sites to activate the *FOXF1* promoter, following partial methylation of the CpG island (P=0.015) (Fig. 4B). The function of the methylated GLI-binding region did not significantly increase in the presence of the excess of GLI2 following the co-transfection of HPMEC with pCS2Gli2 (Fig. 4B).

Discussion

We identified overlapping deletions, located upstream to *FOXF1* in seven unrelated patients with ACD/MPV, defining a putative distant regulatory region (SDR) for *FOXF1*. This region, located 257 kb 5` to *FOXF1*, is 75 kb in size and contains two smaller segments that are distinguished by high evolutionary conservation among land vertebrates and high 7x regulatory potential (Fig. 2). The ~ 1.5 kb part of Segment 1 (Fragment 1a) encompasses several GLI-binding sites (Supplemental Fig. S5). GLI2 is essential for lung development in humans and mice[14, 23-25] (. Using ChIP-chip assay, we have confirmed binding of GLI2 to the Fragment 1a in HPMEC (Supplemental Fig. S6). The 0.6 kb GLI2-binding portion of Fragment 1a significantly increased transcription from the *FOXF1* promoter in our reporter assay, further indicating that it may regulate *FOXF1 in vivo* (Fig. 3A,B). The increase of *FOXF1* transcription occurred not only in *cis* but also *in trans*, suggesting that this activation could occur *in vivo* through chromatin looping that juxtaposes the *FOXF1* promoter and SDR. The CTCF binding within the SDR (Supplemental Fig. S8) may contribute to the formation of this looping architecture[29]. Both *cis* and *trans* activations of the *FOXF1* promoter seem to be specific, since (1) substitution of the 0.6 kb sequence from the 1.5 kb Fragment 1a with a 1.4 kb fragment of Segment 2, harboring sites for transcription factors other than GLI, *in trans* or *in cis* in its original genomic orientation had no effect on the promoter activity, and (2) the 3 kb empty vector pGEM-T Easy did not activate the *FOXF1* promoter.

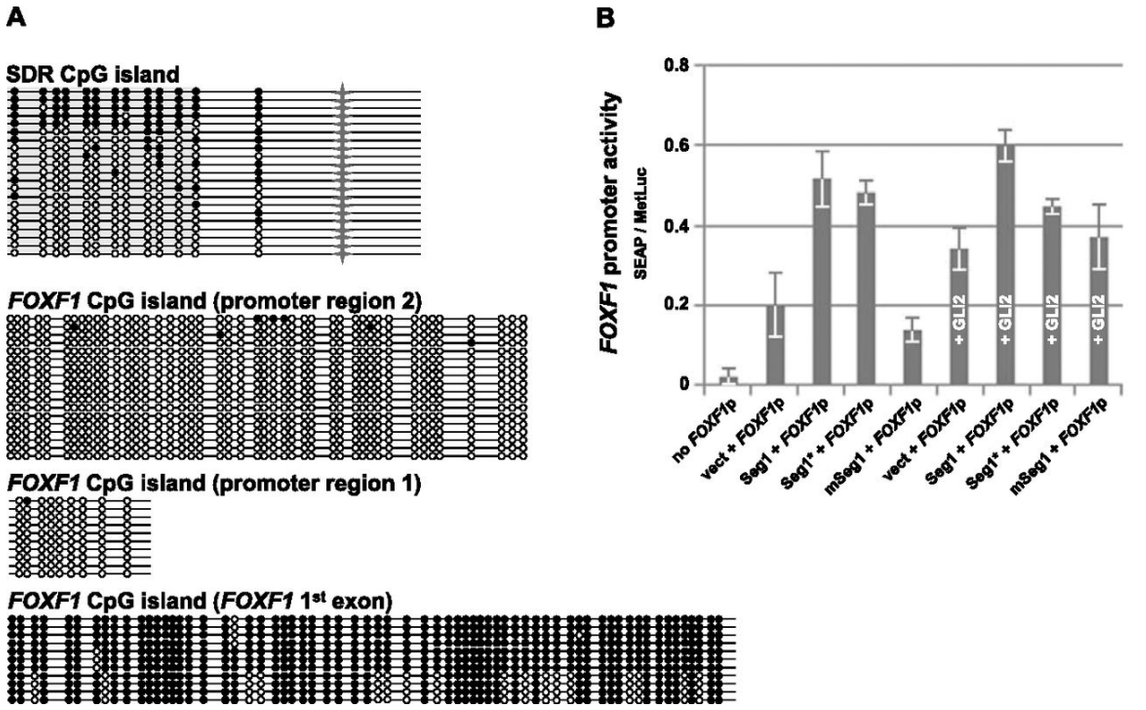

Figure. 4. Methylation of SDR CpG island decreases SDR potential to activate the *FOXF1* promoter. The presence of additional GLI2, constitutively expressed from pCS2Gli2 vector, is denoted as + GLI2. (*A*) Methylation status of the *FOXF1* promoter and SDR CpG island in normal fetal lung (similar CpG methylation pattern was observed in DNA isolated from blood). Each lane represents a separate clone. (●) Methylated CpGs, (○) unmethylated CpGs. Location of a cluster of GLI-binding site variants is higlighted in pink. Position of the consensus "core" GLI-binding site is shown in red. Genomic coordinates of the bisulfite sequenced regions are: SDR CpG island: chr16:86,232,367-86,232,979, *FOXF1* promoter 2: chr16:86,542,223-86,542,807 and promoter 1: chr16:86,543,777-86,543,907, *FOXF1* exon 1: chr16:6,544,458-86,545,037. (*B*) Methylation of the SDR CpG island overlaping with GLI-binding sites dramatically decreases the ability of this regulatory region to activate the *FOXF1* promoter. The presence of additional GLI2 constitutively expressed from pCS2Gli2 vector is denoted by + GLI2. Abbreviations: *FOXF1*p, *FOXF1* promoter cloned in pSEAP2Basic; vect, pGEM-T Easy; Seg1, GLI-binding region (Fragment 1a) cloned in pGEM-T Easy; Seg1*, mock control for mSeg1; mSeg1, GLI-binding region of Fragment 1a with *in vitro* methylated CpG island cloned in pGEM-T Easy.

We supported the chromatin looping hypothesis via 4C studies, in which we identified physical interaction of the SDR with the *FOXF1* promoter around the *Pol*II binding site (Fig. 3C). This interaction brings the distant GLI2-binding sites to the proximity of the *FOXF1* promoter. Although only one of the distant GLI-binding sites of SDR perfectly matches the consensus "core" sequence, the clustering of nine of these sites within this segment may have synergistic effects on GLI2 binding. Moreover, the variant GLI-binding sites with relatively low affinity have been recently shown to strongly induce transcription when present in native promoters [32]. Since we did not detect interactions

between SDR and the *FOXF1* promoter in lymphoblasts, where *FOXF1* is not expressed, we propose that chromatin looping between SDR and *FOXF1* allows GLI2 to increase *FOXF1* activity specifically in lung endothelium.

Importantly, we have also identified putative lncRNA genes within the SDR region (Fig. 2A). LncRNAs play a crucial role in embryonic development [33]. Moreover, lncRNAs have been proposed to play a role in imprinting and both in *cis* and *trans* regulating gene expression by acting as scaffolds for chromatin-modifying complexes and nuclear bodies, as enhancers as well as mediators of long-range chromatin interactions[2, 26, 28, 34, 35]. Thus, it seems tempting to speculate that some of the lung-expressed lncRNAs may be needed for tissue-specific interaction of the *FOXF1* promoter with the distant regulatory regions. LncRNAs can also guide chromatin-modifying complexes to specific genomic loci to regulate gene expression [2]. For instance, Khalil et al. [3] showed that a 34 kb lncRNA TCONS_00024240 (mRNA AK091834), mapping 1.7 kb upstream to *FOXF1*, binds the polycomb repressive complex PRC2. Of note, this lncRNA has the highest expression in human lungs (http://genome.ucsc.edu) (see also Supplemental Fig. S1), partially overlaps with the *FOXF1* promoter, and its locus is largely deleted in the ACD/MPV case 41.4. Our preliminary experiments on *lncFOXF1* knock-down using siRNAs showed that ~50% decrease of *lncFOXF1* expression in HPMEC resulted in a weak increase of *FOXF1* transcription (Supplemental Fig. S1), supporting its suggested function as a transcriptional repressor [3]. However, in case 41.4, in addition to *lncFOXF1* gene, also a GLI-binding consensus sequence is deleted, likely leading to a net decrease of *FOXF1* expression.

Intriguingly, all of the deletions for which we were able to determine the parental chromosome origin arose on maternal chromosome 16, strongly suggesting that *FOXF1* could be paternally imprinted ($p<0.001$, n=14). The paternal imprinting of *FOXF1* has been also predicted bioinformatically [19]. Even so, any paternal imprinting is probably incomplete, because paternal uniparental disomy of chromosome 16 (UPD16pat) has been reported in a child with only pre- and postnatal growth retardation [36]. Our data showing differential expression of *FOXF1* in the lungs supports partial imprinting hypothesis (Supplemental Fig. S10). We also reported a paternally inherited no-stop mutation (last codon) in *FOXF1* in a patient with ACD/MPV [9]. We hypothesize that the extended *FOXF1* RNA transcript might have escaped nonsense mediated decay and negatively interacted with the wild-type copy, resulting in ACD/MPV.

We have investigated in detail CpG methylation as a potential mechanism of *FOXF1* imprinting. The *FOXF1* promoter resides within a large CpG island (Lo et al. 2010), which is not methylated in normal lung tissue (Fig. 4A). Using genome-wide differential methylation array hybridization and bisulfate sequencing, we found that a small CpG island, located within the Fragment 1a of SDR and overlaping with GLI-binding sites, is differentially methylated. Using *in vitro* methylation and reporter assays, we showed that the methylation of this distant CpG island compromises the ability of its GLI-binding region to activate the *FOXF1* promoter *in vitro* (Fig. 4B). Although only two of the GLI-binding site variants include CpGs within their sequence, and thus can be directly affected by CpG methylation, it is likely that methylation of the CpG island changes the local DNA environment, and hence affinity of its GLI-binding sites for GLI2. Interestingly, methylation of CpGs has been also shown to inhibit interaction of the GLI-type zinc-finger factor YYI with its own binding sites [37].

Position effect or *cis*-ruption disorders are defined as conditions arising due to disruption of the *cis* regulatory genomic architecture of a disease gene locus [38, 39]. To date, this phenomenon has been reported for over 40 gene loci; however, the underlying mechanism remains unknown. Of interest, a 7.4 kb *cis*-regulatory deletion disrupting conserved non-coding sequences and their interaction with the promoter of another *FOX* gene, *FOXL2*, mapping over 280 kb apart, has been described as pathogenic for blepharophimosis, ptosis, and epicanthus inversus (BPES, OMIM 110100) [40]. Recently, Guttman et al. [41] successfully knocked-down genome-wide 147 (out of 226) lncRNAs using shRNAs and observed gene dysregulation both in *trans* and *cis*, with *cis* effects ranging up to 300 kb. Our findings further suggest that lncRNA-based chromatin interactions could be responsible for the position effect phenomenon in humans and potentially cause many other disorders of development.

In summary, we propose a model of long-range regulation of *FOXF1* expression in the lungs that includes CpG methylation-controlled GLI2 binding at a distant tissue-specific enhancer, whereby chromosomal looping, likely utilizing lncRNAs and CTCF, juxtapose this enhancer and the *FOXF1* promoter.

Methods

Subject recruitment

DNA samples were collected from nine unrelated probands clinically and histopathologically diagnosed with ACD/MPV. All patients died from severe pulmonary insufficiency and hypertension in the first month of life.

DNA and RNA isolation, and DNA sequencing

Peripheral blood DNA was extracted using Gentra Puregene Blood Kit (Qiagen). DNA from frozen lung tissues was extracted using DNeasy Blood & Tissue Kit (Qiagen). DNA from FFPE lung tissues was isolated following manufacturer's protocol (Agilent Technologies). RNA from cultured HPMEC was extracted using RNeasy Protect Mini Kit (Invitrogen). Lung RNA was isolated from frozen tissues using Illustra TriplePrep Kit (GE Healthcare).

PCR products and plasmid DNA were sequenced by the dye-terminator cycle sequencing method (Lone Star Labs). Before cycle sequencing, PCR products were treated with ExoSAP-IT (USB) to remove unincorporated nucleotides and primers.

DNA sequence analysis

Reference sequences were downloaded from the UCSC Genome Browser (NCBI build 37/hg19, http://.genome.ucsc.edu). DNA sequence similarities were analyzed using BLAT (http://genome.ucsc.edu). Evolutionary conservation and regulatory potential of the analyzed sequences were assessed using the UCSC "Conservation" and "7x Reg potential" tracks, respectively. The 7x regulatory potential scores are computed from alignments of human, chimpanzee, macaque, mouse, rat, dog, and cow genomes using log-ratios of transition probabilities from two variable order Markov models[42, 43]. Sequences were assembled using Sequencher v4.8 (GeneCodes). Repetitive sequences were identified using RepeatMasker (http://repeatmasker.org). GC content was determined using CpGPlot (http://ebi.ac.uk/Tools/emboss/cpgplot). CTCF-binding hot spots were identified using the University of Washington CTCF binding site track from the UCSC genome browser (ChIP-Seq determined) using a threshold value of 500. *P*-values of finding microdeletion with both breakpoints in *Alu* family repetitive elements by chance were calculated using an approach similar to the one presented by de Smith et al. [44](see also Supplemental Methods).

Screening FOXF1 for mutations

The entire coding region of *FOXF1* (two exons) and its splicing sites were amplified for each patient in partially overlapping fragments using GoTaq Flexi DNA Polymerase (Promega) in the presence of 8% DMSO. PCR conditions were: 25 cycles of incubation at 94°C for 30 sec, 56°C for 30 sec, and 72°C for 30 sec. Following purification, PCR products were directly sequenced. DNA sequences were then compared with the *FOXF1* reference sequence (NM_001451.2).

Genomic copy-number analysis

CNVs were identified by aCGH with either custom-designed 16q24.1 region-specific (1 Mb flanking *FOXF1*) 180K oligonucleotide microarrays (Agilent Technologies) or 4.2M genome-wide oligonucleotide microarrays (Roche-NimbleGen). 4x180 microarrays were scanned using Agilent DNA Microarray Scanner, and the data were analyzed using an Agilent Genomic Workbench software. 4.2M microarrays were scanned on a Roche-NimbleGen MS 200 Microarray Scanner. Scanned images of the arrays were processed using NimbleScan v2.5 (Roche NimbleGen) and analyzed using SignalMap v1.9 (Roche NimbleGen).

Amplification and sequencing of microdeletion breakpoints

PCR primers flanking each microdeletion were designed using Primer3 v0.4.0 software (Supplemental Table S4). Amplification of a 5-12 kb junction fragments for sequencing was performed using LA Taq polymerase (TaKaRa Bio USA).

Parental origin of microdeletions

Parental origin of microdeletions was determined following identification of informative SNPs in parental and patients' chromosomes. Regions containing the potentially informative SNPs were amplified by PCR and directly sequenced. Primers used in this analysis were designed based on the location of known SNPs.

ChIP-chip analysis

Chromatin immunoprecipitation with anti-GLI2 antibody was conducted in HPMEC (ScienCell Research Laboratories). Expression of *FOXF1* and *GLI2* in these cells was confirmed by RT-PCR using SuperScript One-Step RT-PCR Kit (Invitrogen) (Supplemental Fig. S7, Supplemental Table S5). The reference DNA was extracted from skin fibroblasts that do not express *FOXF1*. HPMEC were cultured to confluence in EMS medium supplemented with 10% FBS, ECGS and 1% penicillin-streptomycin (ScienCell Research

Laboratories) at 37°C with 5% CO_2. Fibroblasts were cultured in DMEM medium (Invitrogen), supplemented with 10% FBS and 1% penicillin-streptomycin, at 37°C with 10% CO_2. Because of low levels of endogenous GLI2 expression, cells were transfected, 24 hrs before the assay, with GLI2 expressing plasmid pCS2Gli2 [45](Addgene) (1µg of DNA/60 mm plate) using Lipofectamine LTX and Plus Reagent (Invitrogen). Protein-chromatin interactions were captured by incubating cells with 1% formaldehyde for 10 min at room temperature. Cells were then lysed in SDS lysis buffer and sonicated. Fragmentation of the DNA to ~ 1 kb fragments was verified by agarose gel electrophoresis. Immunoprecipitation was performed using EZ ChIP kit (Millipore) and anti-GLI2 antibody (R&D Systems). 5µl of the anti-GLI-2 antibody was added to 1 ml of the sample and the GLI2-DNA complex was immunoprecipitated overnight at 4°C. 20 µl of Normal Sheep IgG (R&D Systems) was used as a negative control. Anti-mono/di/tri methyl-histone H3 (lys4) clone AW304, rabbit monoclonal IgG (Upstate) was used as a positive control. MinElute PCR Purification Kit (Qiagen) was used to purify and concentrate DNA to final volume of 10 µl. Immunoprecipitated DNA was then amplified using GenomePlex Complete Whole Genome Amplification Kit (Sigma-Aldrich). After amplification, samples were labeled with Cyanine dUTP (Agilent Technologies) using Agilent Genomic DNA Enzymatic Labeling Kit and used for array hybridization at 65°C for 24 hrs on custom-designed 16q24.1 region-specific 4x180K oligonucleotide microarray (Agilent Technologies).

Reporter assay constructs

The -5.5 kb *FOXF1* promoter region (chr16: 86,538,679-86,544,175), directly upstream to the *FOXF1*'s AUG initiation codon, was amplified by PCR from genomic DNA isolated from the blood sample of a normal control individual. This region contained a 43 bp-long 5'-untranslated part of the *FOXF1* first exon. Primers used for the amplification were Foxp2F2: 5`CTAGCTAGCACATTTCCTCATATTCTGTGTAGAGAGCACCT3` and FoxAUG2R: 5`TTGCGCCGATTCGAACGGGTGGCTGCTG3`, and included restrictions sites for *Nhe*I and *Bst*BI, respectively. PCR was done using LA Taq DNA polymerase (TaKaRa Bio) in the presence of 6% DMSO, applying 25 cycles of 94°C for 30 sec and 68°C for 5 min. Following digestion with *Nhe*I and *Bst*BI (NEB), the PCR product was cloned between *Nhe*I and *Bst*BI sites of multiple cloning site of the promoterless SEAP reporter vector, pSEAP-Basic (Clontech), generating a plasmid p5.5FoxSEAP. The pSEAP2-Control vector, containing the constitutive SV40 early promoter with its enhancer, was used as a positive control in the assay. The plasmid p41.4Δ4, containing the truncated version of the 5.5 kb *FOXF1* promoter region (devoid of its 5' portion deleted in an

ACD/MPV case 41.4) was constructed by digesting the p5.5FoxSEAP with *Kpn*I and religating the remaining portion of the vector. A 0.6 kb portion of the Segment 1a (chr16:86,232,261-86,232,908), bearing CpG island and GLI2-binding sites, was amplified from genomic DNA of normal control individual using Taq DNA polymerase (Invitrogen) and the following primers upsdelCGliF: 5`GTGCTAGCATGGTGTAAATCACCCTGACAGTAAACACT3` and upsdelCGliR1: `GTGCTAGCGATCTATTCTGCTATCTACCCTCTGCTTTTCC3`, both encompassing the *Nhe*I site. PCR conditions were 25 cycles of 94°C for 30 sec, 58°C for 30 sec and 72°C for 1 min. This 0.6 kb fragment was then cloned both into pGEM-T Easy (Promega) by T-vector cloning to create pGliTEasy, and, following digestion with *Nhe*I, into the *Nhe*I site of p5.5FoxSEAP, upstream to the 5.5 kb *FOXF1* promoter region, generating pGli-5.5FoxSEAP. The 1.4 kb putative enhancer region of the Segment 2 (chr16:86,256,619-86,258,038) was amplified by PCR using Pfu Ultra DNA polymerase (Stratagene) with the primers upsdelEnhF: 5`AAACTTCATCTCTCCCTGCAGCTTCTCTGT3` and upsdelEnhR: 5`GTTGGACTCTGAAATCAGTGCCTTCAACAT3`. The blunt-ended PCR product was A-tailed using GoTaq Flexi DNA polymerase and cloned in pGEM-T Easy vector to generate plasmid pEnhTEasy. This new vector was then cut with *Nhe*I and the putative enhancer-containing fragment was subcloned in *Nhe*I site of p5.5FoxSEAP, generating a plasmid pEnh5.5FoxSEAP.

Cell transfection and reporter assay

The HPMEC were cultured as described above. Transient transfection of HPMEC with reporter gene constructs was done by electroporation using the NEON transfection system (Invitrogen). The electroporation parameters were as follow: pulse voltage, 1,350 V; pulse width, 30 ms; pulse number, 1; cell density, $1.0x10^7$ cells/ml. The transfection efficiency was between 20% and 30%. 0.5 μg of pSEAP2Basic (no promoter), pSEAP2Control (Clontech), pEnhTEasy, pUpstdelGliTEasy, or 1 μg of p5.5FoxSEAP, pGli5.5FoxSEAP, or pEnh5.5FoxSEAP were co-transfected with 0.1 μg of pMetLuc-Control (Clontech) (internal control for transfection efficiency) and with or without 0.4 μg of pCS2Gli2.

In an *in vitro* methylation experiment, 3 μg of linearized, methylated, or nonmethylated pUpstdelGliTEasy vector were co-transfected with p5.5FoxSEAP and pMetLucControl. One set of transfections included also pCS2Gli2. Each well of 24-well plates was seeded with $1.5x10^5$ cells. Activities of SEAP and *Metridia* luciferase (MetLuc) were measured 55 hrs later by a luminescence assays using Ready-to-Glow Dual Secreted

Reporter System (Clontech) according to manufacturer's protocol. Luminescence was recorded using microplate LUMIstar Omega luminometer (BMG Labtech, Durham, NC). SEAP activities of individual transfections were normalized for *Metridia* luciferase activities. Results are shown as the mean of three independent experiments. Statistical significance of the observed differences was estimated applying unpaired *t* test.

4C analysis

3C libraries were generated according to the protocol described by Dostie and Dekker [46] using $1x10^7$ HPMEC or lymphoblasts. Crosslinked DNA was digested with *Eco*RI overnight, and religated with T4 ligase for 4 hrs at 16°C at low DNA concentration. The 3C library was then processed according to the procedure described by Simonis et al. [47]. The ligation junctions were trimmed by digestion of 3C library with *Taq*$^{\alpha}$I overnight, followed by phenol/chloroform extraction, religation at low DNA concentration for 4 hrs at 16°C, and DNA linearization with *Hind*III at 37°C overnight. The 4C template was created by linear amplification of DNA using the Expand Long Template PCR System (Roche Applied Science). 200 ng of 4C template was amplified per reaction. Primers used for amplification were: 5`GGCAGGAAGTTTACAGGGTTTAACG3` and 5`TGTGTGTGCTAATGTGTGGACAAGA3`. The primers were designed within the *Eco*RI-*Taq*$^{\alpha}$I fragment, containing *Pol*II binding site. PCR cycling conditions were: 94°C for 30 sec, 55°C for 1 min, and 68°C for 3 min. Sixteen reactions were conducted for each template. Pooled PCR products were purified using Qiagen PCR cleanup kit (Qiagen).As a reference DNA for hybridization, we used uncrosslinked DNA isolated from HPMEC or lymphoblasts with Puregene Blood Core Kit A, and digested overnight with *Eco*RI and *Taq*$^{\alpha}$I.

Custom designed 720K microarrays covering 2 Mb regions flanking *FOXF1* were designed and produced by Roche-NimbleGen. Labeling and hybridization of 4C DNA libraries was done according to manufacturer's instructions (Roche-NimbleGen). The arrays were scanned on NimbleGen 200 Microarray Scanner, and the data were analyzed using NimbleScan v2.5 and SignalMap v1.9. The 4C peak data were generated from the scaled $\log_2$ ratio data. Peaks were detected by searching four or more probes whose signals are above the specified cutoff values, using a 500 bp sliding window. Each peak was assigned a false discovery rate (FDR) score. Scores ≤ 0.2 are indicative of an interaction site.

Bisulfite sequencing

Bisulfite modification of lung and blood DNA for sequencing was performed using EpiTect Bisulfite Kit (Qiagen). Primers for methylation PCR were designed using MethPrimer software (Li and Dahiya 2002) (http://urogene.org/methprimer/index1.html) (Supplemental Table S6). PCR was performed in 25 µl reaction mixture containing 200 ng of bisulfite-treated DNA, 0.5 nmol of each primer, and 1.25 U of HotStarTaq polymerase (Qiagen). Cycling conditions were 95°C for 1 min, 50°C to 62°C for 30 sec, and 72°C for 1 min. PCR products, ranging from 0.2 to 0.6 kb were purified from unincorporated primers and nucleotides using QIAquick PCR Purification Kit (Qiagen), and T-vector cloned in DH5α cells using pGEM-T Easy Vector System. Plasmid minipreps were prepared from 25 randomly selected transformant colonies using Perfectprep Plasmid Mini kit (Eppendorf) and sequenced using T7 and SP6 promoter universal primers.

In vitro methylation of CpG island

For the *in vitro* methylation assay, the pUpstdelGliTEasy vector was first linearized using *Sph*I restriction enzyme (NEB), and then treated with *Hpa*II methyltransferase (NEB) that methylates the internal C residue of the sequence CCGG. The 25 µl reaction mixture contained 1 µg of the DNA, 1 U *Hpa*II methylase, 50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 5 mM 2-mercaptoethanol, and 80 µM S-adenosylmethinine. Methylation was performed at 37°C for 1 hr. In the mock-methylation reaction, S-adenosylmethinine was omitted. Plasmid DNA was subsequently extracted with phenol/chlorophorm, precipitated with isopropanol and analyzed for the degree of methylation by digestion with *Hpa*II restriction nuclease (NEB). Completely methylated plasmid DNA was then used for transfections with an unmethylated DNA used as a control. The stability of methylation following transfection of the methylated construct was demonstrated recently in the similar experiments reported by Ilan and Katzav (2012) and Matousková et al. (2012).

Analysis of FOXF1 expression from individual parental chromosomes

To determine whether there is a difference in *FOXF1* expression from maternal versus paternal chromosomes, we first looked for heterozygous SNPs in *FOXF1* exons in genomic DNA extracted from normal newborn lung tissue, from which we also extracted RNA. We identified heterologous SNP G>G/A (chr16:86,547,496) within *FOXF1* exon 2. We then performed two sets of RT-PCR, using RNA with reverse primer in one setting ending in G nucleotide (FoxE2RnaRG: CAGAAAGTTTACAGTAGAGGTTGGG) and, in the other setting, in A nucleotide (FoxE2RnaRA:

CAGAAAGTTTACAGTAGAGGTTGGA). The forward primer was the same in both settings (FoxE2RnaF1: GTCTCCCTTTAGAGCCGTCTTTTG). Both pairs of primers were checked for their equal efficiency in priming PCR using genomic DNA. Semi-quantitative RT-PCR was done using SuperScript One-Step RT-PCR Kit. cDNA was synthesized at 50°C for 30 min. Temperature profile of PCR was: 19 cycles of 30 s at 94°C, 45 s at 56°C, and 45 s at 72°C, and final extension for 5 min at 72°C.

Data access

The aCGH, ChIP-chip, and 4C data sets can be accessed through the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nih.gov/geo/) under accession number GSE39258.

Acknowledgements

We thank C. Brown for the access to electroporation system, M. Withers for technical assistance, C. Dugauquier for histopathological analysis, and J. Bondaruk, A. Gambin, T. Gambin, V.V. Kalinichenko, J.R. Lupski, A. Rodriguez, and,C. Shaw-Smith for helpful discussion. Part of this work was supported by NIH grant 1RO1HL101975-01 to P. Stankiewicz, and a NORD grant and Pilot Project Award from the Texas Children's Hospital to P. Sen.

References

1. Birney, E., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.
2. Mattick, J.S., *Linc-ing Long noncoding RNAs and enhancer function.* Dev Cell, 2010. **19**(4): p. 485-6.
3. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11667-72.
4. Chang, V.W. and Y. Ho, *Structural characterization of the mouse Foxf1a gene.* Gene, 2001. **267**(2): p. 201-11.
5. Bishop, N.B., P. Stankiewicz, and R.H. Steinhorn, *Alveolar capillary dysplasia.* Am J Respir Crit Care Med, 2011. **184**(2): p. 172-9.
6. Langston, C., *Misalignment of pulmonary veins and alveolar capillary dysplasia.* Pediatr Pathol, 1991. **11**(1): p. 163-70.
7. Eulmesekian, P., et al., *Alveolar capillary dysplasia: a six-year single center experience.* J Perinat Med, 2005. **33**(4): p. 347-52.
8. Sen, P., et al., *Expanding the phenotype of alveolar capillary dysplasia (ACD).* J Pediatr, 2004. **145**(5): p. 646-51.
9. Stankiewicz, P., et al., *Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations.* Am J Hum Genet, 2009. **84**(6): p. 780-91.
10. Lai, E., et al., *Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10421-3.
11. Pierrou, S., et al., *Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending.* Embo j, 1994. **13**(20): p. 5002-12.
12. Kaestner, K.H., W. Knochel, and D.E. Martinez, *Unified nomenclature for the winged helix/forkhead transcription factors.* Genes Dev, 2000. **14**(2): p. 142-6.
13. Mahlapuu, M., et al., *The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm.* Development, 2001. **128**(2): p. 155-66.
14. Maeda, Y., V. Dave, and J.A. Whitsett, *Transcriptional control of lung morphogenesis.* Physiol Rev, 2007. **87**(1): p. 219-44.
15. Bozyk, P.D., et al., *Mesenchymal stromal cells from neonatal tracheal aspirates demonstrate a pattern of lung-specific gene expression.* Stem Cells Dev, 2011. **20**(11): p. 1995-2007.
16. Mahlapuu, M., S. Enerback, and P. Carlsson, *Haploinsufficiency of the forkhead gene Foxf1, a target for sonic hedgehog signaling, causes lung and foregut malformations.* Development, 2001. **128**(12): p. 2397-406.
17. Kalinichenko, V.V., et al., *Defects in pulmonary vasculature and perinatal lung hemorrhage in mice heterozygous null for the Forkhead Box f1 transcription factor.* Dev Biol, 2001. **235**(2): p. 489-506.
18. Kim, I.M., et al., *Functional characterization of evolutionarily conserved DNA regions in forkhead box f1 gene locus.* J Biol Chem, 2005. **280**(45): p. 37908-16.
19. Luedi, P.P., et al., *Computational and experimental identification of novel human imprinted genes.* Genome Res, 2007. **17**(12): p. 1723-30.
20. Friedrichsen, M., et al., *Carboxylesterase 1 gene duplication and mRNA expression in adipose tissue are linked to obesity and metabolic function.* PLoS One, 2013. **8**(2): p. e56861.
21. Lee, J.A., C.M. Carvalho, and J.R. Lupski, *A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders.* Cell, 2007. **131**(7): p. 1235-47.
22. Hastings, P.J., et al., *Mechanisms of change in gene copy number.* Nat Rev Genet, 2009. **10**(8): p. 551-64.
23. Rutter, M., et al., *Gli2 influences proliferation in the developing lung through regulation of cyclin expression.* Am J Respir Cell Mol Biol, 2010. **42**(5): p. 615-25.
24. Zhang, M., et al., *Expression of SHH signaling pathway components in the developing human lung.* Histochem Cell Biol, 2010. **134**(4): p. 327-35.
25. Madison, B.B., et al., *FoxF1 and FoxL1 link hedgehog signaling and the control of epithelial proliferation in the developing stomach and intestine.* J Biol Chem, 2009. **284**(9): p. 5936-44.
26. Mattick, J.S., *The genetic signatures of noncoding RNAs.* PLoS Genet, 2009. **5**(4): p. e1000459.
27. Orom, U.A., et al., *Long noncoding RNAs with enhancer-like function in human cells.* Cell, 2010. **143**(1): p. 46-58.
28. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Mol Cell, 2011. **43**(6): p. 904-14.
29. Phillips, J.E. and V.G. Corces, *CTCF: master weaver of the genome.* Cell, 2009. **137**(7): p. 1194-211.
30. Bartolomei, M.S., *Genomic imprinting: employing and avoiding epigenetic processes.* Genes Dev, 2009. **23**(18): p. 2124-33.
31. Lo, P.K., et al., *Epigenetic inactivation of the potential tumor suppressor gene FOXF1 in breast cancer.* Cancer Res, 2010. **70**(14): p. 6047-58.

32. Winklmayr, M., et al., *Non-consensus GLI binding sites in Hedgehog target gene regulation*. BMC Mol Biol, 2010. **11**: p. 2.
33. Ulitsky, I., et al., *Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution*. Cell, 2011. **147**(7): p. 1537-50.
34. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals*. Nature, 2009. **458**(7235): p. 223-7.
35. Nagano, T. and P. Fraser, *No-nonsense functions for long noncoding RNAs*. Cell, 2011. **145**(2): p. 178-81.
36. Kohlhase, J., et al., *First confirmed case with paternal uniparental disomy of chromosome 16*. Am J Med Genet, 2000. **91**(3): p. 190-1.
37. Kim, J., et al., *Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3*. Hum Mol Genet, 2003. **12**(3): p. 233-45.
38. Kleinjan, D.A. and V. van Heyningen, *Long-range control of gene expression: emerging mechanisms and disruption in disease*. Am J Hum Genet, 2005. **76**(1): p. 8-32.
39. Kleinjan, D.J. and P. Coutinho, *Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease*. Brief Funct Genomic Proteomic, 2009. **8**(4): p. 317-32.
40. D'Haene, B., et al., *Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promotor: implications for mutation screening*. PLoS Genet, 2009. **5**(6): p. e1000522.
41. Guttman, M., et al., *lincRNAs act in the circuitry controlling pluripotency and differentiation*. Nature, 2011. **477**(7364): p. 295-300.
42. Kolbe, D., et al., *Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat*. Genome Res, 2004. **14**(4): p. 700-7.
43. King, D.C., et al., *Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences*. Genome Res, 2005. **15**(8): p. 1051-60.
44. de Smith, A.J., et al., *Small deletion variants have stable breakpoints commonly associated with alu elements*. PLoS One, 2008. **3**(8): p. e3104.
45. Roessler, E., et al., *A previously unidentified amino-terminal domain regulates transcriptional activity of wild-type and disease-associated human GLI2*. Hum Mol Genet, 2005. **14**(15): p. 2181-8.
46. Dostie, J. and J. Dekker, *Mapping networks of physical interactions between genomic elements using 5C technology*. Nat Protoc, 2007. **2**(4): p. 988-1002.
47. Simonis, M., J. Kooren, and W. de Laat, *An evaluation of 3C-based methods to capture DNA interactions*. Nat Methods, 2007. **4**(11): p. 895-901.

PART *3*

Genetic studies in discordant monozygotic twins

3.1

Copy Number Variation in monozygous twins

Adapted from:

Copy Number Variation in monozygous twins

E. Brosens, K.G. Snoek, D. Veenma, H. Eussen, D. Tibboel, A. de Klein

Book chapter in "Genome-Wide Association Studies: From Polymorphism to Personalized Medicine" Cambridge University Press; *in press*

Abstract

Since the end of the 19th century twin studies have successfully been used to determine the contribution of environmental and genetic factors in traits and diseases. Several methods exist to study the genetic component or heritability of any given trait or disease. One approach is a Genome-Wide Association Study (GWAS), a method which can link human diseases and traits to specific haplotypes using Single Nucleotide Polymorphisms (SNP). In GWAS large groups of cases are compared to large groups of controls. With a similar approach, using the same genotyping platforms, traits and diseases can be associated with large DNA gains or losses.

Comparing genetics and phenotypes of concordant and discordant twins can be a powerful addition to standard association studies. Nowadays, large twin registries exist, enabling follow-up of phenotype development over time. Recent advantages in genetic technology allow for screening of almost the entire human genome. Comparing the DNA of twins is now possible up to the base pair level and allows for mapping of disease genes in cohorts of concordant and discordant monozygotic and dizygotic twins, and intra-twin comparisons in concordant and discordant monozygotic twins. The genetic contribution of a trait or disease can subsequently be inferred by comparing the genetic variance with the environmental variance; in monozygous twins any phenotypical discordance should be attributed to an environmental exposure later in life.

Caution has to be taken by the interpretation of these twin-study results since environmental and genetic components are not always exactly identical in monozygous twins. Maternal blood, including nutrients, oxygen, CO2, metabolites and stem cells can be transferred from one twin to the other via placental anastomoses. Unequal access these factors exposure to waste products can result in environmental exposure differences. Moreover, DNA changes can occur de novo at any time during human life span. These somatic changes can be either absent or present in a different frequency in monozygotic twins and even in individual tissues depending on the timing of twin separation and mutation during pregnancy. It is important to keep in mind that these somatic changes could not only explain twin discordance, but also highlight the importance of screening for genetic aberrations in different tissues.

Introduction

Micro-array technology in GWAS and Copy Number Variations

Genome-wide association studies (GWAS) are used to associate human disease and traits with a certain chromosomal locus or loci. This is done by whole genome genotyping of many cases and controls using so-called tag-SNP probes containing a relative common SNP associated with a region in linkage disequilibrium. If a certain tag-SNP has a statistical significant higher frequency in cases compared to controls, the studied trait or disease is said to be associated with that specific locus. This method has proven to be successful in many occasions [1, 2] and has increased our knowledge of disease etiology and human development significantly. Genotyping has mainly been done using micro-arrays with up to several millions allele specific oligo-probes. These micro-arrays contain information on the relative amount of DNA at a given locus. Using both genotype and quantitative information these SNP based GWAS studies can be extended to include DNA gains or losses. These segmental variations in DNA copy number can arise *de novo* or be inherited in a Mendelian matter. Many of the characteristics of other types of genomic variation are shared: they can be ancestry specific, are driven by selection pressure and can influence the expression of genes by altering their copy number or affecting gene regulatory regions.[3, 4] Losses can result in haploinsufficiency of one or several genes or truncated proteins whilst gains can increase gene expression or can also lead to altered protein structure, reduced protein levels or function.

Copy Number Variation; common, rare and de novo

Many gains and losses are rather common and most likely represent the normal population variance.[4, 5] These recurrent DNA variations have an allele frequency of over 1% and are called *common* Copy Number Polymorphisms (CNPs). CNPs account for a significant proportion of the healthy human genome.[6, 7] They often arise after non-allelic homologous recombination of misaligned DNA segments due to the presence of low copy repeats.[5] In general, common CNV are not associated with severe congenital anomalies but could influence human traits like height [8], aging[9] or age of menarche[10] and raise disease susceptibility for a limited number of complex diseases as age related macular degeneration [11], Crohns disease[12] , obesity [13], diabetes and rheumatoid arthritis[14]. If recurrent gains and losses have an allele frequency under <1%, they are called *rare* Copy Number Variations (CNVs). [15] These CNV are proposed to arise after replication errors e.g. non-homologues end joining, fork stalling and template switching or micro homology mediated break induced replication.[5] These low population frequency

CNVs can contribute to a wide range of diseases for instance in congenital heart disease.[16, 17] Interpretation of the significance of a rare CNV can sometimes be difficult since CNV can be ancestry specific.[18]

Common CNPs and rare CNVs can be used in genome wide association studies [19] and could explain some of the heritability of common diseases or traits.[20] However it has to be expected that many associated loci are likely to have been previously detected by tag-SNPs from "regular" GWAS studies.[14] In contrast, often large (>500kb) *de novo* CNV, affecting many genes and regulatory sites, are related to pathogenic conditions. For instance in congenital anomalies as Congenital Diaphragmatic Hernia [21, 22] Sotos syndrome, craniosynostosis and congenital anomalies of the kidney and urinary tract [23],tumor genesis[24] autism [25] and intellectual disability. [26] It can be difficult to pinpoint the causal gene in these large *de novo* CNVs since many sites are affected. However, overlapping CNVs in similar patients can result in a reduction of the "region of interest". It can be difficult to pinpoint the causal gene in these large *de novo* CNVs since many sites are affected. However, overlapping CNVs in similar patients can result in a reduction of the "region of interest". Summarizing, CNPs are usually associated with traits and not to severe early onset diseases. Rare inherited CNVs can be associated with traits, common and rare diseases with differences in timing and penetrance in individuals. Finally, large *de novo* CNV can be causal in both early and late onset diseases, congenital anomalies and tumorgenesis. (figure 1)

The prevalence of spontaneous multiple offspring in a single gravidity, twinning, is about 1 in 40 pregnancies, although large regional differences are present.[27] The incidence of twin pregnancies has increased in the developed world due to the increasing use of fertility treatments, increased maternal age, weight and height.[28-30] More congenital anomalies are observed in twin pregnancies. Twins can originate either from one oocyte (also called monozygous, ~30%) or from multiple oocytes (di- or polyzygotic, ~70%).[31]. The fetal placental membrane, the chorion, starts to develop around day 3-4 and the amnion forms between day 6 to 8.[27] Depending on time of separation, monozygous twins either have independent amnions, chorions and placentas or share one or more of these structures.[31] In almost all dizygotic pregnancies each embryo develops an individual amnion, chorion and placenta.

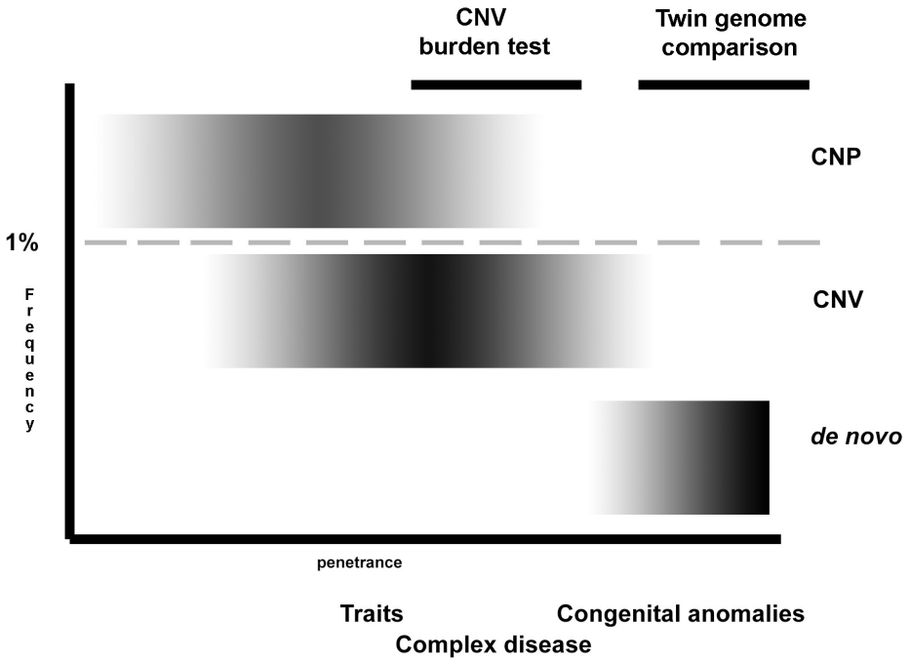

Figure 1 Copy Number Variation frequency and association to traits and their diseases. DNA gains and losses affect large parts of the otherwise diploid human genome. Variation recurring with a population frequency over 1% are called common Copy Number Polymorphisms usually have little impact on health but could influence human traits. Copy Number Variations have an allele frequency under <1 and could raise disease susceptibility for a limited number of sporadic or common complex diseases. Congenital anomalies are mostly associated with de novo CNV, although again rare CNV could contribute often in a multifactorial disease model.

There are however, rare incidences of other mechanisms.[32] Dichorionic diamnionic monozygous (DCDA) twins separate before day three, have two placentas and represent about 20-30% of monozygous twins. Monochorionic diamniotic (70-75%) twins are roughly thought to separate between day 5 and 7 after fertilization. In contrast, monochorionic monoamniotic twins (1-2%) have one placenta and separate after day 8 whilst division at or after day 13 results in conjoined twins.[27, 33] These placenta's can be joined or connected via vascular anastomoses, just as this can happen in dizygotic twins.[27] Genetic factors, in vitro fertilization and increased concentrations of follicle stimulating hormone are factors associated with dizygous twinning, whilst monozygous twinning etiology is less clear. It has been suggested that inner cell mass damage, breaks in the zona pelucida, genetic factors or skewed X-inactivation could be mechanisms resulting in monozygous twinning.[27] Ultrasonographic assessment of chorionicity is optimally performed in the first or early second trimester.[34, 35]

Classical twin design; nature vs. nurture

Using cohorts of concordant and discordant monozygotic and dizygotic twins researchers are able to determine the impact of environment and genetic factors in traits and common or complex diseases.[36] Two advantages of twin-studies are that twins share an intra-uterine environment and are perfect age-matched controls. Moreover, the genetic contribution of a trait or disease can be inferred by comparing the genetic variance with the environmental difference. Since monozygotic twins are (mostly) genetically identical any phenotypical discordance should be attributed to an environmental exposure in-utero or later in life-span. Dizygotic twins only share on average half of their DNA, i.e. like common siblings, and a greater variance between dizygotic versus monozygotic twins is supportive of an increased genetic contribution. Comparing the DNA of twins is now possible up to the base pair level and allow for the mapping of disease genes in dizygotic twins and intra-twin comparisons in monozygotic twins.[37]

	CNP-inherited events	**CNV-rare or private, inherited from affected parent**	**CNV-rare or private, inherited from unaffected parents**	**CNV-rare or private,** *de novo*
Population frequency	high	low or even absent	low or even absent	low or even absent
Disease or trait	May be associated to common late onset diseases (-susceptibility) and traits	May be associated in rare diseases if relevant genes are affected	Possible modifying or predisposing factors in rare diseases	Likely causal in rare diseases if relevant genes are affected
In Case-Control studies	Enrichment in case-control studies	Associated with disease	Almost absent in controls and enriched in cases; absolute frequency is low	Virtually absent in controls and enriched in cases; absolute frequency is low
MZ twin phenotype is concordant	Associated if parent with CNP has trait or disease (-susceptibility) is detected in case-control studies	Associated if monozygous twins are both affected	Possible modifying or predisposing factors in rare diseases	Likely causal in rare diseases if relevant genes are affected
MZ twin phenotype is discordant	CNP not associated with trait or disease (-susceptibility) and traits	Possible role in multifactorial etiology with reduced penetrance[a]	Possible role in multifactorial etiology with reduced penetrance if enriched in cases.[a]	Possible role in multifactorial etiology with reduced penetrance[a]

Table 1 Copy Number alterations in traits and disease. CNP; Copy Number Polymorphism, CNV; Copy Number Variation; [a] additional functional studies are necessary to prove or exclude causality. Partly adapted from [15]

Associations of CNV and traits or diseases

Enrichment of CNVs in classical case-control studies can result in a statistical significant association to a given phenotype. There is a huge increase in the number of publications of GWAS studies establishing associations of CNV with a specific disease or trait, e.g. in osteoporotic fractures[38], colorectal cancer[39], short stature[40], congenital heart disease [16], age of menarche [10], alcohol dependence [41] and schizophrenia.[42] In these studies CNV profiles of cases were compared with those of controls. Enrichment of phenotypical concordance in monozygotic twin cohorts compared to their dizygotic

counterparts hints to a more genetic background. Moreover, genetic discordance in discordant monozygotic twins can give strong clues about genetic background. (as depicted in table 1) The identification of specific CNV more prevalent in affected monozygous twins and their virtual absence in unaffected dizygous twins could identify causal CNV in traits or diseases. CNV-twin studies can be either additive to "regular" GWAS or "classical" twin studies, or used as an independent study protocol. Therefore, monozygotic twins, either concordant or discordant for a given phenotype or trait can be very helpful in these CNV association studies.

