

ARTICLE

Unraveling genetic predisposition to familial or early onset gastric cancer using germline whole-exome sequencing

Ingrid P Vogelaar¹, Rachel S van der Post², J Han JM van Krieken², Liesbeth Spruijt¹, Wendy AG van Zelst-Stams¹, C Marleen Kets¹, Jan Lubinski³, Anna Jakubowska³, Urszula Teodorczyk³, Cora M Aalfs⁴, Liselotte P van Hest⁵, Hugo Pinheiro^{6,7}, Carla Oliveira^{6,7,8}, Shalini N Jhangiani^{9,10}, Donna M Muzny^{9,10}, Richard A Gibbs^{9,10}, James R Lupski^{9,10}, Joep de Lig¹, Lisenka ELM Vissers¹, Alexander Hoischen¹, Christian Gilissen¹, Maartje van de Vorst¹, Jelle J Goeman^{11,12}, Hans K Schackert¹³, Guglielmina N Ranzani¹⁴, Valeria Molinaro¹⁴, Encarna B Gómez García¹⁵, Frederik J Hes¹⁶, Elke Holinski-Feder¹⁷, Maurizio Genuardi¹⁸, Margreet GEM Ausems¹⁹, Rolf H Sijmons²⁰, Anja Wagner²¹, Lizet E van der Kolk²², Inga Bjørnevoll²³, Hildegunn Høberg-Vetti²⁴, Ad Geurts van Kessel¹, Roland P Kuiper¹, Marjolijn JL Ligtenberg^{1,2,25} and Nicoline Hoogerbrugge^{*,1,25}

Recognition of individuals with a genetic predisposition to gastric cancer (GC) enables preventive measures. However, the underlying cause of genetic susceptibility to gastric cancer remains largely unexplained. We performed germline whole-exome sequencing on leukocyte DNA of 54 patients from 53 families with genetically unexplained diffuse-type and intestinal-type GC to identify novel GC-predisposing candidate genes. As young age at diagnosis and familial clustering are hallmarks of genetic tumor susceptibility, we selected patients that were diagnosed below the age of 35, patients from families with two cases of GC at or below age 60 and patients from families with three GC cases at or below age 70. All included individuals were tested negative for germline *CDH1* mutations before or during the study. Variants that were possibly deleterious according to *in silico* predictions were filtered using several independent approaches that were based on gene function and gene mutation burden in controls. Despite a rigorous search, no obvious candidate GC predisposition genes were identified. This negative result stresses the importance of future research studies in large, homogeneous cohorts.

European Journal of Human Genetics (2017) 25, 1246–1252; doi:10.1038/ejhg.2017.138; published online 6 September 2017

INTRODUCTION

Annually, almost one million people develop gastric cancer (GC) and ~723 000 people die of this disease worldwide.¹ This makes GC the fifth most common malignancy and the third leading cause of cancer-related mortality worldwide.¹ In Western Europe, the incidence of gastric cancer (GC) is 8.8 per 100 000 persons for men and 4.3 per 100 000 persons for women.¹

GC is a multifactorial disease in which both genetic and environmental factors are involved. The main environmental factor is

infection with *Helicobacter pylori*, which increases the risk of developing GC about six-fold.² The World Health Organization (WHO) classified *H. pylori* as a class I carcinogen in 1994.^{3,4}

GC is a heterogeneous disease and can be roughly divided into three main types; diffuse-type GC, intestinal-type GC and a remaining group composed of mixed and indeterminate GC types.⁵ Diffuse-type GC (DGC) consists of poorly cohesive single cells without gland formation. Due to the frequent presence of signet ring cells, this type of GC is often referred to as signet ring cell carcinoma. Intestinal-type