Similar as in these studies, enrichment of phenotypical concordance in monozygotic twin cohorts compared to their dizygotic counterparts hints to a more genetic background. Moreover, genetic discordance in phenotypical discordant monozygotic twins can give strong clues about genetic background. (table 1) Specific CNV more prevalent in affected monozygous twins and their virtual absence in unaffected dizygous twins could identify causal CNV in traits or diseases. CNV-twin studies can be either additive to "regular" GWAS or "classical" twin studies, or used as an independent study protocol. Therefore, monozygotic twins, either concordant or discordant for a given phenotype or trait, can be very helpful in these CNV association studies.

Surprisingly, CNV-association studies in twins are even rarer than regular CNV-GWAS studies. Few studies focus on single genes or locus e.g. *CES1* gene duplication with susceptibility to type 2 diabetes.[43] In this study mRNA expression levels and *CES1* gene copy number of 295 monozygous and 170 dizygous twins discordant for obesity were compared. A positive association was found between *CES1* mRNA levels and body mass index, insulin resistance, insulin levels, fasting glucose levels and triglyceride levels. Moreover, gene duplication was associated with insulin sensitivity, glucose tolerance and negatively associated with insulin resistance. Interestingly, *CES1* mRNA levels were not associated with *CES1* Copy Number; most likely other factors also contribute to the associations found in this study.

In a Genome-wide CNV association study [44], the copy number profiles of 20 sporadic patients with type 1 diabetes was compared with 20 healthy controls. The resulting 39 CNVs were subsequent validated in a discordant monozygous twin pair cohort (n=10). Five of these CNV, enriched in the twin cohort, pointed to susceptibility loci for type 1 diabetes. Nag and co-workers studied intra-ocular pressure in 992 twins enlisted in the UK twin registry and replicated their findings in two independent twin cohorts (n=467 and

n=1620). They found an association with a CNV on chromosome 5q21.[45] With a CNV burden test associations were obtained between specific CNV and attention deficit disorder. Although Ehli and co-workers detected two *de novo* CNV; these were concordant in the monozygous twins.[46] However, CNV burden in patients was significantly higher and the affected individuals had larger, more often rare and *de novo*, CNVs compared to unaffected individuals. These rare CNVs could contribute to the attention deficit disorder. This is in line with the "common disease, rare variant" hypothesis or rare allele model, in which recently emerged (in the population) rare alleles have a large effect on disease etiology. In this model penetrance may be influenced by other factors.[47-49]

Identical twins; exceptions to the dogma

There are some exceptions to consider before interpreting twin study results. First of all, the intra-uterine environmental conditions are not always equal. For example, twin-to-twin transfusion syndrome (TTTS), one of the most serious complications of monochorionic multiple gestations, can occur. TTTS develops as discordant loss of arteriovenous anastomoses resulting in asymmetrical flow resistance.[50] In monozygotic twins often the placenta is shared. Blood including nutrients, oxygen, CO2, metabolites and stem cells can be transferred from one twin to the other via placental anastomoses due to increasing hydrostatic and osmotic forces; the resulting disproportional blood supply causes developmental delay and decreased urinary output in the "donor" twin (and oligohydramnion) and increased urinary output and resulting polyhydramnios in the "recipient" twin. The increased blood flow can result in cardiac anomalies in the later twin. The unequal access to maternal blood, transfusion and waste of nutrients, and oligohydramnios and polyhydramnios result in differences in environmental exposure. TTTS develops in 10 to 15 percent of monochorionic twins and is associated with several congenital anomalies. For instance, TTTS and undiagnosed loss of a twin fetus in early stage of pregnancy could be a substantial factor contributing to congenital anomalies as VACTERL association or gastro-intestinal atresia.[33, 51]

Second, DNA changes, mutations or Copy Number Variations, can arise *de novo* at any time during the human life span. Somatic DNA variations, occurring early in development or later in life can be either absent or present in a different frequency in monozygotic twins depending on the timing of egg division and mutation. (figure 2) Clear examples are those CNV that are involved in tumorgenesis [52] although these DNA changes can occur much earlier and be responsible for less dramatic somatic differences between newborns.[53]

It has even been proposed that DNA mutations cause cells in the blastocyst to recognize each other as foreign resulting in separation of these cells and causing the twinning process.[27]

Current twin study design and discordant twin studies

Improvements in genetic and epidemiological insights enable researchers to go beyond the classical twin design. The genomes of parents, mono- and dizygous twins are characterized and very large twin registries enable us to follow the phenotype of interest over time. Several experimental set-ups are now possible. For instance, the DNA of phenotypical discordant monozygotic twins can now be compared and any discrepancy could point to disease causing mutations. Subsequent screening of large cohorts of cases and controls could confirm the impact of this mutation on disease etiology. In a classical twin design the concordance rate of monozygotic and dizygotic twins is compared. Another approach could be to compare CNV burden of affected concordant monozygous twins and compare this to cohorts of affected concordant dizygotic twins. (figure 2) If a specific type of CNV is more present in the monozygous cohort it is likely to be associated with the trait or disease studied. An essential aspect of twin studies, which is independent of study design, is the need for proper zygosity determination.

Experimental procedures

Determination of zygosity

Zygosity can be determined in a number of ways, with variable accuracy in different study populations e.g. blood group, gender, chorionicity, dermatoglyphics and phenotypical appearances. Selection of same-sex pairs combined with questionnaires could be accurate enough in adult twin studies. Although relatively reliably in adults, zygosity typing based on gender, chronicity and appearances alone is not so reliable in newborns. For instance, Mono-chorionicity does not always mean that twins are monozygous [54], same-sex pairs are not always monozygous [55] nor does having an opposite gender confirms dizygosity.[56] DNA based test give an almost hundred percent certainty e.g. comparing the length of several polymorphic tetra-nucleotide short tandem repeat loci, using several SNPs (not in linkage disequilibrium) or genome-wide SNP-arrays.[52, 55, 57]

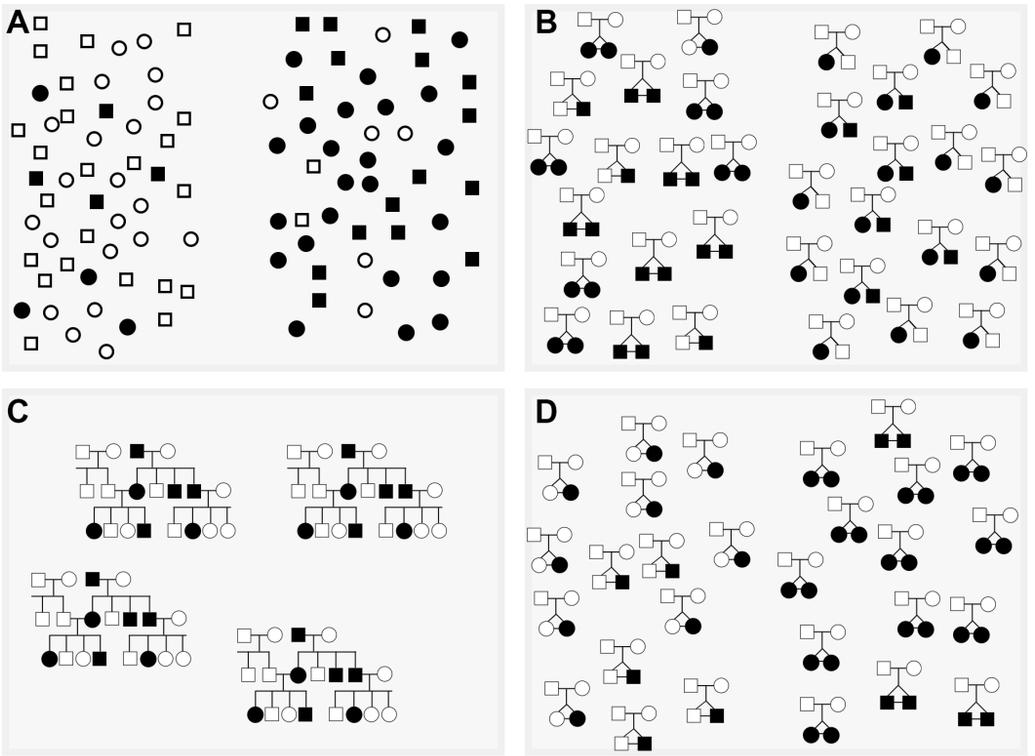

Figure 2 Examples of experimental set-ups to determine the genetic basis or contribution of a disease or trait A. With Genome Wide Association Studies in sporadic cases enrichment of a DNA variant in cases (right) and a significant over-representation of a trait or disease in these cases (■ and ●) compared to controls (left, □ and ○) points to a genetic contribution of a specific locus. **B** If the concordance rate in monozygotic twins is higher compared to dizygous twins it is likely that genetic factors contribute to the trait or disease. **C** A trait or disease can be hereditary. Pedigree and linkage analysis can pinpoint inheritance pattern and locus of the contributing or causal genetic factor. **D** Comparing the genomes of monozygotic twins can aid in finding the causal genetic factor of a trait or disease. In discordant monozygous twins (left) finding a discordant genetic variant would provide first evidence of this causal factor. Finding *de novo* variants in concordant monozygous twin with healthy parents could also provide this evidence. Finding variants that disturb protein function, structure or expression level in several independent families or twin-parent combinations strengthens this evidence for a causal gene; subsequent functional studies would prove causality.

Copy Number Variation profiling

Determining the genome wide segmental copy number can be done with DNA-array using Single Nucleotide Polymorphism (SNP) based probes or by using oligonucleotides in a comparative genomic hybridization (CGH) experiment. SNP based micro-array has the advantage that they rely on two types of information; allelic and segmental copy number state. Therefore, these kinds of arrays are less influenced by DNA input variations. Most laboratories use SNP-array for genotyping and GWAS, homozygosity mapping and CNV analysis. Using these types of arrays there is also no need for zygosity testing in twin births

since zygosity and gender can easily be determined within the genotyping analysis. CNV size and number are dependent on platform used, amount and spacing of probes, analysis settings and CNV calling algorithms and software. [58-60] Throughout the years the resolution to distinguish copy number variation from background noise has increased significantly: current micro-array technologies allow detecting of genomic imbalances down to only a few Kb in size.

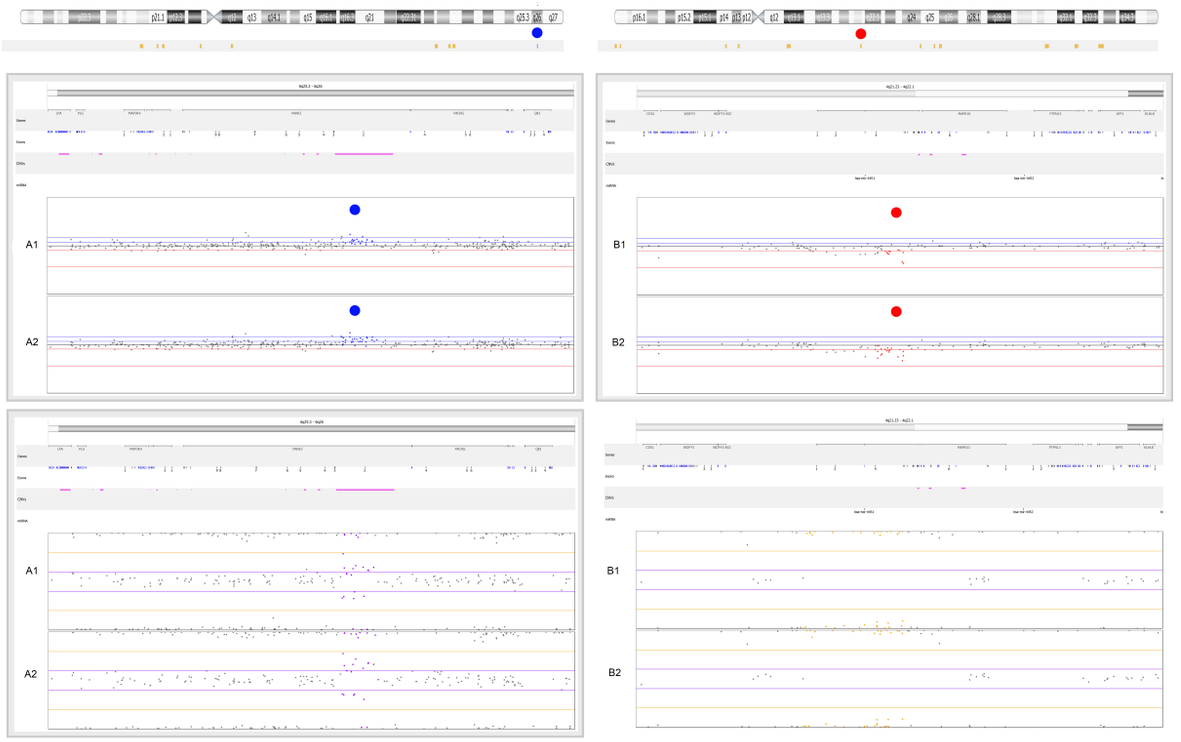

Figure 3 Two examples of concordant copy number variations in discordant twin-pairs. **A1-A2** On the left the SNP array results, visualized with Nexus Copy Number v7 (Biodiscovery inc., Hawthorne,CA,USA), of a monozygotic twin pair discordant for esophageal atresia. Both the affected (upper panel A1 and lower panel A1) and healthy twin (upper panel A2 and lower panel A2) have a 201kb gain affecting exon 2 of the *PARK2* gene. In the upper panel the $\log_2$R ratio, representing the copy number status, shows a raise (blue dots) of several probes. This raise is accompanied by a shift in allelic frequency (B-allele frequency, lower panel) indicative for the over-representation of one allele over the other. The combination of raise in $\log_2$R ratio and shift in B-allele frequency of several allele specific SNP-probes is indicative for a gain in DNA quantity of that specific locus. Both twins have exactly the same gain and this gain was inherited maternally, which was confirmed with qPCR. **B1-B2** On the right the SNP array results of a monozygous twin pair discordant for Congenital Diaphragmatic Hernia. The affected twin (upper panel B1 and lower panel B1) and unaffected sib (upper panel B2 and lower panel B2) both have a 474kb deletion deleting several exons of the *ARHGAP24* gene. The drop in $\log_2$R ratio (upper panel, red dots) and loss of one allele type (yellow dots, lower panel) are indicative of a loss of one copy of DNA at that locus. qPCR confirmed that both twins had the same deletion and that this deletion was inherited maternally. [55]

Confirmation of results with a second technique

It is always advisable to confirm Copy Number events of interest with a second technique, preferably as dissimilar to the initial technique as possible. For instance, using Fluorescence in situ hybridization (FISH) and counting signals on metaphase slides, melting curve analysis or quantitative PCR. Ideally, the effect of the discordant chromosomal aberration is measured e.g. with RNA or protein expression based studies in relevant tissues. The need for validation has several reasons: 1) determination of CNV using micro-array is a relative new technique, each platform has its own sensitivity and specificity. 2) With micro-array the *relative* copy number is determined i.e. compared to one or more reference samples or to a large reference cohort. 3) Moreover, it determines the average amount of DNA at each segment, gains and losses at the same locus result in a diploid copy number and 4) micro-array gives limited information on the location or orientation of a given CNV. 5) Also technical variation can influence experimental outcome e.g. micro-array relies on amplification of DNA and measurements using (fluorescence) intensity.

Considerations in monozygous twins

In general, monozygous twins with a concordant pathogenic genetic aberration have a concordant phenotype. The presence of a newly associated pathogenic CNV in both twins strengthens the implication of the CNV pathogenicity. For instance, two monozygous twin boys with autism both had the newly identified pathogenic 16p11 micro duplication.[61] Surprisingly, monozygous twins with concordant chromosomal anomalies, pathogenic CNV or mutations can also have a discordant phenotype. Even monozygous twins with known genetic syndromes can have a highly variable phenotype.[62] For instance the phenotypical spectrum of the 22q11.2 deletion syndrome [63, 64] is extremely variable. One of the proposed explanations for the discordance is an undetected difference in deletion size.[65] Others could be epigenetic differences, differences in blood perfusion, somatic changes, single nucleotide changes and other genetic abnormalities below the detection threshold on the other allele. Phenotypical discordant monozygous twin studies are an excellent model for studying the phenotypical spectrum of known genetic diseases.[62] Critical to consider in some, often late onset, diseases is that initial phenotypical discordant twins can become concordant e.g. in familial amyloidotic polyneuropathy type 1 [66], schizophrenia [67] or spinocerebellar ataxia.[67] Various factors could contribute to these age of onset differences. For instance, environmental components were suspected to modulate the penetrance of a concordant repeat expansion in a monozygotic twin pair with Huntington Disease.[68]

Timing matters; twin separation and mutation

Phenotypical discordance between monozygotic twins at birth can have several reasons i.e. differences in epigenetic modifications, environmental exposure, stochastic factors but also genetic differences. Here we would like to discuss these DNA differences. This sounds a bit counter-intuitive since by definition the DNA of monozygous twins is the same as they originate from a single zygote. However, there is increasing evidence that there are exceptions and several studies have undertaken the endeavor of comparing their genomes. The rationale behind these types of studies is that discordant somatic mutations arisen during development or lifespan are responsible for the phenotypical discordance. These discordant twin studies are performed at different resolutions. Karyotyping (see Table 2) is able to detect numerical and large structural chromosomal anomalies. CNV analysis (see Table 3) is able to detect DNA gains and losses around the gene level resolution. Finally, whole genome Next Generation Sequencing (NGS) which can not only measure SNPs and small DNA gains and losses called Insertion/ Deletions (InDel) but also CNV at the exon level, inversions and translocations. Moreover, this genome wide screening is not dependent on platform design as is the case with array technology. Here, we will discuss two types of copy number discordance; hetero-karyotypical discordance and Copy Number Variation.

Mutations or non-disjunction during meiosis or early mitosis can lead to either complete concordant mutated genomic DNA or, via unequal distribution and sub-sequent rescue mechanisms, to differences in somatic mosaicism in monozygotic twins. Post-zygotic DNA mutations will lead to somatic mosaicism if the cells survive the damage and the mutations is not repaired. These mutations can either be single base-pair changes, deletions and insertions of several bases or CNVs ranging from dozens of base-pairs to gains and losses of several megabases in length. Depending on the developmental stage and cell types affected, CNV can be present throughout the body, or be present at certain frequencies in specific tissues. If germ cells are affected, the mutation can be transmitted to subsequent generations. Differences in timing of separation and mutation result in differences in genetic manifestations; depending on the number of cells with and without the mutation and their distribution across the twins, tissue specific somatic mosaicism can vary accordingly. Therefore, it is advisable to test different tissues for the presence of a genetic aberration e.g. to search for placental mosaicism or the presence of low-grade mosaicism in dermal fibroblasts.

A source of false positive results when testing for genetic defects in blood or saliva immediately after birth are those differences caused by transfer of lymphocytes via twin-to-twin transfusion. After several weeks each twin has their "own" lymphocytes in its circulation, although transfer of hematopoietic stem cells can't be excluded.[69]

Germline non-disjunction, aneuploidy and somatic mosaicsm

For example, karyotyping of cultured cells derived after amniocentesis of two monochorionic diamniotic twin boys revealed discordance for trisomy 18. Both twins had trisomy 18 mosaicism of blood and buccal smear samples after birth. However, the placenta turned out to be mosaic for trisomy 18, indicative for a mosaicism in the early zygote. Further testing revealed only a low grade (5%) mosaicism in skin fibroblasts in one of the sibs.[70] Therefore, most likely non-disjunction during meiosis II followed by an unequal distribution of normal /affected cells to the twin sibs led to this discordant karyotype. Analysis of the placenta of MCDA twins, in which one had a non-mosaic deletion of the short arm of chromosome X and the other twin had non-mosaic Turner syndrome, revealed a 45,X/46,X, del(X)(p11.1) karyotype in the placenta. [71] Loss of the p-arm of the X-chromosome has led to complete loss of the aberrant X chromosome during early mitotic events.

Aneuploidy prior to conception, i.e. in the germ cells, can result in hetero-karyotypical discordance between monozygotic twins. During meiosis or early post-zygotic mitosis, non-disjunction of sister-chromatids (meiosis II or mitosis) or homologues chromosomes (meiosis I) can lead to an unequal distribution of chromosomes; the resulting cells have either 0 or 2 copies (meiosis I and II nondisjunction) or 1 and 3 copies (mitotic non-disjunction error). Depending on timing of the twin separation and division of cell lineages the monozygous twins can be complete discordant for the aneuploidy or have a certain discordant somatic mosaicism. For example, karyotypes from blood derived of a MCDA twin pregnancy hinted at mosaicism for trisomy 21 in both sibs. However, cheek and skin biopsies revealed two distinct karyotypes; 46, XY in the healthy sib and 47, XY, +21 in the affected sib.[72] This a clear example of the effect of TTTS responsible for the admixture of 47,XY,+21 cells of the affected to the healthy twin. Also, since both twins were mosaic in blood, there was blood flow from the healthy twin to the affected twin effectively diluting the 47,XY,+21 cells with 46,XY cells. Since the trisomy was present in skin and blood, non-disjunction during meiosis is a plausible cause of this discordant aneuploidy.

Genetic abnormality	location	Tested for monozygosity	Technique	Confirmation method	Discordant phenotype	Ref.
Trisomy	13	Short tandem repeat profiling with 7 markers and one gender marker	Quantitative PCR	karyotyping	NA; radiofrequency ablation of affected twin at 18 weeks of gestation	[73]
Trisomy	18	Chorionicity (MCDA) and Short tandem repeat	Karyotyping after amniocentesis of both sacs	Karyotyping of fibroblasts and chorionic villi taken from different sites of placenta	MCA, 47,XY,+18 versus healthy twin with low grade (5%) +18 mosaicism in fibroblasts	[70]
Trisomy	21	Chorionicity and gender	MSS in 1th trimester and nuchal translucency scan	Karyotyping in one case; others not described	21 cases discordant for Down syndrome	[72]
Trisomy	21	Chorionicity and short tandem repeat profiling with 10 markers	Karyotyping amniotic fluid cells	Short tandem repeat with 7 STR markers on Chr. 21	Placenta was mosaic trisomy 21; 46,XY sib spontaneous aborted. Termination of pregnancy of affected fetus. Twins discordant for Down syndrome	[74]
Monosomy	X	Chorionicity and short tandem repeat profiling with 11 markers	Karyotyping amniotic fluid cells	Affected twin 45,X [24] and healthy twin low grade mosaic 45,X [1]/46, XY [29] in fibroblast cultures	Termination of pregnancy; discordant for Turner syndrome ; placenta was mosaic 45,X/46,XY	[74]
Monosomy	X	Chorionicity	Karyotyping and Quantitative PCR	Fluorescence in Situ Hybridization; karyotyping of fibroblasts	45,X in one and 46,X, del(X)(p11.1)	[71]
Mosaic XY	X, Y	SNP-array	SNP-array	Fluorescence in Situ Hybridization	45,X/46,XX in one and 45,X/46,XX, 46, XY in the other twin	[56]

Table 2. Examples of hetero-karyotypic twinning in discordant monozygotic twins. MSS; Maternal serum screening: β-Human Chorionic Gonadotropin (β-HCH) and pregnancy associated plasma protein A levels in first trimester screening and β-HCH , unconjugated oestriol (μE3, Alpha FetoProtein (AFP) and Inhibin-A levels in the second trimester screening

Genetic discordance; somatic mosaicism and de novo CNV

Another source of genetic discordance are *de novo* CNVs. Such CNV could arise during meiosis in the parental germ cells, during early zygotic mitotic divisions or later in life. Therefore, it is recommendable to compare the CNV profiles of twins to that of their parents and siblings in addition to the comparison with the traditional cohorts of ancestry matched affected and non-affected individuals. The *de novo* CNV rate per generation is much higher than the mutation rate; estimates for CNV are 100-10.000 times higher than the single base-pair mutation rate per generation i.e. ~ 2 x 10^{-8} versus ~1-2 10^{-6} to 10^{-4}. [4] Ehli et al. describe 26 *de novo* CNV, measured with SNP-micro-array, in 25 attention deficient concordant and discordant monozygous twin pairs.[46] Some of these *de novo* CNVs were recurrent; there were 18 unique targets for a validation experiment.

Quantitative PCR primers could be developed on 11 of these target regions. Two out of 17 tested *de novo* CNVs could be validated with qPCR and one could not be proven nor excluded i.e. 13 seemingly *de novo* CNV were in fact inherited from one of the parents. One of the *de novo* CNV that was present was duplication in chromosomal band 15q11. This duplication was present in a male concordant unaffected twin-pair and was a pre-twinning event. The other, a post-twinning, *de novo* CNV was a large 1.3 Mb deletion located on band 4q35 and only observed in the oldest sib of a male concordant affected twin pair. The genotype-phenotype correlation in both instances is difficult since both unaffected twins had a *de novo* duplication and only one of the affected twins had a *de novo* deletion. Somatic mosaicism could not be excluded nor confirmed since only buccal swab derived DNA was available.

Congenital anomalies are a major contributor to prenatal and postnatal mortality and childhood morbidity. Although some of these anomalies can be completely explained by chromosomal anomalies, other genetic defects, teratogens or malnutrition, the majority of patients still have an unknown disease etiology. Several explanations have been put forward for these unknown etiologies e.g. yet undiscovered genetic or epigenetic changes, environmental or nutritional factors, disturbances in biological processes, mechanical explanations or otherwise and the lack of systematic data collection of the affected child and the parents. Twin studies can help in assessing the contribution of environmental and genetic factors in many of these congenital anomalies.

Several studies specifically searched for differences in DNA copy number between discordant monozygous twins with congenital anomalies. Breckpot *et al* compared the copy number profiles of six monozygous twin pairs discordant for congenital heart defects. In female twin pair 1 a 186 Kb large duplication was identified in the affected sib. In male pair 3 two small 2kb losses were detected at chromosomal loci 1p33 and 6p25 in the unaffected sib (all false positive differences) and in a male twin pair 6 three confirmed and three false positive CNVs were detected in the affected male sib; the 4kb gain on chromosomal 12p13, a gain on Xp11.23 and a gain on Xq28 were confirmed with qPCR. [75] In a monozygous twin pair discordant for cleft lip and palate and growth restriction Array-CGH on DNA derived from uncultured amniocytes revealed a terminal deletion of chromosome 18q in the affected sib of a MCDA female twin-pair. Subsequent FISH confirmation on cultured cells confirmed a non-mosaic 46, XX, del(18)(q21qter) karyotype in the affected twin while all cells from the healthy sib were normal. However, fetal blood, skin, tendons and lung biopsies were mosaic for the chromosome 18q deletion in varying frequencies in both twins.[53]

Both fetal blood samples had a 25% mosaicism whilst skin, tendon and lung were low grade mosaic in one twin and high grade mosaic in the other. Unequal distribution of affected cells prior to twin separation had resulted in tissue specific differences in mosaic frequency. Twin-to-twin transfusion resulted in mixture of normal and affected cells in both twins; it diluted the 18q21 deletion with normal cells in one and transferred affected cells to the other twin.

An intriguing example of somatic copy number variation mosaicism was published recently. Rio and coworkers [76] describe two DCDA monozygous twin sisters in which one girl has facial dysmorphisms, delay in fine motor skills, language skills and mental development. Her sister did not have facial dysmorphisms and only has delayed speech and unilateral single palmar crease. In the more severely affected twin a *de novo* non-mosaic deletion in chromosomal band 2p23 was detected with array-CGH and confirmed with locus specific FISH while in the mildly affected twin normal cells, 2p25 deletion and 2p25 duplication cells were detected using FISH only. Array-CGH in the latter was normal since it is an average of all cells, the 2p25 duplicated balanced the signal of the 2p25 deleted cells to a seemingly 2n state. Post zygotic non-allelic recombination has led to these three distinct cell populations; normal 46,XX cells, 46,XX dup(2)(p25.3) and 46,XX, del(2)(p25.3) cells. This post zygotic event had to be prior to separation making the drastic differential distribution of three cell lineages possible. Although several studies report on CNV differences in monozygotic twins discordant for specific congenital anomalies [53, 75, 76] others have not found these differences. [55, 77, 78] In our study the DNA of phenotypical discordant twins for esophageal atresia (seven pairs) and congenital diaphragmatic hernia (four pairs) were compared using SNP-array. Genotyping differences were validated with Sanger sequencing and CNVs confirmed with qPCR. No SNP nor (mosaic) CNV differences were present in these twins.[55] Several concordant CNPs and CNVs were observed. Although no shared rare inherited CNV were observed in this limited number of patients, few inherited CNV could be modulators or predisposing factors which, in combination with other factors, could have relevance to the disease phenotype. For instance the duplication observed in both sibs of an esophageal atresia discordant monozygotic twin pair affects the PARK2 gene, this gene is also duplicated in patients with low-grade dysplasia of the esophagus.[79] The deletion of (part of) the ARHGAP24 gene in both the affected and unaffected members of a monozygous twin pair discordant for congenital diaphragmatic hernia by disturbing normal angiogenesis.[55] Baudisch and coworkers performed a similar experiment in four monozygous twins discordant for urorectal malformations. They also did not find CNV differences between twins nor did they find *de novo* CNV in both twins.[77]

Genetic abnormality	Location	Tested for monozygosity	Technique	Confirmation method	Discordance	Reference
CNV difference detected						
CNV	12p13; Xp11 and Xq28	Same sex pairs; determination of zygosity not described	Array-CGH	qPCR	CHD discordance	[75]
Mosaic CNV (mosaic; deletion)	11q and 4q	SNP-test	BAC and SNP-array	Melting curve analysis and PCR based array	CLL discordance	[52] [a]
CNV	2q	SNP-array	BAC and SNP-array	Melting curve analysis and PCR based array	Both sibs were healthy	[52] [a]
Mosaic CNV (mosaic; deletion)	18q21.2qter	STR-profiling with 8 markers	Array-CGH	FISH	Discordant for Cleft lip and palate phenotype; 18q21-qter mosaicism in both twins with different frequencies	[53]
CNV (mosaic; deletion)	2p25.3	STR-profiling with 25markers	Array-CGH	FISH	Mosaicism; del(2)(p25.3)/dup92)(p25.3) and normal cells (1:1:1) in mildly affected sib; only deleted cells in severely affected sib	[76]
CNV	multiple	SNP-array	SNP-array	qPCR*	Discordant for Schizophrenia	[80]
CNV	multiple	SNP-array	SNP-array	-[b]	Discordant for amyotrophic lateral sclerosis	[81]
CNV (loss-*de novo*)	19p13.3	SNP-array	Array-CGH	Custom high density CGH-array	Discordant for Multiple System Atrophy	[82]
CNV (loss-*de novo*)	22q11	SNP-array?	SNP-array; FISH	-	Discordant 22q11 phenotype deletion size	[65]
no CNV difference detected						
CNV	NA	STR-profiling with 16 markers and SNP-array	SNP-array	qPCR	Discordant for urorectal malformations	[77]
CNV	NA	-	BAC -array	-[b]	Discordant for cono-truncal malformations	[83]
CNV	NA	SNP-test	Array-CGH	qPCR	Discordant for Schizophrenia or bipolar disorder	[84]
CNV	22q11	Red cell antigens; STR-profiling	FISH	-	Twin A; anal atresia, tetralogy of Fallot, pulmonary atresia and characteristic facial features; twin B only characteristic facial features	[63]
CNV	NA	SNP-genotyping	SNP and CGH-array	DNA-sequencing	Discordant for cleft lip and/or palate	[78]
CNV	NA	SNP-array	SNP-array	qPCR, DNA-sequencing	Discordant for schizophrenia	[85]
CNV	NA	STR-profiling with 16 markers and SNP-array	SNP-array; NGS-Exome CN	qPCR, FISH	Discordant for CDH or EA/TEF	[55]

Table 3. Examples of Copy Number variation measurements in discordant monozygotic twins. CNV; Copy Number Variation, STR; Short Tandem Repeat, CGH; Comparative Genomic Hybridization;, FISH; Fluorescence In Situ Hybridization, CDH; Congenital Diaphragmatic Hernia, EA; Esophageal Atresia, CHD; Congenital Heart Defect, SNP; Single Nucleotide Polymorphism. [a] Bruder at al. [52] describes several more somatic differences detected with both 32K BAC-array and 300K SNP-array. However, these were not described to be validated with a second technique. [b]no validation experiments addressed in article

Post twinning mutation and tissue specific somatic mosaicism

Somatic changes can be responsible for slight genomic differences between monozygotic twins but also result in differences within different body tissues of the same individual. For instance, Liang *et al.* analyzed the genomes of clones from mouse embryonic stem cells and measured differences in copy number at several loci in those clones. These CNV arose during the mitotic cell divisions. If similar events take place in human stem cells we could have many somatic CNV mosaicism in our tissues.[86] There can be accumulation of somatic CNV in the adult with an increased incidence over time. In a large cohort of monozygotic twins, Forsberg could detect mosaicism for megabase size CNV in 3.4% of individuals of 60 years and older whilst these were absent in the younger study cohort (33-55 years).[87] Five of these, monozygous twin pairs the CNV profiles were determined over time; a 5q deletion increased in frequency in female MZ sib with dysthymic disorder, cholelithiasis and cerebral infarction. In other, unrelated, patients with several forms of malignancies specific CNV accumulation was present. In fact, in two of these patients the CNV were indicative of, but at that time not yet diagnosed, myelodysplastic syndrome. In an apparently healthy individual Piotrowski and co-workers describe somatic CNV differences between cardiac and cerebellar cortex tissue derived DNA.[88] In fact, human neuronal cells can differ greatly in DNA content [89] and discordant CNVs are measured in single cell assays.[90] Accumulation of these genomic imbalances could be an underlying mechanism of late onset neuropathies.[89, 90] Perhaps it is advisable to screen individuals not only several times during their life, but also to evaluate multiple tissues.

CNVs that arise after the separation of the monozygous twins are completely discordant between those twins. After zygosity testing, Bruder *et al* compared the genome of nine Parkinson or Parkinson-like discordant monozygotic twins and ten healthy monozygotic twins.[52] They found (and confirmed with additional techniques) three somatic intra-twin differences. In an 80 year old male twin pair, in the sib with Parkinson, an approximately 20% mosaicism was detected for two large deletions; one chromosome 11q and the other on chromosome 4q. This aberration is often seen in patients with chronic lymphocytic leukemia and this Parkinson patient was previously diagnosed with this condition. In addition to these tumorgenesis related CNVs another discordant CNV was detected in a concordant healthy 60 year old female monozygous twin pair. This CNV, a 70%-80% mosaic deletion of chromosome 2p23 was only present in one of the sisters and was confirmed with high-resolution melting curve analysis and pyro-sequencing

Sasaki compared the CNV profile of a 67 year old monozygotic twin pair.[82] The affected sib was diagnosed with Multiple System Atrophy (MSA) at the age of 57 years; atrophy of the brainstem and cerebellum was confirmed by MRI. His brother did not have MSA symptoms. Three discordant CNVs, present in the affected twin only, were identified with array-CGH: loss in chromosomal bands 2p25, loss on 4q35 and a loss on 19p13. Two independent array experiments (SNP-array and a custom CGH-array) confirmed the presence of the 19p13 loss in 10 additional MSA patients. This loss was absent in 125 controls. These authors used successfully a twin-CNV discordance experiment to find a possible disease associated CNV and confirmed the pathogenicity of this CNV with a burden test in a case control-study.

CNV profiling in both concordant and discordant monozygotic twins can be a used to identify causal or contributing genetic factors to specific traits and diseases, although caution has to be taken by the interpretation of these twin-study results. Twin pregnancy may not be representative of a singleton pregnancy[27] and environmental and genetic components are not always exactly identical in monozygous twins. TTTS can cause intra-uterine environmental differences; somatic mutations can lead to genetic mosaicism in different tissues within one individual and between twins. It is advisable to screen for genetic abnormalities at several time-point during a person's life-span and in different tissues. Accumulation of somatic changes could be responsible for a significant proportion of human disease. Moreover, determining the zygosity status in all pregnancies is advisable. Undiagnosed "vanishing" twins and TTTS could cause congenital malformations, but MZ twin pregnancies could also result in masking genetic aberrations in blood cells. Comparing different tissues, if possible, could be crucial in determining genetic aberrations in congenital anomalies. Comparing the genomes of large cohorts of twins and their parents may reveal not only *de novo* genetic changes but perhaps also shed light on predisposing factors modulating penetrance in traits and diseases in combination with environmental or other factors.

Future prospects

Epigenetic studies in discordant monozygotic twins

Epigenetic mechanisms like X-inactivation, DNA-methylation, histone modifications or other mechanisms influence chromatin structure, maintenance or remodeling and by so influence the expression of genes. These epigenetic signatures can be heritable, vary over time and are responsive to environmental cues.[91] Although these epigenetic signatures in

monozygous twins are mostly very similar, they can differ between monozygous twins. For example, DNA methylation differences can develop over time in monozygous twins, possible caused by environmental influences. [92] Humans exposed peri-conceptional, not late gestational, to malnutrition conditions during the Dutch "hunger winter" in World War II had decreased methylation of the *IGF2* gene 60 years later in life. *IGF2* is maternally imprinted and a key factor in growth and development. Hypomethylation will lead to bi-allelic expression of *IGF2*.[93] These epigenetic changes have been studied in twins and differences in epigenetic signature have been observed in phenotypical discordant monozygous twin pairs. Examples are monozygous twins discordant for systemic lupus erythematosus [94], RETT syndrome (cerebroatrophic hyperammonemia) [95] and Beckwith-Wiedeman syndrome.[96]

Monozygous twins with the same *de novo* pathogenic frame shift mutation on the paternal derived chromosome X in the *MECP2* gene were discordant for RETT syndrome.[95] Weksberg et al describes differences in methylation in monozygous twins discordant for Beckwith-Wiedeman syndrome.[96] There was no skewed X-inactivation, whole genome sequencing revealed no SNP, InDel nor CNV differences. There were no *MECP2* expression or methylation differences in fibroblasts. However, a genome wide methylation survey revealed several differentially methylated regions and the corresponding genes were differentially expressed. Although the *MECP2* locus was unaffected it is perhaps the combined effect, altered expression of a set of genes and the pathogenic *MECP2* mutation that result in RETT syndrome in one of the sibs. Linkage analysis in Beckwith Wiedeman syndrome families points to chromosome band 11p15. In this region sporadic cases of translocations or CNV are described. 20% of patients have uniparental disomy and it is known that genes in this region have maternal or paternal specific expression. [96, 97] Weksberg et al describes differences in methylation in monozygous twins discordant for Beckwith-Wiedeman syndrome. Genetically, this syndrome is characterized by altered expression of imprinted genes on chromosome locus 11p15.5. This expression can be altered by several mechanisms. In th monozygous twins studied, hypomethylation of CpG island KvDMR1 located within the *KCNQ1* gene in the affected twin resulted in bi-allelic expression of the *KCNQ1OT1* ncRNA.[96]

It has been proposed that epigenetic changes could be a major contributor to monozygotic twin discordance [98] and studying these epigenetic differences in monozygotic twins could bridge the gap between environment and our genome.[99] Large scale studies in discordant monozygous twins could be a successful strategy to identify

epigenetic variation responsible complex traits and diseases.[100] Technically, epigenetic profiling is, like copy number measurements, in transition from single gene and genome-wide array based techniques to high throughput techniques based on NGS.[101] So many more examples of discordant methylation patterns will follow in the near future.

Next Generation sequencing based Copy Number profiling and GWAS

Using whole genome NGS data both copy number and nucleotide sequence can be determined. Currently, whole genome sequencing is relatively expensive. In the near future, when prices will drop, large cohorts of sequenced cases can probably be compared to sequenced controls. GWAS on both SNP and CNV data could be performed on this data and not the *associated* but the *responsible* genetic variant could be identified. Whilst variant calling on nucleotide resolution has seen tremendous progress, InDel and CNV calling are still in their infancy. Current NGS based methods use read depth, the number of unique reads that map to a given locus in the genome, to determine the Copy Number Status at that locus. Comparing the read counts of individual reference samples (paired analysis) or using a normalized cohort of reference samples and compare this with patient samples will give a fairly good estimate of the relative copy number. This information can be combined with allelic frequency data obtained from the variant files, much like it is done with SNP-array analysis. Several tools exist, each with its advantages and disadvantages such as reliability to accurately detect relative gains or losses.[102-104] Challenges in accurately detecting copy number from sequencing data are aligning sequence reads in repeat region and issues that arise from capturing segments of the DNA during exome-NGS. [105] It is likely that SNP-array based GWAS studies will be replaced by NGS based methods in the upcoming years. Although GWAS have proven their value in medical research, publications in CNV based association studies in twins are relatively scarce. Burden studies in monozygous twins; both concordant and discordant for a given phenotype could prove to be a valuable addition to the phenotype-genotype toolset. Tissue specific whole genome and epigenome profiling in large cohorts of concordant and discordant monozygous twins could be a promising key to unraveling the impact of gene, environment and gene-environment interactions. CNP and CNV affect much more our genome than SNPs do and could perhaps explain much of the "missing heritability" observed in traits and disease. CNP and CNV affect much more our genome than SNPs do and could perhaps explain much of the "missing heritability" observed in traits and disease.

References

1. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.
2. Manolio, T.A., *Bringing genome-wide association findings into clinical use.* Nat Rev Genet, 2013. **14**(8): p. 549-58.
3. Schlattl, A., et al., *Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions.* Genome Res, 2011. **21**(12): p. 2004-13.
4. Zhang, F., et al., *Copy number variation in human health, disease, and evolution.* Annu Rev Genomics Hum Genet, 2009. **10**: p. 451-81.
5. Stankiewicz, P. and J.R. Lupski, *Structural variation in the human genome and its role in disease.* Annu Rev Med, 2010. **61**: p. 437-55.
6. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.
7. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome.* Science, 2004. **305**(5683): p. 525-8.
8. van Duyvenvoorde, H.A., et al., *Copy number variants in patients with short stature.* Eur J Hum Genet, 2013.
9. Iakoubov, L., et al., *A common copy number variation (CNV) polymorphism in the CNTNAP4 gene: association with aging in females.* PLoS One, 2013. **8**(11): p. e79790.
10. Liu, Y.Z., et al., *Genome-wide copy number variation association analyses for age at menarche.* J Clin Endocrinol Metab, 2012. **97**(11): p. E2133-9.
11. Cantsilieris, S., et al., *Comprehensive analysis of Copy Number Variation of genes at chromosome 1 and 10 loci associated with late age related macular degeneration.* PLoS One, 2012. **7**(4): p. e35255.
12. McCarroll, S.A., et al., *Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease.* Nat Genet, 2008. **40**(9): p. 1107-12.
13. Wheeler, E., et al., *Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity.* Nat Genet, 2013. **45**(5): p. 513-7.
14. Craddock, N., et al., *Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls.* Nature, 2010. **464**(7289): p. 713-20.
15. McCarroll, S.A., *Extending genome-wide association studies to copy-number variation.* Hum Mol Genet, 2008. **17**(R2): p. R135-42.
16. Soemedi, R., et al., *Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease.* Am J Hum Genet, 2012. **91**(3): p. 489-501.
17. Hitz, M.P., et al., *Rare copy number variants contribute to congenital left-sided heart disease.* PLoS Genet, 2012. **8**(9): p. e1002903.
18. Oei, L., et al., *A genome-wide copy number association study of osteoporotic fractures points to the 6p25.1 locus.* J Med Genet, 2013.
19. McCarroll, S.A. and D.M. Altshuler, *Copy-number variation and association studies of human disease.* Nat Genet, 2007. **39**(7 Suppl): p. S37-42.
20. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.
21. Wat, M.J., et al., *Genomic alterations that contribute to the development of isolated and non-isolated congenital diaphragmatic hernia.* J Med Genet, 2011. **48**(5): p. 299-307.
22. Scott, D.A., et al., *Genome-wide oligonucleotide-based array comparative genome hybridization analysis of non-isolated congenital diaphragmatic hernia.* Hum Mol Genet, 2007. **16**(4): p. 424-30.
23. Southard, A.E., L.J. Edelmann, and B.D. Gelb, *Role of copy number variants in structural birth defects.* Pediatrics, 2012. **129**(4): p. 755-63.
24. Stadler, Z.K., et al., *Rare de novo germline copy-number variation in testicular cancer.* Am J Hum Genet, 2012. **91**(2): p. 379-83.
25. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.
26. Hehir-Kwa, J.Y., et al., *De novo copy number variants associated with intellectual disability have a paternal origin and age bias.* J Med Genet, 2011. **48**(11): p. 776-8.
27. Hall, J.G., *Twinning.* Lancet, 2003. **362**(9385): p. 735-43.
28. Russell, R.B., et al., *The changing epidemiology of multiple births in the United States.* Obstet Gynecol, 2003. **101**(1): p. 129-35.
29. Reddy, U.M., A.M. Branum, and M.A. Klebanoff, *Relationship of maternal body mass index and height to twinning.* Obstet Gynecol, 2005. **105**(3): p. 593-7.
30. Basso, O., et al., *Risk of twinning as a function of maternal height and body mass index.* Jama, 2004. **291**(13): p. 1564-6.