¹Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands; ²Department of Pathology, Radboud university medical center, Nijmegen, The Netherlands; ³Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland; ⁴Department of Clinical Genetics, Academic Medical Centre, Amsterdam, The Netherlands; ⁵Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands; ⁶Expression Regulation in Cancer Group, Instituto de Investigação e Inovação em Saúde, Porto, Portugal; ⁷Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal; ⁸Department of Pathology and Oncology, Faculty of Medicine, University of Porto, Al Prof Hernâni Monteiro, Porto, Portugal; ⁹Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX, USA; ¹⁰Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX, USA; ¹¹Department for Health Evidence, Radboud university medical center, Nijmegen, The Netherlands; ¹²Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands; ¹³Department of Surgical Research, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; ¹⁴Department of Biology and Biotechnology, University of Pavia, Pavia, Italy; ¹⁵Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands; ¹⁶Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; ¹⁷Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München, München, Germany; ¹⁸Institute of Genomic Medicine, Catholic University of the Sacred Heart, Rome, Italy; ¹⁹Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; ²⁰Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; ²¹Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; ²²Family Cancer Clinic, The Netherlands Cancer Institute, Amsterdam, The Netherlands; ²³Department of Medical Genetics and Pathology, St. Olavs University Hospital, Trondheim, Norway; ²⁴Western Norway Familial Cancer Center, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway

*Correspondence: Professor N Hoogerbrugge, Department of Human Genetics, Radboud university medical center, P.O. Box 9101, Nijmegen 6500 HB, The Netherlands. Tel: +31 24 36 66205; Fax: +31 24 36 68752; E-mail: nicoline.hoogerbrugge@radboudumc.nl

²⁵These authors contributed equally to the work.

Received 21 November 2016; revised 7 July 2017; accepted 18 July 2017; published online 6 September 2017

GC (IGC) is composed of glandular or tubular components with various degrees of differentiation.⁶

In both low and high GC incidence countries, around 8–30% of patients with GC have a familial history of GC.^{7–11} Germline *CDH1* pathogenic mutations, predisposing to hereditary diffuse gastric cancer (HDGC), have been encountered in a subset of GC families.^{12–19} The International Gastric Cancer Linkage Consortium has recently broadened the *CDH1* testing criteria with the aim to identify as many *CDH1* mutation carriers as possible.²⁰

Families in whom no germline *CDH1* mutation can be identified remain genetically unexplained and may carry pathogenic mutations in other, yet unknown, GC susceptibility genes. Recently, DGC families with mutations in *CTNNA1*^{16,21} and *MAP3K6*²² have been described, but the exact contribution of these genes to GC predisposition remains unclear until more families with mutations in these genes are reported. In families with IGC exhibiting an autosomal dominant inheritance pattern, genetic susceptibility genes may also play a role, but no genes have yet been associated with this type of GC.

The aim of the current study was to identify novel candidate GC susceptibility genes using whole-exome sequencing of germline DNA isolated from the blood of patients suspected of genetic predisposition for GC, but without *CDH1* mutations.

MATERIALS AND METHODS

Patient selection for exome sequencing

In our exome sequencing cohort, 54 patients from 53 families meeting one of the following criteria were included: one gastric cancer diagnosed below the age of 35 years, two GC cases diagnosed in first- or second-degree relatives at or below the age of 60 years (index diagnosed at or below the age of 50 years) or three cases of GC in first- or second-degree relatives diagnosed at or below 70 years of age. The majority of the patients ($n=33$) had previously been proven negative for *CDH1* mutations. For each family a single patient was included, with the exception of one family for which two patients were tested. Patient characteristics are shown in Table 1. This study was approved by the medical ethics committee of the Radboud university medical center, reference number 2013/201 and the Institutional Review Board of the Baylor College of Medicine.