31. Weber, M.A. and N.J. Sebire, *Genetics and developmental pathology of twinning*. Semin Fetal Neonatal Med, 2010. **15**(6): p. 313-8.
32. Souter, V.L., et al., *A case of true hermaphroditism reveals an unusual mechanism of twinning*. Hum Genet, 2007. **121**(2): p. 179-85.
33. Pharoah, P.O., *Causal hypothesis for some congenital anomalies*. Twin Res Hum Genet, 2005. **8**(6): p. 543-50.
34. Shetty, A. and A.P. Smith, *The sonographic diagnosis of chorionicity*. Prenat Diagn, 2005. **25**(9): p. 735-9.
35. Lee, Y.M., et al., *Antenatal sonographic prediction of twin chorionicity*. Am J Obstet Gynecol, 2006. **195**(3): p. 863-7.
36. van Dongen, J., et al., *The continuing value of twin studies in the omics era*. Nat Rev Genet, 2012. **13**(9): p. 640-53.
37. Tan, Q., et al., *Dissecting complex phenotypes using the genomics of twins*. Funct Integr Genomics, 2010. **10**(3): p. 321-7.
38. Fernandez-Rozadilla, C., et al., *A genome-wide association study on copy-number variation identifies a 11q11 loss as a candidate susceptibility variant for colorectal cancer*. Hum Genet, 2014. **133**(5): p. 525-34.
39. Yang, R., et al., *Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3*. Carcinogenesis, 2014. **35**(2): p. 315-23.
40. Zahnleiter, D., et al., *Rare copy number variants are a common cause of short stature*. PLoS Genet, 2013. **9**(3): p. e1003365.
41. Lin, P., et al., *Copy number variations in 6q14.1 and 5q13.2 are associated with alcohol dependence*. Alcohol Clin Exp Res, 2012. **36**(9): p. 1512-8.
42. Bergen, S.E., et al., *Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder*. Mol Psychiatry, 2012. **17**(9): p. 880-6.
43. Friedrichsen, M., et al., *Carboxylesterase 1 gene duplication and mRNA expression in adipose tissue are linked to obesity and metabolic function*. PLoS One, 2013. **8**(2): p. e56861.
44. Grayson, B.L., et al., *Genome-wide analysis of copy number variation in type 1 diabetes*. PLoS One, 2010. **5**(11): p. e15393.
45. Nag, A., et al., *Copy number variation at chromosome 5q21.2 is associated with intraocular pressure*. Invest Ophthalmol Vis Sci, 2013. **54**(5): p. 3607-12.
46. Ehli, E.A., et al., *De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems*. Eur J Hum Genet, 2012. **20**(10): p. 1037-43.
47. Nelson, M.R., et al., *An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people*. Science, 2012. **337**(6090): p. 100-4.
48. Tennessen, J.A., et al., *Evolution and functional impact of rare coding variation from deep sequencing of human exomes*. Science, 2012. **337**(6090): p. 64-9.
49. Gibson, G., *Rare and common variants: twenty arguments*. Nat Rev Genet, 2011. **13**(2): p. 135-45.
50. Sebire, N.J., D. Talbert, and N.M. Fisk, *Twin-to-twin transfusion syndrome results from dynamic asymmetrical reduction in placental anastomoses: a hypothesis*. Placenta, 2001. **22**(5): p. 383-91.
51. Pharoah, P.O., S.V. Glinianaia, and J. Rankin, *Congenital anomalies in multiple births after early loss of a conceptus*. Hum Reprod, 2009. **24**(3): p. 726-31.
52. Bruder, C.E., et al., *Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles*. Am J Hum Genet, 2008. **82**(3): p. 763-71.
53. Essaoui, M., et al., *Monozygotic twins discordant for 18q21.2qter deletion detected by array CGH in amniotic fluid*. Eur J Med Genet, 2013. **56**(9): p. 502-5.
54. Smeets, D., et al., *Monochorionic dizygous twins presenting with blood chimerism and discordant sex*. Twin Res Hum Genet, 2013. **16**(4): p. 799-801.
55. Veenma, D., et al., *Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts*. Eur J Hum Genet, 2012. **20**(3): p. 298-304.
56. Razzaghian, H.R., et al., *Post-zygotic and inter-individual structural genetic variation in a presumptive enhancer element of the locus between the IL10Rbeta and IFNAR1 genes*. PLoS One, 2013. **8**(9): p. e67752.
57. Machin, G., *Non-identical monozygotic twins, intermediate twin types, zygosity testing, and the non-random nature of monozygotic twinning: a review*. Am J Med Genet C Semin Med Genet, 2009. **151c**(2): p. 110-27.
58. Tucker, T., et al., *Comparison of genome-wide array genomic hybridization platforms for the detection of copy number variants in idiopathic mental retardation*. BMC Med Genomics, 2011. **4**: p. 25.
59. Haraksingh, R.R., et al., *Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms*. PLoS One, 2011. **6**(11): p. e27859.
60. Pinto, D., et al., *Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants*. Nat Biotechnol, 2011. **29**(6): p. 512-20.
61. Tabet, A.C., et al., *Autism multiplex family with 16p11.2p12.2 microduplication syndrome in monozygotic twins and distal 16p11.2 deletion in their brother*. Eur J Hum Genet, 2012. **20**(5): p. 540-6.

62. Zwijnenburg, P.J., H. Meijers-Heijboer, and D.I. Boomsma, *Identical but not the same: the value of discordant monozygotic twins in genetic research.* Am J Med Genet B Neuropsychiatr Genet, 2010. **153b**(6): p. 1134-49.
63. Yamagishi, H., et al., *Phenotypic discordance in monozygotic twins with 22q11.2 deletion.* Am J Med Genet, 1998. **78**(4): p. 319-21.
64. Vincent, M.C., et al., *22q11 deletion in DGS/VCFS monozygotic twins with discordant phenotypes.* Genet Couns, 1999. **10**(1): p. 43-9.
65. Halder, A., et al., *Chromosome 22q11.2 microdeletion in monozygotic twins with discordant phenotype and deletion size.* Mol Cytogenet, 2012. **5**(1): p. 13.
66. Munar-Ques, M., et al., *Two pairs of proven monozygotic twins discordant for familial amyloid neuropathy (FAP) TTR Met 30.* J Med Genet, 1999. **36**(8): p. 629-32.
67. O'Reilly, R., et al., *Monozygotic twins with early-onset schizophrenia and late-onset bipolar disorder: a case report.* J Med Case Rep, 2013. **7**(1): p. 134.
68. Friedman, J.H., et al., *Monozygotic twins discordant for Huntington disease after 7 years.* Arch Neurol, 2005. **62**(6): p. 995-7.
69. Erlich, Y., *Blood ties: chimerism can mask twin discordance in high-throughput sequencing.* Twin Res Hum Genet, 2011. **14**(2): p. 137-43.
70. Reuss, A., et al., *Monozygotic twins discordant for trisomy 18.* Ultrasound Obstet Gynecol, 2011. **38**(6): p. 727-8.
71. Gentilin, B., et al., *Discordant prenatal phenotype and karyotype of monozygotic twins characterized by the unequal distribution of two cell lines investigated by different methods: a review.* Twin Res Hum Genet, 2008. **11**(3): p. 352-6.
72. Egan, E., et al., *The Outcome of Twin Pregnancies Discordant for Trisomy 21.* Twin Res Hum Genet, 2013: p. 1-7.
73. Dixit, A., et al., *Monozygotic twins discordant for trisomy 13: counselling and management issues.* J Perinatol, 2012. **32**(8): p. 639-41.
74. Nieuwint, A., et al., *'Identical' twins with discordant karyotypes.* Prenat Diagn, 1999. **19**(1): p. 72-6.
75. Breckpot, J., et al., *Differences in Copy Number Variation between Discordant Monozygotic Twins as a Model for Exploring Chromosomal Mosaicism in Congenital Heart Defects.* Mol Syndromol, 2012. **2**(2): p. 81-87.
76. Rio, M., et al., *Monozygotic twins discordant for submicroscopic chromosomal anomalies in 2p25.3 region detected by array CGH.* Clin Genet, 2013. **84**(1): p. 31-6.
77. Baudisch, F., et al., *CNV analysis in monozygotic twin pairs discordant for urorectal malformations.* Twin Res Hum Genet, 2013. **16**(4): p. 802-7.
78. Kimani, J.W., et al., *Search for genomic alterations in monozygotic twins discordant for cleft lip and/or palate.* Twin Res Hum Genet, 2009. **12**(5): p. 462-8.
79. Gu, J., et al., *Genome-wide catalogue of chromosomal aberrations in barrett's esophagus and esophageal adenocarcinoma: a high-density single nucleotide polymorphism array analysis.* Cancer Prev Res (Phila), 2010. **3**(9): p. 1176-86.
80. Maiti, S., et al., *Ontogenetic de novo copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia.* PLoS One, 2011. **6**(3): p. e17125.
81. Pamphlett, R. and J.M. Morahan, *Copy number imbalances in blood and hair in monozygotic twins discordant for amyotrophic lateral sclerosis.* J Clin Neurosci, 2011. **18**(9): p. 1231-4.
82. Sasaki, H., et al., *Copy number loss of (src homology 2 domain containing)-transforming protein 2 (SHC2) gene: discordant loss in monozygotic twins and frequent loss in patients with multiple system atrophy.* Mol Brain, 2011. **4**: p. 24.
83. Jensen, H., E. Kjeldsen, and V.E. Hjortdal, *Could submicroscopical chromosomal imbalances cause conotruncal malformations in twins?* Congenit Heart Dis, 2012. **7**(2): p. 170-7.
84. Bloom, R.J., et al., *Comprehensive analysis of copy number variation in monozygotic twins discordant for bipolar disorder or schizophrenia.* Schizophr Res, 2013. **146**(1-3): p. 289-90.
85. Ono, S., et al., *Failure to confirm CNVs as of aetiological significance in twin pairs discordant for schizophrenia.* Twin Res Hum Genet, 2010. **13**(5): p. 455-60.
86. Liang, Q., et al., *Extensive genomic copy number variation in embryonic stem cells.* Proc Natl Acad Sci U S A, 2008. **105**(45): p. 17453-6.
87. Forsberg, L.A., et al., *Age-related somatic structural changes in the nuclear genome of human blood cells.* Am J Hum Genet, 2012. **90**(2): p. 217-28.
88. Piotrowski, A., et al., *Somatic mosaicism for copy number variation in differentiated human tissues.* Hum Mutat, 2008. **29**(9): p. 1118-24.
89. Fischer, H.G., et al., *Changes in neuronal DNA content variation in the human brain during aging.* Aging Cell, 2012. **11**(4): p. 628-33.
90. McConnell, M.J., et al., *Mosaic copy number variation in human neurons.* Science, 2013. **342**(6158): p. 632-7.

91. Bonasio, R., S. Tu, and D. Reinberg, *Molecular signals of epigenetic states.* Science, 2010. **330**(6004): p. 612-6.
92. Wong, C.C., et al., *A longitudinal study of epigenetic variation in twins.* Epigenetics, 2010. **5**(6): p. 516-26.
93. Heijmans, B.T., et al., *Persistent epigenetic differences associated with prenatal exposure to famine in humans.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17046-9.
94. Javierre, B.M., et al., *Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.* Genome Res, 2010. **20**(2): p. 170-9.
95. Miyake, K., et al., *Comparison of Genomic and Epigenomic Expression in Monozygotic Twins Discordant for Rett Syndrome.* PLoS One, 2013. **8**(6): p. e66729.
96. Weksberg, R., et al., *Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome.* Hum Mol Genet, 2002. **11**(11): p. 1317-25.
97. Soejima, H. and K. Higashimoto, *Epigenetic and genetic alterations of the imprinting disorder Beckwith-Wiedemann syndrome and related disorders.* J Hum Genet, 2013. **58**(7): p. 402-9.
98. Czyz, W., et al., *Genetic, environmental and stochastic factors in monozygotic twin discordance with a focus on epigenetic differences.* BMC Med, 2012. **10**: p. 93.
99. Zhang, D., et al., *Twin-based DNA methylation analysis takes the center stage of studies of human complex diseases.* J Genet Genomics, 2012. **39**(11): p. 581-6.
100. Bell, J.T. and T.D. Spector, *A twin approach to unraveling epigenetics.* Trends Genet, 2011. **27**(3): p. 116-25.
101. Huss, M., *Introduction into the analysis of high-throughput-sequencing based epigenome data.* Brief Bioinform, 2010. **11**(5): p. 512-23.
102. Guo, Y., et al., *Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control.* Biomed Res Int, 2013. **2013**: p. 915636.
103. Duan, J., et al., *Comparative studies of copy number variation detection methods for next-generation sequencing technologies.* PLoS One, 2013. **8**(3): p. e59128.
104. de Ligt, J., et al., *Detection of clinically relevant copy number variants with whole-exome sequencing.* Hum Mutat, 2013. **34**(10): p. 1439-48.
105. Teo, S.M., et al., *Statistical challenges associated with detecting copy number variations with next-generation sequencing.* Bioinformatics, 2012. **28**(21): p. 2711-8.

3.2

Copy Number Detection in Discordant Monozygotic Twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) Cohorts

Adapted from:

Copy Number Detection in Discordant Monozygotic Twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) Cohorts.

D. Veenma, E. Brosens, E. de Jong, C. van de Ven, C. Meeuwissen, T. Cohen-Overbeek, M. Boter, H. Eussen, H. Douben, D. Tibboel and A. de Klein. Eur J Hum Genet. 2012 Mar;20(3):298-304. doi: 10.1038/ejhg.2011.194

Abstract

The occurrence of phenotypic differences between monozygotic twins is commonly attributed to environmental factors, assuming that monozygotic twins have a complete identical genetic make-up. Yet, recently several lines of evidence showed that both genetic and epigenetic factors could play a role in phenotypic discordance after all. A high occurrence of copy number variation differences was observed within monozygotic twin pairs discordant for Parkinson disease, thereby stressing on the importance of post-zygotic mutations as disease-predisposing events.

In this study, the prevalence of discrepant copy number variations was analysed in discordant monozygotic twins of the Esophageal Atresia and Congenital Diaphragmatic Hernia cohort in the Netherlands. Blood-derived DNA from 11 pairs (seven Esophageal Atresia and four Congenital Diaphragmatic Hernia) was screened using high-resolution SNP arrays.

Results showed an identical copy number profile in each twin pair. Mosaic chromosome gain or losses could not be detected either with a detection threshold of twenty percent. Some of the germ-line structural events demonstrated in five out of eleven twin pairs could function as a susceptible genetic background. For example, the 177Kb loss of chromosome 10q26 in CDH pair-3 harbours the TCF7L2 gene (Tcf4 protein), which is implicated in the regulation of muscle fiber type development and maturation.

In conclusion, discrepant copy number variations are not a common cause of twin discordancy in these investigated congenital anomaly cohorts.

Introduction

Monozygotic (MZ) twin comparisons have been used for many decades to specify contributions of both nature (heredity) and nurture (environment)[1]. Normally the study design is based on the presumption that monozygotic twins come from one fertilized egg and therefore have complete identical genetic make-ups. Yet, recently several lines of evidence suggested that genetic and epigenetic factors could play a role in MZ phenotypic variances after all [2-6]. Using a BAC array platform, Bruder *et al.*[6] demonstrated that discordance in their monozygotic Parkinson's disease (PD) twin cohort of nine individuals could be the result of Copy Number Variation (CNV) differences. However, Baranzini *et al.* [7] could not reproduce this high intra-twin pair variability of structural variants using both array and next-generation sequencing in three twin pairs discordant for Multiple Sclerosis. We investigated whether discrepant CNVs could cause discordance in MZ twin pairs of the Dutch Esophageal Atresia (EA [MIM 189960]) and Congenital Diaphragmatic Hernia (CDH [MIM 142340]) cohort. Blood-derived DNA from 11 (7 EA and 4 CDH) pairs of MZ twins was screened using high-resolution SNP arrays.

EA generally presents at birth with a defective formation of the esophagus with or without a fistulous tract to the trachea. Although not lethal in most cases, long-term morbidity plays a significant role in these patients. CDH is a more severe birth defect characterised by defective formation of the diaphragm, lung hypoplasia and pulmonary hypertension. Despite medical advances mortality for isolated cases is 20% and for none-isolated cases up to 60%. Both EA and CDH are presumed to have a multifactorial etiology and the identification of chromosomal aberrations and knockout animal models provide strong evidence for a genetic component [8]. In contrast, both anomalies present with low twin concordance rates, 10.7% and 15.6% for EA and CDH respectively, and sibling recurrence rates are low (1-2%) as well.

Shaw-Smith [9] already pointed out that the incidence of twinning in EA is 2.6 times higher than statistically expected. 206 pairs are described in literature up until now, however information on zygosity is less thorough [9-16]. Orford *et al.*[15] stated that at least 80% of reported EA twins are same-sex pairs. In total, 22 out of these 206 twin pairs are concordant for the EA phenotype. In literature, 77 twin pairs have been described for CDH of which 53 were recognized as monozygotic [17-20]. 12 pairs were concordant for the CDH phenotype.

The rationale of this study was to investigate whether CNVs in the affected twin sibling could account for phenotypic discordance of either Esophageal Atresia or Congenital Diaphragmatic Hernia MZ twin siblings. Although results showed no such proof, germ-line structural events were detected and these could represent a susceptible genetic background as seen in other genetic anomalies. Results are discussed in the context of earlier MZ twin reports.

Materials and methods

Ethics statement

Research involving human participants has been approved by the "Medical Ethical Committee (METC) at Erasmus-MC, which specifically approved for blood withdrawal of both twins and their parents. Informed consent forms were obtained for the index case and his/her parents at once and for the healthy twin separately.

*P*atients

The 11 affected twin samples were collected from the congenital anomaly cohort in Rotterdam (Erasmus MC Sophia's Hospital, the Netherlands) in which 541 EA- and 626 CDH- patients are currently registered. Of these, 22 CDH patients (14 dizygotic, five MZ, three not tested) and 35 EA patients (six dizygotic, nine MZ, 20 not tested) were the result of a twin pregnancy. Included were those twin samples with a written parental informed consent, quality material of both siblings and confirmed monozygosity by STR profiling (AmpFISTR identifiler PCR amplification kit, Applied Biosystems, Foster City, CA). Another exclusion criterion was the identification of a genetic abnormality, most commonly an aneuploidy.

DNA isolation

Automated DNA extraction from peripheral blood (or skin fibroblasts in case of two affected CDH twins) was performed using local standard protocols. DNA quality and concentration were checked with the Quant-iT™ PicoGreen ® dsDNA Kit (Invitrogen Corporation, Carlsbad, California, USA).

Whole-genome high-resolution SNP array

SNP analysis was carried out using the Illumina HumanCytoSNP-12 bead chip version 2.2 (Illumina, San Diego, CA, USA). This chip includes 220.000 of the most informative SNPs en markers with a median physical distance of 6.2 Kb. DNA samples were processed according to the manufacturer's protocol. The call rate of this array batch was above 0.98, except for 1 sample.

SNP array analysis

Data for each bead chip were self-normalized in Genomestudio GT® (Illumina, San Diego, CA, USA) using information contained within the array. Copy number estimates for each individual sample were determined by comparison to a common reference set of 200 CEU samples from the HapMAP project [www.hapmap.org/downloads/raw_data] (supplied by Illumina, manifest files) and visualized in the Nexus software program (version five, Biodiscovery, El Segundo, CA, USA) as log2 ratios. Analysis settings were as follow: both SNP-FASST and SNP-Rank segmentation methods were executed independently with significance thresholds ranging from $1x10^{-4}$ to $1x10^{-6}$ and log-ratio thresholds of 0.18 and -0.18 for duplication and deletions respectively. The max contiguous probe spacing was 1000 Kbp and the minimum number of probes per segment was set to three, limiting CNV detection to sizes above 18.6 Kb. Subsequently, only CNVs above 50 Kb were validated. Paired analysis for deletions and duplications was performed in each affected twin versus its healthy co-twin.

As described recently high-resolution (SNP) arrays are suitable for detection of both germ-line and mosaic CNVs [21-25]. Mosaic copy number aberrations are hallmarked by a concomitant change of log2 intensity signal and a shift in b-allele frequency. The detection limit (sensitivity) of the Nexus SNP-FASST algorithm for mosaic CNVs is 20 percent using a heterozygous imbalance threshold of 0.45 [22]. To review functionality of each putative CNV at once, occurrence frequencies in a qualified normal pediatric cohort of 2026 individuals[26] (CHOP [http://cnv.chop.edu/]) and in the DGV [http://www.tcag.ca] were uploaded in the Nexus program as well. Since these populations display various ethnic backgrounds, comparison to an in-house normal reference set was performed as well. Additionally, possible intra-twin pair genotype differences (with respect to all SNP-markers presented on the array) were evaluated in Genomestudio GT® using the paired analysis settings.

Pair EA	GA (wks)	Birth order (Patient)	(twin-sib)	Obstetric history	EA	Fistel	Type of additional anomalies
EA1	37,3	1	2	Breech presentation	+	+	Dysmorphic Auricular tags, Cleft uvula Abnormal dermatoglyphics Heart ASD Lung Lunghypoplasia right Neurologic/skeletal Scoliosis Fusion of vertebrae Hemivertebrae IUGR
EA2	36	2	1	Breech presentation Maternal medication: sintrom	+	+	Heart VSD Lung Lunghypoplasia
EA3	NA	NA	NA	NA	+	+	Heart cardiac situs Dextrocardia healthy twin
EA4	NA	NA	NA	NA	+	-	-
EA5	33,5	1	2	Breech presentation Fever durante partu	+	+	Dysmorphic Triangular face, Deep set eyes, Palpebral fissures slant down Small mandible Thin fingers,hypoplastic thumbs Proximal placement of thumb Hypoplastic or absent radii Sacral hemangioma healthy twin
EA6	34, 4	NA	NA	Polyhydramnion Maternal medication: corticosteroids	+	+	Heart VSD Tricuspid incompetence
EA7		NA	NA	-	+	+	-

Table 1A. Clinical features EA Cohort. The following abbreviations are used: ASD; Atrial Septal defect, EA; esophageal atresia, GA; Gestational Age, IUGR; Intra Uterine growth Retardation, VSD; Ventrical Septal Defect. Unfortunately, for a few EA subjects detailed clinical data is unavailable

Pair CDH	GA (wks)	Birth order (Patient)	(Healthy twin)	Obstetric history	CDH	Type of additional anomalies
CDH1	35,3	2	1	Sectio Caesarea Breech presentation	left	-
CDH2	33,4	1	2	Sectio Caesarea >24hrs ruptured membrane	left	-
CDH3	38,5	2	1	-	right	Urogenital Inguinal hernia Hydrocele testis
CDH4	34,1	2	1	Sectio Caesarea	left	Dysmorphic Small mandibula IUGR

Table 1B. Clinical features CDH Cohort. The following abbreviations are used: ASD; Atrial Septal defect, CDH; Congenital Diaphragmatic Hernia, GA; Gestational Age, IUGR; Intra Uterine growth Retardation, VSD; Ventrical Septal Defect.

Validation using Fluorescent In-Situ Hybridization and relative-quantitative PCR analysis

Confirmation of each CNV with quantitative real time PCR and/or FISH was executed in the twin-siblings and their parents according to local standard protocols with minor modifications [22, 27]. For FISH, BAC clones were selected from the UCSC genome browser [http://genome.ucsc.edu/], purchased at BACPAC resources centre (Oakland, California, USA) and labelled (Random Prime labelling system Invitrogen Corporation, Carlsbad, California, USA)) with Bio-16-dUTP or Dig-11-dUTP (Roche applied science, Indianapolis, USA). After validation on control metaphases, the chromosome 22 BAC clones RP11-62K15 and RP1-66M5 were used for confirmation in EA-pair-I.

Primer pairs for quantitative real-time PCR were designed from unique sequences within the minimal deleted or duplicated regions of each copy number change using Primer Express software v2.0 (Applied Biosystems, Carlsbad, California, USA). The nucleotide-nucleotide BLAST algorithm at NCBI [http://www.ncbi.nlm.nih.gov/BLAST/] was used to confirm that each PCR amplification product was unique. Quantitative PCR analyses were performed using an ABI7300 Real-time PCR system in combination with KAPA-SYBR fast master mix (KapaBiosystems, Woburn, MA, USA). Experiments were designed with a region of the *C14ORF145* gene serving as a control locus as previously described [27].

Results

Clinical characterization and monozygosity screening of twin pairs

Clinical features of each twin pair are summarized in **Table 1.** Briefly, seven out of 11 pairs were discordant for the phenotype of EA **(Table1A)** and four out of 11 for CDH **(Table 1B)**. Four out of eleven EA-affected patients harbored (major) additional anomalies. Considering CDH; there is a variable expression of left and right CDH with all persons (as expected) featuring lung hypoplasia. We are dealing with an isolated CDH cohort since most anomalies in pairs 3 and 4 are minor. Finally, zygosity status of each twin pair was confirmed (data not shown) by STR profiling using the commercially available STR identifiler kits of Applied Biosystems.

Paired CNV analysis of discordant monozygotic twins

Results of the paired CNV analysis of each MZ twin couple are summarized in **Table 2** showing no evidence of pathogenic CNV discordance in both congenital anomaly cohorts. In order to detect mosaic (somatic) aberrations, specific attention was payed to b-allele frequency changes as well. In the EA cohort a total of ten germline CNVs were identified. Seven concerned common Copy Number Polymorphisms (CNP) defined by the occurrence of

the CNV in at least five individuals of qualified normal pediatric cohorts in literature. The remaining three events were present in both the twin and at least one healthy parent and are therefore less likely to be pathogenic as well. For example, the 666 Kb sized chromosome 22q deletion in EA pair-1 (**Figure 1**) was found both in the healthy twin and his mother and partly overlaps with CNVs catalogued in control cohorts.

Existence of inherited CNVs was detected in the CDH cohort as well. A total of three CNVs were distinguished of which two are not prevalent in normal cohorts. All three events were present in the healthy twin as well. **Figure 2** represents the 177Kb loss of chromosome 10q26 in CDH pair-3 and harbors the *TCF7L2* gene (*Tcf4* protein), which is mainly known for its involvement in blood glucose homeostasis as a result of Wnt signaling changes**.** Not ruled out in this study are the presence of balanced genomic alterations and small (<50kB) or very-low mosaic (<20%) chromosome aberrations beyond the detection level of our experimental approach.

SNP genotype analysis monozygotic twin cohorts

SNP genotype differences between the affected and unaffected twin siblings were evaluated for each SNP on the Illumina® bead chip. After removal of less accurately called SNPs, genotyping analysis showed concordance for almost all SNPs (n=299671) within each MZ pair. A total of five SNPs in three EA-pairs were dissimilar and three SNPs in two CDH pairs **(Table 3)**. CDH pair-3 showed discrepancy for 99 SNPs, which could be attributed to less overall genotyping accuracy and therefore was not analyzed further**.** Until now, only rs2824374 (which is closely linked to the *CXADR* gene) could be associated with embryonic (mal) development, however literature only reports on effects to the kidneys and cochlea [28, 29]. None of the other identified discordant intra-twin SNP loci are directly linked to a phenotype.

Figure 1

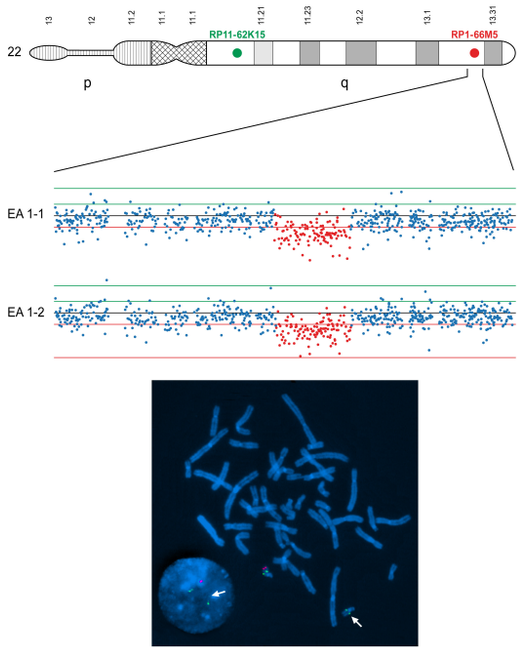

Figure 2

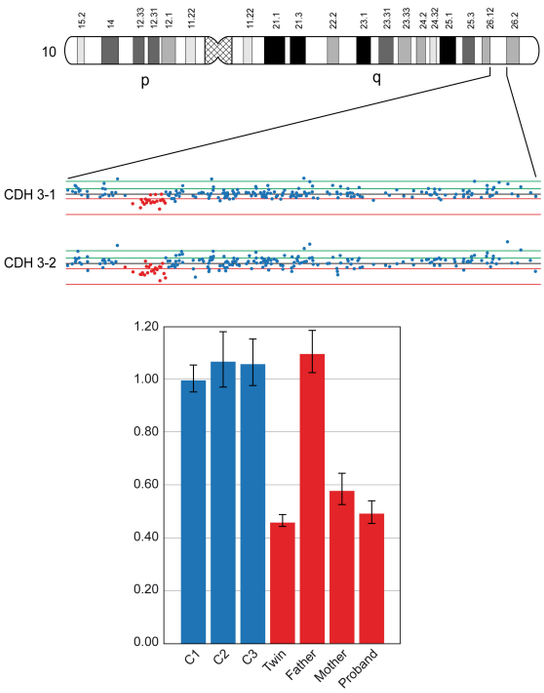

Figure 1 (left panel) SNP and Fluorescent In-Situ Hybridization results of inherited chromosome 22-CNV in EA pair-I Nexus results (Top) of the 666 Kb deletion on chromosome 22q13.3 in both individuals of EA pair-I showing a clear drop in log2 intensity signal validated by FISH (Bottom) on metaphase chromosomes of the affected EA twin-1. Probes: control: RP11-62K15 (green) and target: RP11-66M5 (red). Parental analysis (results not shown) demonstrated that this genomic event is inherited from the mother and therefore less likely pathogenic. In addition, no gene is allocated to this region neither are any miRNA transcripts hampering the identification of functional elements in this region as well.

Figure 2 (right panel) SNP and relative Q-PCR results of inherited chromosome 10-CNV in CDH pair-3. Nexus result (Top) of the chromosome 10q26 deletion event in CDH pair-3 showing a clear drop in log2 intensity signal, which was confirmed by relative q-PCR (bottom) in the affected proband, the unaffected twin and the mother.

EA pair	Chromosomal location (bp)	CNV type	Length (bp)	Gene symbols	Discordant twin	Normal Pediatric Cohort #	Validation
1	chr22:46,163,818-46,870,578	CN Loss	666192	*FLJ46257 (hypothetical)*	yes	yes (0/2026, 0/370, overlap 25 DGV)	FISH, inherited maternally
2	chr7:75,420,580-75,471,147	CN Loss	50567	*POR, STYXL1, TMEM120A*	yes	yes (0/2026, 0/370, overlap 2 DGV)	q-PCR, inherited (both parents heterozygous deletion)
3	chr6:162,638,827-162,840,229	CN Gain	201402	*PARK2*	yes	yes: 5/2026	q-PCR, inherited maternally
	chr17:41,522,684-41,646,903	CN Gain	124219	*KIAA1267*	yes	yes: 69/2026	Copy Number Polymorphism
	chr18:1,895,191-1,960,898	CN Loss	65707	-	yes	yes: 26/2026	Copy Number Polymorphism
4	chr14:19,283,777-19,479,370	CN Gain	195593	*OR4K1, OR4K2, OR4K5, etc*	yes	yes: 35/2026	Copy Number Polymorphism
5	chr12:7,892,014-8,027,862	CN gain	135848	*SLC2A14,SLC2A3*	yes	yes: 39/2026	Copy Number Polymorphism
6	chr12:31,146,084-31,303,651	CN gain	157567	*DDX11*	yes	yes: 20/370	Copy Number Polymorphism
	chr2:89,728,406-89,885,025	CN gain	156619		yes	yes: 14/2026	Copy Number Polymorphism
	chr8:2,322,561-2,577,455	CN gain	254894		yes	yes: 32/2026	Copy Number Polymorphism
7	-	-	-	-	-	-	-

CDH pair	Chromosomal location (bp)	CNV (type)	Length (bp)	Gene symbols	Discordant twin	Normal Cohort #	Validation
1	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-
3	chr10:114,660,279-114,838,014	CN Loss	177735	*TCF7L2* [22]	yes	no	q-PCR, inherited maternally
4	chr4:86,595,913-87,070,050	CN Loss	474137	*ARHGAP24* [23-24]	yes	yes (0/2026, 0/370, overlap 27 DGV)	q-PCR, inherited maternally
	chr4:57,743,262-57,795,280	CN Gain	52018	-	yes	yes (2/2026, 0/370, overlap 15 DGV)	Copy Number Polymorphism

Table 2. Inherited CNVs detected in MZ twins of the Rotterdam congenital anomaly cohort. The following abbreviations are used: CN; Copy Number, bp; basepairs, EA; Esophageal Atresia, CDH; Congenital Diaphragmatic Hernia.# Normal Cohort; CHOP [http://cnv.chop.edu/] and DGV [http://tcag.ca].& Decipher; [http://decipher.sanger.ac.uk/application/].Percentage of overlap with CNVs in control databases are designated if not 100% aligned.

Discussion

A high occurrence of copy number variants that differed between siblings discordant for Parkinson's disease was recently suggested [6]. However, intra-twin pair variability for germ-line CNVs could not be detected in our subset of EA and CDH MZ twins. Within the limitations of the used experimental approach, structural variants in mosaic form (above 20%) could neither be demonstrated. Application of next-generation sequencing methods will allow for an easier and more sensitive calling of the smallest mosaic aberrations in the near future and will add up to the (scarce) data generated recently on this topic by some other groups [30-33].

Various causes could account for the discrepancy in CNV findings between our congenital anomaly twin cohort and the Parkinson cohort. First of all: an age factor. The rather high prevalence of mosaic CNVs in PD twins could have been generated during lifetime. This was suggested by a small study of the group of Dumanski *et al.*[34, 35], who identified mosaic aberrations in a wide-range of tissues of three phenotypically normal individuals. This hypothesis would imply that age-accumulated (tissue-specific) CNV events could play a role in diseases developing symptoms later in life. Consequently, they are expected to contribute less to congenital disorders. Secondly, differences in CNV prevalence between our study and the Parkinson study could be based on methodological differences such as choice of platform. Although Bruder *et al.* [6] presented confirmative evidence for a few of their CNVs using a different platform, detailed confirmation of most CNVs was lacking. On the other hand, structural DNA variation might play a minor role in EA- and CDH- pathophysiology, suggesting that in these congenital cohorts the focus should be widened on environmental and epigenetic factors.

Two recent studies [2, 7] revealed a (significant) proportion of epigenetic variability between MZ twins in investigated tissues. However, in the Multiple Sclerosis twin cohort study these changes could not account for disease discordance. A similar study for EA, CDH or other congenital anomalies is difficult to perform, since the target tissues cannot be obtained from the healthy co-twin for obvious reasons. Structural variations restricted to the affected esophagus and diaphragm tissue could represent another cause for twin discordancy, yet was not excluded in this monozygotic cohort due to unavailability of the affected material.

Pair EA	Discordant SNPs	Chromosomal location (dbSNP build 130) (bp)	Gene symbols
1	rs11573502	chrX:24888693-24889193	*POLA1*
2	-	-	-
3	-	-	-
4	rs10125846	chr9:2821641-2822141	*KIAA0020*
	rs1744767	chr20:35200380-35200880	*LOC140699* isoform 3
	rs438895	chr1:8260692-8261192	
5	-	-	-
6	-	-	-
7	rs1576026	chr9:25,453,089-25,453,589	-

Pair CDH	Discordant SNPs	Chromosomal location (dbSNP build 130) (bp)	Gene symbols
1	rs17730982	chr8:134851934-134852434	
	rs2824374	chr21:17879252-17879752	*CXADR*; possible involvement kidney & cochlear development [25,26]
2	-	-	-
4	rs6571064	chr6:103400463-103400963	-
3	Not evaluated; SNP Quality rate < 0.95	-	-

Table 3. Discordant SNPs in MZ twin pairs of the Rotterdam Congenital anomaly cohort. The following abbreviations are used:bp; basepairs, EA; Esophageal Atresia, CDH; Congenital Diaphragmatic Hernia.

Finally, although our results showed no prove for CNV contribution to phenotypic MZ-discordance, germ-line structural events were detected in both cohorts and these events could represent a so-called susceptible genetic background. In five out of eleven twin pairs germline CNVs were identified. These were rarely detected in a specific pediatric normal population [26] and/or our in-house control cohort and could therefore represent an increased susceptibility to congenital anomalies by means of a dosage responsive- or position- effect. For example, the 177Kb loss of chromosome 10q26 in CDH pair-3 might be of functional importance. A recent report demonstrated Tcf4 (alias *Tcf7l2*) expression in connective tissue fibroblasts during development and suggested its role in the regulation of muscle fiber type development and maturation [36]. Additionally, certain polymorphisms and mutations in *TCF7L2* are linked to an increased risk of type 2 diabetes[37]. This implies that loss of one functional *TCF7L2* allele might be associated with (super) normal glucose tolerance. Indeed we observed evidence of increased serum glucose (a Glucose of 12.2 mmol/l was identified within 24 hours postnatal) in the affected individual of twin pair 3.

However, also 1 normal glucose level (Glucose 3.6 mmol/l) was determined within the same time window and since this patient was critically ill and died shortly thereafter no absolute conclusions can be drawn from these results. The healthy twin had an unremarkable medical record so far. Similarly, the haploinsufficient *ARHGAP24* gene in CDH pair-4 (encoding a vascular, cell-specific GTPase-activating protein) could confer genetic susceptibility for CDH by means of its function in modulating angiogenesis and through its interaction with filamin-A [37, 38]. Girirajan et al.[39] recently demonstrated that a second hit may elicit a severe phenotype in offspring of healthy CNV-carriers. Hypothetically this second hit can constitute another CNV in the same or associated disease pathway as well as a pathogenic SNP. These results underline the importance of archiving all genomic events (also those with a "benign" nature at first sight) in a freely accessible database such as initiated by the ISCA consortium [https://www.iscaconsortium.org]. Detailed and unbiased phenotyping is crucial for the understanding of the more complex genotype-phenotype correlations as well.

In summary, we investigated whether the existence of discrepant CNVs could be causal to the phenotypic discordance in MZ twin pairs of the Esophageal Atresia and Congenital Diaphragmatic Hernia cohort in Rotterdam and found no such proof. Prospective collection of DNA material in various MZ twin cohorts is warranted to evaluate the possibility of such genetic factors contributing to human phenotypic variability in general and to twin-discordance specific. We feel that the use of high-resolution SNP arrays and sequencing based methods are more suitable in these designs than BAC arrays. Finally, phenotypic correlations can only be made after proper analysis in normal cohorts as well.

Acknowledgments

The authors would like to thank all patients, specifically all twins and their parents for making this study possible. Also, we thank Tom de Vries-Lentsch for his help in preparing the figures.

Conflict of interest statement

This study was funded by the Sophia's Children's Hospital Foundation for Medical Research, Rotterdam, The Netherlands (SSWO 493 and 551). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Hall, J.G., *Twinning*. Lancet, 2003. **362**(9385): p. 735-43.
2. Kaminsky, Z.A., et al., *DNA methylation profiles in monozygotic and dizygotic twins*. Nat Genet, 2009. **41**(2): p. 240-5.
3. Kimani, J.W., et al., *X-chromosome inactivation patterns in monozygotic twins and sib pairs discordant for nonsyndromic cleft lip and/or palate*. Am J Med Genet A, 2007. **143A**(24): p. 3267-72.
4. Machin, G.A., *Some causes of genotypic and phenotypic discordance in monozygotic twin pairs*. Am J Med Genet, 1996. **61**(3): p. 216-28.
5. Streit, A. and R.J. Sommer, *Genetics: Random expression goes binary*. Nature. **463**(7283): p. 891-2.
6. Bruder, C.E., et al., *Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles*. Am J Hum Genet, 2008. **82**(3): p. 763-71.
7. Baranzini, S.E., et al., *Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis*. Nature, 2010. **464**(7293): p. 1351-6.
8. Holder, A.M., et al., *Genetic factors in congenital diaphragmatic hernia*. Am J Hum Genet, 2007. **80**(5): p. 825-45.
9. Shaw-Smith, C., *Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology*. J Med Genet, 2006. **43**(7): p. 545-54.
10. Sunagawa, S., et al., *Dichorionic twin fetuses with VACTERL association*. J Obstet Gynaecol Res, 2007. **33**(4): p. 570-3.
11. Depaepe, A., H. Dolk, and M.F. Lechat, *The epidemiology of tracheo-oesophageal fistula and oesophageal atresia in Europe. EUROCAT Working Group*. Arch Dis Child, 1993. **68**(6): p. 743-8.
12. Farquhar, J., R. Carachi, and P.A. Raine, *Twins with oesophageal atresia and the CHARGE association*. Eur J Pediatr Surg, 2002. **12**(1): p. 56-8.
13. Khan, R.A. and K.L. Narashimhan, *Discordance in twins for esophageal atresia with tracheoesophageal fistula*. Indian J Pediatr, 2009. **76**(3): p. 334-5.
14. Ohno, K., et al., *Esophageal atresia with tracheoesophageal fistula in both members of monozygotic twins*. Pediatr Surg Int, 2008. **24**(10): p. 1137-9.
15. Orford, J., et al., *Oesophageal atresia in twins*. Pediatr Surg Int, 2000. **16**(8): p. 541-5.
16. Trobs, R.B., et al., *Congenital oesophageal atresia discordant for tracheo-oesophageal fistula occurring in a set of dizygotic twins*. Eur J Pediatr Surg, 2006. **16**(4): p. 260-4.
17. Donahoe, P.K., K.M. Noonan, and K. Lage, *Genetic tools and algorithms for gene discovery in major congenital anomalies*. Birth Defects Res A Clin Mol Teratol, 2009. **85**(1): p. 6-12.
18. Lau, S.T., et al., *Fraternal twins with Morgagni hernias*. J Pediatr Surg, 2005. **40**(4): p. 725-7.
19. Machado, A.P., et al., *Concordance for bilateral congenital diaphragmatic hernia in a monozygotic dichorionic twin pair - first clinical report*. Fetal Diagn Ther. **27**(2): p. 106-9.
20. Pober, B.R., et al., *Infants with Bochdalek diaphragmatic hernia: sibling precurrence and monozygotic twin discordance in a hospital-based malformation surveillance program*. Am J Med Genet A, 2005. **138A**(2): p. 81-8.
21. Conlin, L.K., et al., *Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis*. Hum. Mol. Genet., 2010. **19**(7): p. 1263-75.
22. Veenma, D., et al., *Comparable low-level mosaicism in affected and non affected tissue of a complex CDH patient*. PloS one, 2010. **5**(12): p. e15348.
23. Shaffer, L.G., et al., *Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases*. The Journal of pediatrics, 2006. **149**(1): p. 98-102.
24. Ballif, B.C., et al., *Detection of low-level mosaicism by array CGH in routine diagnostic specimens*. American journal of medical genetics. Part A, 2006. **140**(24): p. 2757-67.
25. Xiang, B., et al., *Analytical and clinical validity of whole-genome oligonucleotide array comparative genomic hybridization for pediatric patients with mental retardation and developmental delay*. American journal of medical genetics. Part A, 2008. **146A**(15): p. 1942-54.
26. Shaikh, T.H., et al., *High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications*. Genome research, 2009. **19**(9): p. 1682-90.
27. Scott, D.A., et al., *Genome-wide oligonucleotide-based array comparative genome hybridization analysis of non-isolated congenital diaphragmatic hernia*. Human molecular genetics, 2007. **16**(4): p. 424-30.
28. Excoffon, K.J., et al., *The Coxsackievirus and Adenovirus Receptor: a new adhesion protein in cochlear development*. Hear Res, 2006. **215**(1-2): p. 1-9.
29. Raschperger, E., et al., *The coxsackie and adenovirus receptor (CAR) is required for renal epithelial differentiation within the zebrafish pronephros*. Dev Biol, 2008. **313**(1): p. 455-64.

30. Razzaghian, H.R., et al., *Somatic mosaicism for chromosome X and Y aneuploidies in monozygotic twins heterozygous for sickle cell disease mutation.* Am J Med Genet A, 2010. **152A**(10): p. 2595-8.
31. Hasler, R., et al., *Genetic control of global gene expression levels in the intestinal mucosa: a human twin study.* Physiol Genomics, 2009. **38**(1): p. 73-9.
32. Notini, A.J., J.M. Craig, and S.J. White, *Copy number variation and mosaicism.* Cytogenet Genome Res, 2008. **123**(1-4): p. 270-7.
33. Yurov, Y.B., et al., *Aneuploidy and confined chromosomal mosaicism in the developing human brain.* PLoS One, 2007. **2**(6): p. e558.
34. Piotrowski, A., et al., *Somatic mosaicism for copy number variation in differentiated human tissues.* Hum Mutat, 2008. **29**(9): p. 1118-24.
35. De, S., *Somatic mosaicism in healthy human tissues.* Trends Genet April 13. [Epub ahead of print], 2011.
36. Mathew, S.J., et al., *Connective tissue fibroblasts and Tcf4 regulate myogenesis.* Development, 2011. **138**(2): p. 371-84.
37. Nakamura, F., et al., *Molecular basis of filamin A-FilGAP interaction and its impairment in congenital disorders associated with filamin A mutations.* PLoS One, 2009. **4**(3): p. e4928.
38. Su, Z.J., et al., *A vascular cell-restricted RhoGAP, p73RhoGAP, is a key regulator of angiogenesis.* Proc Natl Acad Sci U S A, 2004. **101**(33): p. 12212-7.
39. Girirajan, S., et al., *A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay.* Nature genetics, 2010. **42**(3): p. 203-9.