Exome sequencing, variant annotation and exclusion of normal variation

Detailed information on the sequencing statistics of individual samples can be found in Supplementary Table 1 Online. Whole-exome sequencing of genomic DNA extracted from peripheral blood cells of the patient was performed using the 5500XL SOLiD platform (Life Technologies, Bleiswijk, The Netherlands) for 26 samples and on the Illumina HiSeq (2x100bp paired end; Illumina, Inc., San Diego, CA, USA) for 13 samples (BGI, Copenhagen, Denmark). Exome enrichment was performed using either the human SureSelect All Exon 50 Mb kit ($n=11$) or the human SureSelect All Exon V4 kit ($n=28$), targeting the coding regions of ~21 000 human genes (Agilent Technologies, Santa Clara, CA, USA). Reads were mapped to the Human Genome Reference Assembly GRCh37/hg19 using LifeScope software (Life Technologies) for samples sequenced on the SOLiD instrument and variants were called with the DiBayes algorithm. Exomes that were sequenced on the Illumina instrument were mapped using BWA and variants were called with GATK. All variants were annotated using an in-house annotation pipeline, as described previously.^{23,24}

For 15 patients, exome sequencing was performed through the Human Genome Sequencing Center at Baylor College of Medicine, according to previously described methods.^{25,26} Sequencing was performed on the Illumina HiSeq 2000 platform (Illumina, Inc.). Subsequently, reads were mapped and aligned to the Human Genome Reference Assembly GRCh37/hg19 using the BCM-HGSC Mercury pipeline.²⁷ Variant calling was performed with the Atlas²⁸ and SAMtools²⁹ algorithms; variant annotation was performed with an in-house developed annotation pipeline³⁰ based on ANNOVAR.³¹ Custom scripts were used incorporating multiple databases to retrieve more information on identified variants.

From our total set of variants we selected high-confidence (≥ 5 variant reads or $\geq 20\%$ variant reads) non-synonymous variants that were absent from dbSNP or had a dbSNP (v132) frequency $< 1\%$ and which occurred at most once in our in-house variant database (2096 exomes, the majority of which are from European ancestry).²³

Enrichment of truncating variants compared to controls

The number of different truncating variants (nonsense variants, indels leading to a frameshift and variants in canonical splice sites) per gene was established for our data set and an independent in-house database containing 2329 exomes. Also, the number of genes with a given number of variants was determined for the combined datasets. The Fisher's exact test (incorporated in the IBM SPSS Statistics software package version 20, IBM Corporation, Armonk, NY, USA) was used to determine whether the number of variants for a certain gene in our set was enriched compared with the control data set. To correct for multiple testing we used the modified Bonferroni procedure for discrete data³² based on the number of genes with the same number of truncating mutations observed in the gene of interest. All genes with a P -value < 0.05 after multiple testing correction were included for further analysis.

Variant prioritization based on gene function

Missense variants with a PhyloP ≥ 3 in selected pathways (see below) were analyzed using the Alamut 2.0 software package (Interactive Biosoftware, Rouen, France), which incorporates SIFT,³³ PolyPhen-2,³⁴ Align GVGD³⁵ and dbSNP (build 135). Missense variants that were predicted deleterious/damaging by at least two of these programs were considered possibly deleterious. Data was analyzed using Alamut software between September 2013 and December 2014.

These possibly deleterious missense variants and the truncating variants were prioritized based on gene function using the following criteria. The first criterion used included variants in known hereditary (gastric) cancer predisposing genes. For this analysis we used an in-house generated list of 113 genes (Supplementary Table 2 Online). In addition, we assessed the recently described GC-predisposing genes *CTNNA1*^{16,21} and *MAP3K6*²² for variants. Second, we selected genes putatively involved in GC development. A gene list (Supplementary Table 2 Online) for this category was composed by combining a list of known tumor-suppressor genes³⁶ with genes from the following KEGG pathways:³⁷ regulation of actin cytoskeleton (entry 04810), adherens junction (entry 04520), focal adhesion (entry 04510), epithelial cell signaling in *H. pylori* infection (entry 05120) and pathways in cancer (entry 05200). Third, based on the detection of a homozygous putatively deleterious variant in *MYD88* in one of the patients of this cohort,³⁸ we used the Resource of Asian Primary Immunodeficiencies (RAPID) gene list,³⁹ an in-house generated candidate gene list and three KEGG pathways (JAK-STAT pathway (entry 04630), NF κ B pathway (entry 04064) and TLR pathway (entry 04620)³⁷ to select variants known to predispose to immunodeficiencies. As a fourth criterion, we selected genes with a high expression in the stomach (based on data from the Tissue-specific Gene Expression and Regulation (TiGER) database⁴⁰). The combined gene list for the categories mentioned above can be found in Supplementary Table 2 Online.