3.3

Exon level Copy Number and Sequence comparison in Congenital Diaphragmatic Hernia and Esophageal Atresia discordant monozygous twin cohorts

Adapted from:

Exon level Copy Number and Sequence comparison in Congenital Diaphragmatic Hernia and Esophageal Atresia discordant monozygous twin cohorts

E. Brosens, R.W.W Brouwer, T. Brands, D. Veenma , H. Eussen, W.F.J van IJcken, A. de Klein.

Manuscript in preparation

Abstract

Congenital Hernia of the Diaphragm (CDH) and Esophageal Atresia with/or without Trachea-esophageal Fistula (EA/TEF) are congenital anomalies that can either be present as isolated anomaly or in association with other birth defects. Both anomalies likely have a multifactor etiology, are associated with known (genetic) syndromes and can occur in combination with specific chromosomal aberrations, Copy Number Variations (CNV) or mutations. When evaluating the genetic component of a disease, twin studies can help to elucidate potential causal or predisposing genetic factors. Monozygotic (MZ) twins are believed to have the same genetic content and share the same environment during development. We hypothesize that de novo mutations arisen early in embryonic development could explain the phenotypical differences in discordant MZ twins.

In total of six EA/TEF and four CDH discordant MZ twins are characterized with standard SNP-array, exome bead array and exome-NGS in order to detect, complete or mosaic, DNA discrepancies. As described previously, we could not detect any CNV (mosaicism) differences in these twins with micro-array. Genotyping with SNP-array did not result in the detection of twin-differences. Standard exome-NGS variant calling with the Genome Analysis ToolKit revealed numerous discrepant SNPs and InDels. Visual inspection of hundreds of these events with the Broad institute's IGV indicated that these discrepancies were actually false positive differences due to technical limitations, analysis settings (thresholding) or limitations of the variant calling. By comparing different alignment techniques, variant callers, statistics and analysis strategies we could reduce the number of likely candidate differences significantly. Interestingly, at a Bonferonni level of statistical significance variants were real "NGS" differences and not the result of for instance thresholding or low coverage. However, we could not confirm the remaining differences with Sanger sequencing. With millions of measuring points, there are bound to be differences. These differences are in general not biological, but technical in nature. Differences detected in these experiments reflect the accuracy and limitations of current sequencing technologies and variant calling pipelines.

Introduction

About one in forty pregnancies, is a twin pregnancy and one in three of these twin pairs are monozygous (MZ); siblings originating from one oocyte.[1] The later implies that MZ twins are genetically identical. Usually monozygotic twins are also phenotypically very similar. However, MZ twins with concordant chromosomal anomalies, pathogenic CNV or mutations can have a discordant disease phenotype.[2, 3] This phenotypical discordance at birth could be the result of for instance differences in epigenetic modifications or, surprisingly, environmental exposure differences.

However, recently it has been shown that not all MZ twins have exactly the same genome. For instance Bruder and co-workers identified three somatic intra-twin Copy Number Variation (CNV) differences in a cohort of nine Parkinson-like discordant and ten healthy monozygotic twins.[4] It has been suggested that DNA changes could causing twinning as the blastocyst recognizes these mutated cells as foreign resulting in splitting of the blastocyst.[1] Recently, CNV differences were also reported in monozygotic twins discordant for certain congenital anomalies [5-7] and both Voigt *et al* and Kaplan *et al* describe postzygotic somatic mosaicism in monozygotic twins discordant for neurofribomatosis type 1.[8, 9] However, Baranzini and coworkers could not find evidence for sequence differences in MZ twins discordant for Multiple Sclerosis[10], neither could Solomon in a twin pair discordant for VACTERL association.[11]

Previously, we could not detect genotype or (mosaic) CNV differences in phenotypical discordant twins for esophageal atresia (EA) and congenital diaphragmatic hernia (CDH) with SNP-array[12] EA and CDH are severe developmental defects and patients often have additional associated anomalies.[13, 14] EA is characterized by a blind ending esophageal tube with often a faulty connection of the distal esophagus to the trachea, a trachea-esophageal fistula (TEF). The hallmark of CDH is a hole in the diaphragm and patients often also have lung hypoplasia and pulmonary hypertension. Both diseases are thought to have a multifactorial etiology. EA and CDH are variable features in many genetic syndromes.[13, 14]The twinning incidence is 2.6 times higher in EA pregnancies compared to the general background.[15] Twin concordance rates are relatively low for both conditions, ranging between 10-15%.[12]

In this current study we compare the DNA of these twins at a much higher resolution. Using Whole-Exome Sequencing (WES) and WES-Copy Number profiling we have compared the genetic sequence and Copy Number at the exon level. We hypothesize that discordant somatic mutations are responsible for the phenotypical discordance in these Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) discordant monozygous twin cohorts.

Methods

The Erasmus MC-Sophia cohort of congenital anomalies

The Medical Ethical Review Board of Erasmus MC - Sophia Children's Hospital approved this study. Parental informed consent included the genetic studies in both siblings and their parents. The medical charts were reviewed and clinical and and follow-up data registered to an interactive database. Discordant monozygotic twin were selected from the Erasmus University MC-Sophia TE-cohort (n=582) [14, 16] and CDH cohort (n=703) and most have been described previously.[13] DNA was extracted from peripheral blood when the twins were at least one year of age to avoid contamination of with sibling DNA resulting from transfer of lymphocytes via twin-to-twin transfusion. Unfortunately, transfer of hematopoietic stem cells can't be excluded.[17] We used DNA derived from skin fibroblasts in CDH patients CDH01 and CDH02. Mozozygosity was determined with short tandem repeat profiling (AmpFISTR identifiler PCR amplification kit, Applied Biosystems, Foster City, CA, USA) and later confirmed with SNP-array. Patients did not have confirmed genetic syndromes or *de novo* chromosomal anomalies prior to analysis.

Copy Number Variation profiling and genotyping

Genotyping and sub sequent Copy Number Variation profiling was done with Illumina HumanCytoSNP-12 version 2 or the Illumina Human Omni Express-12 version 1.0 Bead Chips. (Illumina, San Diego, CA, USA) Additionally, all twin-pairs were genotyped with the Illumina Exomev1.1 genotyping chip. Arrays were processed according to their manufacturer's standard protocol using Genome Studio software version 2011.1 (Illumina, Inc., San Diego, USA). We used CoNVEX version 0.6 [18] and the normalized whole exome sequencing reads to determine exon level Copy Number. Each exon captured with the target enrichment kit was divided in three virtual probes. Each of these probes needed to be in the same segment to be called as an CNV. An "B-allele" frequency track was made using the GATK unified genotyper variants. This track was used to support the loss and gains calls made by the segmentation algorithm in the Nexus CN7.1 software used to visualize CNV profiles. (Biodiscovery Inc, El Segundo, CA, USA) Analysis settings ,CNV

profiling, analysis and confirmation using Fluoresence in situ Hybridization, MAQ-assay (Multiplicom Inc., Niel, Belgium) and qPCR have been described previously.[19]

Whole Exome Capture, sequencing, quality control and analysis

Genomic DNA was fragmented (Covaris, Inc. Woburn, Massachusetts, USA) and yield and fragment size determined with the Bioanalyser 2100 bioanalysis chip and Agilent DNA 1000 Kit.(Agilent BioAnalyzer, Santa Clara, CA). We used the SureSelect Human All Exon 50 Mb Targeted exome enrichment kit v2 (Agilent Technologies, Inc., Santa Clara, California) and Illumina TruSeq version 4 paired end 2x 101 bp sequence procedure on the Hiseq2000 sequencer (Illumina, Inc., San Diego, USA)for all twin pairs. Sub sequent de-multiplexing, alignment to the hg19 reference genome with Burrows-Wheeler Aligner version 0.6.2 [20], generation of chromosome sorted BAM-files with SAMtools version 0.1.12a [21] and quality control is automated in the NARWHAL pipeline[22]. The quality control parameters include total reads, aligned reads on target, mean coverage of the target region, target with at least 20X coverage. These are listed in supplementary table 1.

Genotyping and InDel calling was done with the Bayesian genotyper incorporated in the genome analysis toolkit version 1.2.9 [23], SAMtools mpileup, or in house developed callers. Variants were annotated with ANNOVAR version 2013-feb-21.[24] Moreover, raw sequencing read alignment, variant calling and quality control was also performed using CLC-bio.(Qiagen Inc., Venlo, the Netherlands) We used both variant calling methods of this software tool: a quality based and probabilistic method. Next the data was annotated in the same pipeline as GATK and (using ANNOVAR and both Refseq and ENSEMBL gene annotation) Variant confirmation is done using Sanger sequencing, Illumina's Exomev1.1 genotyping chip (Illumina, Inc., San Diego, USA) or one of the other targeted sequencing methods used in EA twin pair 1 and 2. Variant filtering was done with TIBCO Spotfire version 5.5.0.36 (TIBCO, Boston, MA, USA) and Cartagenia Bench NGS version 3.04 (Cartagenia Inc, Boston, MA, USA)

Whole Exome variant analysis and twin comparison

We pair-wise compared the variants detected in the Illumina's Exomev1.1 genotyping chip, GATK pipeline, CLC-bio quality based and probabilistic variant calling pipeline and screened for variants present in one and absent in the other sibling. Moreover, we compared the variants of the Illumina's Exomev1.1 genotyping with those called with SAMtools Mpileup and probabilistic differentiator. First, non-covered base-pairs are labeled with "N". Next, using first binomial and second Fisher exact statistics, we screened the exome variants detected in the SAMtools Mpileup pipeline for statistical significant ($p \leq 1.0 \times 10^{-8}$) differences, taken into account coverage and variant frequency. The top ranking

significant differences were validated with Sanger sequencing in twins and their parents. In addition to a twin comparison we used two gene panels, one for trachea-esophageal anomalies and VACTERL associated features and one for Congenital Diaphragmatic Hernia. (supplementary figure 2)

In supplementary table 3a through 3j are the (predicted) splicing, frameshift, stopgain and stoploss mutations in candidate genes detected with the GATK unified genotyper, CLC-bio probabilistic and the CLC-bio quality based variant callers analysed with the Cartagenia software. Visual inspection of the discordant variants was done in the Broad institutes IGV. As a final step we exported all rare or private exonic variants, detected in the SAMtools mpileup pipeline, with an allele frequency ≤ 0.001 in either 1000 Genomes and/or ESP6500 to Nexus CN to search for overlap with the CNVs detected in both the WS-CN and SNP-array experiments. An example is given in figure 1.

Results

The Erasmus MC-Sophia cohort of congenital anomalies

7 EA pairs (labeled EA1 to and EA 7) and 4 CDH twin pairs (labeled CDH 1 to CDH 4) are described previously.[12] Numbering and phenotypical descriptions are kept consistent. We did not have enough DNA material for whole exome sequencing of twin pair EA4 and EA7. However, we were able to include an additional discordant EA pair, EA 8.

Exon level copy number comparison

In previous experiments, we excluded large *de novo* CN and concluded that if a rare or private CNV was present in the affected twin, it was also present in the unaffected sibling and inherited from a healthy parent.[12] We had detected several genotyping mismatches, validation with Sanger sequencing confirmed that these were all technical errors and not actual biological differences. SNP-array genotyping with the Illumina HumanExomev1.1 genotyping chip -containing more than 200,000 probes specific for relatively rare variants selected from 12,000 sequences from several large sequencing projects- did not result in the detection of additional genotyping discrepancies.

WGS-CN resulted in the detection of hundreds of small exon level copy number variations. Interestingly, both the overall probe level profile and normalized copy number profile of each twin pair was almost perfectly similar. Visual inspection of the remaining discrepancies of the automated pairwise comparison resulted in few variants possible discordant between twins. Most of this discordant variation was not present in the affected sib of the twins. In these instances, only the healthy twin deviated from the diploid state.

Most of these regions contained segmental duplications and were known Copy Number Polymorphisms. (see table 2) Confirmation of the differences affecting non polymorphic regions; the rare and private CNV, is ongoing.

Pair	Gestational age	Congenital anomalies in addition to EA/TEF or CDH	DNA source
EA 1-1	37.3	Dysmorphic features, auricular tags,celft uvula, abnormal dermatoglyphics, atrial septal defect, rightsided lung hypoplasia, neurological anomalies, scoliosis, fusion of vertebrae, hemivertebra, intrauterine growth restriction	Blood/blood
EA 2-1	36	Ventricular septal defect, lunghypoplasia	Blood/blood
EA 3-1	?	Cardiac situs inversus	Blood/blood
EA 5-1	33.5	Dysmorphic features, palpebral fissures slant down, deep-set eyes, triangular face, micrognathia, thin fingers, hypoplastic proximal placed thumbs, hypoplastic radii and a sacral hemangioma in the healthy twin	Blood/blood
EA 6-1	34, 4	Ventricular septal defect, tricuspid incompetence	Blood/blood
EA 8-1	?	Dysmorphic features, auricular tags,celft uvula, abnormal dermatoglyphics, atrial septal defect, rightsided lung hypoplasia, neurological anomalies, scoliosis, fusion of vertebrae, hemivertebra, intrauterine growth restriction	Blood/blood
CDH 1-1	35.3	-	Dermal fibroblasts/Blood
CDH 2-1	33.4	-	Blood/blood
CDH 3-1	38.5	Urogenital malformations, ingenuil hernia and hydrocele testis	Blood/blood
CDH 4-1	34.1	Dysmorphic features, small mandibula and intraunterine growth restriction	Dermal fibroblasts/Blood

Table 1. Phenotype descriptions twin pairs. Affected (-1) in EA monozygous twins all have an Esophageal atresia as well as a tracheo-esophageal fistula. CDH affected twins all have Congenital Diaphragmatic Hernia. Siblings (-2) are healthy or indicated otherwise. The sib of EA 3-1 (EA 3-2) has dextrocardia.

Whole Exome variant analysis and twin comparison

Pair-wise comparison of variants detected with the GATK unified genotyper, CLC-bio quality based and probabilistic variant callers resulted in several hundred to thousands of inter-twin differences. Visual inspection of hundreds of these differences in the Broad institute's IGV indicated that most discrepancies were actually false positive differences. Most of the variation was present in both twin-sibs and did not differ much in their variant frequency. In other instances, the region was not (sufficiently) covered in one of the twins. These detected differences were a result of cut-off values in variant calling threshold settings, limitations of the variant calling, or sequence coverage differences.

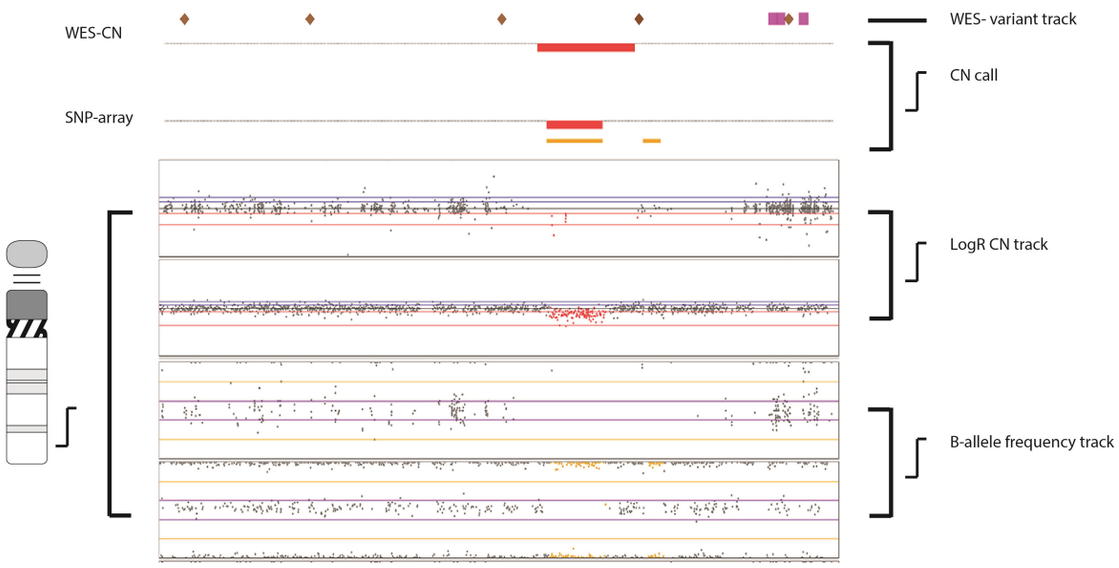

Figure 1. Example of combined analysis of rare loss of function mutations, SNP-array CNV profiling and WES-CN. In the upper panel, rare variants are depicted with specific symbols matching each variant type. The next two tracks, WES-CN and SNP-array CN calls depict the region in which a call is made. In the LogR tracks, relative copy number status of WES data (upper panel) and SNP array are depicted, In the bottom two panels the zygosity status is depicted as a B-allele frequency plot, with the upper panel describing the variation seen in the WES-variant files and the bottom panel the SNP-array genotyping results. In this example rare variants surround the deletion seen in both exon level WES-CN (red dots) and SNP-array (red dots). Since this is a deletion, the region is homozygous. This maternal inherited deletion was present in both twin-sibs and is described previously.[12]

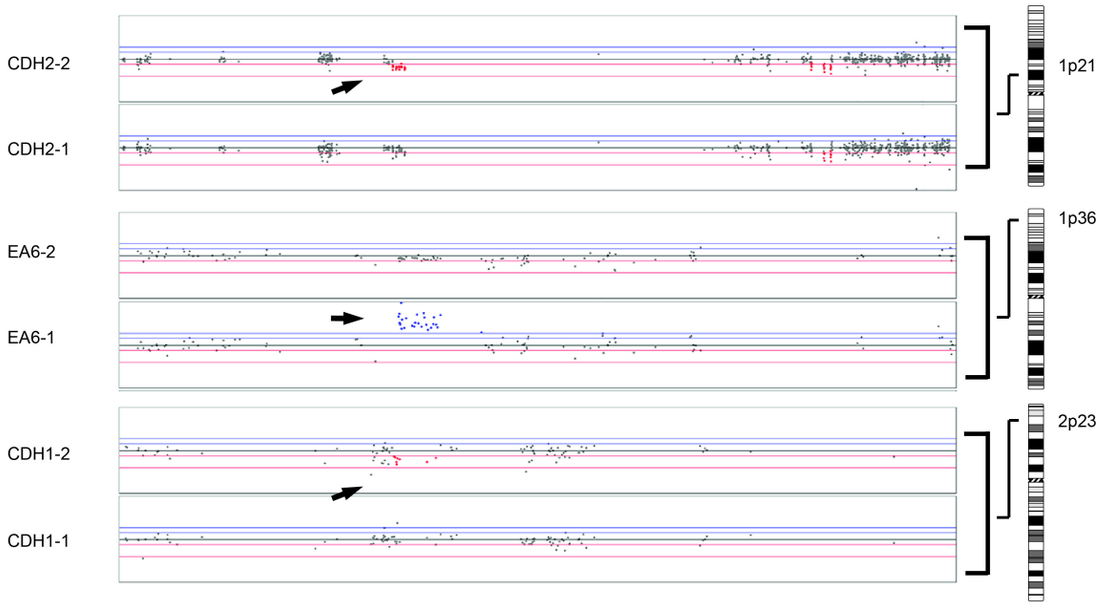

Figure 2. Examples of exon-level copy number twin differences (arrow) Depicted are the differences between healthy twin CDH 2-2 (loss, in red) and affected twin CDH 2-1 (diploid). Other examples are the gain on 1p36 seen in patient EA 6-1 (blue) and its absence in the healthy twin EA 6-2. The loss (red) on 2p23 is seen in healthy sib of twin pair CDH 1-2, yet absent in affected twin CDH 1-1.

To circumvent this time consuming visual inspection, we mined raw sequencing reads, to determine read depth and allele counts at positions with a variant in each twin. Next, we used a statistical approach, a negative binomial model, to rank the detected differences. Using this approach, with a p-value cut-off of 10^{-8}, we reduced the number of significant candidates differences to workable numbers for each twin-pair. (see table 3)

Moreover, we detected 7 genes (*ADAM3A*, *PGKB*, *EXOSC10*, *HOMEZ*, *PR2C2*, *SYN3* and *TRMT44*) with a discrepant variant in two twin pairs. Of these the variants seen in *ADAM3A* and *HOMEZ* were in the coding region. Primers were designed to determine if these top-ranking differences were actual genotype differences. However, almost all statistically significant discordant variant were located in or near a highly repetitive sequence. This complicated primer design. Since Sanger sequencing of candidate differences for which primers could be designed did not confirm the observed differences, also these most likely represent technical artefacts.

Next we used a different statistical approach and filtered variant in or near repetitive regions. Using Fisher exact statistics and a p-value cut-off of 10^{-8} we reduced the number of candidate discordant variants and designed primers on the top-ranking variants for each twin pair. None of the statistical significant exonic variants with the Fisher exact test were shared between the twin pairs. Two exonic variants had overlap between statistical tests: a snp (rs1924654) in *NBPF10* gene and a nonsynonymous variant in *NRG3*. Validation of these twin differences with Sangers sequencing is ongoing, although preliminary results indicate that also these are false positive differences. Using the Gene panels for trachea-esophageal anomalies and VACTERL association and the gene panel for Congenital Diaphragmatic Hernia we detected many loss of function mutations in candidate genes, visual inspection in the Broad institutes' IGV revealed that all these were shared between twins. (supplementary table 3)

Discussion

On average 30.000 variants were detected in the targeted protein coding and ncRNA coding segments of the genome in each patient. We detected several loss of function mutations, present in both twins, even in the genes present in the candidate gene panels. Overall, the vast majority of variants were shared between twins, especially taken into account that not all of the targeted sequences are covered with the same amount of reads.

Twin Pair (-individual)	Chromosome Region	Length	Event	Cyto-band	Gene Count	Gene Symbols	Count of DGV	Count of segdup	CNV/CNP
EA 1-2	chr2:179,516,818-179,530,927	14,109	Loss	q31.2	3	MIR548N, TTN	13	2	CNV
EA 1-2	chr22:18,671,397-21,536,612	2,865,215	Loss	q11.21	1	AK129567, AK302545, GGT3P, LOC729444, USP41	129	11	CNP
EA 1-2	chr22:24,237,925-24,399,444	161,519	Loss	q11.23	18	LOC284889 – GSTTP2	519	19	CNP
EA 2-1	chr11:433,750-455,481	21,731	Gain	p15.5	3	ANO9, Metazoa_SRP, PTDSS2	19	0	CNV
EA 3-2	chr7:63,193,176-63,224,926	31,750	Loss	q11.21	2	DQ599768, DQ574660	65	22	CNP
EA 5-1	-	0	-	-	-	-	-	-	-
EA 6-1*	chr1:16,883,061-16,947,163	64,102	Gain	p36.13	3	AB1, NBPF1, CROCCP2	312	75	CNP
EA 6-1*	chr4:8,974,559-9,377,934	403,375	Gain	p16.1	0	USP17L10 - LOC728405	7	116	CNP
EA 6-1*	chr5:34,181,127-34,206,027	24,900	Gain	p13.2	4	DQ587763, DQ571461, DQ578105, DQ598168	169	45	CNP
EA 6-1*	chr10:135,376,371-135,534,747	158,376	Gain	q26.3	10	SYCE1 - DUX4L6	419	63	CNP
EA 6-2	chr13:53,100,276-53,114,495	14,219	Nullizygous	q14.3	1	TPTE2P3	15	1	CNV
EA 6-1*	chr13:57,715,524-57,977,845	262,321	Gain	q21.1	5	PRR20A, PRR20B, PRR20D, PRR20C, PRR20E	531	2	CNP
EA 6-1*	chr18:44,506,918-44,557,893	50,975	Gain	q21.1	5	TCEB3CL, LOC100506888, TCEB3C, TCEB3CL2, KATNAL2	103	2	CNP
EA 6-1*	chr20:27,500,000-29,635,688	2,135,688	Gain	q11.1 - q11.21	2	Y_RNA, FRG1B	98	45	CNP
EA 8-1	chr1:16,890,738-16,948,559	57,821	Gain	p36.13	2	NBPF1, CROCCP2	313	73	CNP
EA 8-1	chr9:140,063,526-140,066,387	2,861	Gain	q34.3	2	MIR3621, LRRC26	8	0	CNV
EA 8-1	chr17:36,263,896-36,347,903	84,007	Gain	q12	8	DQ587906 - LOC440434	208	46	CNP
CDH 1-2	chr10:135,475,929-135,534,747	58,818	Gain	q26.3	7	DUX4L2 - DUX4L6	154	51	CNP
CDH 1-2	chr13:57,137,164-58,048,498	911,334	Gain	q21.1	5	PRR20A, PRR20B, PRR20D, PRR20C, PRR20E	566	2	CNP
CDH 2-2	chr1:104,161,260-106,434,412	2,273,152	Loss	p21.1	8	AMY2A - BC043293	1016	26	CNP
CDH 3-2	chr1:6,268,575-6,269,853	1,278	Gain	p36.31	2	RNF207, RNF207	1	0	CNV
CDH 3-1	chr19:1,009,067-1,012,410	3,343	Loss	p13.3	3	GRIN3B, FLJ00277, C19orf6	11	0	CNP
CDH 3-2	chr8:86,651,927-86,839,906	187,979	Gain	q21.2	0	-	125	72	CNP
CDH 4-2	chr22:24,367,999-24,390,248	22,249	Gain	q11.23	3	LOC391322, GSTT1, GSTTP2	168	6	CNP

Table 2 WES-CN twin differences. * high signal to noise ratio resulting in waving probe pattern, -1 = affected twin, -2= healthy twin, CNV = rare or private Copy Number Variation, CNP = Copy Number Polymorphism common in the general population.

Twin-Pair	GATK unified genotyper	CLC-bio quality based	CLC-bio probabilistic	negative binomial statistics	Fisher exact test + repeat filter
EA 01	184	978	412	14	3
EA 02	176	2525	751	12	3
EA 03	238	3900	816	11	14
EA 05	705	2422	1226	296	71
EA 06	169	3511*	5415*	5	1
EA 08	146	634	1039	3	6
CDH 01	284	776	1208	27	102
CDH 02	200	419	3341	28	1
CDH 03	117	876	1470	2	1
CDH 04	145	612	1486	15	23

Table 3 WES-variant twin differences. Twin-sib differences at ≥ 20 X coverage with at least 2 variants, MAF ≤ 0.001, ncRNA and exonic regions, splicing, deletions, substitutions, stopgain, stoploss and nonsynonymous changes. If applicable including a statistical model and a repeat filter. CLC-bio variant calling allowed for multiple variant types at one location. * high frequency of several different variants per specific location in the affected sibling.

Visual inspection of the WES-CN profiles revealed a remarkable similarity in the general profile. Comparing the DNA of monozygous twins is as performing a duplicate experiment. With millions of measuring points, there are bound to be differences. These differences are in general not biological, but technical in nature. Differences detected in these experiments reflect the accuracy and limitations of current sequencing technologies, its downstream processes and variant analysis. Importantly, these experiments provide a better perspective regarding the strengths and limitations of this new technology.

In general, results in genome wide association studies are only significant when corrected for multiple testing, often with a Bonferonni correction ($p \leq 10^{-8}$). If we perform this statistical correction ($p \leq 0.05$ divided by the number of bases in the target, $50*10^6$) we achieve the same required order of significance, 10^{-8}. Of interest is that only when applying this order of significance, the discordant variants detected in high coverage, low repeat regions pat the visual examination of the reads in the IG viewer. The statistical difference test used is applicable on twins, but also on families and may be a good QC control parameter in NGS experiments.

Exon level copy number comparison

Using WES-CN we determined the exon level copy number profile of each individual. Dividing the each exon in three virtual probes decreased the noise and resulted in 100-150 exon-level copy number variants in each twin. These events are not validated with additional techniques. However, visual inspection of the probe profiles, both the logR ratio and the B-allele frequency revealed a remarkable similarity between the two twin samples. Even the more noisy sample (EA 06-1) had a similar waving as its twin duplicate. Most differences measured within the software, were actually marginal threshold differences. Only a few regions seemed to be really different. Of these more than half were only present in the healthy twin and completely overlap common Copy Number Polymorphisms. Validation of the remaining five discordant Copy Number Variations is ongoing.

Whole Exome variant analysis and twin comparison

At 20X coverage around 2-3 percent of protein coding and ncRNA variants are not measured in both twin-sibs. Marking bases not sequenced with "N" enables differentiation between "base not sequenced" and " base sequenced but equal to the reference" it can lead to over represention of the number of "N" in samples of reduced quality. This is clearly the case in EA 6-1. Most of the observed differences are the result of limitations of the capturing methods, sequencing technology, thresholds settings, coverage differences and limitations of the variant callers. A combined statistical and filtering approaches can rank these differences into two categories: 1) the difference has a high chance of being an artefact, most likely due to limits of the technologies and algorithms used or 2) The differences is likely to be real. When we finally used additional filtering steps and another variant calling method: SAM tools mpileup, a probabilistic method based on the Fisher exact test which takes into account allelic depth and allelic frequency and sub sequent filtering of repetitive regions. This resulted in only a few discordant variants left. Validation of the top 10 ranking significant differences for each twin-pair is ongoing, preliminary results indicate that the differences detected with exome sequencing and exome level copy number variation profiling are neither *de novo* nor different between twins.

Although no discordant loss of function mutations were detected using SAM tools mpileup, although there were numerous rare and private nonsynonymous SNV in candidate genes. Both CLC-bio callers detected possible discordant variants in candidate genes. However, visual inspection revealed that they were all present in the healthy sibling. Inheritance of variants from a healthy parent and the presence of these variants in the

healthy sibling does not exclude their pathogenicity. As with other low penetrance diseases, there are monozygous twins with the same genetic syndrome, yet a discordant phenotype. [2] Variants could contribute in a multifactorial manner, environmental or epigenetic factors could tilt the balance in the affected twin from normal to abnormal development. For instance, EA patient 3-1 has cardiac situs inversus and trachea-esophageal anomalies. Interestingly, the twin-sib (3-2) has dextrocardia. It is possible that a pathogenic mutation causes more severe anomalies in one twin and a somewhat less complex phenotype in the other. EA patient five has features (thumb and limb malformations) which could fit the observed *SALL1* mutation (Townes Brocks) and *FANC* mutation (Fanconi Anemia). Reduced penetrance and variable expressivity has been described for Fanconi anemia[25], although neither of the siblings has the characteristic skin pigmentations. Also Townes-Brocks syndrome has a variable phenotype, although most patients do have anorectal malformations and ear anomalies.[26] Patient CDH1-1 and CDH3-1 have a mutation in the Slit3 gene, mutations in this gene cause congenital diaphragmatic hernia in mice.[27]

Many variants were present multiple times in these, and other, samples. Variants in *VANGL1*, *NOTCH2*, *FANCD2*, *EFTUD2* and *RIPK4* were shared between EA twins and *KIF7*, *PTEN* and *RECQL4* between CDH twins. These most likely represent technological artefacts as they seem NGS pipeline specific. A burden test, using similar sequencing technologies, read mappers and variant callers on different patient cohorts could give more clues about the true nature and relevance of many pathogenic variants in candidate genes. Screening of large in-house cohorts of patients and healthy individuals using different capturing, sequencing and analysis pipelines seems a prerequisite to either filter out these likely technological errors in order to determine the "true" variant burden of candidate genes. Previously, we could not detect differences in CNV profiles in these two cohorts of discordant monozygotic twins.[12] In these current experiments we detected many differences using whole exome sequencing and WES-CN, although to date none of them could be confirmed with Sanger sequencing. Elucidating genetic factors by comparing the DNA of discordant monozygotic twin has proven successful in other occasions. [4, 7, 28] Absence of detected and validated discordant variants does not mean that they are not actually there. It could merely mean that we could not detect them using current technologies, software and experimental set-up. However, current results indicate that underlying etiological factors, at least in these twin pregnancies, are more likely to be non-genetic.

It is known that environmental components can differ in monozygous twin pregnancies and that twin pregnancies are perhaps not representative of regular singleton growth conditions.[1] Twin-to-twin transfusion syndrome (TTTS), one of the most serious complications of monochorionic multiple gestations, can result in these exposure differences. TTTS develops in 10 to 15 percent of monochorionic twins as a disbalance in arteriovenous anastomoses lead to asymmetrical flow resistance.[29] TTTS is associated with several congenital anomalies and can result in loss of the fetus. TTTS and undiagnosed loss of a twin fetus in early stage of pregnancy, a so called vanishing twin, is hypothesized to be an important etiological component of congenital anomalies as VACTERL association or gastro-intestinal atresia.[30, 31] Phenotypical descriptions of twins suffering from TTTS include IUGR and cardiac defects, which are present in EA01, EA08 and CDH04. TTTS is also described to be contributing to congenital anomalies as VACTERL association or gastro-intestinal atresia.[30, 31]

Genetic factors likely contribute to congenital anomalies as esophageal atresia and congenital diaphragmatic hernia. However, we could not detect discordant variants between affected and unaffected monozygotic twins. Although other factors - the twinning process itself, environmental factors, TTTS, or epigenetic modifications- could be the underlying condition responsible for the malformations observed in these twin pregnancies, care has to be taken to definitively exclude DNA discrepancies. At 20X coverage only 80 percent of the coding region was covered and non-coding DNA variation was not taken into account in this targeted approach. Furthermore, determining mosaic differences with Sanger sequencing is difficult, if not practically impossible. If a discordant heterozygous variant is *inherited* and present in both twins, albeit in different frequencies, this mosaic difference is most likely an technical artefact. However, confirmation of the *de novo* nature of the mutation, absence in the parental DNA, proves the relevance of the detected difference. If a *de novo* variant is detected, further studies are needed to investigate whether, and at what frequency, the *de novo* variant is present in different tissues of both twins.

Acknowledgements

The authors would like to thank M. Nellist for helpful discussion and C. Kockx, R. Veeris and L. van Steen for their technical assistance.

References

1. Hall, J.G., *Twinning*. Lancet, 2003. **362**(9385): p. 735-43.
2. Zwijnenburg, P.J., H. Meijers-Heijboer, and D.I. Boomsma, *Identical but not the same: the value of discordant monozygotic twins in genetic research*. Am J Med Genet B Neuropsychiatr Genet, 2010. **153b**(6): p. 1134-49.
3. van Dongen, J., et al., *The continuing value of twin studies in the omics era*. Nat Rev Genet, 2012. **13**(9): p. 640-53.
4. Bruder, C.E., et al., *Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles*. Am J Hum Genet, 2008. **82**(3): p. 763-71.
5. Breckpot, J., et al., *Differences in Copy Number Variation between Discordant Monozygotic Twins as a Model for Exploring Chromosomal Mosaicism in Congenital Heart Defects*. Mol Syndromol, 2012. **2**(2): p. 81-87.
6. Essaoui, M., et al., *Monozygotic twins discordant for 18q21.2qter deletion detected by array CGH in amniotic fluid*. Eur J Med Genet, 2013. **56**(9): p. 502-5.
7. Rio, M., et al., *Monozygotic twins discordant for submicroscopic chromosomal anomalies in 2p25.3 region detected by array CGH*. Clin Genet, 2013. **84**(1): p. 31-6.
8. Kaplan, L., et al., *Monozygotic twins discordant for neurofibromatosis 1*. Am J Med Genet A, 2010. **152a**(3): p. 601-6.
9. Vogt, J., et al., *Monozygotic twins discordant for neurofibromatosis type 1 due to a postzygotic NF1 gene mutation*. Hum Mutat, 2011. **32**(6): p. E2134-47.
10. Baranzini, S.E., et al., *Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis*. Nature, 2010. **464**(7293): p. 1351-6.
11. Solomon, B.D., et al., *Exome Sequencing and High-Density Microarray Testing in Monozygotic Twin Pairs Discordant for Features of VACTERL Association*. Mol Syndromol, 2013. **4**(1-2): p. 27-31.
12. Veenma, D., et al., *Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts*. Eur J Hum Genet, 2012. **20**(3): p. 298-304.
13. Veenma, D.C., A. de Klein, and D. Tibboel, *Developmental and genetic aspects of congenital diaphragmatic hernia*. Pediatr Pulmonol, 2012. **47**(6): p. 534-45.
14. Brosens, E., et al., *Locus and phenotype heterogeneity in patients with trachea-esophageal malformations and associated anomalies*. Journal of Medical Genetics, 2014.
15. Shaw-Smith, C., *Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology*. J Med Genet, 2006. **43**(7): p. 545-54.
16. de Jong, E.M., et al., *Non-VACTERL-type anomalies are frequent in patients with esophageal atresia/tracheo-esophageal fistula and full or partial VACTERL association*. Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 92-7.
17. Erlich, Y., *Blood ties: chimerism can mask twin discordance in high-throughput sequencing*. Twin Res Hum Genet, 2011. **14**(2): p. 137-43.
18. Amarasinghe, K.C., J. Li, and S.K. Halgamuge, *CoNVEX: copy number variation estimation in exome sequencing data using HMM*. BMC Bioinformatics, 2013. **14 Suppl 2**: p. S2.
19. Brosens, E., et al., *Whole-exome mutation and CNV analysis in familial esophageal atresia* manuscript in preparation, 2014.
20. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform*. Bioinformatics, 2010. **26**(5): p. 589-95.
21. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
22. Brouwer, R.W., et al., *NARWHAL, a primary analysis pipeline for NGS data*. Bioinformatics, 2012. **28**(2): p. 284-5.
23. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. Genome Res, 2010. **20**(9): p. 1297-303.
24. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data*. Nucleic Acids Res, 2010. **38**(16): p. e164.
25. Giampietro, P.F., et al., *The need for more accurate and timely diagnosis in Fanconi anemia: a report from the International Fanconi Anemia Registry*. Pediatrics, 1993. **91**(6): p. 1116-20.
26. Kohlhase, J., *Townes-Brocks Syndrome*, in *GeneReviews*, R.A. Pagon, et al., Editors. 1993, University of Washington, Seattle. All rights reserved.: Seattle (WA).
27. Yuan, W., et al., *A genetic model for a central (septum transversum) congenital diaphragmatic hernia in mice lacking Slit3*. Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5217-22.
28. Sasaki, H., et al., *Copy number loss of (src homology 2 domain containing)-transforming protein 2 (SHC2) gene: discordant loss in monozygotic twins and frequent loss in patients with multiple system atrophy*. Mol Brain, 2011. **4**: p. 24.
29. Sebire, N.J., D. Talbert, and N.M. Fisk, *Twin-to-twin transfusion syndrome results from dynamic asymmetrical reduction in placental anastomoses: a hypothesis*. Placenta, 2001. **22**(5): p. 383-91.
30. Pharoah, P.O., S.V. Glinianaia, and J. Rankin, *Congenital anomalies in multiple births after early loss of a conceptus*. Hum Reprod, 2009. **24**(3): p. 726-31.
31. Pharoah, P.O., *Causal hypothesis for some congenital anomalies*. Twin Res Hum Genet, 2005. **8**(6): p. 543-50.

PART 4

Next generation sequencing in familial and consanguineous patients

4.1

Molecular studies in familial and consanguineous patients with anomalies of the esophagus or other VACTERL associated anomalies

Adapted from:

Molecular studies in familial and consanguineous patients with anomalies of the esophagus or other VACTERL associated anomalies

E. Brosens, T. Brands, R.W.W. Brouwer, Y van Bever, A.S. Brooks, H. Eussen, R.M.H Wijnen, W..F.J. van IJcken, D. Tibboel, A. de Klein. *Ongoing research*

Abstract

Esophageal Atresia (EA) with or without Tracheo-Esophageal Fistula (TEF) is a relative common congenital anomaly whose cause is unknown in the majority of patients. These trachea-esophageal (TE) anomalies can be present either as an isolated defect or in association with other developmental defects: e.g. as one of the core features of the VACTERL (Vertebral, Anal, Cardiac, Tracheo-Esophageal, Renal and Limb anomalies) association. Furthermore, is TE a variable features in several known single gene disorders.

The hypothesis that genetic defects contribute to EA/TEF and VACTERL disease etiology is further supported by the high concordance rate in monozygotic twins compared to dizygotic twins and murine knockout models. TE and VACTERL association are usually sporadic findings and the familial recurrence rate of non-syndromic is low (1-3%).

We have collected 5 families with 2 patients with TE anomalies and 3 consanguineous trio's. Evaluation of a possible phenotype genotype association in these families will improve our understanding of the heterogeneous TE phenotype. Using high density SNP-array familial relationships were confirmed and their Copy Number Variations profiles and Runs of Homozygosity determined.. Using first a candidate gene approach and second a gene discovery approach Whole Exome Sequencing (WES) and Whole Exome Sequencing Copy Number profiling (WES-CN) identified possible pathogenic mutations and several candidate genes and loci.

Introduction

Esophageal Atresia (EA) with or without Tracheo-Esophageal Fistula (TEF) (MIM 189960) is a developmental defect of the foregut. This loss of continuity of the esophagus and connection of the trachea to the esophagus has a prevalence ranging between 1 and 4.5 per 10.000 live births.[1, 2] Males are more likely to be born with this condition than girls, this 3:2 gender disparity is hypothesized to be confounded by genetic and environmental factors.[3, 4] These Tracheo-Esophageal (TE) anomalies can be the sole malformation, although in approximately half of patients TE anomalies are associated with other congenital defects. Frequently the associated malformations are those of the VACTERL spectrum of Vertebral, Anorectal, Cardiac, Tracheo-esophageal, Renal or urinary tract and Limbs malformations.[5, 6] Other anomalies are also common e.g. microcephaly, duodenal atresia, single umbilical artery, micrognathia, pyloric stenosis and genitourinary malformations.[5, 7]

EA/TEF is variable feature in more than 70 genetic syndromes.[8] In some syndromes a TE defectis a frequent observed phenomenon, for instance in Microphthalmia and esophageal atresia syndrome (*SOX2*), Feingold syndrome (*MYCN*), CHARGE (Coloboma, Heart anomaly, choanal Atresia, Retardation, Genital and Ear anomalies) syndrome (*CHD7*) and in trisomy 13. However, in most syndromes, EA and or TEF are rather incidental findings. Further evidence for a genetic background for EA/TEF are the elevated probandwise concordant rate in monozygotic twins ion comparison that of dizygotic twins[9], murine knockout models[10] and the presence of *de novo* genetic aberrations of variable size in sporadic EA/TEF. [10-13] Familial recurrence rate is low (1-3%) and TE and VACTERL association are usually sporadic findings.[14, 15] Therefore, most likely dominant *de novo* or X-linked recessive mutations are responsible for most of the patients with a genetic etiology although autosomal recessive forms as Fanconi anemia cannot be excluded.

In this study we aim to identify causal genetic factors in a specific subset of patients with TE-anomalies. By evaluating of a possible phenotype genotype association in the familial TE-patients and TE patients born from consanguineous marriages we hope to identify genes or loci to explain the heterogeneous TE phenotype seen in these families.
We have collected 5 families with TE anomalies and three patient-parent trio's from consanguineous marriages. Using high density SNP-array familial relationships were confirmed and their Copy Number Variations profiles and Runs of Homozygosity (ROH)

determined. Homozygosity mapping is a method which can be used to detect recessive mutations in both inbred[16] and outbred[17] populations. With an online prioritarization tool, "the genomic oligoarray and SNP-array evaluation toolv2.0"[18], we searched for known autosomal recessive disorders in ROH regions. In addition we have used a Next generation sequencing and by using both a candidate gene approach or a gene discovery approach we applied Whole Exome Sequencing (WES) and Whole Exome Sequencing Copy Number profiling (WES-CN) to identified possible pathogenic mutations in candidate genes and loci.

Methods

Familial and consanguineous TE-patients in the Erasmus MC-Sophia TE cohort

This study was approved by the Medical Ethical Review Board of Erasmus MC - Sophia Children's Hospital. Clinical- and follow-up data were extracted from medical charts. Unless otherwise indicated, none of the included patients had a previous confirmed genetic syndrome, known chromosomal anomaly, pathogenic point mutation nor was there evidence for an association with an environmental components. Since 1988 onwards, patients with TE are included in our Erasmus University MC-Sophia TE-cohort (n=582). An overview of this cohort has been described previously.[5, 8]

In the Erasmus MC-Sophia TE cohort there are ten familial TE patients (1.9%) in five families (FAM1-5). Informed consent and DNA material was available for genetic studies of eight of these patients (see figure 2) . Here, we define "familial EA" as those families in which an index patient has another reported family member, either deceased or alive, with EA and/ or TEF. No genetic syndrome should be present, which could explain the phenotype segregating through the pedigree.

Although suspected in more patients, there were six confirmed consanguineous patients in our cohort (1%). Informed consent and parental DNA material for genetic studies was available for four patients. One consanguineous patient had Fanconi anemia, a diagnosis confirmed with chromosome breakage studies and mutation analysis. This patient was excluded from further analysis. The three remaining patients and their related parents were analyzed (FAM C1-3)

Copy Number Variation profiling with SNP-array

Using high density SNP-array genotype information familial relationships were confirmed and their Copy Number Variation profiles and Regions of Homozygosity (ROH)

determined. Micro-arrays were processed using the Illumina Human Omni Express-12 version 1.0 Bead Chip (Illumina, Inc., San Diego, USA) according to their manufacturer's standard protocol. Normalized output was generated with Illumina's Genome Studio program version 2011.1 (Illumina, San Diego, CA, USA) and visualized in Nexus CN7.1 (Biodiscovery Inc, El Segundo, CA, USA)

In general, inheritance of CNVs was determined if they were larger than 30kb, contained genes and were either absent from or had a low frequency in the database of genomic variation.[19] Variants absent from this database were marked as "private CNV" and if present 5 times or less as "rare CNV", all others as common polymorphisms. This 5 count cut-off is indicative of an allele frequency below 0.0025. Regions were considered to be homozygous and marked as ROH when they contained 10 probes or more and were larger than 2Mb in size. If private and rare CNV were seemingly absent in the parental CNV profiles, they were sub sequentially validated with a second technique; either MultiPlex Amplicon Quantification (MAQ) or WES-CN.

Copy Number Variation profiling with Exome-NGS

Whole-Exome Copy Number (WES-CN) was determined using CoNVEX version 0.6.[20] This program uses a Hidden Markov model, similar to SNP-array, to determine the copy cumber status from the normalized depth of coverage (DOC) relative that of a cohort of reference samples. Exonic DOC was sub divided in three virtual probes and segmentation was performed in Nexus CN7.1. Allelic information derived from GATK-variant calling was transformed to a B-allele frequency track to increase reliable copy number calling. The Nexus CN7.1 software was used to manually compare SNP-array results with the WGS-CN results Moreover, we compared the WGS-CN events to the exome sequencing variants and searched for overlap of these events with loss of function mutations with an allele frequency ≤ 0.001.

MultiPlex Amplicon Quantification

Putative de novo private and rare CNVs detected with SNP-array are validated with the MAQ-assay. (Multiplicom Inc., Niel, Belgium) as described before[21] With the MAQ-S software the fluorescent intensities (translated into peak heights) of patient and reference sample are compared and visualized in dosage plots.

Whole Exome Capture and Sequencing

First, genomic DNA was sheared using Covaris adaptive focussed acoustic technology. (Covaris, Inc. Woburn, Massachusetts, USA) Fragmentation yield and fragment

size distributions are measured with the Bioanalyser 2100 bioanalysis chips and Agilent DNA 1000 Kit. (Agilent BioAnalyzer, Santa Clara, CA). The protein coding regions of the DNA were captured with SureSelect Human All Exon 50 Mb Targeted exome enrichment kit v4 (Agilent Technologies, Inc., Santa Clara, California) and sequenced using Illumina TruSeq version 4 paired end 2x 101 bp sequence procedure on the Hiseq2000 (Illumina, Inc., San Diego, USA).

Whole exome Sequencing quality control, alignment and variant calling

At least 5 Giga-bases of raw sequence data was mapped to the reference genome (hg19) using two independent methods: the NARWHAL pipeline[22] and CLC-bio readmapper. (Qiagen Inc., Venlo, the Netherlands) NARWHAL automates de-multiplexing, alignment to the hg19 reference genome with Burrows-Wheeler Aligner version 0.6.2 [23] and curation of low quality reads. Chromosome sorted BAM-files are generated with SAMtools version 0.1.12a [24] and quality control is performed on these sorted BAM files. The mean coverage over the target was at least 50X with over 80% of bases >20X. Most samples however, were >70X with >90% of target bases covered >20X.

Genotyping was done with the Bayesian genotyper incorporated in the genome analysis toolkit version 1.2.9 [25] and SAMtools mpileup in combination with an inhouse probabilistic differentiator, variants were annotated with ANNOVAR version 2013-feb-21.[26] In the second method, reads are aligned and quality control is performed using CLC-bio(Qiagen Inc., Venlo, the Netherlands). Variants are called and annotated using a quality based and a probabilistic method with variant callers and annotation tools incorporated in the CLC-bio package. Variant filtering was done with Cartagenia Bench NGS version 3.04 (Cartagenia Inc, Boston, MA, USA) and TIBCO Spotfire version 5.5.0.36 (TIBCO, Boston, MA, USA).

Whole Exome Sequencing: candidate gene approach

First step in variant analysis is a candidate gene approach using a gene panel specific for TE-anomalies and VACTERL association. Candidate genes were derived from human diseases, animal models and developmental pathways [27] using online databases. [28-30] These genes and the gene panel are described previously [8, 31] and a summary is listed in table 1. Variant filtering and prioritization settings are described in figure 1 and the gene discovery approach section. Private or rare (At least 20X coverage, MAF≤0.001) de novo or X-linked protein disturbing (splicing, frameshift, stopgain or loss) variants in these candidate genes were evaluated. We did not evaluate other rare possible protein disturbing

inherited variants in candidate genes in this first phase of the analysis. These variants, if fitting inheritance models, would also be detected in the gene discovery approach. Rare and private variants in the gene panel are listed in supplementary table 4 and are available on request.

Candidate genes	Tracheo-esophageal anomalies and VACTERL association	Ref
Diagnostic set of genetic syndromes	*CHD7, MYCN, SALL1, MID1, EFTUD2*	[8, 31]
Research Genetic syndromes	Over 140 genes from genetic syndromes in which TE anomalies have been described or with overlap with VACTERL association (see supplementary figure 1)	[8, 31]
Animal models	*IFT172, Rarα/ Rarβ , Nkx2.1, Rab25, Hoxc4 , Chrd, Ctnnb1, Dync2h1, Efnb2, Foxp4, Fuz, Lec, Sox4, Wdr35, Foxp2, Foxp1*	[8, 30]
Pathways	*Bmp7, Barx1, Rhou, Wnt7b, fgf10, Bmp4, tp63, Keap1, Wny5a, Wnt11, Wnt2, Wnt2b*	[27]
De novo CNV in sporadic patients	*LPP, GTPBP5, EPPK1, PLEC-1, PARP10, AATF, TADA2A,GSTP1,MAP2, HNF1B,EFNB2*	[32-39]

Table 1 Candidate gene panel for Tracheo-esophageal anomalies and associated VACTERL spectrum anomalies

Whole Exome Sequencing: gene discovery approach

If no obvious pathogenic mutation was detected in the candidate gene panels, the entire exome was studied. Variant filter settings were: coverage ≥ 10X, MAF ≤ 0.1% in 1000G or ESP6500, absence in dbSNP135, inheritance model, presence or absence in healthy/affected family members, possible splicing effects, exonic or ncRNA exonic variant location and finally the function consequence of the mutation. (ncRNA affected, nonsynonymous, frameshift, stopgain, stoploss) Prioritization steps include whether the detected variants were there present in relevant genes, absent in segmental duplications and their population frequency (MAF ≤ 0.001) We also evaluated their predicted deleteriousness and conservation using GERP (>3.0), PhyloP (>0.95), SIFT (>0.95), Polyphen2 (>0.85) and mutation taster (A or D).

All inheritance models were performed, if multiple family members were affected, the variation had to be shared and had to fit logical inheritance models. Fisrt shared, loss of function mutations following recessive, compound heterozygous or X-linked models were considered. Next dominant de novo mutations in individual patients were considered. These variants were also compared with the exon level CNV to exclude unmasking of these variants. Finally, nonsynonymous variants involved in relevant biological processes were considered if fitting logical inheritance models. Variation was compared with animal

phenotypes from the MGI database, human syndromes in the OMIM database and gene function or presence in relevant pathways.

De novo, homozygous recessive and X-linked rare and private variants predicted to be conserved are displayed in the supplementary figures. For displaying purposes only those compound heterozygous variants are displayed of which at least one was conserved and predicted to be pathogenic by all prediction tools. These in silico prediction tools were used to prioritize variants and not as an exclusion criterion for possible follow up analysis. A scheme of candidate gene and gene discovery approach is depicted in figure 1. All inheritance model filtering strategies were performed on every family e.g. homozygous recessive and compound heterozygous mutations, all distinct *de novo* mutations as 'well as X-linked inheritance in male patients.

Candidate variant confirmation

Variants of interest were confirmed with illumina's Exomev1.1 genotyping chip or Sanger sequencing. Primers were developed using the online primer3 software package, available on http://frodo.wi.mit.edu/. After Touch-Down PCR the amplicons were purified using ExoSAP-IT (USB, Staufen, Germany) and sequenced using BigDye Terminator chemistry v3.1 on the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The resulting sequence reads were aligned to the hg19 reference sequence using SeqScape software version 2.6 (Applied Biosystems). Primer sequences are available on request.

Results

Familial and Consanguineous TE-patients in the Erasmus MC-Sophia TE cohort

In the first family two brothers were born from seemingly healthy parents (Pedigrees are given in figure 2.). The oldest brother (patient 1) has isolated EA/TEF with and a patent ductus arteriosus. His more severely affected younger brother(patient 2) has in addition to EA/TEF, hemivertebrae, cryptorchid testes, mild ectasia pyelum, inguinal hernia, Ureteropelvic junction stenosis, and a sacral dimple. In family two both patients have EA/TEF. The nephew (patient 4) of a girl (patient 3) with an isolated EA/TEF has hypoplastic proximal paced thumbs in addition to his esophageal atresia.

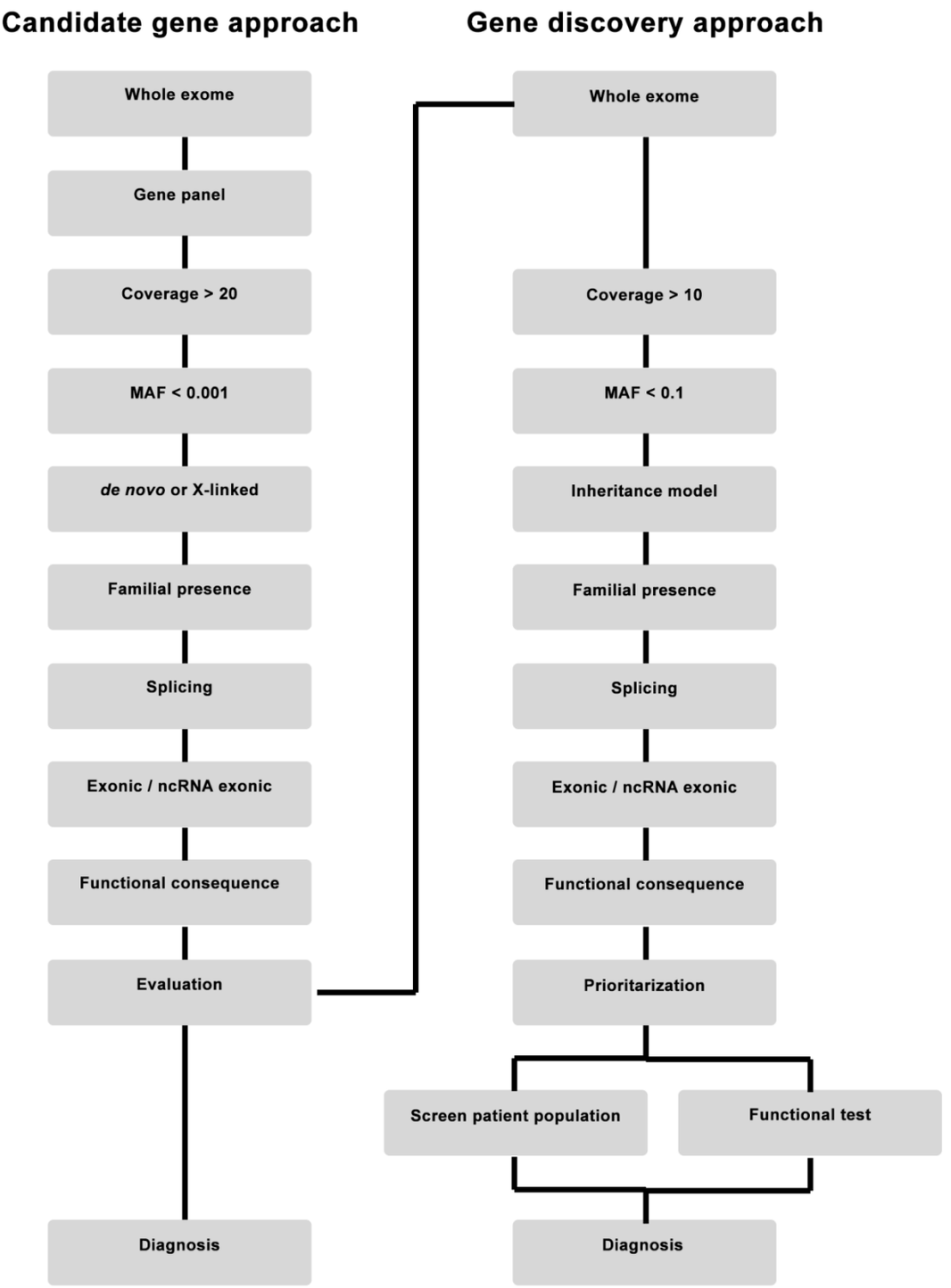

Figure 1. NGS variant analysis, filtering and prioritization scheme.

The patients in family three and four have deceased, more distant related, family member with TE anomalies. No information was currently available on congenital malformations other than the tracheo-esophageal anomalies. The boy(patient 5) from family three has EA/TEF, ventricular septal defect, an abnormal connection between the bronchus and his trachea, sandal gap of toes, radial deviation of his hand and hypoplastic proximal placed thumbs. The affected girl (patient 6) from family four has EA/TEF, a right-sided aortic arch and a ventricular septal defect.

Family 5 has two affected family members: One boy (patient 7) with EA/TEF, anal atresia and a ventricular septal defect and his uncle (patient 8)is nephew has EA/TEF, anal atresia, low-set ears, hypospadias, a dislocated dysplastic hip, renal agenesis of the left kidney and dysplasia of the right kidney, double ureters, urinary reflux, Urethral fistulae and urethral stenosis. No genetic syndromes were confirmed in these patients prior to analysis. The 22q11 duplication in family 5 did not segregate with the Esophageal atresia phenotype.

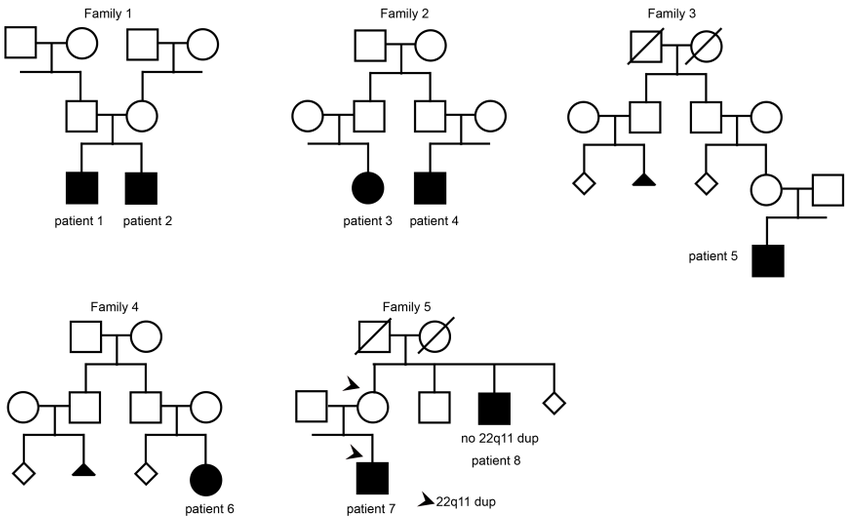

Figure 2 Pedigrees of familial EA/TEF patients

Patient 9 is a male born from consanguineous parents and with EA/TEF, abnormal dermatoglyphic patterns and plantar creases, clinodactyly, hypospadias, pronounced nipples, hypoplastic toe nails and a short chin and neck. His parents are second degree relatives. Patient 10 is a boy born from consanguineous parents (2^{nd} degree relatives) with hemiverteabrae, tethered cord, horshoe kidneys, atrial and ventricular septal defect. Patient 11 is a boy and his parents have a common ancestor 5 -6 generations back. He has EA/TEF, anal atresia, Meckel diverticulum, an abnormal sacrum with a sacral dimple, coronal clefts of vertebrae, an asymmetric square shaped face, prominent forehead,

anomalous venous return, coarctation, patent ductus arteriosus and an abnormal arterial supply to his right lung. Pedigrees of these consanguineous families are given in figure 3

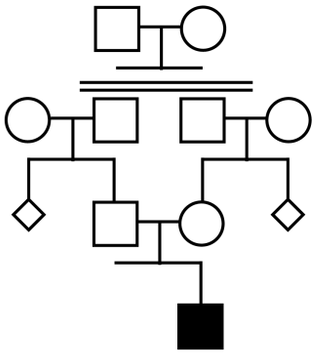

Figure 3 Example pedigrees of a consanguineous EA/TEF patient

Copy Number Variation profiling

Genotyping confirmed parental relationships. Several rare and private CNV were identified in each familial patient using SNP array but all of these CNV . were inherited from unaffected parents (see supplementary table 2). In patient 11 one private CNV covering a gene from the gene panel, FGFR2, was detected. This CNV was present in both his father and his mother. In addition to SNP-array CNV profiling, WES-CN revealed numerous smaller CNV events. Most events were also present in one of the parents, although numerous exon level events seemed de novo. None of the small seemingly de novo exon level WES-CN events covered genes from the candidate gene panel. WGS-CN did however confirm the CNVs containing genes seen in the SNP-array. The number of these putative de novo events (≥ 50-100 per exome on average) suggests that many represent background noise or thresholding issues. The later seems likely since, the segmentation patterns are very constant and also seen in other non EA affected individuals.. More experience and refinement of thresholds, analysis settings and perhaps ConVEX algorithm is needed to allow for hypothesis free exon level WES-CN analysis. Of note, a *de novo* polymorphism on chromosome 15q11 was detected in patient 9. (see figure 4)

There were more variants detected in genes involved in rare and private CNVs. In all members of family 1 a nonsynonymous mutation was detected in exon 2 of the Deoxyribose-Phosphate Aldolase (*DERA*) gene. This gene ispredicted to have an effect on splicing has 9 exons. Patients 1 and 2 have a maternally inherited private loss of the loss of the last three exons of this gene. However, this mutation was present in a low frequency in

all sequenced samples and likely represents a technical artifact. Patient 3 and both her parents possess a rare variant in the Maestro Heat-Like Repeat Family Member 5 (*MROH5*) gene.. In patients 3, this gene is involved in a rare, maternaly inherited, gain in DNA copy number.

There were no Runs of homozygosity larger than 5Mb in the familial patients. (see supplementary table 3) However, these ROH events were abundant in the consanguineous patients. No large (≥ 2 Mb) ROH was detected in patient eleven, likely reflecting the multiple generations to the common ancestor. Relative large (>2Mb) runs of homozygosity and are depicted in supplementary table 3. The ROH regions larger than 2Mb detected with SNP-array were inserted genomic oligoarray and SNP-array evaluation toolv2.0[18] and screened for recessive candidate syndrome genes using Human Phenotype Ontology terms esophageal atresia or tracheo-esophageal fistula. No autosomal recessive genetic disorders were highlighted in patient nine, one gene was highlighted in patient ten: *WNT7a* on chromosome 3p25 associated with the autosomal recessive Al-Awadi/Raas-Rothschild (AARR) syndrome. (#OMIM 276820) but no disease causing variants were detected in this gene.

Whole Exome Sequencing: candidate gene approach

Interestingly, the use of two independent read-mapping methods, the NARWAL pipeline and CLC-bio readmapper and different variant calling algorithms (GATK Unified genotyper, CLC-bio®, SAMtools pileup) resulted in the detection both concordant and discordant cross-platform variants. Candidate variants from all pipelines were combined and used for further analysis.

The candidate gene approach resulted in the identification of numerous rare and private putative protein disturbing variants in these candidate genes. (listed in table 2 and supplementary table 4) Almost all of them were heterozygous and inherited from an unaffected parent. Many heterozygous variants were also present in both parents and even present in all samples and could reflect technical artifacts. No possible *de novo* or X-linked variants in these genes could be identified in family 1,3 or 4. In family 2, two independent putative *de novo* variants are identified. In patient 3 a nonsynonymous variant in the *PAX2* gene (NM_000278:c.A380C:p.D127A) and in patient 4, a putative *de novo* variant in the *MYCN* gene. (NM_005378:c.C1177T:p.R393C). Also for the consanguineous patients, the candidate gene approach also resulted in the identification of numerous rare and private possible protein disturbing variants in candidate genes. (listed in supplementary table 4) All

of them were heterozygous, mostly inherited and oftenpresent in both parents. No rare or private homozygous recessive *WNT7A* mutations were detected in patient 10. However, a putative de novo mutation was detected in the Fibrillin 2 (*FBN2*) gene (NM_001999:c.A8014C:p.T2672P). resequencing of patient 7 and 11 is ongoing.

Whole Exome Sequencing: gene discovery approach

We did not detect any homozygous recessive variants (table 4 and supplementary table 5) in these esophageal atresia patients, which we can associate with trachea-esophageal development. Closer inspection of most detected heterozygous *de novo* variants in the broad institutes IG viewer indicated that these were in fact inherited or possible technical artefacts. We did however detect rare and private heterozygous inherited variants which could influence the patients phenotype following compound heterozygous and X-linked recessive models. Little homozygous recessive candidate variants were detected in the consanguineous patients. After stringent filtering, only one of the homozygous recessive variants was in the large homozygous stretches (table 3). This was a homozygous stopgain mutation in the prenylcysteine oxidase 1 (*PCYOX1*) gene in patient 10. There was shared compound heterozygous variation in dynein genes *DNAH2*, *DNAH3* and *DNAH11* in familial patients 1 and 2, DNAH17 and DYNC2H1 in patient 5 and DNAH17 in patient 6. Consanguineous patients 9 (DNAH8 and9) and 10 (*DRC1*, *DNAH1*, *DNAH5*, *DNAH9*)are compound heterozygous for several dynein proteins. These variants did not pass the conservation/pathogenic prediction filter, but all had a MAF below 0.001.

In summary, using high density SNP-array and WES-CN we detected rare or unique CNV in each patient. Large ROH, absent from parental samples, was observed in the consanguineous patients. The candidate gene approach identified a de novo *MYCN* mutation in one of two affected family members of family 2. Moreover, many inherited, heterozygous rare and private nonsynonymous variants were present in genes from the candidate gene panel. We could at this point not identify overlap of loss of function mutations and exon level CNV in genes from the candidate gene panel. In general, shared compound heterozygous or recessive variation between the affected family members was rare. In the consanguineous patients homozygous candidate variants were detected, although most of them were located outside the large ROH regions detected with SNP-array.

patient	Gene	AAChange	AVSIFT	LJB_PhyloP	LJB_PhyloP_Pred	LJB_SIFT	LJB_SIFT_Pred	LJB_PolyPhen2	LJB_PolyPhen2_Pred	LJB_LRT	LJB_LRT_Pred	LJB_MutationTaster	LJB_MutationTaster_Pred	LJB_GERP..	ref	alt	inheritance
3	*PAX2*	NM_000278:c.A380C:p.D127A	0	0.999084	C	1	D	0.993	D	1	D	0.999991	D	5.58	A	C	de novo
3	*GLI3*	NM_000168:c.T2614C:p.S872P	0.04	0.997188	C	0.958516	NA	0.998	D	0.999999	D	0.999258	D	4.5	A	G	paternal
4	*MYCN*	NM_005378:c.C1177T:p.R393C	0	0.998798	C	1	D	1	D	1	D	0.999992	D	4.86	C	T	de novo
4	*FOXP4*	NM_001012426:c.C76A:p.Q26K	0.09	0.998545	C	0.8	T	0.99	D	0.999948	D	0.99554	D	5	C	A	paternal
4	*LPP*	NM_001167672:c.G431C:p.S144T	0.09	0.992553	C	0.75	T	0	B	0.998635	N	0.00024	N	4.22	G	C	paternal
5	*FRAS1*	NM_001166133:c.4566delT:p.P1522fs	-	-	-	-	-	-	-	-	-	-	-	-	C	-	paternal
10	*FBN2*	NM_001999:c.A8014C:p.T2672P	0.01	0.997812	C	0.99	D	0.966	D	0.999995	D	0.944498	D	4.69	T	G	de novo
10	*PARP10*	NM_032789:c.C2446G:p.P816A	0	0.999308	C	0.99	D	0.6	P	0.955811	N	0.138334	N	4.27	G	C	maternal

Table 2. Potential pathogenic variants in the candidate gene approach. Five variants were inherited from unaffected parents and two variants were de novo. All variants except one were nonsynonymous. (*splicing). Variants present in both parents are listed in the full table (supplementary table 4)

Patient	Gene	AAChange	AVSIFT	LJB_PhyloP	LJB_PhyloP_Pred	LJB_SIFT	LJB_SIFT_Pred	LJB_PolyPhen2	LJB_PolyPhen2_Pred	LJB_LRT	LJB_LRT_Pred	LJB_MutationTaster	LJB_MutationTaster_Pred	LJB_GERP..	ref	alt	Inheritance
9	ELMO2	NM_182764:c.C1470A:p.N490K	0.05	0.999525	C	0.46	T	0.001	B	0.99993	D	0.978091	D	4.53	G	T	homozygous recessive
9	FNIP1	NM_001008738:c.C2702G:p.T901S	0.29	0.96794	C	0.8	T	0.418689	NA	0.850073	NA	0.00002	N	3.42	G	C	homozygous recessive
9	REL	NM_002908:c.A206G:p.N69S	0.09	0.976613	C	0.91	T	0	B	1	N	0.846676	D	3.56	A	G	homozygous recessive
10	PCYOX1*	NM_016297:c.C334T:p.Q112X	0.87	0.937199	N	0.901013	NA	0.701616	NA	0.865987	N	1	A	3.89	C	T	homozygous recessive

Table 3 . Homozygous recessive variants in consanguineous patients (gene discovery approach) All variants except one were nonsynonymous. (*splicing). Full table in supplementary table 5

Patient	Gene	AAChange	AVSIFT	LJB_PhyloP	LJB_PhyloP_Pred	LJB_SIFT	LJB_SIFT_Pred	LJB_PolyPhen2	LJB_PolyPhen2_Pred	LJB_LRT	LJB_LRT_Pred	LJB_MutationTaster	LJB_MutationTaster_Pred	LJB_GERP..	ref	alt	Inheritance
3	ATP11C	NM_001010986:c.G2267T:p.G756V	0	0.998709	C	1	D	0.999	D	1	D	0.999986	D	5.12	C	A	de novo
3	TTC16	NM_144965:c.C1013T:p.A338V	0	0.998688	C	0.99	D	0.991	D	0.99999	D	0.634657	D	5.11	C	T	de novo
4	BCL6	NM_001130845:c.T1682C:p.L561P	0	0.999115	C	1	D	0.998	D	1	D	0.999982	D	5.29	A	G	compound heterozygous
4	BCL6	NM_001134738:c.A34C:p.T12P	0	0.998462	C	0.99	D	0.908	D	1	D	0.999385	D	5.26	T	G	compound heterozygous
4	FHDC1	NM_033393:c.C409A:p.L137I	0	0.980859	C	0.96	D	0.993	D	1	D	0.824498	D	3.92	C	A	de novo
4	HERC4	NM_015601:c.C2962T:p.R988C	0	0.999708	C	1	D	1	D	1	D	0.999999	D	5.43	G	A	de novo
4	IDH3G	NM_004135:c.A560G:p.E187G	0	0.998164	C	1	D	1	D	1	D	0.986759	D	5.4	T	C	compound heterozygous
4	IDH3G	NM_004135:c.T557G:p.V186G	0	0.998881	C	1	D	0.999	D	1	D	0.994419	D	5.4	A	C	compound heterozygous
4	IDH3G	NM_004135:c.T554G:p.V185G	0	0.998881	C	1	D	1	D	1	D	0.866292	D	5.4	A	C	compound heterozygous
4	IDH3G	NM_004135:c.C160G:p.R54G	0	0.999058	C	0.99	D	0.881	D	1	D	0.999548	D	4.56	G	C	compound heterozygous
5	PPCS	NM_024664:c.T110G:p.V37G	0	0.968379	C	1	D	0.988	D	0.999993	D	0.999266	D	3.92	T	G	compound heterozygous
5	PPCS	NM_024664:c.T113G:p.V38G	0	0.969663	C	1	D	0.967	D	0.999935	D	0.999973	D	3.93	T	G	compound heterozygous
9	FOXI1	NM_012188:c.A506C:p.N169T	0	0.998242	C	1	D	0.998	D	1	D	0.999995	D	4.42	A	C	compound heterozygous
9	FOXI1	NM_012188:c.T508C:p.S170P	0	0.997227	C	1	D	0.997	D	1	D	0.999994	D	4.42	T	C	compound heterozygous
9	PLXNA2	NM_025179:c.T2494G:p.C832G	0.01	0.998588	C	1	D	1	D	1	D	0.999919	D	5.17	A	C	compound heterozygous
9	PLXNA2	NM_025179:c.T2476G:p.C826G	0	0.998588	C	1	D	1	D	1	D	0.999914	D	5.17	A	C	compound heterozygous
9	PLXNA2	NM_025179:c.T1609G:p.C537G	0	0.996938	C	1	D	1	D	1	D	0.999978	D	3.88	A	C	compound heterozygous
9	SGSH	NM_000199:c.A685C:p.T229P	0.01	0.995812	C	0.99	D	0.995	D	1	D	0.999247	D	4.01	T	G	compound heterozygous
9	SGSH	NM_000199:c.A146C:p.H49P	0.04	0.989649	C	0.96	D	0.998	D	0.999993	D	0.993561	D	3.59	T	G	compound heterozygous
9	SRPX	NM_001170750:c.C409T:p.R137W	0	0.96319	C	0.99	D	0.986	D	0.999949	D	0.997251	D	3.29	G	A	X-linked
9	WDFY3	NM_014991:c.A2759C:p.H920P	0	0.99849	C	0.96	D	0.999	D	1	D	0.999915	D	5.6	T	G	compound heterozygous
9	WDFY3	NM_014991:c.A2750C:p.H917P	0	0.99849	C	1	D	0.999	D	1	D	0.999987	D	5.6	T	G	compound heterozygous
10	GRAMD4	NM_015124:c.T1412C:p.L471P	0	0.994126	C	1	D	0.996	D	0.999963	D	0.969509	D	4.18	T	C	de novo
10	SLC18A1	NM_003053:c.G59T:p.R20L	0	0.982259	C	0.99	D	0.972	D	0.999999	D	0.972412	D	3.93	C	A	de novo

Table 4 . Variants, predicted in conserved loci and predicted to be pathogenic by all prediction tools. (gene discovery approach) Full table in supplementary table 5

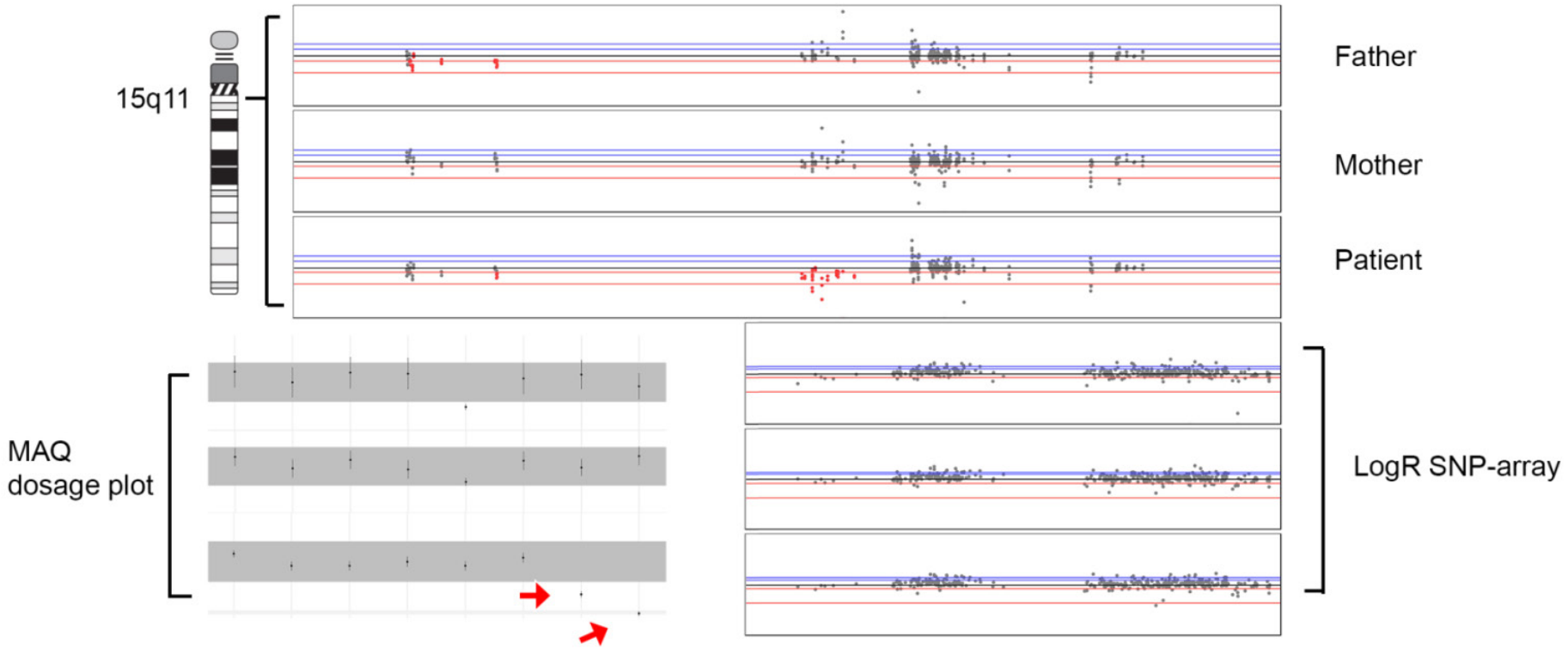

Figure 2. One polymorphism , in the 15q11 region was detected with WES-CN. In the upper panel red dots indicate loss, two MAQ assay probes (red arrow) inside the deleted region, confirm this loss. SNP-array resolution was not enough to call this loss. One rare variant (brown dot, lower panel) is present in this loss, however it is present in all samples.

Discussion

Candidate gene approach

The candidate gene approach, using a gene panel containing the genes from syndromes with TE features and knockout models is a first tier approach to identify variants that are most likely involved in the disease phenotype. This approach resulted in the identification of numerous rare and private possible protein disturbing variants in candidate genes (see table 2 and supplementary table 4). We started with validating the most intriguing genes and these genes and CNVs will be discussed below, further validation is ongoing)

All except the putative *de novo* *PAX2*, *MYCN* and *FBN2* mutations are inherited, nonsynonymous, heterozygous and often predicted to be deleterious. Although patients with mutations in *PAX2* and *FBN2* can have EA/TEF[8], the EA/TEF frequency in patients with papillorenal syndrome, CAKUT (*PAX2*) or Beals syndrome (*FBN2*) is low . Both affected members of family 2 had a putative de novo variant, albeit in a different gene from the candidate gene panel. Patient 3, had a *de novo* mutation in the paired box 2 gene (*PAX2*). PAX2 is primarily expressed in the developing nervous system and kidneys.[40] Heterozygous mutations in this gene can cause papillorenal syndrome or renal hypoplasia. This mutation is predicted to be pathogenic by all prediction tools and affects a conserved sequence. However, patient3 has isolated EA/TEF and no eye or renal anomalies, characteristic of papillorenal syndrome or renal hypoplasia. She also has a paternal inherited *GLI3* mutation, although GLI3 is a transcriptional repressor of PAX2.[41] and most likely would enhance *PAX2* expression and not inhibit normal expression.

However,Q-RT PCR or other *in vivo* studies are needed to exclude any (synergetic) effect of these two mutations. Iin her nephew, patient 4 a de novo variants was identified in the *MYCN* gene. Mutations in *MYCN* are known to cause the autosomal dominant Feingold syndrome, characterized by microcephaly, duodenal atresia, mental retardation and all VACTERL spectrum anomalies including esophageal atresia and hypoplastic thumbs [42] Two of these features are also present in patient 4r: hypoplastic proximal paced thumbs and trachea-esophageal malformations , which makes a Feingold diagnosis in this patient plausible. This MYCN mutation is not present in his niece suggesting that these two TE cases have no common etiology...

In male patient 5 an paternal inherited *FRAS1* mutation was identified. This splicing mutation had no clinical effect in his father since only homozygous or compound heterozygous mutations in this gene cause Fraser syndrome. Moreover, there was little overlap with the typical Fraser syndrome phenotype (no eye abnormalities, syndactyly or genitourinary anomalies).

Patient 7 has a maternal inherited 22q11 microduplication. This patient has a severe phenotype; all 6 VACTERL anomalies were present . In contrast, his mother, had only mild dysmorphic features. Adenovo 22q11 duplication as possible cause of VACTERL has been e described before either as a micro duplication [43] or in case with an unbalanced translocation.[44] An altered expression of the *TBX1* gene, responsible for the phenotype in DiGeorge syndrome [MIM 188400][45] could explain the patients phenotype. However, this duplication does not segregate with the TE-phenotype in this family, since it is absent in his uncle, patient 8 . Sequenc analysis and conformation of observed common variant in f this family is ongoing.

Patient 9 is compound heterozygous for the *PLEC* gene. A previously reported patient with esophageal atresia, vertebral , anorectal, cardiac and urogenital anomalies has a *de novo* duplication of the plectin gene region[33], an impact on trachea-esophageal development from these compound heterozygous mutations needs further investigation.

In consanguineous patient 11 one private CNV involving part of the intronic region of the *FGFR2* gene was detected. Homozygous mutations in this gene cause autosomal recessive Antley-Bixler syndrome without genital anomalies or disordered steroidogenesis (OMIM #207410). However it has also been suggested that also heterozygous mutations could cause this syndrome.[46] These patients have craniosynostosis, facial anomalies, limb malformations, choanal stenosis or atresia and joint contractures.[47] Incidentally, patients with Antley-Bixler syndrome have esophageal atresia.[48] This CNV was inherited and present in both parents. However, this gain was present in both the unaffected, parents. Sequencing is ongoing and the analysis whether this duplication causes an altered or loss of FGFR2 protein remains to be confirmed. One common CNV polymorphism, loss of the 15q11 region was detected with WES-CN. This loss, not detected with SNP-array, is implicated in patients with developmental and congenital anomalies including esophageal atresia and trachea-esophageal fistula.[49] This region is deleted in three additional esophageal atresia patients in our cohort.[8] However, it is packed with repeat elements e.g.

LINE, SINE and others making validation difficult this could be like the more common 16p13 region a recurrent CNV predisposing for developmental anomalies (ref giririja)

Gene discovery approach

Familial TE

In family 1 no genes from the candidate gene panel were affected by rare or private variations. Since there is no information regarding other affected family members in this family where both sons are affected and both parents are healthy, several inheritance models could be plausible. Therefore, we considered X-linked recessive, homozygous recessive and compound heterozygous as plausible inheritance models. One rare X-linked recessive variants was present, a rare nonsynonymous SNV in the retinoic acid induced 2 (*RAI2*) gene. This gene is involved in X-linked Nance-Horan syndrome. Patients affected by this syndrome have cataracts, dental anomalies and dysmorphic features, symptoms absent in our family 1 patients. In patients 5 , a homozygous recessive mutation in the (*ATPC2C2*) gene, and in patient 6 a compound heterozygous mutations in the Phosphopantothenoylcysteine Synthetase gene (*PPCS*) and a putative *de novo* mutation in the (*PFAS*) were identified. But these mutations c, could not be associated with disruptions in foregut development. Two separate de novo mutations were identified in family two. Moreover, there was no shared rare or private variation following compound heterozygous or homozygous recessive models.

Consanguineous TE- patients

Autosomal recessive diseases arise if a detrimental mutation is inherited from both parents and therefore present in a homozygous state in affected offspring. Although these parental mutations could have arisen independently, they are often inherited from a common ancestor and the parents share a haplotype in this identical by descent (IBD) region.

Using high density SNP-array Runs of Homozygosity can be detected indicative for uniparental dysomy or homozygous IBD regions containing homozygous recessive mutations.[50] Different types of definitions exist regarding the length of ROH region to use in homozygosity mapping. Runs of Homozygosity are present in every human genome, vary in length across individuals, and can be ancestry specific. Therefore, a ROH should have certain characteristics to increase the probability that the ROH region studied contains the detrimental allele of interest. Auton and coworkers define highly homozygous regions as

stretches of at least 50 SNPs in a 1 cM region, with SNPs having an minor allele frequency (MAF) of at least 5%.[51] Although heavily dependent on region specific recombination rate this is roughly 500kb-2000kb containing at least 50 SNPs.[52] Pemberton et al. have classified ROH in short (tens of kb), medium (hundreds of kb to few Mb) and long ROH (several Mb in length) with the medium and long ROH corresponding to shared haplotypes resulting from common ancestry.[53] Szpiech and coworkers used the prediction tool Polyphen2[54] to classify variants found with exome sequencing in 27 individuals and find enrichment of variants predicted to be damaging in long Runs of Homozygosity.[55]

Patient 9 did not have large ROH regions, absent in his parents, in which the *in silico* genomic oligo-array and SNP-array evaluation toolv2.0[18] predicted autosomal recessive syndromes in which patients were reported with trachea-esophageal anomalies. One syndrome was highlighted in the homozygosity enrichment analysis of patient 10. Homozygous mutations in the *WNT7a* on chromosome 3p25 are associated with the autosomal recessive Al-awadi/Raas-Rothschild syndrome (#OMIM 276820). This syndrome is characterized by limb malformations, hypoplastic pelvis, Mullerian hypoplasia and abnormal genitalia. Esophageal atresia has been described in patients with this syndrome.[56] However, no rare or private homozygous recessive *WNT7A* mutations were detected in patient 10 . The only rare or private homozygous recessive variant present in in patent 10 was , was a stopgain mutation in the prenylcysteine oxidase 1 gene.but no link with foregut development could be made.

There were no new ROH regions in patient 11. This was expected since his parents have a familial relationship several generations back and recombination would have caused heterozygosity throughout the genome. Sequencing of this patient is ongoing.

To conclude, we identified the causal mutation in one affected family member, a *MYCN* mutation in a patient with features of Feingold syndrome. Since this mutation was *de novo*, and therefore absent in his niece, it ruled out familial TE in this family. The chance occurrence of two independent events leading to a similar phenotype with an incidence of about 2.5 in 10.000 seemed unlikely. However, it happened and gives a different perspective of familial TE. The absence of other Feingold specific symptoms in this patient can be due to variable specificity of phenotypical features. Screening patients with TE-anomalies on *MYCN* and other syndromic TE anomalies, even if only a few characteristic features are present, seems warranted.

Identifying the causal variant out of the hundreds of candidate variants is difficult in heterogeneous conditions as EA/TEF and VACTERL. If variant filtering was less stringent, excluding conservation and prediction tools as a criterion, and only the allele frequency (MAF ≤ 0.001) many cilia related variants were present. There was shared compound heterozygous variation in dynein genes *DNAH2*, *DNAH3*,*DNAH11* ,*DNAH11 and DYNC2H1* in familial patients and consanguineous patients nine and ten are compound heterozygous for several dynein proteins (*DRC1*, *DNAH1*, *DNAH5*, *DNAH8*, *DNAH9*). Three of these are causal genes for ciliary dyskinesia (*DRC1* and *DNAH5/11*). Homozygous mutations in *DRC1* , seen in patient ten, cause autosomal recessive primary ciliary dyskinesia[57], as do compound heterozygous or homozygous mutations in the *DNAH5* gene.[58, 59]

These variants are intriguing since there is increasing evidence suggesting ciliopathies as a causal factor in specific subsets of VACTERL patients. These include for instance patients with VACTERL-hydrocephalus type of syndrome[60] and asphyxiating thoracic dystrophy [61] patients with craniofacial, limb and rib defects[62, 63]. Moreover, compound heterozygous mutations could be responsible, as has been shown for Sensenbrenner syndrome.[64] Perhaps the combined effect of all variants disrupts the dynein complex structure or proper functioning. Testing if the cilia function properly, for instance measuring cilia number, length and structure[65] in dermal fibroblast could prove the involvement of altered cilia functioning in these patients. A word of caution however, there are thousands of cilia related genes and perhaps merely this number is reflected in the observed frequency..

Possible *pathogenic* variants were detected in genes from genetic syndromes of which TE anomalies are a variable feature. Not all of the characteristic features of those syndromes are present in these patients. However, incomplete penetrance of the phenotype could mask the causal genetic syndrome. Screening patient cohort for pathogenic mutations in genes from known syndromes or genes from animal knockout models with VACTERL (-) like phenotypes could aid in diagnosing many of previously undiagnosed patients.

Acknowledgements

The authors would like to thank Zeliha Ozgur, M. Ploeg, D. Huigh and H.A. Nastiti for their technical assistance.

References

1. Oddsberg, J., Y. Lu, and J. Lagergren, *Aspects of esophageal atresia in a population-based setting: incidence, mortality, and cancer risk.* Pediatr Surg Int, 2012. **28**(3): p. 249-57.
2. Pedersen, R.N., et al., *Oesophageal atresia: prevalence, prenatal diagnosis and associated anomalies in 23 European regions.* Arch Dis Child, 2012. **97**(3): p. 227-32.
3. Cui, W., et al., *Sex differences in birth defects: a study of opposite-sex twins.* Birth Defects Res A Clin Mol Teratol, 2005. **73**(11): p. 876-80.
4. Lisi, A., et al., *Sex and congenital malformations: an international perspective.* Am J Med Genet A, 2005. **134a**(1): p. 49-57.
5. de Jong, E.M., et al., *Non-VACTERL-type anomalies are frequent in patients with esophageal atresia/tracheo-esophageal fistula and full or partial VACTERL association.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 92-7.
6. Shaw-Smith, C., *Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology.* J Med Genet, 2006. **43**(7): p. 545-54.
7. Stoll, C., et al., *Associated malformations in patients with esophageal atresia.* Eur J Med Genet, 2009. **52**(5): p. 287-90.
8. Brosens, E., et al., *Locus and phenotype heterogeneity in patients with trachea-esophageal malformations and associated anomalies.* Journal of Medical Genetics, 2014.
9. Schulz, A.C., et al., *Nine new twin pairs with esophageal atresia: a review of the literature and performance of a twin study of the disorder.* Birth Defects Res A Clin Mol Teratol, 2012. **94**(3): p. 182-6.
10. Felix, J.F., et al., *Genetics and developmental biology of oesophageal atresia and tracheo-oesophageal fistula: lessons from mice relevant for paediatric surgeons.* Pediatr Surg Int, 2004. **20**(10): p. 731-6.
11. de Jong, E.M., et al., *Etiology of esophageal atresia and tracheoesophageal fistula: "mind the gap".* Curr Gastroenterol Rep, 2010. **12**(3): p. 215-22.
12. Felix, J.F., D. Tibboel, and A. de Klein, *Chromosomal anomalies in the aetiology of oesophageal atresia and tracheo-oesophageal fistula.* Eur J Med Genet, 2007. **50**(3): p. 163-75.
13. Bednarczyk, D., M.M. Sasiadek, and R. Smigiel, *Chromosome Aberrations And Gene Mutations In Patients With Esophageal Atresia.* J Pediatr Gastroenterol Nutr, 2013.
14. Choinitzki, V., et al., *Second study on the recurrence risk of isolated esophageal atresia with or without trachea-esophageal fistula among first-degree relatives: No evidence for increased risk of recurrence of EA/TEF or for malformations of the VATER/VACTERL association spectrum.* Birth Defects Res A Clin Mol Teratol, 2013.
15. Robert, E., et al., *An international collaborative study of the epidemiology of esophageal atresia or stenosis.* Reprod Toxicol, 1993. **7**(5): p. 405-21.
16. Otto, E.A., et al., *Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy.* Nat Genet, 2010. **42**(10): p. 840-50.
17. Hildebrandt, F., et al., *A systematic approach to mapping recessive disease genes in individuals from outbred populations.* PLoS Genet, 2009. **5**(1): p. e1000353.
18. Wierenga, K.J., et al., *A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents.* Genet Med, 2013. **15**(5): p. 354-60.
19. Macdonald, J.R., et al., *The Database of Genomic Variants: a curated collection of structural variation in the human genome.* Nucleic Acids Res, 2014. **42**(1): p. D986-92.
20. Amarasinghe, K.C., J. Li, and S.K. Halgamuge, *CoNVEX: copy number variation estimation in exome sequencing data using HMM.* BMC Bioinformatics, 2013. **14 Suppl 2**: p. S2.
21. Fonseca, D.J., et al., *A de novo 14q12q13.3 interstitial deletion in a patient affected by a severe neurodevelopmental disorder of unknown origin.* Am J Med Genet A, 2012. **158A**(3): p. 689-93.
22. Brouwer, R.W., et al., *NARWHAL, a primary analysis pipeline for NGS data.* Bioinformatics, 2012. **28**(2): p. 284-5.
23. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589-95.
24. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
25. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
26. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
27. Jacobs, I.J., W.Y. Ku, and J. Que, *Genetic and cellular mechanisms regulating anterior foregut and esophageal development.* Dev Biol, 2012. **369**(1): p. 54-64.
28. Hamosh, A., et al., *Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.* Nucleic Acids Res, 2005. **33**(Database issue): p. D514-7.