The National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP; 6503 exomes) database (hereafter referred to as EVS⁴¹), which contains sequencing data of ~6500 individuals of European and African descent was used to assess whether selected variants were present in individuals selected for other diseases than cancer. Furthermore, we used a second in-house database containing 2329 exomes with high coverage to exclude common variants. Finally, the database from the Exome Aggregation Consortium (ExAC⁴²) was used to obtain the frequency of specific variants in a larger control population.

For all truncating variants affecting genes not represented in the selected pathways presented above, the possible relation of the affected gene with GC tumorigenesis was evaluated based on the known function of the gene.

Variants described in this manuscript and Supplementary Tables Online are submitted to the Leiden Open Variant Database (LOVD, ID numbers 103989–104041).

Table 1 Patient characteristics exome sequencing cohort

Subject (UPN)	Country of inclusion	Age at diagnosis GC	Family history of GC	GC Criterium ^a	GC histology reviewed	Lauren	WHO	GC histology original pathology reports	Histology of GC cases in family	Other remarks
005A	The Netherlands	50	FDR DGC 50, FDR GC 48	3	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
018A	The Netherlands	26	SDR GC 27, TDR GC 55	3	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
023A	The Netherlands	43	FDR DGC 58	2	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
025A	The Netherlands	32	No	1	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
036A	The Netherlands	23	No	1	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	Consanguineous
042A	The Netherlands	35	FDR GC 50	2	No			Diffuse (poorly cohesive)	DGC	
049A	The Netherlands	66	FDR IGC 54, FDR GC 60, SDR IGC 83, SDR GC 50, SDR GC 58, SDR GC 60	3	Yes	Intestinal	Tubular	Intestinal (tubular)	IGC	
059A	The Netherlands	35	FDR DGC 51, FDR GC 22, FDR GC 42	3	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
059B	The Netherlands	51	FDR DGC 35, FDR GC 22, FDR GC 42	3	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
107A	The Netherlands	69	FDR GC 57, FDR GC 66, FDR GC 72, FDR GC 81	3	Yes	Intestinal	Tubular	Not specified 'adenocarcinoma'	IGC	
112A	The Netherlands	34	No	1	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
119A	The Netherlands	31	No	1	Yes	Intestinal	Tubular	Intestinal (tubular)	IGC	
120A	The Netherlands	31	No	1	Yes	Intestinal	Tubular	Not specified 'adenocarcinoma'	IGC	Consanguineous
121A	The Netherlands	27	No	1	No			Diffuse (poorly cohesive)	DGC	
124A	The Netherlands	54	FRD IGC 51, FDR IGC 53, TDR IGC 55, TDR GC 35	3	Yes	Intestinal	Mucinous	Intestinal (mucinous)	IGC	
125A	The Netherlands	33	FDR DGC 56	2	No			Diffuse (poorly cohesive)	DGC	
135A	The Netherlands	46	FDR GC 52	2	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	Patient also had ductal breast carcinoma at age 51
162A	The Netherlands	33	FDR IGC 42	2	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	Combined DGC and IGC	
164A	The Netherlands	36	No	1	Yes	Diffuse	Poorly cohesive	Not specified 'adenocarcinoma'	DGC	
167A	The Netherlands	31	No	1	Yes	Diffuse	Poorly cohesive	Diffuse (poorly cohesive)	DGC	
509A	Portugal	49	FDR GC, SDR GC, SDR GC	3	No			Diffuse (poorly cohesive)	DGC	
510A	Portugal	41	FDR DGC 21, FDR DGC 40	3	No			Diffuse (poorly cohesive)	DGC	
511A	Portugal	34	SDR DGC 54	2	No			Diffuse (poorly cohesive)	DGC	
512A	Portugal	48	FDR DGC, SDR DGC	3	No			Diffuse (poorly cohesive)	DGC	
513A	Portugal	30	Familial history of DGC, incomplete information	1	No			Diffuse (poorly cohesive)	DGC	
516A	Germany	29	No	1	No			Diffuse (poorly cohesive)	DGC	
520A	Germany	28	No	1	No			Diffuse (poorly cohesive)	DGC	