29. Guest, S.S., C.D. Evans, and R.M. Winter, *The Online London Dysmorphology Database.* Genet Med, 1999. **1**(5): p. 207-12.
30. Eppig, J.T., et al., *The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D881-6.
31. Solomon, B.D., et al., *Clinical geneticists' views of VACTERL/VATER association.* Am J Med Genet A, 2012. **158a**(12): p. 3087-100.
32. Solomon, B.D., et al., *De novo deletion of chromosome 20q13.33 in a patient with tracheo-esophageal fistula, cardiac defects and genitourinary anomalies implicates GTPBP5 as a candidate gene.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(9): p. 862-5.
33. Hilger, A., et al., *De novo microduplications at 1q41, 2q37.3, and 8q24.3 in patients with VATER/VACTERL association.* Eur J Hum Genet, 2013. **21**(12): p. 1377-82.
34. Arrington, C.B., et al., *Haploinsufficiency of the LIM domain containing preferred translocation partner in lipoma (LPP) gene in patients with tetralogy of Fallot and VACTERL association.* Am J Med Genet A, 2010. **152a**(11): p. 2919-23.
35. Smigiel, R., et al., *Oesophageal atresia with tracheoesophageal fistula and anal atresia in a patient with a de novo microduplication in 17q12.* Eur J Med Genet, 2014. **57**(1): p. 40-3.
36. Ferreira de Almeida, T. and D.R. Bertola, *Microdeletion 11q13.1.q13.2 in a patient presenting with developmental delay, facial dysmorphism, and esophageal atresia: possible role of the GSTP1 gene in esophagus malformation.* Birth Defects Res A Clin Mol Teratol, 2013. **97**(7): p. 463-6.
37. van Binsbergen, E., et al., *A fetus with de novo 2q33.2q35 deletion including MAP2 with brain anomalies, esophageal atresia, and laryngeal stenosis.* Am J Med Genet A, 2014. **164a**(1): p. 194-8.
38. Faguer, S., et al., *A 17q12 chromosomal duplication associated with renal disease and esophageal atresia.* Eur J Med Genet, 2011. **54**(4): p. e437-40.
39. Dworschak, G.C., et al., *De novo 13q deletions in two patients with mild anorectal malformations as part of VATER/VACTERL and VATER/VACTERL-like association and analysis of EFNB2 in patients with anorectal malformations.* Am J Med Genet A, 2013. **161a**(12): p. 3035-41.
40. Heller, N. and A.W. Brandli, *Xenopus Pax-2 displays multiple splice forms during embryogenesis and pronephric kidney development.* Mech Dev, 1997. **69**(1-2): p. 83-104.
41. Hu, M.C., et al., *GLI3-dependent transcriptional repression of Gli1, Gli2 and kidney patterning genes disrupts renal morphogenesis.* Development, 2006. **133**(3): p. 569-78.
42. van Bokhoven, H., et al., *MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome.* Nat Genet, 2005. **37**(5): p. 465-7.
43. Schramm, C., et al., *De novo microduplication at 22q11.21 in a patient with VACTERL association.* Eur J Med Genet, 2011. **54**(1): p. 9-13.
44. Prieto, J.C., et al., *Phenotypic expansion of the supernumerary derivative (22) chromosome syndrome: VACTERL and Hirschsprung's disease.* J Pediatr Surg, 2007. **42**(11): p. 1928-32.
45. Shaikh, T.H., et al., *Low copy repeats mediate distal chromosome 22q11.2 deletions: sequence analysis predicts breakpoint mechanisms.* Genome Res, 2007. **17**(4): p. 482-91.
46. Chun, K., et al., *FGFR2 mutation associated with clinical manifestations consistent with Antley-Bixler syndrome.* Am J Med Genet, 1998. **77**(3): p. 219-24.
47. McGlaughlin, K.L., et al., *Spectrum of Antley-Bixler syndrome.* J Craniofac Surg, 2010. **21**(5): p. 1560-4.
48. Feigin, E., et al., *Antley-Bixler syndrome and esophageal atresia in a patient with trisomy 21.* Clin Genet, 1995. **47**(1): p. 53-5.
49. Wong, D., et al., *Expanding the BP1-BP2 15q11.2 Microdeletion Phenotype: Tracheoesophageal Fistula and Congenital Cataracts.* Case Rep Genet, 2013. **2013**: p. 801094.
50. Kearney, H.M., J.B. Kearney, and L.K. Conlin, *Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations.* Clin Lab Med, 2011. **31**(4): p. 595-613, ix.
51. Auton, A., et al., *Global distribution of genomic diversity underscores rich complex history of continental human populations.* Genome Res, 2009. **19**(5): p. 795-803.
52. Yu, A., et al., *Comparison of human genetic and sequence-based physical maps.* Nature, 2001. **409**(6822): p. 951-3.
53. Pemberton, T.J., et al., *Genomic patterns of homozygosity in worldwide human populations.* Am J Hum Genet, 2012. **91**(2): p. 275-92.
54. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
55. Szpiech, Z.A., et al., *Long runs of homozygosity are enriched for deleterious variation.* Am J Hum Genet, 2013. **93**(1): p. 90-102.
56. Lonardo, F., et al., *Al-Awadi/Raas-Rothschild syndrome: two new cases and review.* Am J Med Genet A, 2007. **143a**(24): p. 3169-74.

57. Wirschell, M., et al., *The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans*. Nat Genet, 2013. **45**(3): p. 262-8.
58. Olbrich, H., et al., *Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry*. Nat Genet, 2002. **30**(2): p. 143-4.
59. El Zein, L., H. Omran, and P. Bouvagnet, *Lateralization defects and ciliary dyskinesia: lessons from algae*. Trends Genet, 2003. **19**(3): p. 162-7.
60. Friedland-Little, J.M., et al., *A novel murine allele of Intraflagellar Transport Protein 172 causes a syndrome including VACTERL-like features with hydrocephalus*. Hum Mol Genet, 2011. **20**(19): p. 3725-37.
61. Halbritter, J., et al., *Defects in the IFT-B component IFT172 cause Jeune and Mainzer-Saldino syndromes in humans*. Am J Hum Genet, 2013. **93**(5): p. 915-25.
62. Ashe, A., et al., *Mutations in mouse Ift144 model the craniofacial, limb and rib defects in skeletal ciliopathies*. Hum Mol Genet, 2012. **21**(8): p. 1808-23.
63. Fehrenbach, H., et al., *Mutations in WDR19 encoding the intraflagellar transport component IFT144 cause a broad spectrum of ciliopathies*. Pediatr Nephrol, 2014.
64. Gilissen, C., et al., *Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome*. Am J Hum Genet, 2010. **87**(3): p. 418-23.
65. Kheradmand Kia, S., et al., *RTTN mutations link primary cilia function to organization of the human cerebral cortex*. Am J Hum Genet, 2012. **91**(3): p. 533-40.

4.2

Whole exome resequencing reveals recessive mutations in *TRAP1* in individuals with CAKUT and VACTERL association

Adapted from:

Whole exome resequencing reveals recessive mutations in TRAP1 in individuals with CAKUT and VACTERL association

Pawaree Saisawat, Stefan Kohl, Alina C. Hilger, Daw-Yang Hwang, Heon Yung Gee, Gabriel C. Dworschak, Velibor Tasic, Tracie Pennimpede, Sivakumar Natarajan, Ethan Sperry, Danilo S. Matassa, Radovan Bogdanovic, Ivo de Blaauw, Carlo L.M. Marcelis, Charlotte H.W. Wijers, Enrika Bartels, Eberhard Schmiedeke, Dominik Schmidt, Stefanie Märzheuser, Sabine Grasshoff-Derr, Stefan Holland-Cunz, Michael Ludwig, Markus M. Nöthen, Markus Draaken, Erwin Brosens, Hugo Heij, Dick Tibboel, Bernhard G. Herrmann, Ben D. Solomon, Annelies de Klein, Iris A.L.M. van Rooij, Franca Esposito, Heiko M. Reutter, and Friedhelm Hildebrandt. *Kidney Int. 2013 Oct 23. ; Epub ahead of print*

Abstract

Congenital abnormalities of the kidney and urinary tract (CAKUT) account for approximately 50% of children with chronic kidney disease and they are the most frequent cause of end-stage renal disease in children in the US. However, its genetic etiology remains mostly elusive. VACTERL association is a rare disorder that involves multiple organs including the kidney and urinary tract in up to 60% of the cases.

By homozygosity mapping and whole exome resequencing combined with high-throughput mutation analysis by array-based multiplex PCR and next-generation sequencing, we identified recessive mutations in the gene TNF receptor-associated protein 1 (TRAP1) in two families with isolated CAKUT and three families with VACTERL association. TRAP1 is a heat shock protein 90 (HSP90)-related mitochondrial chaperone possibly involved in anti-apoptotic and ER-stress signaling. Trap1 is expressed in renal epithelia of developing mouse kidney E13.5 and in the kidney of adult rats, most prominently in proximal tubules and in thick medullary ascending limbs of Henle's loop.

We thus identified mutations in TRAP1 as highly likely causing CAKUT or CAKUT in VACTERL association.

Introduction

Congenital abnormalities of the kidney and urinary tract (CAKUT) occur in 3-6 per 1,000 live births. CAKUT are the most frequent cause for chronic kidney disease in children (~50%)[1, 2] in the US. The acronym "CAKUT" comprises heterogeneous malformations involving the kidney (e.g. renal agenesis, hypodysplasia), and the urinary tract (e.g. vesicoureteral reflux, ureteropelvic junction obstruction)[3]. These congenital anomalies are related because a part of their pathogenesis is an impaired co-development of nephrogenic tissues derived from the metanephric mesenchyme and the ureteric bud[4]. Twenty monogenic causes of isolated CAKUT in humans have been published to date as reviewed recently by Yosypiv[5]. However, they only account for ~10% - 20% of all cases indicating a broad genetic heterogeneity of CAKUT. A recent study on copy number variations (CNVs) in a large cohort of individuals with CAKUT and two publications identifying novel monogenic causes of CAKUT bring further evidence that CAKUT is a condition of extensive genetic heterogeneity[6-8]. CAKUT most frequently occur isolated, but might be associated with extra-renal phenotypes, for instance with VACTERL association (MIM [#192350]).

The acronym "VACTERL" describes the combination of at least three of the following congenital anomalies: vertebral defects (V), anorectal malformations (A), cardiac defects (C), tracheoesophageal fistula with or without esophageal atresia (TE), renal malformations (R), and limb defects (L). VACTERL association is a rare disease that occurs mostly sporadic in 1/10,000-40,000 live births[9]. Its etiology is enigmatic, although animal models suggest an involvement of Sonic hedgehog signaling[10]. In humans, *ZIC3* mutations are the cause of a closely related non-classic VACTERL condition (VACTERL-X, MIM [#314390])[11, 12]. Additionally, there are six case reports published of individuals with VACTERL association in conjunction with mitochondrial dysfunction as summarized recently by Siebel and Solomon[13]. In order to identify new recessive genes that cause isolated CAKUT or CAKUT in VACTERL association, we performed homozygosity mapping and whole exome resequencing in 24 affected individuals with CAKUT from 16 families, and in 4 individuals with CAKUT in VACTERL.

Results

Whole exome resequencing identifies a homozygous mutation in TRAP1 in CAKUT and in VACTERL association

By homozygosity mapping in a family of two sibs (A3403) with unilateral and bilateral vesicoureteral reflux (VUR) III°, respectively **(Figure 1A, B and Table 1)**, we identified a short 5.2 Mb segment of homozygosity on chromosome 5 **(Figure 1C)**, indicating distant consanguinity of the parents. This finding suggested that in this family CAKUT are most likely caused by a homozygous recessive mutation in an unknown CAKUT gene. We performed whole exome resequencing in individual A3403-21 as described previously by the authors[14, 15]. In order not to miss either a homozygous mutation in a short run of homozygosity or a compound heterozygous mutation (which, as in this case, cannot be excluded *a priori* in families with remote consanguinity[16]), we considered variants not only in the homozygosity peak but within regions of genetic linkage for both sibs (coverage ≥ 4; minor variant frequency, MVF ≥ 0.2). Following variant filtering we retained 38 variants in 13 genes for Sanger confirmation and segregation analysis (**Supplementary Table S1 online**). Only a single homozygous missense mutation (R469H) in the gene *TRAP1* on chromosome 16p13.3 survived the variant filtering process and segregation analysis (**Figure 1D**). This homozygous variant in *TRAP1* in A3403-21 and -22 was positioned in a ~1.5 Mb run of apparent homozygosity that was not detected by homozygosity mapping **(Figure 1C)**, because the threshold for detection of "homozygosity peaks" is 2.1 Mb[17].

In family A4252 with CAKUT in VACTERL we performed whole exome resequencing in an affected individual (A4252-21). This girl was born with a right double kidney and duplex ureter, left VUR, esophageal atresia type IIIb, and anal atresia with a vestibular fistula (**Figure 1E, F and Table 1**). Although there was no consanguinity reported in this family, homozygosity mapping showed unusually broad homozygosity peaks on chromosome 16 on the p-terminus and q-terminus (5.5 and 9.6 Mb, respectively) **(Figure 1G)**. In this case, we hypothesized that CAKUT in VACTERL is caused by a homozygous mutation within these homozygous regions.

When evaluating whole exome resequencing data in this individual, the 512,733 variants initially detected (MVF ≥ .55; coverage ≥ 2) were reduced to only 11 variants within the "homozygosity peaks" on chromosome 16 and 18 (**Supplementary Table S2 online**).

Family -Individ. (sex)	Ethnic origin	Nucleotide alteration[a]	Deduced protein change	Continuous amino-acid sequence conservation	MutT[b]	Poly-Phen2[c]	SIFT[d]	MAF in EVS[e]	Exon (state; segregation)	Urinary tract phenotypes	Other phenotypes
A3403 -21 (F) -22 (F)	Serbian	c.1406G>A	p.R469H	*E. coli* (*C. elegans* has L)	0.99	0.997	0.00	0.77%	13 (Hom; Fa, Mo)	-21: VUR-III° R -22: VUR-III° R and L	None
A4252 -21 (F)	Central European	c.1406G>A	p.R469H	*E. coli* (*C. elegans* has L)	0.99	0.997	0.00	0.77%	13 (Hom; Mo; partial maternal isodisomy)	Double kidney R VUR L	VACTERL association including esophageal atresia IIIb, anal atresia, vestibular fistula
A3051 -21 (M)	Macedonian	c.127_137dup c.1324G>A	p.R46Sfs*75 p.E442K	NA *D. rerio*	NA 0.99	NA 0.003	NA 0.3	Absent 0.08%	2 (het; Mo) 12 (het; Fa)	MCDK L	None
A4884 -21 (F)	Dutch	c.757A>G c.1573C>T	p.I253V p.L525F	*E. coli* (*X. tropicalis* has V, *S. cerevisiae* has L) *E. coli*	0.99 0.99	0.433 0.942	0.00 0.00	0.91% Absent	7 (het; Mo) 14 (het; Fa)	Renal agenesis R	VACTERL association including cervical/thoracic hemivertebrae, 5 dysplastic short ribs R, anal atresia with rectoperineal fistula, ASD type II, esophageal atresia, abnormal position of thumbs
EA1717 -21 (F)	Dutch	c.1330T>A c.1663G>A	p.Y444N p.V555I	*C. elegans* *C. intestinalis*	0.99 0.99	0.985 0.115	0.03 0.39	0.91%[f] Absent	12 (het; Fa) 14 (het; Mo)	Pyelectasis and VUR L	VACTERL association including anal atresia, esophageal atresia, ASD, VSD, hypoplastic/absent humerus, persistent L vena cava superior, cloaca

Table 1. Mutations of TRAP1 in five families with isolated CAKUT or CAKUT in VACTERL association Abbreviations: ASD, atrial septum defect; CAKUT, congenital abnormalities of the kidney and urinary tract; cDNA, complementary DNA; E. coli, Escherichia coli; EVS, Exome Variant Server; F, female; Fa, mutation segregating from the father; L, left; NA, not applicable; M, male; MAF, minor allele frequency; MCDK, multicystic dysplastic kidney; Mo, mutation segregating from the mother; MutT, MutationTaster; R, right; SIFT, sorting intolerant from tolerant; TRAP1, tumor necrosis factor (TNF) receptor–associated protein 1; VSD, ventricular septum defect; VUR-III1, vesicoureteral reflux third degree. aTRAP1 cDNA mutations are numbered according to human cDNA reference sequence NM_016292.2, where þ1 corresponds to the A of ATG start translation codon. bMutationTaster score. Range: 0–1.0, 1.0 being most deleterious. cPolyPhen2 (HumVar) score. Range: 0–1.0, 1.0 being most deleterious. dSIFT score. Range: 0–1.0, 0 being most deleterious. eMinor allele frequency in 8600 alleles of Americans of European descent. fOne individual is homozygous for this allele.

The only variant that was confirmed by Sanger sequencing and that altered a conserved amino acid residue was *TRAP1* R469H, the same allele as in family A3403. By comparison of SNPs in the affected girl and her parents, we demonstrated that partial maternal isodisomy of chromosome 16 with two recombinants (one located on the p-arm and one located on the q-arm) was the underlying cause of homozygosity for *TRAP1* R469H (**Figure 1G-J**).

The *TRAP1* allele c.1406G>A, p.R469H alters an evolutionary highly conserved amino acid residue and it is predicted to be deleterious for protein function by publically available software programs (**Table 1 and Supplementary Figure S1 online**). In the Exome Variant Server (EVS) database, R469H has a minor allele frequency (MAF) of 0.9% in Americans of European descent. In our cohort of 675 individuals with CAKUT, most of them European, the MAF is 1.9%. The three affected individuals from two unrelated

families with homozygous *TRAP1* R469H, as well as 6 additional heterozygous carriers share haplotypes at the *TRAP1* locus (**Figure S2 online**) which speaks for *TRAP1* R469H being a European founder mutation.

Mutation analysis reveals three additional families with TRAP1 mutations

We subsequently analyzed the coding sequence of *TRAP1* in a cohort of 675 individuals with isolated CAKUT **(Supplementary Table S3 online)** and 300 individuals with classic VACTERL association (i.e. VACTERL-X and other related disorders have been excluded) using a barcoded multiplex PCR approach and consecutive next generation sequencing as described previously by the authors[18]. As a control group, we included 800 individuals with the distinct renal phenotype of nephronophthisis.

We detected six additional recessive mutations in *TRAP1* in a compound heterozygous state in three additional unrelated families with CAKUT or CAKUT in VACTERL (**Table 1, Figure 1K, L, M, Supplementary Figure S1, and S3 online**). In individual A3051-21 with a left-sided multicystic dysplatic kidney (MCDK), we found a maternally inherited protein-truncating frame-shift mutation (c.127_137dup, p.R46fs*75). This mutation abrogates the N-terminal mitochondrial targeting sequence of *TRAP1*, which makes this a null allele. The second allele was a missense mutation (c.1324G>A, p.E442K) which segregated from the father.

In individual A4884-21 with CAKUT in VACTERL, including right renal agenesis, vertebral malformations, anal atresia with a rectoperineal fistula, atrial septum defect type II, esophageal atresia, and abnormal position of the thumbs (**Table 1 and Supplementary Figure S4 online**), we detected compound heterozygous missense mutations in *TRAP1* located in the ATPase-domain (c.757A>G, p.I253V) and in the HSP90-domain (c.1573C>T, p.L525F) (**Figure 1L**). In individual EA1717 with CAKUT in VACTERL, including pyelectasis, left VUR, a complex anorectal malformation including anal atresia and persistent cloaca, esophageal atresia, cardiac defects, limb defects and, persistent left vena cava superior (**Table 1**), we detected compound heterozygous missense mutations which are both located in the HSP90-domain of *TRAP1* (c.1330T>A, p.Y444N and c.1663G>A, p.V555I).

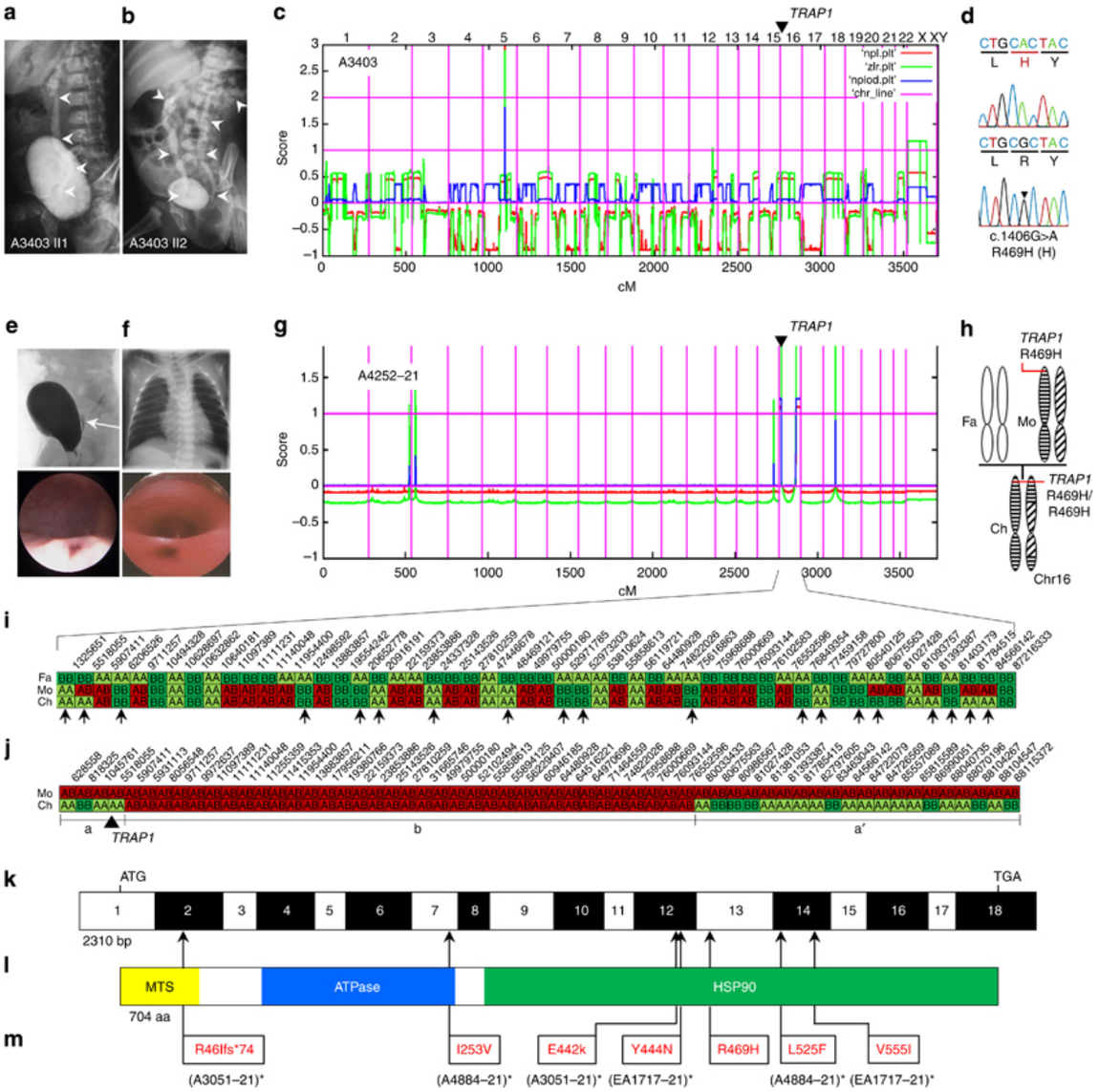

Figure 1 (legend on page 224)

In order to exclude the presence of recessive mutations in controls, we sequenced the *TRAP1* coding sequence in 800 individuals with the distinct renal phenotype of nephronophthisis (NPHP). We detected the *TRAP1* allele I253V seven times (MAF 0.87%), T444N twice (MAF 0.25%), and R469H twice (MAF 0.025%), all of them as single heterozygous alleles. *TRAP1* R46Sfs*75, E442K, L525F, and V555I were absent from our control cohort. Furthermore, no other possibly deleterious variants were present in a homozygous or compound heterozygous state in 800 individuals with NPHP.

Figure 1 | Homozygosity mapping and whole-exome resequencing identifies mutations in tumor necrosis factor (TNF) receptor–associated protein 1 (TRAP1) as causing congenital abnormalities of the kidney and urinary tract (CAKUT) or VACTERL association. (a, b) Voiding cysturethrograms (VCUG) of CAKUT siblings A3403-21 and A3403-22 showing unilateral vesicoureteral reflux (VUR) grade III and bilateral VUR, respectively (white arrow heads). (c) Nonparametric LOD (NPL) scores across the human genome in two affected siblings. X axis represents Affymetrix 250k StyI array single-nucleotide polymorphism (SNP) positions across human chromosomes concatenated from the p-terminus (left) to the q-terminus (right). Genetic distance is given in cM. A single peak indicates distantly related parents. (d) Chromatogram of newly identified homozygous missense mutation (arrow head) in the gene encoding TNF receptor–associated protein 1 (TRAP1) over wild-type control. (e) VCUG (upper panel) and cystoscopy (lower panel) demonstrating VUR and a dilated ureteral orifice, respectively. (f) Chest X-ray (top panel) and esophagoscopy (bottom panel) showing esophageal atresia and esophagotracheal fistula in individual A4252-21 with CAKUT in VACTERL association. (g) NPL score in an individual A4252-21 with VACTERL association. Two maximum peaks indicate homozygosity at the p-terminus and q-terminus of chromosome 16. (h) Panel on the right illustrates maternal heterodisomy of chromosome 16 and partial uniparental isodisomy (p-ter and q-ter) of the child (Fa, father; Mo, mother; Ch, child). (i) Partial haplotypes of selected markers and their physical positions across chromosome 16 in the Fa, the Mo, and the affected Ch of CAKUT family A4252. Selected markers (biallelic SNPs; minor allele frequency (MAF)¼0.496–0.5) homozygous in the father are shown in green (alleles AA) and light green (alleles BB). The fact that for 19 of 52 alleles there is paternal noncontribution in the child strongly suggests maternal heterodisomy of chromosome 16. No paternal noncontribution was observed in the child on any other chromosome (data not shown). (j) Selected markers (biallelic SNPs; MAF¼0.497–0.5) heterozygous in the Mo of family A4252 are shown for alleles coded in red (AB; phase unknown). Note that in the central segment (b), separated by vertical lines, the Ch haplotype is identical to the mother's haplotype. In the p-ter (a) and q-ter (a') segments (a, a'), the child is homozygous, indicating maternal isodisomy in these segments. (k) Exon structure of human TRAP1 complementary DNA. Positions of start codon (ATG) and of stop codon (TGA) are indicated. (l) Domain structure of the TRAP1 protein. HSP, heat-shock protein; MTS, mitochondrial targeting sequence. (m) Translational changes of detected mutations are shown relative to their positions in TRAP1 complementary DNA (see l) and TRAP1 protein (see m) for affected individuals with CAKUT or CAKUT in VACTERL association with recessive TRAP1 mutations. Family numbers are shown in parentheses. *Denotes an individual with compound heterozygous mutations in TRAP1.

Trap1 is expressed in developing and adult kidney

In order to determine whether *TRAP1* has a function during kidney development, we analyzed *Trap1* expression in developing kidney in mouse embryos 13.5 dpc. *Trap1* expression seemed to be expressed at this stage in renal vesicles according to *Trap1* transcription assays publically available through the Gudmap project. By *in-situ* hybridization in E13.5 mouse embryos, we found *Trap1* to be strongly expressed in kidney, adrenal gland, and gonad. *Trap1* expression specifically localized to renal epithelia (**Fig. 2**).

In order to characterize TRAP1 localization in adult kidney, we performed immunofluorescence stainings in rat using a monoclonal TRAP1 antibody in conjunction with established renal markers (**Figure 3**). TRAP1 is present most prominently in peanut-lectin-marked proximal tubules in the renal cortex (**Figure 3A-B**). In renal medulla, we detected TRAP1 in peanut-lectin-negative tubular segments and in NKCC-marked ($Na^+K^+2Cl^-$ co-transporter) thick ascending limbs of Henle's loop (**Figure 3C-D**). TRAP1 co-localizes with mitochondrial marker MTCO1 in renal cortex and medulla.

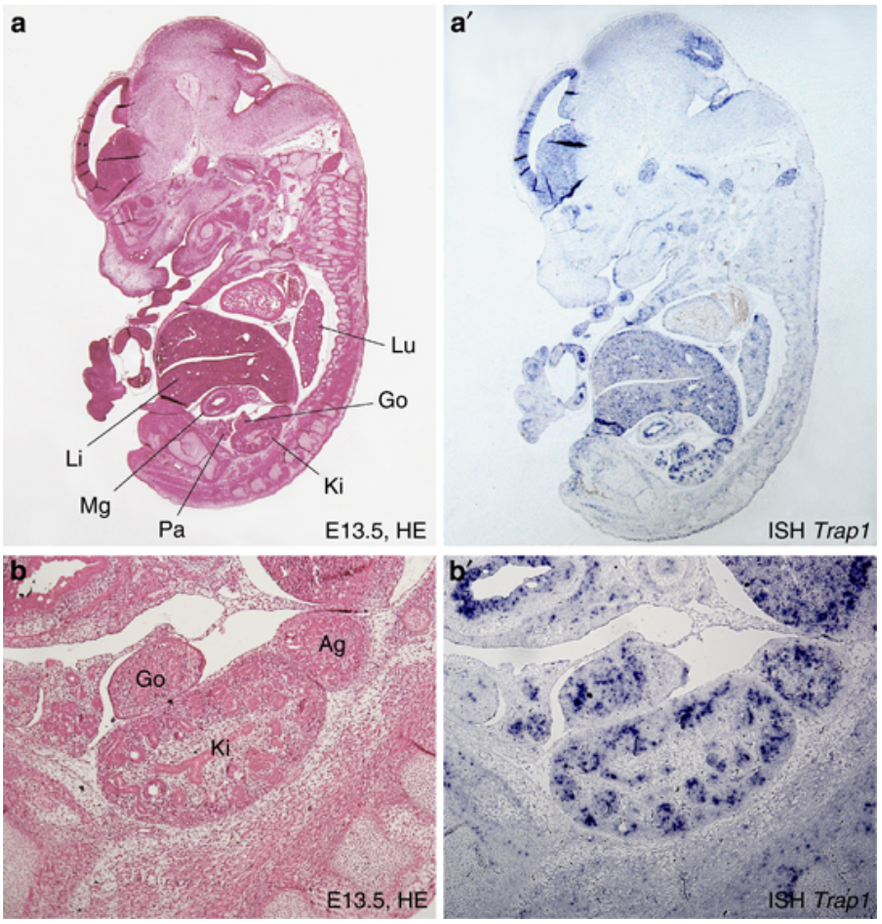

Figure 2 Trap1 is highly expressed in the renal epithelia of E13.5 mouse embryos. The upper panel shows a HE-stained sagittal section (a) and a Trap1-ISH (a') in consecutive sections of a mouse embryo E13.5. Note the prominent Trap1 expression in the developing kidney (marked 'Ki' in the left panel). The lower panel shows higher magnifications of E13.5 mouse kidney. (b) HE staining, (b') Trap1-ISH. The Trap1-ISH staining pattern is consistent with Trap1 being expressed specifically in renal epithelia (b'). Ag, adrenal gland; Go, gonad; HE, hematoxylineosin; ISH, in situ hybridization; Ki, kidney (i.e., metanephros); Li, liver; Lu, lung; Mg, midgut; Pa, pancreatic primordium; TRAP1, tumor necrosis factor (TNF) receptor–associated protein 1.

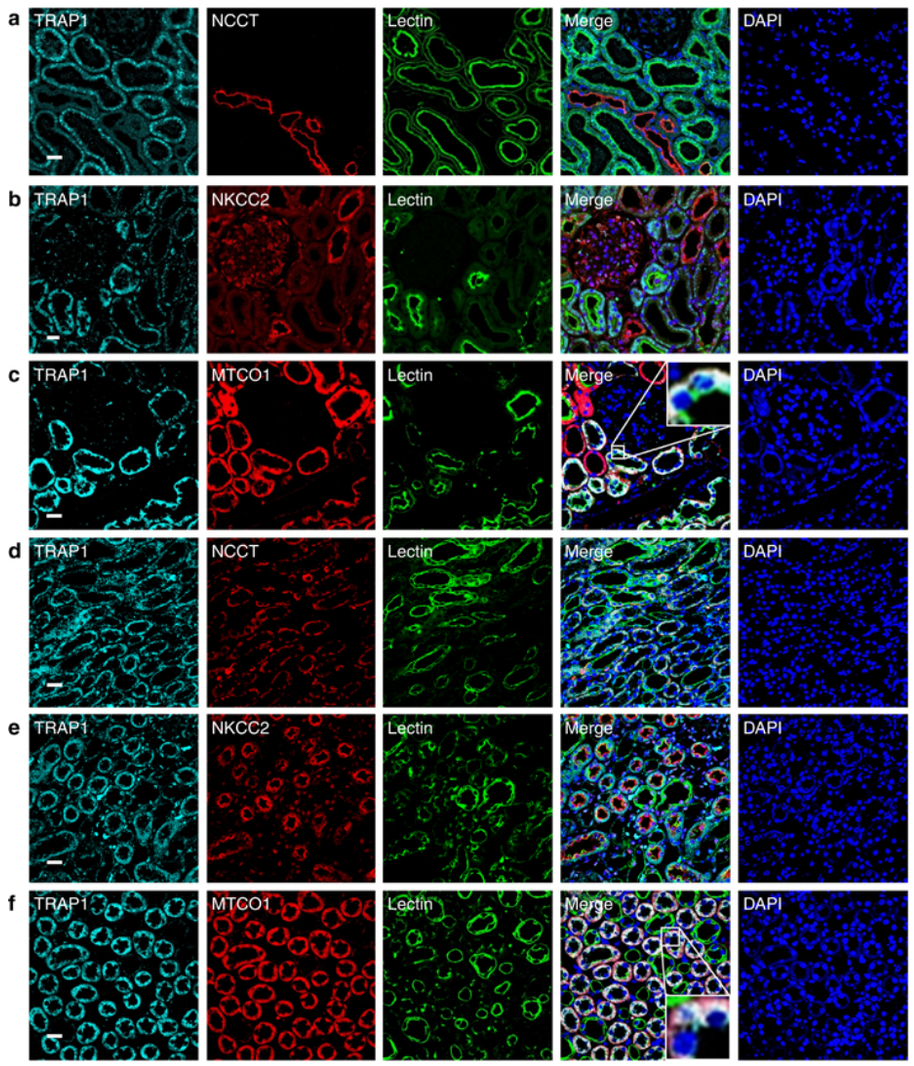

Figure 3 | Renal tubular segmental localization of TRAP1 by immunofluorescence microscopy in adult rat kidney. (a–c) Renal cortex. (a) TRAP1 (blue–green) is located in proximal tubules (marked green by peanut lectin) and is absent from distal convoluted tubules (DCT;marked red by NCCT-ab). (False color addition of blue–green and green appears as light green.) (b) TRAP1 is mostly absent from cortical thick ascending limb (TAL) of Henle's loop (marked red by NKCC2-ab). (c) TRAP1 colocalizes with mitochondrial marker MTCO1 (marked red by MTCO1-ab) in proximal tubules. (False color addition of blue–green and red appears as white.) (d–f) Renal medulla. TRAP1 is absent from peanut-lectin-positive tubular segments. (d) TRAP1 expression is present in DCT (marked red by NCCT-ab). (e) TRAP1 is expressed in medullary thick ascending limb (mTAL) of Henle's loop (marked red by NKCC2-ab). (f) TRAP1 localizes to mitochondria (marked red by MTCO1-ab) of peanut-lectin-negative tubular segments. Scale bar: 201 m. ab, antibody; DAPI, 40,6-diamidino-2-phenylindole; MTCO1, mitochondrially encoded cytochrome c oxidase 1; NCCT, Naþ-Cl- co-transporter; NKCC2, Naþ-Kþ-2Cl- co-transporter; TRAP1, tumor necrosis factor (TNF) receptor–associated protein 1.

Discussion

In the present study, we identified by whole exome resequencing and high-throughput mutation analysis five unrelated families with CAKUT or CAKUT in VACTERL association with recessive mutations in *TRAP1*. Two sibs with CAKUT had a homozygous missense mutation (R469H), which segregated from a common ancestor of their parents. A girl with VACTERL association had the identical homozygous mutation due to maternal isodisomy of chromosome 16 p-ter and q-ter. In a cohort of 675 individuals with CAKUT and 300 individuals with classic VACTERL association we identified 3 additional individuals carrying compound heterozygous mutations in *TRAP1*. Homozygous or compound heterozygous deleterious variants were absent from 800 control individuals. By ISH and IF, we showed that *Trap1* is expressed in early mouse renal epithelia whereas the Trap1 protein is present only in defined segments of developed nephrons in rat.

In 6,500 individuals recorded in the EVS server there are several non-synonymous variants present in *TRAP1*, including heterozygous truncating variants in 11 individuals. However, deleterious alleles in recessive disease-genes, unlike in dominant disease-genes, do not underlie direct negative selection through evolution. Consequently the presence of rare deleterious variants in recessive disease genes in a large cohort is an expected finding.

The allele *TRAP1* Y444N, detected as compound heterozygous mutation in an individual with CAKUT in VACTERL, is present homozygously in a single individual of the ESP cohort of 6,500 healthy Americans. However, in the context of CAKUT, this does not necessarily mean the variant is non-pathogenic. CAKUT frequently remain completely asymptomatic. For instance, a double-kidney or unilateral renal agenesis typically are an "accidental finding" in renal ultrasound.

The fact that the homozygous mutation *TRAP1* R469H was found in an individual with CAKUT and an individual with VACTERL association is surprising. However, in CAKUT and in VACTERL association intra-familial phenotypic variability is very common[19-21]. Even in a single individual different CAKUT phenotypes may be present, for instance left renal agenesis and right VUR.

The frequencies of individuals with recessive *TRAP1* mutations in our cohorts (0.15% in CAKUT, 0.6% in CAKUT with VACTERL) suggest that mutations in *TRAP1* are a rare cause of these conditions. Similarly, mutations in two recently identified CAKUT-causing genes, *WNT4* and *DSTYK*, are rare causes of CAKUT[7, 8]. These findings in humans, along with numerous CAKUT-mouse models, indicate that CAKUT are a common clinical phenotype arising from a multitude of different single-gene causes.

In conclusion, we propose that recessive mutations in *TRAP1* are a novel rare cause of isolated CAKUT and the first recessive cause of the VACTERL association.

Subjects and Methods

Human subjects.

We obtained blood samples and pedigrees following informed consent from individuals with CAKUT and from individuals with VACTERL association. Approval for human subjects research was obtained from the University of Michigan Institutional Review Board and other institutions involved. The diagnosis of CAKUT and VACTERL association was based on published clinical criteria[9].

Homozygosity mapping.

We performed homozygosity mapping as described previously[17].

Whole exome resequencing (WER).

Exome library preparation and massively parallel resequencing was conducted using the SeqCap EZ Exome v2 (Nimblegen) and Genome Analyzer II (Illumina). Subsequent variant detection, filtering and analysis have been described previously by the authors[14, 15]. All detected variants were confirmed by Sanger sequencing.

Immunofluorescence microscopy (IF).

IF was performed as previously described by the authors[14] using a Leica SP5X system with an upright DM6000 compound microscope and images were processed with the Leica AF software suite. Antibodies used: TRAP1 (Abcam, [TRAP1-6], Cat# ab2721), MTCO1 (Abcam Cat# ab45918), NKCC2 (LSBio Cat# LS-C150446), NCCT (Millipore Cat# AB3553). Specificity of the anti-TRAP1 antibody for rat TRAP1 was confirmed in immunoblot (**Figure S5 online**).

In-situ hybridization (ISH).

ISH was conducted on sections of wildtype mouse embryos with an NMRI background at embryonic day 13.5. Mouse embryos were dissected into ice cold phosphate buffered saline (PBS), fixed overnight in 4% paraformaldehyde/PBS, and then processed into paraffin wax. ISH was performed on paraffin sections (5µm) using antisense probes generated by PCR from an E11.0 total embryo cDNA library, and specific staining was verified using a sense probe. PCR products contained *3'* T7 and *5'* T3 RNA polymerase binding sites for in vitro transcription and probes were purified using G-50 sephadex columns (GE Healthcare). The 779bp probe for Trap1 spans exons 13-17 (Accession: NM_026508.2). ISH was performed according to the protocol from (Chotteau-Lelievre et al., 2006) with minor modifications, and detection of AP activity was visualized using BM Purple (Roche Diagnostics). Following staining, slides were quickly dehydrated in 80% and then 100% ethanol, cleared twice for 1 min in xylene (Roth) and coverslips were mounted with Entellan mounting medium (Merck). Photographs were obtained using AxioVision software (Zeiss) with a Zeiss AxioCam and SteREO Discovery.V12 microscope. Three sections from at least 2 different embryos were analyzed.

Bioinformatics.

NGS reads alignment and variant detection was done with Genomics Workbench software (CLC Biotech). Mapping parameter: Global alignment, length fraction = 0.9, and similarity fraction = 0.9. Genetic location is according to the assembly of the Genome Reference Consortium GRCh37.

Web Resources

1000 Genomes Browser, http://browser.1000genomes.org;
Ensembl Genome Browser, http://www.ensembl.org;
Exome Variant Server, http://evs.gs.washington.edu/EVS;
Mutation Taster, http://www.mutationtaster.org;
Gudmap (GenitoUrinary Molecular Anatomy Project), http://www.gudmap.org;
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org;
Polyphen2, http://genetics.bwh.harvard.edu/pph2;
Sorting Intolerant From Tolerant (SIFT), http://sift.bii.a-star.edu.sg;
The Human Protein Atlas, http://www.proteinatlas.org;
UCSC Genome Browser, http://genome.ucsc.edu/cgi-bin/hgGateway.

Acknowledgements

The authors thank the physicians and families for participating in this study, the German self-help organization for people with anorectal malformations (SoMA e.V.), and all participating physicians, nurses, research assistants, laboratory analysts, and project members of AGORA (Aetiologic research into Genetic and Occupational/Environmental Risk Factors for Anomalies in Children) for their support in building this biobank. FH is an Investigator of the Howard Hughes Medical Institute, a Doris Duke Distinguished Clinical Scientist, and a Frederick GL Huetwell Professor. ACH, GCD, EB, ES, DS, SG-D, SM, SH, SH-C, MMN, ML, HR, MD are members of the "Network for the Systematic Investigation of the Molecular Causes, Clinical Implications and Psychosocial Outcome of Congenital Uro-Rectal Malformations" (CURE-Net).

This research was supported by grants from the National Institutes of Health (to FH; R01-DK045345 and R01-DK088767), by the March of Dimes Foundation (6FY11-241), by the Division of Intramural Research, by the National Human Genome Research Institute (NHGRI), by the National Institutes of Health and Human Services, by the Bundesministerium für Bildung und Forschung, (grant 01GM08107), by the BONFOR program of the University of Bonn (to EB; grant O-149.0099, and to GCD; grant O-149.0096), by Sophia Scientific Research Foundation (to EB; grant SSWO S13/9), by the associazione Italiana per la Ricerca sul Cancro (AIRC) (to FE; grant IG13128), and by the Italian Ministry of Health (to FE; grant GR-2010-2310057).

Disclosure

The authors report no conflicts of interest.

References

1. NAPRTCS, *2011 Annual Dialysis Report*. 2011, https://web.emmes.com/study/ped/annlrept/annualrept2011.pdf.
2. Smith, J.M., et al., *Contributions of the Transplant Registry: The 2006 Annual Report of the North American Pediatric Renal Trials and Collaborative Studies (NAPRTCS)*. Pediatric Transplantation, 2007. **11**(4): p. 366-73.
3. Pope, J.C.t., et al., *How they begin and how they end: classic and new theories for the development and deterioration of congenital anomalies of the kidney and urinary tract, CAKUT*. Journal of the American Society of Nephrology : Jasn, 1999. **10**(9): p. 2018-28.
4. Ichikawa, I., et al., *Paradigm shift from classic anatomic theories to contemporary cell biological views of CAKUT*. Kidney Int, 2002. **61**(3): p. 889-98.
5. Yosypiv, I.V., *Congenital anomalies of the kidney and urinary tract: a genetic disorder?* Int J Nephrol, 2012. **2012**: p. 909083.
6. Sanna-Cherchi, S., et al., *Copy-number disorders are a common cause of congenital kidney malformations*. American Journal of Human Genetics, 2012. **91**(6): p. 987-97.
7. Vivante, A., et al., *Renal hypodysplasia associates with a WNT4 variant that causes aberrant canonical WNT signaling*. Journal of the American Society of Nephrology : Jasn, 2013. **24**(4): p. 550-8.
8. Sanna-Cherchi, S., et al., *Mutations in DSTYK and Dominant Urinary Tract Malformations*. The New England Journal of Medicine, 2013.
9. Solomon, B.D., *VACTERL/VATER Association*. Orphanet J Rare Dis, 2011. **6**: p. 56.
10. Vaze, D., S. Mahalik, and K.L. Rao, *Novel association of VACTERL, neural tube defect and crossed renal ectopia: sonic hedgehog signaling: a point of coherence?* Congenit Anom (Kyoto), 2012. **52**(4): p. 211-5.
11. Wessels, M.W., et al., *Polyalanine expansion in the ZIC3 gene leading to X-linked heterotaxy with VACTERL association: a new polyalanine disorder?* Journal of Medical Genetics, 2010. **47**(5): p. 351-5.
12. Chung, B., et al., *From VACTERL-H to heterotaxy: variable expressivity of ZIC3-related disorders*. American journal of medical genetics. Part A, 2011. **155A**(5): p. 1123-8.
13. Siebel, S. and B.D. Solomon, *Mitochondrial Factors and VACTERL Association-Related Congenital Malformations*. Molecular Syndromology, 2013. **4**(1-2): p. 63-73.
14. Chaki, M., et al., *Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling*. Cell, 2012. **150**(3): p. 533-48.
15. Zhou, W., et al., *FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair*. Nature genetics, 2012. **44**(8): p. 910-5.
16. Ten Kate, L.P., et al., *Consanguinity sans reproche*. Human genetics, 1991. **86**(3): p. 295-6.
17. Hildebrandt, F., et al., *A Systematic Approach to Mapping Recessive Disease Genes in Individuals from Outbred Populations*. PLoS Genet, 2009. **5**(1): p. e1000353.
18. Halbritter, J., et al., *High-throughput mutation analysis in patients with a nephronophthisis-associated ciliopathy applying multiplexed barcoded array-based PCR amplification and next-generation sequencing*. Journal of Medical Genetics, 2012. **49**(12): p. 756-67.
19. Solomon, B.D., et al., *Evidence for inheritance in patients with VACTERL association*. Hum Genet, 2010. **127**(6): p. 731-3.
20. Solomon, B.D., et al., *Evidence for inheritance in patients with VACTERL association*. Human genetics, 2010. **127**(6): p. 731-3.
21. Hilger, A., et al., *Familial occurrence of the VATER/VACTERL association*. Pediatric Surgery International, 2012. **28**(7): p. 725-9.

PART 5

General discussion and Summary

5.1

General Discussion

General Discussion

In the last few years genetic research has taken a significant leap forward, both in resolution to detect genetic variation as in our understanding of the impact of this genetic variation on human health and development. It has become increasingly clear that our genomes are packed with variation. Of all known types of DNA variation, Copy Number Variation (CNV) affects largest part of our genome, Single Nucleotide Variation (SNV) is most abundant in total variant numbers.[1, 2] Hundreds of these variants in each genome are considered "pathogenic", private variants affecting gene splicing, causing frame shift or introducing stop gains or losses.[3, 4] CNV causing deletions and duplications of sometimes more than a MB (1000000bp), holding many genes.[5] Yet, all this variation is often inherited from parents not affected by the often severe conditions seen in their children. Discriminate the causal variant(s) in the midst of these thousands of candidate sequencing variants requires well planned filtering strategies, which include knowledge on inheritance patterns, allele frequencies in the general population, type and functional impact of the variation and information necessary to prioritize the affected genes. This last information not only includes gene specific information but also well documented and structured phenotype information.[6, 7]

CNV can be inherited in a Mendelian manner and if they have an allele frequency under <1% in control cohorts, they are called *rare* Copy Number Variations (CNVs). [8] CNV are proposed to arise after replication errors e.g. fork stalling and template switching, non-homologues end joining or micro homology mediated break induced replication. [2, 5] Replication errors resulting in a large *de novo* CNV occur approximately once in a hundred genomes per generation.[9, 10] The majority will result in decreased fecundity and extinct, others have little effect and are inherited across generations. CNV are under selective pressure [11, 12] and could very well be one of the driving forces of human evolution. These large *de novo* CNV have frequently been associated with congenital anomalies.[13, 14] Especially rare CNVs have shown to impact human disease in a variety of ways. For instance, they are involved in complex disease [15], can be the second hit unmasking a mutation, as has been shown for TAR syndrome [16, 17] or work in a two-hit model.[18] There are diagnostic guidelines for the use of micro-array [19] and CNV interpretation.[20] The interpretation of a rare CNV, inherited from unaffected parents, is difficult since the resulting phenotype can be highly variable and even a sub-clinical phenotype.[21, 22]

The impact of aneuploidies, structural chromosomal aberrations and de novo CNV

We describe the genetic variation in our large cohort of patients. EA/TEF is a rare condition, with an incidence of 2.5 in 10.000 live births. Consequently, each year about 10-15 patients are born or admitted to our hospital. We have collected patient material, clinical and genetic material since 1988. Our cohort currently has 600 patients, in about half of these patients cytogenetic studies at a 30kb resolution has been performed. Aneuploidies and structural chromosomal anomalies have been described at many chromosomal loci in patients with EA/TEF, some even recurrent such as deletions on chromosome 2q37, 4q35, 6q13-q15 and duplications on 3p25-pter and 5q34-qter.[23, 24] There are also chromosomal anomalies with a link to a syndrome in the London Dysmorphology Database, chromosome 10 - paternal disomy, chromosome 17q22-q23.2 - submicroscopic deletion, chromosome 6q27 - submicroscopic deletion and mosaic trisomy 16 all have trachea-esophageal anomalies as variable features.[25]

Large CNVs are rare in our genome. Approximately 9% of the individuals in control cohorts have a CNV larger than 500kb and 3% has a CNV larger than 1 Mb with a *de novo* rate of 1.2 x 10-2 per generation.[9, 10] Assuming variation increases with decreasing variation size, these numbers are somewhat larger focusing on CNV ≥ 30kb. In the Erasmus MC-Sophia TE-cohort on average 1 rare or private CNV larger than 30kb was present in each TE patient, ~6% of which d*e novo.*

Some of these variations affect genes with a direct link to the observed phenotype. Other CNV affect genes with no obvious association to foregut abnormalities and could very well be non-causal. Chromosomal anomalies, excluding those seen in genetic syndromes such as trisomy 13, 18 or 21 are rare. Large structural events do not seem to impact EA/TEF and VACTERL disease burden as much as it does in other developmental disorders such as Congenital Diaphragmatic Hernia[26] or intellectual disability.[27] The *de novo* CNV seen in TE and VACTERL patients is also non-recurrent, with the exception of CNV overlapping the causative locus of a genetic syndrome.[28] It is nonetheless important to screen newborns with congenital anomalies, including patients with EA/TEF and VACTERL associated features to exclude large pathogenic variation. Mega-base chromosomal anomalies and aneuploidies affect hundreds of genes and the malformations are and will be often severe. Sub-sequent parental counseling and treatment choice can be more specific and better fit current and future needs. The impact of *de novo* CNV (6%), aneuploidies (3.6%) and structural chromosomal anomalies (0.3%) on the total cohort is limited. Yet, for an *individual* patient molecular karyotyping can be vital.

Micro-array technology has replaced high resolution GTG-banded karyotyping as a first tier diagnostic procedure.[29] Large, unbalanced, deletions and duplications will be continued to be detected. However, balanced translocations and inversions will be missed by micro-array, since this technique gives only information about the relative copy number, not the location of DNA segments. Similarly, inherited, balanced translocations will be missed by micro-array alone. For example, Cetinkaya and co-workers [30] describe an inherited (1;13)(p8;q12) translocation in a patient with esophageal atresia.

Chromosomal anomalies can give clues to candidate genes and loci; some cover regions harboring candidate genes in TE; deletions on 17q21.3-q24.2 harbor *Nog* and *Tbx4* which cause TE anomalies in animal knockout studies[31-33]; 13q deletions are also seen as de novo CNV in patients with VACTERL association[34] and an unbalanced translocation in fact deleting chromosome 7q34qter and duplicating 8q24qter involves the *SHH* and *HLXB9* loci. Deletions of SHH are rare, remarkably few reports are available describing loss of *SHH* in patients with TE anomalies.[35-39] One of the patients in the Erasmus MC-Sophia TE-cohort has a *de novo* deletion of chromosome 7q36, including the *SHH* gene.

Hilger and co-workers describe three de novo CNV, micro-duplications on chromosome 1q41, 2q37.3 and 8q24.3 in VACTERL association patients. A mutation screening of candidate genes in these regions did not result in additional patients with a shared phenotype and affected gene.[40] However, the 8q24 region is also described to be duplicated in two structural chromosomal aberrations and could be a susceptibility locus.[23] The scarce reports and presence in our cohort of *de novo* CNV in patients with TE anomalies and VACTERL association demonstrate that these factors do not have a high impact on TE disease burden. [39-48]

CNV recurrence; predisposition, reduced penetrance and variable expressivity

CNV can be inherited in a Mendelian manner and behaves much like other genetic variation. There are common polymorphisms and rare or private CNV, the latter likely to be under strong selection pressure.[1] Finding two overlapping CNV, in two unrelated patients in a cohort of patients with a rare disease could be indicative of a susceptibility locus or gene. Almost all patients in our cohort have a rare or private CNV, as does the general population. Recurrence of variation at a specific locus in a rare disease as EA/TEF and VACTERL association is likely to be more than a chance occurrence. These CNV are inherited and their unaffected parents usually do not report any family history of trachea-esophageal or other VACTERL associated anomalies. There are several reasons why this recurrence of are and private CNV could have an impact on foregut development. Variable

expressivity and reduced penetrance is seen in many of the single gene disorders with a TE component such as Feingold[49] and CHARGE[50], in patients with micro-deletion and -duplication syndromes[22, 51, 52] even in patients with whole chromosome duplications.[53]

It is conceivable that this variability could lead to sub-clinical anomalies in unaffected parents or other family members. These unaffected parents are thoroughly inspected for visual characteristics of the phenotype. However, they are not consistently screened for anomalies of the intestinal organs with MRI, X-ray or other imaging technology. Therefore, minor anomalies causing sub clinical phenotypes in the esophagus, heart, kidney or vertebrae could remain unnoticed. Current analysis strategies, mainly based on a *dominant de novo* and penetrant paradigm, of CNV in patients with congenital anomalies do not cover this aspect of CNV pathogenicity. Rare and private CNV, virtually absent from large control databases, are either new in a population and will be common polymorphisms in the future, or they are rare because they are under evolutionary constraint. Caution has to be taken when immediately considering an inherited CNV benign, further parental counseling may be needed. Recent publications describing tissue specific CNV justifies studying CNV in target tissue instead of blood derived DNA to reduce the chance of tissue specific effects.

CNV interpretation greatly benefits from copy number profiling studies in fully screened healthy control populations [54-56] containing several generations of subjects. Inheritance of a CNV is an important exclusion for technical artefacts and a *de novo* CNV can be determined. Moreover, these cohorts should contain people from as many genetic backgrounds as possible, since CNV can be ancestry specific.[57] Large cohorts of concise phenotyped individuals are a common in the GWAS community, therefore patient care would greatly benefit from the combined Copy Number and phenotype information from these studies.

In a response to the discovery of single cell Copy Number Variation[58], Maosko and McCarrol state that perhaps these cells do not transcribe the genes underlying these CNV and therefore replicate these late in cell division, increasing the chance of mistakes.[59] Cells "know" what part of their private genome is essential for their survival and functioning. They state that it is perhaps best to focus first on those parts of the genome that are *never* affected by detrimental DNA changes. Profiling these large cohorts could aid in describing those parts of the genome that never are affected by CNV, these regions likely are so crucial that disrupting them leads to a phenotype severely affecting fecundity. Conceivably, these regions should be enriched for loci crucial for development. A

recently proposed approach to prioritize candidate gens in large CNV is to look at ohnologs, usually dosage sensitive genes retained after ancestral whole genome duplication events since these genes are enriched in pathogenic CNV.[60, 61] Perhaps as important as discovering pathogenic variation and CNVs, is describing those that are *not* causative for a disease. Screening large cohorts of patients and controls is vital in the genome wide association studies (GWAS) and copy number profiling era, but will perhaps be even more essential in current genome and exome wide sequencing projects.

Small Insertions/deletions and single base-pair mutations

Gene mutations play a major role in many of the syndromes of which TE anomalies are a variable feature. Most syndromes are autosomal dominant in which either the syndrome phenotype segregates through the pedigree or occur *de novo*. Nine percent of patients from the Erasmus MC-Sophia cohort have a confirmed genetic syndrome. There are several reasons to suspect that the actual number is in fact higher. The first reason is that patients are screened for a genetic anomaly when their phenotype fits a specific genetic syndrome or syndromes. However, variable expressivity, reduced penetrance, modifying factors, skewed X-inactivation and stochastic effects could lead to difficulties recognizing a specific syndrome. Other reasons are more practical, not all phenotypical characteristics are recorded consistently, genetic analysis is sometimes not performed and registered in house and perhaps most important not all syndromes were known at the moment of genetic research.

The discovery of these genetic syndromes has seen a steep increase in recent years. For instance the gene responsible for AEG-syndrome (*SOX2*) in 2003[62], CHARGE syndrome (*CHD7*) in 2004[63], Feingold syndrome (*MYCN*) has been described in 2005[64], Alveolar capillary dysplasia in 2009[28] and *EFTUD2* mutations in patients with mandibulofacial dysostosis and esophageal atresia in 2013.[65] We have included patients from 1988 onwards, and many of the more "historic" patients may have a known genetic syndrome based on their phenotype.[66] One of the two affected family members in our WES experiment had a de novo *MYCN* mutation, already described in literature.[49] His phenotype was not characteristic for Feingold syndrome, only EA/TEF and thumb anomalies were present. This clearly demonstrates the need for screening the entire cohort, regardless of associated features, for the presence of known genetic syndromes. In fact, research is ongoing in collaboration with the Universitätsklinikum Bonn, Radbout Universiteit Nijmegen, Boston Children's hospital and the Inova Translational Medicine

Institute, in which the unresolved patients are being screened for the most frequent syndromic-TE mutations as *MYCN*, *MID1*, *CHD7*, *SOX2* and others.

Genetic and phenotypical heterogeneity seen in EA/TEF patients increases the need of large sample sizes, since we do not study one, but many separate conditions. It is now evident that only by combining cohorts of patients (and ideally controls) in large collaborative consortiums, statistical sound results can be obtained from genetic, environmental and other epidemiological studies. A clear consistent terminology, technology and experimental approach are needed to allow for the cross-institutional analysis of these large datasets. Collaborative studies have been initiated to unite the research for certain congenital anomalies, VACTERL association included. Patients would benefit from these endeavors, since they are the fastest way to increase and share knowledge on TE-anomalies.

Shift from coding DNA only to whole genome

Currently, most of the research has focused on genetic factors within the coding part of the DNA. Recently, the ENCODE project has described numerous regions of the DNA which have a regulatory function. Regions in the DNA not only code for proteins, but also for a variety of ncRNA molecules which have the potential to regulate gene expression. We describe the deletion of a regulatory lncRNA in patients with TE-anomalies and Alveolar Capillary dysplasia.[67] This focus on the coding part of the DNA also occurs in most of the current sequencing projects. Exome sequencing target enrichment kits do have some ncRNA molecule coverage. Genetic malformations affecting regulatory elements may explain more of the missing heritability seen in TE-anomalies. The deletion of certain specific genes could lead to in-utero death of a fetus due to the severe effect of having severely decreased amounts of protein, whilst less severe effects in protein level could cause a severe, but nowadays not lethal effect as EA and or TEF.

The genetic contribution; re-analysis and evaluation

With patients cohorts being screened with exome sequencing, new associated gene mutations will be described, some leading to new syndromes. Screening "historic" patient cohorts on these new, but also the already known, mutations in genes will give a clearer picture of the actual genetic contribution in TE patients. The same holds true for CNV-profiling. Increasing ability of this technique in detecting smaller and smaller CNV justifies the re-analysis of previously patient-parent trio's for de novo CNV.

Exon level copy number profiling using Next Generation Sequencing data has the potential to discriminate between diploid and other copy number states below the kb level. Combining variant level information with these CN events can potentially lead to the discovery of small CNV unmasking loss of function mutations.

Rodent models have variable relevance for human foregut disease

The responsible genes in many syndromes have previously been first identified using animal studies. For instance, Mahlupuu identified *Foxf1* as a candidate for Alveolar capillary dysplasia and VACTERL association.[68, 69] and years later Stankiewicz et al. identifies the human disease gene *FOXF1* in Alveolar Capillary Dysplasia.[28, 67] Animal models describing trachea-esophageal anomalies after loss of function or gene knockout are plentiful[66, 70] and these are increasingly associated with their human counterparts in disease such as in autosomal recessive Bartsocas-Papas syndrome (*RIPK4*)[71] short rib polydactyly syndrome (*DYNC2H1*)[72], Sensenbrenner syndrome[73] and Jeune and Mainzer-Saldino syndrome.[74] Currently endeavors screening patient populations of isolated and complex TE for gene mutations in these genes are ongoing, possibly identifying new genes involved in human abnormal tracheo-esophageal development.

However, these candidate genes, although giving a clear phenotype in the rodent models, do not always have the same phenotype in man. For instance, mouse double knockouts of retinoic acid receptors *Rarα* and *Rarβ* have EA/TEF and lung hypoplasia or agenesis, *Nkx2-1* knockouts develop TEF, whilst no patients have been described with EA/TEF and defects in these genes.[75] The opposite is also true, whilst many human patients with defects in *MYCN*, *MID1* and *CHD7* have EA/TEF, animal models do not.[76, 77] Animal knockouts of *SHH* have been described and they can have abnormal foregut development.[78] Deletion or mutation of the *SHH* gene have been described in human: patients with loss of function mutations often have autosomal dominant holoprosencephaly (OMIM#142945)[79] Only three reports describe *SHH* deletion in patients with TE and other VACTERL associated anomalies and none of the 396 patients with *SHH* mutations in a holoprosencephaly cohort had trachea-esophageal anomalies.[80]

Gene mutations in *GLI2* can also cause holoprosencephaly (#610829) [81] and in *GLI3* Pallister Hall syndrome (#OMIM146510). [82] Tracheo-esophageal anomalies are only incidentally described in patients with mutations in these three genes, yet have been shown to be crucial in foregut development.[83] Rodent double knockouts of *Gli2* and *Gli3* can have TE-anomalies, but a knockout of only one of these genes is not sufficient to give

the TE phenotype.[75] Absence of a TE phenotype in most of SHH and GLI mutated patients could be indicative of either in-utero death of patients carrying both a pathogenic mutation and having TE-anomalies, or these genes do not impact TE-disease burden as much as we currently believe.

Another example is Adriamycin, a chemotherapeutic agent, interfering with DNA replication and inhibiting DNA and RNA synthesis.[84] VACTERL type of anomalies, including EA/TEF, have been described in the offspring of Adriamycin-treated rats.[85] In humans adriamycin does not effectively cross the placenta in humans[86] and no EA/TEF has been described in exposed fetuses. These teratogenic or genetic altered animal models may have tissue specific protein expression and timing differences compared to humans. In mice, at around day 9 the foregut has formed after which lung bud morphogenesis starts. Human development has a different pace; it reaches the same developmental stage around day 20-22, thus much earlier in embryogenesis.. The esophagus in mice is, compared to humans, relatively shorter, does not have columnar glandular epithelium in the fore stomach, the epithelium is less thick, has a keratin layer and the type of musculature surrounding the esophagus differs.[77]

Therefore it seems evident that not all data from animal models are of relevance for the human situation. Genetic, transcriptomic and protein studies in human patients and controls around week four of development is impossible. A alternative approach circumventing this unavailability of human foregut cells is the use of human induced pluripotent stem cells.[87] An approach proven to be applicable for human anterior foregut.[88, 89] Patient specific stem cells can be generated from fibroblasts by the pluripotency factors - Oct3/4, Sox2, c-MYC and KLF4- and transformed to definite endoderm by high concentrations of activin A. Subsequent removal of Activin A from culture resulted in an increase in *CDX2* and *SOX2* expression. Green and coworkers subsequent tested several combinations of morphogens and inhibitors and the addition of Noggin in a specific time interval resulted in foregut specific expression patterns.[88] Using patient and control fibroblasts and transform these with the iPS technique to an early foregut structure could enableus to compare not only the effect of the genetic alterations in individual patients, but also to compare large groups of patients and controls. Also the influence of teratogens and alterations of key signaling factors, cell structures (cilia) or biological processes can be studied in this way.

TE anomalies as minor and major features

Although most frequently associated major malformations are those of the VACTERL spectrum of anomalies[75], others are also recurrent, such as specific craniofacial anomalies, eye and ear anomalies and malformations of the genitourinary, respiratory and gastrointestinal systems.[90-92] These malformations sometimes seem to cluster together and could have one specific factor as a causative mechanism. For instance, genes downstream of *SHH* and *GLI*, e.g. *SALL1*, *MYCN*, *FOXF1* are the causal genes in genetic syndromes with variable TE-anomaly frequency.

Animal models give clues about other genes or pathways involved in foregut development and have led to the discovery of genes mutated in patients with malformations of foregut-derived organs.[28, 67, 68] Animal knockout models of transcription factors, such as the forkheadbox (*Foxf1*, *Foxp1*, *Foxp2*, *Foxp4)* [28, 67, 93] and homeobox (e.g. *Hoxc4*, *MEOX2*, *NKX2.1*)[94-98] transcription factors give insight in the importance of these molecules in foregut development. Other animal models even provide clues about biological important structures and functioning. Loss of function or knockout models of *Shh*, *Gli2*, *Gli3*, *Bmp4*, *Ift172*, *Dync2h1*, *Fuz and Wdr35*[73, 74, 99-103] all affect cilia morphology, formation and SHH signal transduction.[104] These knockout and loss of function models can be translated to human research with relative ease; screening of patient populations is ongoing. Ciliopathies with Tracheo-esophageal anomalies as a variable feature have been discovered in human and mice.[72, 74, 101, 105]

In fact, trachea-esophageal anomalies have been described in well over 70 genetic syndromes. These anomalies either *are* an associated feature in a genetic syndrome or occur *sporadically* in patients with a certain genetic syndrome. Of key importance here is that, not all patients carrying a pathogenic mutation in these genes have trachea-esophageal anomalies. Only in a small subset of currently identified genetic syndromes are TE-anomalies frequently seen. In the vast remainder, these anomalies are only incidentally described.[66]

Perhaps a distinction has to be made in syndromes in which TE anomalies are part of the characteristic features of that specific syndrome and genetic syndromes in which TE anomalies *can* occur. In this model the first type of genetic defects *actively disturb* foregut morphogenesis and the second type alter the embryonic development in such a way that also foregut development *can* be affected. The presence of a sporadic TE phenotype in these syndromes could be due to a variety of reasons, a chance occurrence included. We could

speculate that not *the* genetic syndrome, but having *a* genetic defect increases the TE-anomaly frequency.

General growth restrictions or other more systemic effects could set in motion a cascade of events leading to the shift from normal to abnormal development. As an example, it has been speculated that part of VACTERL association etiology is not caused by genetic or environmental causes but to having a twin pregnancy.[106, 107] Other more, mechanistically factors could also be important in foregut anomalies. In a rodent model, Adriamycin influences the notochord in such a way that the structure and signals from the notochord are altered thereby influencing foregut development.[108]

Environmental factors in TE-anomalies

Apart fromidentified genetic factors, associated environmental factors are divers. There are general risk factors, as increased maternal age and low maternal parity. There are also risk factors with a more environmental component as exposure to anti-thyroid drugs and caffeine intake.[66, 84] As observed in the genetic component of EA/TEF, type, severity and frequency of congenital anomalies associated with these exposures or risk factors is highly variable. It is also difficult to go beyond association and prove causality due to the methods currently employed. Results from teratogenic animal models, as Adriamycin or cadmium exposed rodent models, are difficult to transform and interpret in humans.

Questionnaires only associate, even if the outcome of well performed epidemiological research is highly significant. Genetic association studies need a functional test to prove causality; other studies associating biological outcomes with (self-) observations do so too. Detecting a "biological distinct fingerprint" in patients, absent in controls and show that this fingerprint alters crucial cellular mechanisms would greatly advance epidemiological studies. Nowadays, there are epigenetic biomarkers for cigarette smoke [109, 110] and it is known that fetal alcohol exposure alters DNA methylation and histone modifications.[111, 112] These types of studies could give a straightforward, unbiased, answer to exposure to other teratogens or other factors suspected to be involved in foregut development. For instance, cadmium exposure can disturb *Shh* and *Wnt* signaling [113, 114] and leaves an epigenic recognizable mark.[115]

Anti-thyroid drugs taken during organogenesis are associated with severe congenital anomalies, including VACTERL associated malformations, in 2-3% of children.[116] Other risk factors for the development of congenital anomalies are maternal diabetes[117] the trans generation effects of maternal in utero DES exposure[118] and exposure to

herbicides. If these factors would leave a distinct epigenetic mark, we could measure this in patients (and their parents). We could determine their impact on TE-disease burden in our cohort and study which gene expression alterations they cause. Differences in epigenetic profile have been shown to exist in phenotypical discordant monozygous twin pairs. [119-121] in which only one twin is affected. These epigenetic differences could also be present in our discordant twin cohort and should be further investigated. Studying epigenetic differences in these twins could give answers to which pathways, genes or even environmental components cause their phenotypical difference.

Combined Transcriptomic end (epi-) genetic studies could give answers to the impact of DNA variation and environmental components on gene expression. Yet, timing of measurement and tissue specificity are variables hampering precise and reliable transcriptomic analysis. However, it could very well be that there is nothing wrong with protein structure or *total* amount of specific proteins but that disturbances in their *relative* amount (compared to factors in their network) or local gradients of factors at crucial time intervals are disturbed, masking the importance of specific factors. Proper foregut development could be such a time dependent and precise balancing act, that deviation from the equilibrium at the wrong moment in development causes a cascade effect leading to the variety of congenital malformations seen in some of these patients.

To conclude, the phenotypical and genetic heterogeneity seen in EA/TEF patients are indicative of several underlying cause(s), each contributing only in a small proportion to the total TE disease burden. (see figure 1)

Future developments

Gene panels

Increasing evidence in patients and animal studies suggest that a wide variety of factors can disturb foregut development. In fact, so many individual factors can be associated with the development of TE-anomalies that single gene approaches are only feasible in the rare syndromes with a high EA/TEF frequency as Feingold or CHARGE syndrome.

Variable expressivity, reduced penetrance and sub-clinical phenotypes can, and do, hamper the recognition of a genetic syndrome. Genome wide approaches give such a huge amount of data that, if present, the causal variant or variants are perhaps even more difficult to discriminate. Using first a panel of disease causing genes and only second a genome wide approach reduces the risk of spending many hours of work and of resources, whilst the causal gene could have been identified in a few minutes.

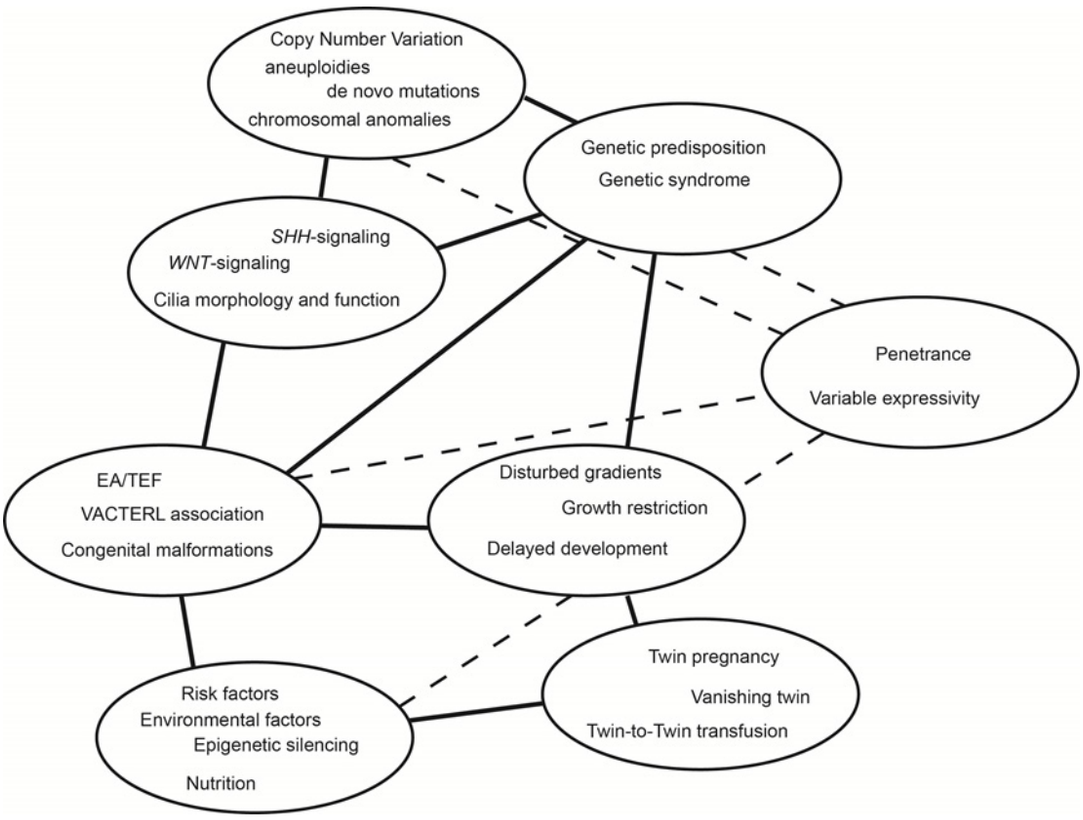

Figure 1. Mind map of etiological factors in trachea-esophageal anomalies and associated features.

Using this candidate gene approach in a parent-child trio can quickly identify loss of function mutations and their inheritance pattern. Perhaps equally important as proper Next Generation Sequencing technology and analysis is describing the patients phenotype in an organized structured and detailed way.[122] Subtle differences are present between some of the syndromes with variable TE-frequency. Precise phenotyping of patient and parents, structuring this information in a way suitable for bioinformatic approaches and updating these phenotypes as time progresses will enhance the number of genotype-phenotype correlations in a significant way.

The wide genetic and phenotypical heterogeneity of TE defects indicate a multifactorial etiology. Next-Generation Sequencing and high resolution SNP arrays now enable us to detect smaller and smaller *de novo* CNV and mutations. Their use in the few familial cases and phenotypically clustered groups of sporadic patients will reveal causal genetic variation and may identify new genetic syndromes. For instance, after the observation that pyloric stenosis has a high frequency in EA/TEF patients, we started to

investigate whether or not the same genetic variant could be responsible for this phenotypical cluster.

To date, the majority of EA/TEF and VACTERL-association patients have been sporadic; many of them caused by yet unknown private *de novo* CNV or mutations. Improved surgical treatment has increased patient survival. Now they are reaching adulthood and may wish to become pregnant or raise a family, these patients are at risk of transmitting their *de novo* mutations to the next generations in an autosomal dominant manner. Genetic counseling and genome profiling of young adults using techniques that were unavailable at their time of their birth may be an important instrument not only for identifying genes involved in disease etiology, but also for predicting high-risk pregnancies.

Current patients might face another kind of co-morbidity. It has become increasingly evident that genes involved in congenital anomalies and childhood syndromes increase the chance of having some kind of cancer later in life.[123] TE-anomalies are seen in patients with Fanconi anemia (*FANC*-genes), loss of function mutations in *SOX2* and *MYCN* cause AEG-syndrome and Feingold syndrome, whilst overexpression of these oncogenes is associated with tumorgenesis.[124, 125] There are reports associating SNPs to genes known to be important in esophageal development to the development of Barrett's esophagus and adenocarcinoma.[126, 127] Patients with EA/TEF often have gastro-esophageal reflux. This reflux and/or these SNPs could result in a predisposition to develop Barrett's and sub sequent adenocarcinoma of the esophagus. Currently, we are evaluating if these SNPs have a higher frequency in patients with EA/TEF.

Tracheo-esophageal anomalies can be part of a broad spectrum of associated anomalies from isolated atresia's or trachea-esophageal fistulas. This heterogeneity can be caused by the differences in penetrance of a limited set of (epi-) genetic and/or environmental factors. Another explanation could be that this heterogeneity is the representation of many individual causes. Regardless what causes this heterogeneity, large patient and control cohorts are needed to associate either many loci of low effect, or to find the second hit of a large effect, low frequency locus. In a rare disease this will only be possible by combining patient cohorts. The effect of multiple rare or private inherited variants detected with NGS and studied with a burden test, also need large sample sizes. Uniform parental questionnaires should be used in multicenter studies, possibly in combination with one specific type of high-density SNP array or sequencing pipeline, thus enabling epidemiological and Genome Wide Association Studies in large cohorts.

The greatest challenge in elucidating EA/TEF and VACTERL association etiology will be to categorize the non-genetic contribution to their etiology. While the use of uniform questionnaires will lead to new associations, proving the causality of these associations will be much more difficult. It may be best first to focus on the effect of known mutations, teratogen and risk factors, possibly by using functional tests to study suspected biological pathways or processes. Patients' various genetic and environmental causality presents challenges with regard to counseling and informing parents on recurrence risk and future co-morbidity, and also the risk of familial recurrence. Greater knowledge of environmental risk factors, genetic predisposition and causal genetic syndromes may even produce preventive strategies and the ability to predict co-morbidity.

References

1. Girirajan, S., C.D. Campbell, and E.E. Eichler, *Human copy number variation and complex genetic disease.* Annu Rev Genet, 2011. **45**: p. 203-26.
2. Stankiewicz, P. and J.R. Lupski, *Structural variation in the human genome and its role in disease.* Annu Rev Med, 2010. **61**: p. 437-55.
3. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nat Rev Genet, 2011. **12**(11): p. 745-55.
4. Tennessen, J.A., et al., *Evolution and functional impact of rare coding variation from deep sequencing of human exomes.* Science, 2012. **337**(6090): p. 64-9.
5. Zhang, F., et al., *Copy number variation in human health, disease, and evolution.* Annu Rev Genomics Hum Genet, 2009. **10**: p. 451-81.
6. Cooper, G.M. and J. Shendure, *Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data.* Nat Rev Genet, 2011. **12**(9): p. 628-40.
7. Gilissen, C., et al., *Disease gene identification strategies for exome sequencing.* Eur J Hum Genet, 2012. **20**(5): p. 490-7.
8. McCarroll, S.A., *Extending genome-wide association studies to copy-number variation.* Hum Mol Genet, 2008. **17**(R2): p. R135-42.
9. Itsara, A., et al., *Population analysis of large copy number variants and hotspots of human genetic disease.* Am J Hum Genet, 2009. **84**(2): p. 148-61.
10. Itsara, A., et al., *De novo rates and selection of large copy number variation.* Genome Res, 2010. **20**(11): p. 1469-81.
11. Gazave, E., et al., *Copy number variation analysis in the great apes reveals species-specific patterns of structural variation.* Genome Res, 2011. **21**(10): p. 1626-39.
12. Sudmant, P.H., et al., *Evolution and diversity of copy number variation in the great ape lineage.* Genome Res, 2013. **23**(9): p. 1373-82.
13. Southard, A.E., L.J. Edelmann, and B.D. Gelb, *Role of copy number variants in structural birth defects.* Pediatrics, 2012. **129**(4): p. 755-63.
14. Vissers, L.E. and P. Stankiewicz, *Microdeletion and microduplication syndromes.* Methods Mol Biol, 2012. **838**: p. 29-75.
15. Cooper, G.M., et al., *A copy number variation morbidity map of developmental delay.* Nat Genet, 2011. **43**(9): p. 838-46.
16. Klopocki, E., et al., *Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome.* Am J Hum Genet, 2007. **80**(2): p. 232-40.
17. Albers, C.A., et al., *Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome.* Nat Genet, 2012. **44**(4): p. 435-9, s1-2.
18. Girirajan, S., et al., *A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay.* Nat Genet, 2010. **42**(3): p. 203-9.
19. Kearney, H.M., et al., *American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities.* Genet Med, 2011. **13**(7): p. 676-9.
20. Kearney, H.M., et al., *American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants.* Genet Med, 2011. **13**(7): p. 680-5.
21. Ou, Z., et al., *Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes.* Genet Med, 2008. **10**(4): p. 267-77.
22. Wentzel, C., et al., *Clinical variability of the 22q11.2 duplication syndrome.* Eur J Med Genet, 2008. **51**(6): p. 501-10.
23. Felix, J.F., D. Tibboel, and A. de Klein, *Chromosomal anomalies in the aetiology of oesophageal atresia and tracheo-oesophageal fistula.* Eur J Med Genet, 2007. **50**(3): p. 163-75.
24. Lurie, I.W., *Structural autosomal imbalance and oesophageal defects: addendum to the article by Felix et al. (2007).* Eur J Med Genet, 2007. **50**(4): p. 322-5.
25. Guest, S.S., C.D. Evans, and R.M. Winter, *The Online London Dysmorphology Database.* Genet Med, 1999. **1**(5): p. 207-12.
26. Wat, M.J., et al., *Genomic alterations that contribute to the development of isolated and non-isolated congenital diaphragmatic hernia.* J Med Genet, 2011. **48**(5): p. 299-307.
27. Hehir-Kwa, J.Y., et al., *De novo copy number variants associated with intellectual disability have a paternal origin and age bias.* J Med Genet, 2011. **48**(11): p. 776-8.
28. Stankiewicz, P., et al., *Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations.* Am J Hum Genet, 2009. **84**(6): p. 780-91.

29. Miller, D.T., et al., *Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies.* Am J Hum Genet, 2010. **86**(5): p. 749-64.
30. Cetinkaya, M., et al., *Oesophageal atresia in a neonate with a familial translocation t(1;13)(p8;q12).* Genet Couns, 2012. **23**(1): p. 73-5.
31. Dallapiccola, B., et al., *Interstitial deletion del(17) (q21.3q23 or 24.2) syndrome.* Clin Genet, 1993. **43**(1): p. 54-5.
32. Marsh, A.J., et al., *Interstitial deletion of chromosome 17 (del(17)(q22q23.3)) confirms a link with oesophageal atresia.* J Med Genet, 2000. **37**(9): p. 701-4.
33. Park, J.P., et al., *A unique de novo interstitial deletion del(17)(q21.3q23) in a phenotypically abnormal infant.* Clin Genet, 1992. **41**(1): p. 54-6.
34. Dworschak, G.C., et al., *De novo 13q deletions in two patients with mild anorectal malformations as part of VATER/VACTERL and VATER/VACTERL-like association and analysis of EFNB2 in patients with anorectal malformations.* Am J Med Genet A, 2013. **161a**(12): p. 3035-41.
35. Speleman, F., et al., *Detection of subtle reciprocal translocations by fluorescence in situ hybridization.* Clin Genet, 1992. **41**(4): p. 169-74.
36. Schinzel, A., *Cyclopia and cebocephaly in two newborn infants with unbalanced segregation of a familial translocation rcp (1;7)(q32;q34).* Am J Med Genet, 1984. **18**(1): p. 153-61.
37. Schinzel, A., *A further case of cyclopia due to unbalanced segregation of a previously reported rcp(1;7)(q32;q34) familial translocation.* Am J Med Genet, 1986. **24**(1): p. 205-6.
38. Schinzel, A., *Catalogue of unbalanced chromosome aberrations in man.* 2 ed. 2001, New York: De Gruyter.
39. Zen, P.R., et al., *Esophageal stenosis in a child presenting a de novo 7q terminal deletion.* Eur J Med Genet, 2010. **53**(5): p. 333-6.
40. Hilger, A., et al., *De novo microduplications at 1q41, 2q37.3, and 8q24.3 in patients with VATER/VACTERL association.* Eur J Hum Genet, 2013. **21**(12): p. 1377-82.
41. Wong, D., et al., *Expanding the BP1-BP2 15q11.2 Microdeletion Phenotype: Tracheoesophageal Fistula and Congenital Cataracts.* Case Rep Genet, 2013. **2013**: p. 801094.
42. Brosens, E., et al., *VACTERL Association Etiology: The Impact of de novo and Rare Copy Number Variations.* Mol Syndromol, 2013. **4**(1-2): p. 20-6.
43. Faguer, S., et al., *A 17q12 chromosomal duplication associated with renal disease and esophageal atresia.* Eur J Med Genet, 2011. **54**(4): p. e437-40.
44. Ferreira de Almeida, T. and D.R. Bertola, *Microdeletion 11q13.1.q13.2 in a patient presenting with developmental delay, facial dysmorphism, and esophageal atresia: possible role of the GSTP1 gene in esophagus malformation.* Birth Defects Res A Clin Mol Teratol, 2013. **97**(7): p. 463-6.
45. Schramm, C., et al., *De novo microduplication at 22q11.21 in a patient with VACTERL association.* Eur J Med Genet, 2011. **54**(1): p. 9-13.
46. Solomon, B.D., et al., *De novo deletion of chromosome 20q13.33 in a patient with tracheo-esophageal fistula, cardiac defects and genitourinary anomalies implicates GTPBP5 as a candidate gene.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(9): p. 862-5.
47. van Binsbergen, E., et al., *A fetus with de novo 2q33.2q35 deletion including MAP2 with brain anomalies, esophageal atresia, and laryngeal stenosis.* Am J Med Genet A, 2014. **164a**(1): p. 194-8.
48. Smigiel, R., et al., *Oesophageal atresia with tracheoesophageal fistula and anal atresia in a patient with a de novo microduplication in 17q12.* Eur J Med Genet, 2014. **57**(1): p. 40-3.
49. Cognet, M., et al., *Dissection of the MYCN locus in Feingold syndrome and isolated oesophageal atresia.* Eur J Hum Genet, 2011. **19**(5): p. 602-6.
50. Jongmans, M.C., et al., *CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene.* J Med Genet, 2006. **43**(4): p. 306-14.
51. Robin, N.H. and R.J. Shprintzen, *Defining the clinical spectrum of deletion 22q11.2.* J Pediatr, 2005. **147**(1): p. 90-6.
52. Tan, T.Y., et al., *Phenotypic variability of distal 22q11.2 copy number abnormalities.* Am J Med Genet A, 2011. **155A**(7): p. 1623-33.
53. Brosens, E., et al., *Structural and numerical changes of chromosome X in patients with esophageal atresia.* Eur J Hum Genet, 2014.
54. Shaikh, T.H., et al., *High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications.* Genome Res, 2009. **19**(9): p. 1682-90.
55. Mills, R.E., et al., *Mapping copy number variation by population-scale genome sequencing.* Nature, 2011. **470**(7332): p. 59-65.
56. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.
57. Oei, L., et al., *A genome-wide copy number association study of osteoporotic fractures points to the 6p25.1 locus.* J Med Genet, 2013.

58. McConnell, M.J., et al., *Mosaic copy number variation in human neurons.* Science, 2013. **342**(6158): p. 632-7.
59. Macosko, E.Z. and S.A. McCarroll, *Genetics. Our fallen genomes.* Science, 2013. **342**(6158): p. 564-5.
60. McLysaght, A., et al., *Ohnologs are overrepresented in pathogenic copy number mutations.* Proc Natl Acad Sci U S A, 2014. **111**(1): p. 361-6.
61. Makino, T. and A. McLysaght, *Ohnologs in the human genome are dosage balanced and frequently associated with disease.* Proc Natl Acad Sci U S A, 2010. **107**(20): p. 9270-4.
62. Fantes, J., et al., *Mutations in SOX2 cause anophthalmia.* Nat Genet, 2003. **33**(4): p. 461-3.
63. Vissers, L.E., et al., *Mutations in a new member of the chromodomain gene family cause CHARGE syndrome.* Nat Genet, 2004. **36**(9): p. 955-7.
64. van Bokhoven, H., et al., *MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome.* Nat Genet, 2005. **37**(5): p. 465-7.
65. Gordon, C.T., et al., *EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia.* J Med Genet, 2012. **49**(12): p. 737-46.
66. Brosens, E., et al., *Locus and phenotype heterogeneity in patients with trachea-esophageal malformations and associated anomalies.* Journal of Medical Genetics, 2014.
67. Szafranski, P., et al., *Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder.* Genome Res, 2013. **23**(1): p. 23-33.
68. Mahlapuu, M., S. Enerback, and P. Carlsson, *Haploinsufficiency of the forkhead gene Foxf1, a target for sonic hedgehog signaling, causes lung and foregut malformations.* Development, 2001. **128**(12): p. 2397-406.
69. Mahlapuu, M., et al., *The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm.* Development, 2001. **128**(2): p. 155-66.
70. Blake, J.A., et al., *The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse.* Nucleic Acids Res, 2014. **42**(1): p. D810-7.
71. Kalay, E., et al., *Mutations in RIPK4 cause the autosomal-recessive form of popliteal pterygium syndrome.* Am J Hum Genet, 2012. **90**(1): p. 76-85.
72. Merrill, A.E., et al., *Ciliary abnormalities due to defects in the retrograde transport protein DYNC2H1 in short-rib polydactyly syndrome.* Am J Hum Genet, 2009. **84**(4): p. 542-9.
73. Gilissen, C., et al., *Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome.* Am J Hum Genet, 2010. **87**(3): p. 418-23.
74. Halbritter, J., et al., *Defects in the IFT-B component IFT172 cause Jeune and Mainzer-Saldino syndromes in humans.* Am J Hum Genet, 2013. **93**(5): p. 915-25.
75. de Jong, E.M., et al., *Etiology of esophageal atresia and tracheoesophageal fistula: "mind the gap".* Curr Gastroenterol Rep, 2010. **12**(3): p. 215-22.
76. Que, J., et al., *Morphogenesis of the trachea and esophagus: current players and new roles for noggin and Bmps.* Differentiation, 2006. **74**(7): p. 422-37.
77. Jacobs, I.J., W.Y. Ku, and J. Que, *Genetic and cellular mechanisms regulating anterior foregut and esophageal development.* Dev Biol, 2012. **369**(1): p. 54-64.
78. Chiang, C., et al., *Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function.* Nature, 1996. **383**(6599): p. 407-13.
79. Roessler, E., et al., *Mutations in the human Sonic Hedgehog gene cause holoprosencephaly.* Nat Genet, 1996. **14**(3): p. 357-60.
80. Solomon, B.D., et al., *Genotypic and phenotypic analysis of 396 individuals with mutations in Sonic Hedgehog.* J Med Genet, 2012. **49**(7): p. 473-9.
81. Roessler, E., et al., *Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13424-9.
82. Kang, S., et al., *Linkage mapping and phenotypic analysis of autosomal dominant Pallister-Hall syndrome.* J Med Genet, 1997. **34**(6): p. 441-6.
83. Litingtung, Y., et al., *Sonic hedgehog is essential to foregut development.* Nat Genet, 1998. **20**(1): p. 58-61.
84. Felix, J.F., et al., *Genetic and environmental factors in the etiology of esophageal atresia and/or tracheoesophageal fistula: an overview of the current concepts.* Birth Defects Res A Clin Mol Teratol, 2009. **85**(9): p. 747-54.
85. Mc Laughlin, D., et al., *Adriamycin-Induced Models of VACTERL Association.* Mol Syndromol, 2013. **4**(1-2): p. 46-62.
86. Germann, N., F. Goffinet, and F. Goldwasser, *Anthracyclines during pregnancy: embryo-fetal outcome in 160 patients.* Ann Oncol, 2004. **15**(1): p. 146-50.
87. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
88. Green, M.D., et al., *Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells.* Nat Biotechnol, 2011. **29**(3): p. 267-72.
89. Hannan, N.R., et al., *Generation of multipotent foregut stem cells from human pluripotent stem cells.* Stem Cell Reports, 2013. **1**(4): p. 293-306.