Table 1 (Continued)

Subject (UPN)	Country of inclusion	Age at diagnosis	Family history	GC			GC histology	Histology of	Other remarks
		GC	of GC	Criterion ^a	reviewed	Lauren	WHO	original pathology reports	
521A	Germany	43	FDR GC 47	2	No		Intestinal (tubular)	IGC	
522A	Germany	23	No	1	No		Not specified 'adenocarcinoma'	AC	
529A	Italy	34	SDR GC 41	2	No		Not specified 'adenocarcinoma'	AC	
530A	Italy	30	No	1	No		Not specified 'adenocarcinoma'	AC	
531A	Italy	34	No	1	No		Diffuse (poorly cohesive)	DGC	Patient also had basal cell carcinoma at age 42 and DGC 53
532A	Italy	30	No	1	No		Not specified 'adenocarcinoma'	AC	
701A	Poland	22	No	1	No		Diffuse (poorly cohesive)	DGC	
703A	Poland	68	FDR GC 42, FDR GC 65	3	No		Not specified 'adenocarcinoma'	AC	
711A	Poland	49	FDR GC 46	2	No		Intestinal (mucinous)	IGC	
716A	Poland	45	FDR GC 50	2	No		Intestinal (tubular)	IGC	
725A	Poland	38	FDR GC 50, SDR GC 56	3	No		Not specified 'adenocarcinoma'	AC	
727A	Poland	42	FDR GC 50, SDR GC 56	3	No		Not specified 'adenocarcinoma'	AC	
729A	Poland	48	FDR GC 40, FDR GC 79	3	No		Intestinal (tubular)	IGC	Patient also had liver cancer age 50
730A	Poland	29	SDR GC 35	2	No		Intestinal (tubular)	IGC	
731A	Poland	28	No	1	No		Not specified 'adenocarcinoma'	AC	
733A	Poland	70	FDR GC 52, FDR GC 62	3	No		Not specified 'adenocarcinoma'	AC	
741A	Poland	29	No	1	No		Diffuse (poorly cohesive)	DGC	
743A	Poland	25	FDR GC 49	2	No		Not specified 'adenocarcinoma'	AC	
755A	Poland	30	No	1	No		Not specified 'adenocarcinoma'	AC	
756A	Poland	50	FDR GC 48	2	No		Mixed	Mixed	
759A	Poland	25	No	1	No		Diffuse (poorly cohesive)	DGC	
760A	Poland	30	FDR GC 70, SDR GC <75, SDR GC <75	1	No		Not specified 'adenocarcinoma'	AC	
762A	Poland	32	FDR GC 46	2	No		Not specified 'adenocarcinoma'	AC	
763A	Poland	45	FDR GC 40	2	No		Not specified 'adenocarcinoma'	AC	
772A	Poland	34	No	1	No		Not specified 'adenocarcinoma'	AC	
774A	Poland	33	No	1	No		Diffuse (poorly cohesive)	DGC	
780A	Poland	36	FDR GC 39	2	No		Diffuse (poorly cohesive)	DGC	

Abbreviations: AC, adenocarcinoma not specified; DGC, diffuse-type gastric cancer; FDR, first-degree relative; IGC, intestinal-type gastric cancer; SDR, second-degree relative; TDR, third-degree relative.