90. de Jong, E.M., et al., *Non-VACTERL-type anomalies are frequent in patients with esophageal atresia/tracheo-esophageal fistula and full or partial VACTERL association.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 92-7.
91. Solomon, B.D., et al., *Analysis of component findings in 79 patients diagnosed with VACTERL association.* Am J Med Genet A, 2010. **152a**(9): p. 2236-44.
92. Solomon, B.D., M.S. Raam, and D.E. Pineda-Alvarez, *Analysis of genitourinary anomalies in patients with VACTERL (Vertebral anomalies, Anal atresia, Cardiac malformations, Tracheo-Esophageal fistula, Renal anomalies, Limb abnormalities) association.* Congenit Anom (Kyoto), 2011. **51**(2): p. 87-91.
93. Shu, W., et al., *Foxp2 and Foxp1 cooperatively regulate lung and esophagus development.* Development, 2007. **134**(10): p. 1991-2000.
94. Garber, E.D., *"Bent-Tail," A Dominant, Sex-Linked Mutation in the Mouse.* Proc Natl Acad Sci U S A, 1952. **38**(10): p. 876-9.
95. Boulet, A.M. and M.R. Capecchi, *Targeted disruption of hoxc-4 causes esophageal defects and vertebral transformations.* Dev Biol, 1996. **177**(1): p. 232-49.
96. Szumska, D., et al., *VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5.* Genes Dev, 2008. **22**(11): p. 1465-77.
97. Mankoo, B.S., et al., *The concerted action of Meox homeobox genes is required upstream of genetic pathways essential for the formation, patterning and differentiation of somites.* Development, 2003. **130**(19): p. 4655-64.
98. Minoo, P., et al., *Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos.* Dev Biol, 1999. **209**(1): p. 60-71.
99. Information submitted by the NHLBI Cardiovascular Development Consortium (CvDC), B.t.B.P., *Modeling the Genetic Basis for Human Congenital Heart Disease in Mice* M.D.D.S. (B2B/CvDC), Editor. 2011.
100. Mill, P., et al., *Human and mouse mutations in WDR35 cause short-rib polydactyly syndromes due to abnormal ciliogenesis.* Am J Hum Genet, 2011. **88**(4): p. 508-15.
101. Friedland-Little, J.M., et al., *A novel murine allele of Intraflagellar Transport Protein 172 causes a syndrome including VACTERL-like features with hydrocephalus.* Hum Mol Genet, 2011. **20**(19): p. 3725-37.
102. Astorga, J. and P. Carlsson, *Hedgehog induction of murine vasculogenesis is mediated by Foxf1 and Bmp4.* Development, 2007. **134**(20): p. 3753-61.
103. Li, Y., et al., *Bmp4 is required for tracheal formation: a novel mouse model for tracheal agenesis.* Dev Biol, 2008. **322**(1): p. 145-55.
104. Ngan, E.S., K.H. Kim, and C.C. Hui, *Sonic Hedgehog Signaling and VACTERL Association.* Mol Syndromol, 2013. **4**(1-2): p. 32-45.
105. Schmidts, M., et al., *Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome) without major polydactyly, renal or retinal involvement.* J Med Genet, 2013. **50**(5): p. 309-23.
106. Pharoah, P.O., *Causal hypothesis for some congenital anomalies.* Twin Res Hum Genet, 2005. **8**(6): p. 543-50.
107. Pharoah, P.O., S.V. Glinianaia, and J. Rankin, *Congenital anomalies in multiple births after early loss of a conceptus.* Hum Reprod, 2009. **24**(3): p. 726-31.
108. Hajduk, P., et al., *Abnormal notochord branching is associated with foregut malformations in the adriamycin treated mouse model.* PLoS One, 2011. **6**(11): p. e27635.
109. Shenker, N.S., et al., *DNA methylation as a long-term biomarker of exposure to tobacco smoke.* Epidemiology, 2013. **24**(5): p. 712-6.
110. Lee, K.W. and Z. Pausova, *Cigarette smoking and DNA methylation.* Front Genet, 2013. **4**: p. 132.
111. Ungerer, M., J. Knezovich, and M. Ramsay, *In utero alcohol exposure, epigenetic changes, and their consequences.* Alcohol Res, 2013. **35**(1): p. 37-46.
112. Perkins, A., et al., *Alcohol exposure during development: Impact on the epigenome.* Int J Dev Neurosci, 2013. **31**(6): p. 391-7.
113. Fujiwara, N., et al., *Disturbance of SHH signalling pathway during early embryogenesis in the cadmium-induced omphalocele chick model.* Pediatr Surg Int, 2013. **29**(2): p. 165-70.
114. Hanson, M.L., et al., *Prenatal cadmium exposure dysregulates sonic hedgehog and Wnt/beta-catenin signaling in the thymus resulting in altered thymocyte development.* Toxicol Appl Pharmacol, 2010. **242**(2): p. 136-45.
115. Sanders, A.P., et al., *Cadmium exposure and the epigenome: Exposure-associated patterns of DNA methylation in leukocytes from mother-baby pairs.* Epigenetics, 2014. **9**(2): p. 212-21.
116. Laurberg, P. and S.L. Andersen, *Antithyroid drug use in early pregnancy and birth defects. Time windows of relative safety and high risk?* Eur J Endocrinol, 2014.
117. Castori, M., *Diabetic embryopathy: a developmental perspective from fertilization to adulthood.* Mol Syndromol, 2013. **4**(1-2): p. 74-86.
118. Felix, J.F., et al., *Esophageal atresia and tracheoesophageal fistula in children of women exposed to diethylstilbestrol in utero.* Am J Obstet Gynecol, 2007. **197**(1): p. 38.e1-5.

119. Javierre, B.M., et al., *Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.* Genome Res, 2010. **20**(2): p. 170-9.
120. Miyake, K., et al., *Comparison of Genomic and Epigenomic Expression in Monozygotic Twins Discordant for Rett Syndrome.* PLoS One, 2013. **8**(6): p. e66729.
121. Weksberg, R., et al., *Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome.* Hum Mol Genet, 2002. **11**(11): p. 1317-25.
122. Hennekam, R.C. and L.G. Biesecker, *Next-generation sequencing demands next-generation phenotyping.* Hum Mutat, 2012. ***33***(5): p. 884-6.
123. Kelleher, F.C., D. Fennelly, and M. Rafferty, *Common critical pathways in embryogenesis and cancer.* Acta Oncol, 2006. **45**(4): p. 375-88.
124. Hussenet, T., et al., *SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas.* PLoS One, 2010. **5**(1): p. e8960.
125. Cheng, J.M., et al., *Preferential amplification of the paternal allele of the N-myc gene in human neuroblastomas.* Nat Genet, 1993. **4**(2): p. 191-4.
126. Levine, D.M., et al., *A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus.* 2013.
127. Dura, P., et al., *Barrett associated MHC and FOXF1 variants also increase esophageal carcinoma risk.* Int J Cancer, 2013. **133**(7): p. 1751-5.

Summary & Samenvatting

Summary

Esophageal atresia (EA) is a relatively rare congenital anomaly in which there is no connection between the proximal esophagus and the stomach. In more than 90% of patients, the distal esophagus has an abnormal connection to the trachea; this is called a trachea-esophageal fistula (TEF).[1-3]

The first chapter" Clinical and etiological heterogeneity in patients with tracheo-esophageal malformations and associated anomalies" is a review of the current knowledge on the etiological aspects of the anomaly, with an emphasis on its phenotypical and causative heterogeneity. Approximately two thirds of patients also have other major malformations, mostly one or more of the types of defects included in the so-called VACTERL association: vertebral, anorectal, cardiac, tracheo-esophageal, renal or urinary tract, and limbs malformations).[4] EA/TEF is a multifactorial condition for which specific risk, environmental and genetic factors have been identified. These are described in chapter one. Furthermore, trachea-esophageal anomalies, including EA/TEF, are a variable feature in over 70 genetic syndromes.[5-7] Examples are anophthalmia-esophageal-genital (AEG) syndrome, Feingold syndrome, CHARGE syndrome, and trisomy 18.

In the Erasmus MC-Sophia cohort of trachea-esophageal anomalies (TE) nine percent of patients have a known genetic syndrome and another 1-2% the condition is strongly associated with an environmental factor. This leaves almost 90% of TE disease burden unexplained. Using two relatively new techniques, SNP-array and whole exome sequencing we aimed to explore genetic variation in EA/TEF and VACTERL association. The second chapter, "Copy Number Variations in patients with EA/TEF and VACTERL associated malformations" describes the chromosomal aneuploidies, structural chromosomal aberrations and copy number variations (CNVs) found in this cohort and described in literature. Trisomies13, 18 and 21 are risk factors for EA/TEF, although it is most often seen in people with trisomy 18. In chapter 2.3 we describe three patients with trisomy X and EA/TEF and exclude the role of non-random X-inactivation. The high frequency of trisomy X in our cohort indicates that this also could be a risk factor for gastro-intestinal anomalies, including EA/TEF [8, 9]

Many structural chromosomal anomalies have been described in patients with TE, but these hardly contribute to the total TE burden, in our cohort up to 0.3%, not including the aneuploidies. In chapter 2.1 and 2.2 we describe *de novo* and rare recurrent inherited

CNVs in patients with TE anomalies and VACTERL association. Large *de novo* CNVs have been associated with syndromes with characteristic features such as EA/TEF and other anomalies of the VACTERL association.[10] Consistent with the literature, *de novo* CNVs in our cohort are rare and non-recurrent. Thus they do not have a high impact on total TE disease burden. *De novo* CNVs can point to candidate genes possibly mutated in other patients. Others have detected one of those candidate genes, *LPP*, in a *de novo* deletion.[11] *LPP* copy number changes and mutations were not found in our large cohort, which is indicative of the low impact of LPP gene aberrations in patients with EA/TEF or VACTERL.[12]

CNVs can be inherited in a Mendelian manner and behave much like other genetic variation. Finding two overlapping rare CNVs in two unrelated patients in a cohort of patients with a rare disease can be indicative of susceptibility locus or gene. We found 23 loci affected more than once by a CNV; all of these CNVs were inherited from unaffected parents. Rare CNVs sometimes affect regions of the genome devoid of genes. However, these regions can still be of importance as we have shown in patients with alveolar capillary dysplasia and VACTERL-association features, including EA/TEF.[13] Two patients in our cohort had an overlapping de novo deletion of the maternal copy of an *FOXF1* upstream lncRNA affecting the expression of the paternal imprinted gene *FOXF1* [13]

In Chapter three, "Genetic studies in discordant monozygotic twins", we evaluate the value of genetic monozygotic twin studies and the presence of CNVs, insertions, deletions and single nucleotide variations in discordant monozygotic twins. Somatic events can result in differences in DNA variation between monozygotic twins [14] and may have contributed to the phenotypical differences in our discordant twin cohort. The detected rare and private CNV were present in both twins at the same frequency.[15] We also measured numerous DNA variation differences with whole exome sequencing (WES) and exon level copy number profiling. Statistical variant comparison reduced the number of putative differences to only a few candidates. Validation is ongoing, but preliminary results regrettably indicate that most of these differences are of a technical nature.

In chapter four, "Next generation sequencing in familial and consanguineous patients", we describe our ongoing efforts to discover genetic factors involved in a small subset of EA/TEF patients. These patients either have a family member with reported tracheo-esophageal anomalies or are born from consanguineous parents. By comparing the DNA variation of unaffected parents and affected family members with DNA of the index patients we identified many *de novo*, compound heterozygous and some X-linked and

homozygous recessive variation which might have an impact on patient phenotype. Further research is needed to evaluate their precise contribution. EA/TEF is a rare disease and is not likely to co-occur in a family twice just by chance. Remarkably, only one of two affected family members had a de novo *MYCN* mutation, earlier described in patients with Feingold syndrome.[16] Interestingly, all patients had rare and private variation in multiple genes from the candidate gene panel and in genes related to cilia structure and functioning. These inherited heterozygous variants mutations might act together in biological relevant processes and in doing so tip the balance from normal to abnormal development. Non-penetrance of syndrome specific characteristics can hinder its recognition. Screening patients for mutations in the most frequently affected syndromal EA/TEF genes (*SOX2*, *MYCN*, *CHD7*, *MID1*) might be warranted.

In chapter 4.2 we propose a valuable approach to explain previously "unsolved" etiology in VACTERL-association patients. Hitherto VACTERL association is a "diagnosis per exclusionem", given only when no other genetic syndrome can be diagnosed. Screening this group of VACTERL and VACTERL-like patients when candidate genes emerge can result in the identification of patients with a shared genotype and phenotype. In a whole exome sequencing experiment in patients with congenital anomalies of the kidney and urinary tract a mutation in the TNF receptor-associated protein-1 gene (*TRAP1*) was found in two patients. Subsequent screening for these TRAP1 mutations in the VACTERL cohort revealed three compound heterozygous mutated patients, all with renal anomalies.[17]

In the general discussion we argue that in view of the phenotypical and etiological heterogeneity in TE-anomalies studies in large patient and control cohorts are needed to elucidate factors involved and improve insight in foregut development. Since EA/TEF is rare, this is only feasible by collaborating in large disease consortia in which physicians, researchers and patients discuss their experiences, results and hypothesis and data and material from genetic and epidemiological studies is shared among institutions. Knowledge of environmental and other risk factors, genetic predisposition and causal genetic syndromes would improve parental and patient counseling and may result in the ability to predict co-morbidity and even to devise and employ preventive strategies.

Samenvatting

Oesophagusatresie (OA) is een zeldzame aangeboren afwijking waarbij een verbinding tussen het bovenste deel van de slokdarm en de maag ontbreekt. Bovendien bestaat er bij meer dan 90% van de gevallen een abnormale verbinding van het onderste deel van de slokdarm met de luchtpijp; de tracheo-oesophageale fistel (TOF).[1-3]

Deel een van dit proefschrift " Klinische en etiologische heterogeniteit bij patiënten met tracheo-oesophageale malformaties en andere geassocieerde afwijkingen" presenteert de huidige kennis over de etiologische aspecten van aangeboren slokdarmafwijkingen. Het blijkt dat er grote verschillen zijn in oorzaken en verschijningsvormen. Twee derde van de patiënten hebben ook andere afwijkingen, vaak afwijkingen die behoren tot de zogenaamde VACTERL-spectrum: vertebrale afwijkingen, anusatresie, cardiale afwijkingen, tracheo-oesophageale afwijkingen, renale afwijkingen en ledemaatsafwijkingen.[4] In dit eerste hoofdstuk beschrijven we de verschillende risico, milieu en genetische genetische factoren die ten grondslag kunnen liggen aan OA/TOF. OA/TOF komt voor bij meer dan 70 verschillende genetische syndromen, echter bij de meeste syndromen incidenteel.[5-7] OA/TOF wordt wel vaak gezien bij het anophthalmia-oesophageaal-genitiaal syndroom, Feingold syndroom, CHARGE syndroom en trisomie 18. Bij ongeveer negen procent van de patiënten uit het Erasmus MC-Sophia OA/TOF-cohort is een genetisch syndroom gediagnostiseerd. Bij 1-2 procent van de patiënten bestaat het sterke vermoeden dat de oorzaak bij specifieke milieufactoren moet worden gezocht. Dit betekent dat bij ongeveer 90 procent van de patiënten de oorzaak nog onbekend is. Met behulp van twee nieuwe technieken, SNP-array en whole exome sequencing, hebben we vervolgens genetisch onderzoek gedaan bij deze laatste groep. In deel twee beschrijven we de aneuploidiëen, structurele chromosomale afwijkingen en zogenaamde "copy number variations" (CNV) die we hebben gevonden. Het bleek dat een trisomie van chromosoom 13, 18 of 21 is geassocieerd met een verhoogd risico op OA/TOF, maar dat de afwijking het meest werd gezien bij patiënten met trisomie 18. Ook patiënten met trisomie X hebben een verhoogd risico hebben op gastro-intestinale malformaties, met inbegrip van OA/TOF.[8, 9] We hebben hierbij uitgesloten dat er een X-inactivatie voorkeur was voor een specifiek chromosoom X.

Chromosomale afwijkingen, aneuploidiëen niet meegeteld, kwamen voor bij slechts 0,3% van de patiënten in ons cohort. De hoofdstukken 2.1 en 2.2 beschrijven andere typen DNA–afwijkingen die we hebben gevonden: zeldzame *de novo* CNVs en overgeërfde CNVs. Grote *de novo* CNVs zijn vaker beschreven bij patiënten met genetische syndromen waar

OA/TOF en andere afwijkingen uit het VACTERL spectrum deel van kunnen uitmaken.[10] In overeenstemming met de literatuur kwamen *de novo* CNVs echter niet vaak voor in ons cohort. *De novo* CNV in ons cohort en beschreven in de literatuur, komt vaak maar een keer op die specifieke plaats in het DNA voor. Dit type variatie heeft derhalve geen grote impact heeft op het ontstaan van OA/TOF. Wel is het zo dat *de novo* CNVs kunnen wijzen op genen die gemuteerd kunnen zijn bij andere patiënten, en anderen hebben gevonden dat *LPP* een goede kandidaat is.[11] Echter, in ons grote cohort vonden we geen CNVs en single base pair mutaties in dit gen, en we vermoeden daarom dat dit gen slechts een kleine rol speelt bij het ontstaan van OA/TOF.[12]

CNVs kunnen een Mendeliaans overervingspatroon hebben maar ook kenmerken van andersoortige DNA-variaties. Wanneer twee zeldzame CNVs, die niet gevonden zijn in zeer grote controlecohorten, meerdere keren voorkomen bij een zeldzame ziekte als OA/TOF kan dit wijzen op een predispositie locus of gen. Wij hebben 23 van dit soort loci gevonden in ons cohort, alle overgeërfd van niet aangedane ouders. Deze zeldzame CNV kunnen genen bevatten, maar ook regio's in het DNA zonder genen kunnen een afwijkend aantal kopieën hebben. Dat deze regio's nog steeds belangrijk kunnen zijn laten we zien in hoofdstuk 2.5. Twee patiënten met alveolaire capillaire dysplasie en VACTERL-spectrum afwijkingen hebben een overlappende *de novo* deletie in de maternale kopie van een regio die een lncRNA bevat welke de expressie van het paternaal geïmprinte gen *FOXF1* reguleert. [13]

In deel drie "Genetische studies in discordante eeneiige tweelingen" evalueren we de waarde van genetische studies bij eeneiige tweelingen en beschrijven we onze studie naar de aanwezigheid van verschillen in single nucleotide variations, CNVs en insertions/deletions in discordante eeneiige tweelingen. Somatische DNA veranderingen kunnen resulteren in verschillen in de DNA-sequentie van eeneiige tweelingen[14] en het is mogelijk dat dit soort veranderingen bijdragen aan de fenotypische verschillen in ons discordante tweelingencohort. We hebben zeldzame CNVs (niet gezien in controlecohorten) gevonden, maar deze waren gelijkelijk aanwezig bij beide helften van een tweeling.[15] Ook hebben we het gehele eiwit-coderende deel van het DNA onderworpen aan "whole exome sequencing" en "exon level copy number profiling". Met een nieuwe statistische methode konden we het aantal gevonden verschillen binnen een tweeling paar terugbrengen tot enkele mogelijk relevante verschillen. Deze laatste worden nu verder onderzocht, maar de voorlopige resultaten laten zien dat ook deze verschillen van technische aard zijn.

Hoofdstuk vier "Next generation sequencing bij familiale en consanguïne patiënten" beschrijft ons lopende onderzoek bij twee specifieke patiëntenpopulaties. Deze patiënten hebben óf andere familieleden met tracheo-oesophagelae afwijkingen, óf zijn kinderen van bloedverwante ouders. Door nu de DNA variatie van de onaangedane ouders en aangedane familieleden te vergelijken met het DNA van de index patiënt denken we genetische factoren te kunnen identificeren die betrokken zijn bij het ontstaan van OA/TOF. We hebben inderdaad al bepaalde *de novo*, compound heterozygoot, X-linked en homozygoot recessieve variaties gevonden die van invloed kunnen zijn op het fenotype. Een van de patiënten had een *de novo* variant in het *MYCN* gen; dit is het gen dat aangedaan is bij patiënten met het Feingold syndroom. De gevonden variant is eenmaal eerder gezien bij een patiënt met dit syndroom. [16] Opmerkelijk is dat slechts een van de twee aangedane kinderen in deze familie deze *de novo MYCN* verandering had. Gezien de zeldzaamheid leek het op voorhand niet waarschijnlijk dat OA/TOF twee keer, onafhankelijk van elkaar, in een familie zou voorkomen. De verminderde penetrantie van sommige genetische syndromen kan de herkenning ervan hinderen. Screenen van de gehele patiëntenpopulatie op pathogene afwijkingen in de meest voorkomende OA/TOF genen (*SOX2*, *MYCN*, *CHD7*, *MID1*) wordt dan ook aanbevolen.

Een andere interessante waarneming is het voorkomen bij iedere patiënt van meerdere zeldzame varianten in genen uit ons kandidaat-genenpanel en in genen gerelateerd aan cilia structuur en werking. Deze varianten zijn overgeërfd van onaangedane ouders, en doordat ze samen van invloed zijn op relevante biologische processen kunnen ze mogelijk de balans verstoren tussen normale en abnormale ontwikkeling. Meerdere factoren samen kunnen zo wellicht zorgen voor een predispositie voor voordarm aandoeningen. Hoofdstuk 4.2 beschrijft een aanpak om de genetische oorzaak van de aangeboren afwijkingen te vinden bij patiënten gediagnostiseerd met "VACTERL–associatie". De diagnose "VACTERL associatie" is een "diagnosis per exclusionem", alleen gegeven wanneer geen andere syndroom kan worden geïdentificeerd. Het screenen van deze patiënten op DNA-veranderingen in relevante genen gevonden in andere studies kan resulteren in het vinden van meerdere patiënten met overeenkomstige aangeboren afwijkingen én met dezelfde genetische afwijking. Een whole exome sequencing experiment bij patiënten met aangeboren afwijkingen van de nieren en de urinewegen resulteerde in de identificatie van twee patiënten met een pathogene afwijking in de TNF receptor-associated protein-1 gene (*TRAP1*). Vervolgens werden in meerdere cohorten, waaronder het Rotterdam OA/TOF cohort, nog eens drie patiënten met afwijkingen in dit gen gevonden, allen met afwijkingen aan de nieren.[17]

In de algemene discussie "Voordarmontwikkeling, een kwestie van evenwicht" wijzen we er op dat onderzoek bij grotere groepen patiënten nodig is om een beter overzicht te krijgen van de factoren die zijn betrokken bij het ontstaan van OA/TOF. Omdat dit een zeldzame ziekte is, kan dit alleen verwezenlijkt worden door samen te werken in grote consortia waarin artsen, onderzoekers en patiënten hun ervaringen, resultaten van onderzoek, nieuwe hypothesen en gegevens en materiaal uit genetische en epidemiologische studies delen. Meer inzicht in milieu- en andere risicofactoren, genetische predispositie en causale genetische syndromen maakt betere begeleiding en voorlichting van ouders en patiënten mogelijk en kan wellicht helpen eventuele co-morbiditeit te voorspellen, of zelfs preventiebeleid mogelijk maken.

References

1. Genevieve, D., et al., *An overview of isolated and syndromic oesophageal atresia.* Clin Genet, 2007. **71**(5): p. 392-9.
2. Spitz, L., *Oesophageal atresia.* Orphanet J Rare Dis, 2007. **2**: p. 24.
3. Bax, K.N., A.M. Roskott, and D.C. van der Zee, *Esophageal atresia without distal tracheoesophageal fistula: high incidence of proximal fistula.* J Pediatr Surg, 2008. **43**(3): p. 522-5.
4. de Jong, E.M., et al., *Etiology of esophageal atresia and tracheoesophageal fistula: "mind the gap".* Curr Gastroenterol Rep, 2010. **12**(3): p. 215-22.
5. Felix, J.F., et al., *Environmental factors in the etiology of esophageal atresia and congenital diaphragmatic hernia: results of a case-control study.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(2): p. 98-105.
6. Felix, J.F., et al., *Genetic and environmental factors in the etiology of esophageal atresia and/or tracheoesophageal fistula: an overview of the current concepts.* Birth Defects Res A Clin Mol Teratol, 2009. **85**(9): p. 747-54.
7. Oddsberg, J., *Environmental factors in the etiology of esophageal atresia.* J Pediatr Gastroenterol Nutr, 2011. **52 Suppl 1**: p. S4-5.
8. Zizka, J., et al., *[Trisomy X in a neonate with esophageal atresia].* Cesk Pediatr, 1977. **32**(12): p. 723-4.
9. Brosens, E., et al., *Structural and numerical changes of chromosome X in patients with esophageal atresia.* European Journal of Human Genetics, 2013.
10. Southard, A.E., L.J. Edelmann, and B.D. Gelb, *Role of copy number variants in structural birth defects.* Pediatrics, 2012. **129**(4): p. 755-63.
11. Arrington, C.B., et al., *Haploinsufficiency of the LIM domain containing preferred translocation partner in lipoma (LPP) gene in patients with tetralogy of Fallot and VACTERL association.* Am J Med Genet A, 2010. **152a**(11): p. 2919-23.
12. Hernandez-Garcia, A., et al., *Contribution of LPP copy number and sequence changes to esophageal atresia, tracheoesophageal fistula, and VACTERL association.* Am J Med Genet A, 2012. **158a**(7): p. 1785-7.
13. Szafranski, P., et al., *Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder.* Genome Res, 2013. **23**(1): p. 23-33.
14. Bruder, C.E., et al., *Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles.* Am J Hum Genet, 2008. **82**(3): p. 763-71.
15. Veenma, D., et al., *Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts.* Eur J Hum Genet, 2012. **20**(3): p. 298-304.
16. Cognet, M., et al., *Dissection of the MYCN locus in Feingold syndrome and isolated oesophageal atresia.* Eur J Hum Genet, 2011. **19**(5): p. 602-6.
17. Saisawat, P., et al., *Whole-exome resequencing reveals recessive mutations in TRAP1 in individuals with CAKUT and VACTERL association.* Kidney Int, 2013.

Appendices

Dankwoord

Met trots en voldoening schrijf ik dit dankwoord. Natuurlijk met het besef dat ik op de schouders van mijn voorgangers, Janine, Liesbeth en Martje, sta. Dankbaar ook voor de inzet en toewijding van collega onderzoekers, analisten, kinderartsen, chirurgen, genetici, bio-informatici, verpleegkundigen 's, co-auteurs en velen anderen. Allen dragen bij aan dit onderzoek en daarmee ook met de totstandkoming van dit manuscript. Mijn dank gaat vooral ook uit aan alle patiënten, ouders en de patiëntenvereniging VOKS; zonder uw medewerking zou dit onderzoek niet mogelijk zijn geweest. Ook zijn de financiële bijdragen van het Erasmus MC (Mrace Grant) en de "Sophia stichting wetenschappelijk onderzoek" onontbeerlijk geweest voor dit onderzoek.

Dank aan mijn promotor, prof.dr. Dick Tibboel, voor de mogelijkheid om mijn promotietraject op de afdeling kinderchirurgie te volgen, de hulp en advies gedurende het project en bij het schrijven van dit manuscript. Mijn co-promotor, dr. Annelies de Klein, hartelijk dank voor de begeleiding en de mogelijkheid om mij zowel vakinhoudelijk als persoonlijk te ontwikkelen. Dank dat je me vrij liet ook zelf richting te geven aan de uitvoering van het onderzoek, en bij te sturen wanneer het nodig was. Annelies, het zijn geen gemakkelijke jaren geweest voor je, ik heb respect voor je kracht om, ondanks alles, zeer gedreven en enthousiast niet alleen te zorgen voor continuïteit binnen je eigen onderzoekslijnen, maar ook nog eens die van anderen in raad en daad bij te staan.

De leden van de leescommissie; prof.dr. Raoul Hennekam, prof.dr. Robert Hofstra en prof.dr.ir. Joris Veltman wil ik bedanken voor het lezen en beoordelen van het manuscript. De leden van de grote promotie commissie prof.dr. Sjaak Philipsen, dr. Titia Cohen, dr. Hanneke IJsselstijn en dr. Robbert Rottier hartelijk dank voor jullie interesse en het zitting nemen in de commissie. Robbert, dank ook voor het mentoraat en verbreden van mijn blik, voorbij de DNA sequentie, en het warme welkom bij de werkbesprekingen en literatuurdiscussies. Prof. dr. Irwin Reiss, bedankt voor uw advies en begeleiding gedurende het mentoraat.

Dank aan mijn paranimfen Bert en Leon. Bert, ik heb veel bewondering voor de niet aflatende energie en passie voor de genetica en alles wat daarbij komt kijken. Je kritische feedback was zeer waardevol en het is fijn om af en toe even stoom bij je te kunnen afblazen. Leon dank voor de vriendschap de afgelopen jaren, ik vind het een hele eer dat je me ook hier bij wilt staan. (help me eraan te herinneren deze keer wel alles mee naar huis te nemen)

Prof.dr. Rene Wijnen, dank voor uw aanwijzingen en hulp, uw enthousiasme werkt aanstekelijk. Kees van de Ven, John Vlot, Conny Meeussen en Pim Sloots bedankt voor jullie hulp bij de inclusie en het afnemen van de biopten. Prof.dr. Joost Gribnau en dr. Stefan Barakat dank voor jullie hulp bij de X-inactivatie experimenten. I would like to thank our national and international collaborators, prof.dr. Hugo Heij, dr. Ben Solomon, dr. Daryl Scott, prof.dr. Heiko Reutter, dr. Pawel Stankiewicz, prof.dr. Friedhelm Hildenbrandt, dr. Carlo Marcelis, dr. Iris van Rooij, dr. Andres Hernández-García, dr. Przemyslaw Szafranski, Pawaree Saisawat, Alina Hilger, Lotte Wijers and Hitisha Zaveri.

In de afgelopen jaren heb ik veel behulpzame en vriendelijke collega's gehad bij de oogheelkunde, kinderchirurgie, biomics en klinische genetica. Mijn dank dan ook voor de administratieve, ICT en praktische ondersteuning, het lab-werk en de analyses, maar vooral ook voor de gezelligheid tijdens het werk en op menig lab-dag, congres, borrel of etentje. Marialuisa, Simone and Christian, I really enjoyed the ASHG in Boston. Atze, Danny, Shimriet, Erik, Celine en Ronald, jammer dat ik dit jaar niet mee kan naar de MGC workshop, ik heb erg genoten in Luxemburg. Maria and Yunia, I enjoy our collaboration on your MMIHS project. Guido, hartelijk dank dat ik bij jou altijd terecht kon voor vragen en uitleg over PCR en Sanger sequencing. Mark, dank voor de adviezen en de scherpe vragen bij de werkbesprekingen. Drs. Yolande van Bever en dr. Alice Brooks, dank voor jullie hulp, input en kritische blik op de manuscripten en uitleg over de vele genetische syndromen. Ook de hulp van de diagnostiek afdelingen op de 20e en 24e waardeer ik erg, Diane en Lorette dank voor het (EBV-) kweekwerk, de DNA-isolatie, Guido en Robert, dank voor het array-werk.

Hannie, Jolanda, Anna, Daan, Tom, Hedwika, Robert, Marjan, Cindy, Ilse, Hanneke, Jacqueline, Thomas, Emine en Serdar jullie inzet , steun, hulp, collegialiteit en vooral ook gezelligheid maakten dat werk toch ook een beetje hobby was. Sanne en Kitty dank voor jullie ondersteuning en hulp bij de inclusie. Erik, Desiree, Daphne en Rhiana bedankt voor het invallen. Collega onderzoekers uit het Sophia, Bram, Marie-Chantal, Dorian, Alexandra,

Miriam, Nicole en Gerbrich, dank voor de samenwerking. Dr.ir. Wilfred van IJcken, Mirjam, Christel, Frank, Walter en Zeliha dank voor jullie collegialiteit, hulp en inzet bij het "NGS" deel van dit onderzoek. Rutger, veel dank voor je hulp en het meedenken bij de analyses. (overigens heb ik nooit geweten dat de metro in Parijs zo duur was) Heleen, Anne, Joshua, Kim, Lalini, Ilona, Yadi en Marjon dank voor de leerzame discussies tijdens de journal club sessies en werkbesprekingen. De studenten Stephanie, Laura, Linda, Rebecca, Daphne en Mirjam, dank voor jullie enorme inzet en hulp bij het onderzoek.

Met dit proefschrift sluit ik een periode van bijna tien jaar af, die begon met het verlangen mezelf meer te verdiepen in wetenschappelijk onderzoek. Sindsdien is menig weekendje, feestje of wandeling gesneuveld om mijn ambitie te verwezenlijken. Ik ben mijn vriendin, (schoon-) familie en vrienden veel dank verschuldigd. Niet alleen voor de raad en steun op de momenten dat ik het nodig had, maar vooral ook voor hun loyaliteit en begrip op de momenten dat ik er niet was. Neeltje, ik ben niet altijd de gezelligste geweest, ik waardeer jouw liefde, zorgzaamheid, geduld en begrip enorm. Ik heb mijn inhaalslag mogen maken.

Erwin

Curriculum Vitae

Erwin Brosens, born in Breda in 1976 finished secondary school at the Mencia de Mendoza lyceum in 1993 and vocational education (Biomedical laboratory techniques, histology/cytology major) in 1999. He received his Bachelor degree in 2002 at the Avans University of Applied Sciences. (Biomedical laboratory Sciences, biotechnology major). During his final year he studied protein-protein interactions of MADS – box transcription factors in Petunia hybrida en Arabidopsis thaliana at Plant research International in Wageningen under the supervision of prof. dr.ir. Richard Immink. After his graduation he started working as a cytogenetics' technician at the Clinical Genetics department of the Erasmus MC-Sophia Children's Hospital in Rotterdam under the supervision of dr. Cokkie Wouters and dr. Pino Poddighe. From 2008-2010 he worked as a research technician in the Uveal Melanoma group of dr. Annelies de Klein at the Rotterdam Eye Hospital and Clinical Genetics department. In the evening hours he obtained his second Bachelor degree (2009; health, nutrition and toxicology major) and his Masters (2010; health, nutrition and toxicology major) in Environmental Sciences at the Open University of The Netherlands. In 2010 he got the opportunity to work on this PhD project under the supervision of Prof. dr. Dick Tibboel and dr. Annelies de Klein at the departments of Pediatric Surgery and Clinical Genetics in Rotterdam.

PhD Portfolio

Summary of PhD training and teaching activities

Name PhD student: **Erwin Brosens**	PhD period: 01/11/2010-17/06/2014
Erasmus MC Department: Paediatric Surgery	Promotor(s): Prof. Dr. D. Tibboel
Research School: MGC	Supervisor: Dr. A. de Klein

1. PhD training

	Year	Workload (ECTS)
General academic skills		
- Biomedical English Writing and Communication	2013	4
- Research Integrity	2012	2
- Project Management	2011	4
- Introduction to Management and Organisation	2011	4
- Guidelines and Policy of Clinical Research ("BROK" Course)	2013	2
Research skills and Methodology		
- Basic Course on Statistics part A	2014	2
- Introduction Course on SPSS	2012	1
- Basic Course on R	2012	
In-depth courses (e.g. Research school, Medical Training)		
- Cell and Developmental Biology	2011	3
- Genetics Course	2011	3
- Biochemistry and Biophysics	2011	3
- Literature Course	2011-2012	2
- Presenting Skills	2012	1
Presentations		
- Wednesday morning meetings	2011-2013	1
- Monthly Bridge Meeting 2012	2012	0.5
- Sophia Scientific Research foundation Grant proposal (resulted in "Jan C. Molenaarprijs" for best scientific presentation)	2012	0.5
- Monthly KG/KC Science Meeting	2013-2014	0.5
- Ledendag Vereniging Ouderen en Kinderen met Slokdarmafsluiting (poster)	2013	0.5
- Landelijk overleg Cytogenetica / DNA diagnostiek	2013	0.5
- Erasmus MC NGS workflow and analysis pipeline meeting	2013	0.5
(Inter-)national conferences		
- Genomic Disorders - The Genomics of Rare Diseases, Cambridge (GBR) (Poster)	2011	1
- Genetica Retraite Rolduc,Kerkrade (NLD) (Presentation)	2012-2014	2
- European Society of Human Genetics Conference, Nurnberg (DEU) (Attendance)	2012	1
- Nederlandse Vereniging voor Humane Genetica, Arnhem (NLD) (Poster)	2012	1
- MGC PhD Workshop, Luxembourg (LUX) (Presentation)	2013	1
- European Society of Human Genetics Conference, Paris (FRA) (Poster)	2013	1
- American Society of Human Genetics Conference, Boston (USA) (Poster)	2013	1
- European Society of Human Genetics Conference, Milan (ITA) (Poster)	2014	1

1. PhD training		
	Year	**Workload (ECTS)**
Seminars and workshops		
- Safely working in the Laboratory	2002	0.5
- Browsing genes and genomes with ENSEMBL	2008	0,5
- The NEXUS training Course	2009+2010	1
- SNP's and Human Diseases Course VII	2010	2
- Next generation Sequencing Course	2011	1
- Molecular Diagnostics Course IV	2009	1
- Biomedical Research Techniques Course III (2011 &2008)	2008+2011	1,5
- Analysis of microarray gene expression data Course	2010	1
- Development, Stem Cells and Disease	2011	1
- CLC-BIO Course	2012	0.5
- A first encounter with next generation sequencing data	2012	1
- EBI Road show 2012	2012	1
- NGS 2013 Oxford: Bioinformatics and Data Analysis	2013	2
- NBIC NGS resequencing course	2013	1
- Career planning course	2014	1
Didactic skills		
- Training "omgaan met groepen voor tutoren"	2012	0.5
Other		
- Reviewing manuscripts	2011-2014	1
- Work and literature discussions	2010-2013	0.5
- Contact and Meetings with Software developers	2010-2013	1

2. Teaching activities		
	Year	**Workload (ECTS)**
Lecturing		
- Biomedical Research Techniques Course XI	2012	1
- Tutor 1th Year medical students	2012	1
Supervising practicals and excursions		
- Junior Science Instructor	2011+2012	3
- Biodiscovery Nexus CN training new users	2011-2013	0.5
Supervising Interns/ Master's theses		
- MLO Intern Stephanie Leeggangers	2011-2012	4
- MLO Intern Laura van 't Sand	2013	1
- BSc (HLO) Intern Tom Brands	2011-2012	2
- BSc (HLO) Intern Linda van Steen	2012-2013	2
- MSc Intern Daphne Huigh	2013	2
- MSc Intern Mirjam Ploeg	2013	2
- BSc (HLO) Intern Rebecca Veeris	2014	1
Other		
- Member Erasmus MC PhD committee	2011-2014	0.5
- Member "stuurgroep Internationalisering Koers 018"	2013-2014	0.5
- Member "Klankbordgroep middelbaar laboratorium onderwijs"	2012-2013	0.5
- Member BSc thesis defense committee	2012-2013	0.5

Publications

Copy Number Variation in monozygous twins

E. Brosens, K.G. Snoek, D. Veenma, H. Eussen, D. Tibboel, A. de Klein

Book chapter in "Genome Wide Association Studies: From Polymorphism to Pernonalized Medicine" Cambridge University Press; in press

Clinical and etiological heterogeneity in patients with tracheo-esophageal malformations and associated anomalies.

Erwin Brosens, Mirjam Ploeg, Yolande van Bever, Anna E. Koopmans, Hanneke IJsselstijn, Robbert J. Rottier, Rene Wijnen, Dick. Tibboel, Annelies de Klein; *Eur.J. Med. Genet. re-submitted 2014*

Increased Incidence of Hypertophic Pyloric Stenosis in Esophageal Atresia patients.

Nicole WG van Beelen, Daphne S Mous, **Erwin Brosens**, Annelies de Klein, Cees P van de Ven, John Vlot, Hanneke IJsselstijn, Rene MH Wijnen; *Eur J Pediatr Surg. 2014 Feb;24(1):20-4. doi: 10.1055/s-0033-1352527. Epub 2013 Aug 27.*

Structural and numerical changes of chromosome X in patients with esophageal atresia.

Erwin Brosens, Elisabeth M. de Jong, Tahsin-Stefan Barakat, Bert H Eussen, Barbara D'haene, Elfride de Baere , Pino P Poddighe, Robert-Jan Galjaard, Joost Gribnau, Alice S Brooks, Dick Tibboel, and Annelies de Klein; *Eur J Hum Genet. 2014 Jan 8. doi: 10.1038/ejhg.2013.295. [Epub ahead of print]*

Whole exome resequencing reveals recessive mutations in *TRAP1* in individuals with CAKUT and VACTERL association

Pawaree Saisawat, Stefan Kohl, Alina C. Hilger, Daw-Yang Hwang, Heon Yung Gee, Gabriel C. Dworschak, Velibor Tasic, Sivakumar Natarajan, Ethan Sperry, Danilo S. Matassa, Radovan Bogdanovic, Ivo de Blaauw, Carlo L.M. Marcelis, Charlotte H.W. Wijers, Enrika Bartels, Eberhard Schmiedeke, Dominik Schmidt, Stefanie Märzheuser, Sabine Grasshoff-Derr, Stefan Holland-Cunz, Michael Ludwig, **Erwin Brosens**, Hugo Heij, Dick Tibboel , Annelies de Klein, Markus M. Nöthen, Markus Draaken, Ben D. Solomon, Iris A.L.M. van Rooij, Franca Esposito, Heiko M. Reutter, and Friedhelm Hildebrandt; *Kidney Int. 2013 Oct 23. doi: 10.1038/ki.2013.417. [Epub ahead of print]*

VACTERL association aetiology: the impact of de novo and rare Copy Number Variations.

E. Brosens, H. Eussen, Y. van Bever, R. van Helm, H. IJselstijn, H.P. Zaveri, R. Wijnen, D.Scott, D. Tibboel, A. de Klein

Mol Syndromol. 2013 Feb;4(1-2):20-6. doi: 10.1159/000345577.

Contribution of *LPP* Copy Number and Sequence Changes in Esophageal Atresia, Tracheoesophageal Fistula, and VACTERL Association

Andres Hernández-García, **Erwin Brosens**, Hitisha P. Zaveri, Elisabeth M de Jong, Zhiyin Yu, Caraciolo J. Fernandes, Anthony Johnson, Maria Blazo, Seema Lalani, Dick Tibboel, Annelies de Klein, Daryl A. Scott
Am J Med Genet A. 2012 Jul;158A(7):1785-7. doi: 10.1002/ajmg.a.35391. Epub 2012 May 25.

Copy number detection in discordant monozygotic twins of Congenital Diaphragmatic Hernia (CDH) and Esophageal Atresia (EA) cohorts

Danielle Veenma, **Erwin Brosens**, Elisabeth de Jong, Kees van de Ven, Conny Meeussen, Titia Cohen-Overbeek, Marjan Boter, Hubertus Eussen, Hannie Douben, Dick Tibboel and Annelies de Klein
Eur J Hum Genet. 2012 Mar;20(3):298-304. doi: 10.1038/ejhg.2011.194. Epub 2011 Nov 9.

Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder

Przemyslaw Szafranski, Avinash V. Dharmadhikari,**Erwin Brosens**, Priyatansh Gurha, Katarzyna E. Kołodziejska,Ou Zhishuo, Piotr Dittwald, Tadeusz Majewski,K. Naga Mohan, Bo Chen, Richard E. Person, Dick Tibboel,Annelies de Klein, Jason Pinner, Maya Chopra, Girvan Malcolm,Gregory Peters, Susan Arbuckle, Sixto F. Guiang III,Virginia A. Hustead, Jose Jessurun, Russel Hirsch,David P. Witte, Isabelle Maystadt, Neil Sebire, Richard Fisher,Claire Langston, Partha Sen and Paweł Stankiewicz
Genome Res. 2013 Jan;23(1):23-33. doi: 10.1101/gr.141887.112. Epub 2012 Oct 3.

Expression profiling with and without Retinoic Acid (RA) in 15q26 deleted Congenital Diaphragmatic Hernia fibroblasts: relative cellular RA deficiency.

Danielle Veenma, **Erwin Brosens**, M.Peters, M. Jhamai, C. Wouters, M. Pescatori, R. Rottier, A. Kremer, Dick Tibboel and Annelies de Klein
PhD Thesis D. Veenma; p151-168

Fine mapping of structural chromosome 3 deletions in uveal melanoma cell lines

Th. van den Bosch, **E. Brosens**, D. Mooijman, J. Vaarwater, M.M. Verbiest, W. van Gils, H.T. Brüggenwirth, E. Kiliç, D. Paridaens, A. de Klein
PhD Thesis Th. van den Bosch; p115-135

Germline and Somatic SNP-array analysis of 117 isolated and complex Congenital Diaphragmatic Hernia patients-the Rotterdam Chohort-

D. Veenma, **E. Brosens**, D. Huigh, A. van Bodegom, H. Douben, B. Eussen, P.Jhamai, A.Brooks, Y.van Bever, R.Galjaard, C.Wouters, A.Uiterlinden, A.Flake, K.Kutsche, T.Schaible, R.Wijnen, D.Tibboel and A.de Klein.
PhD Thesis D. Veenma; p39-58