^aOne gastric cancer diagnosed below the age of 35 years, two GC cases diagnosed in first- or second-degree relatives at or below the age of 60 years (index diagnosed at or below the age of 50 years) or three cases of GC in first- or second-degree relatives of GC diagnosed at or below 70 years of age.

Validation of variants and *CDH1* exon 1 germline mutation analysis by Sanger sequencing

The DNA sequence surrounding the variant was amplified using polymerase chain reaction (PCR, primer sequences and PCR conditions are available on request) and screened for mutations using BigDye terminator sequencing (BigDye Terminators (v 1.1) Applied Biosystems, Foster City, CA, USA). Analysis was performed on an ABI 3730 DNA Analyzer (Applied Biosystems). Subsequently, the data were analyzed using Vector NTI advance v11.0 (Invitrogen Corporations, Paisley, UK) or Chromas Lite (Technelysium, Australia). For mutation analysis of *CDH1* exon 1 in a subset of the patients, primers surrounding the intron-exon boundaries of this exon were used. PCR and sequencing was performed as described for variant validation.

RESULTS

Patient cohort and characteristics

Whole-exome sequencing was performed on germline DNA from 54 patients of 53 families. In this cohort 23 patients below the age of 35 were included (two without a family history of GC), 16 patients had two cases of GC in the family at or below the age of 60 and 15 patients were from families with three or more GC cases at or below 70 years of age; in this group two patients from one family were included. The mean age at diagnosis of all patients included was 37.9 years (SD 11.9, range 22–70).

According to the original pathology reports, 27 patients had DGC, 8 patients had IGC, one GC was mixed-type and 18 tumors were 'adenocarcinoma not otherwise specified'. For 17 cases, we were able to review and confirm the histology of the GC (12 patients with DGC and five with IGC). For the remaining cases no revision could be performed.

Exome sequencing statistics

Three different enrichment kits and two different sequencing platforms were used for exome sequencing. On average, 5.2 Gb of data aligned to targets was generated per sample (range: 2–10.2 Gb), hitting 98.9% of the targets (96–99.98%) with an average coverage of $81.7 \times$ ($36.8\text{--}132 \times$). A coverage of at least 10-fold was reached for 93.3% of

the targets (81.1–99.31%) and 87.7% was covered at least 20-fold (68.9–98.19%). The statistics for individual exome sequencing data can be found in Supplementary Table 1 Online. Approximately 9200 variants from 54 cases remained (average 170, range 87–379). To test our quality settings, we performed Sanger sequencing on a subset of these variants and found that we were able to confirm 93% of the variants.

Variants in previously described gastric cancer predisposition genes

None of the cases carried pathogenic mutations in *CDH1*. Also, no variants were found in *CTNNA1*, which has recently been described as GC-predisposing gene.^{16,21} Three variants were identified in *MAP3K6*; two missense variants (p.Y591C and p.L541P) and one amino acid deletion (p.K1125del). *MAP3K6* has been associated with familial GC.²² However, based on the high number of *MAP3K6* variants found in non-cancerous controls, we did not follow-up on these variants.

Enrichment of truncating variants compared with controls

As the deleterious effect of truncating variants (nonsense variants, indels leading to a frameshift and variants in canonical splice sites) is most prominent, we tested whether the recurrence of truncating mutations in a given gene was different from that in a control exome sequence data set of 2329 individuals. In our data set, in 12 genes two different truncating variants occurred (Supplementary Table 3 Online). After correction for multiple testing, no enrichment of truncating variants was found in these genes compared to the control cohort.

Occurrence of homozygous or compound heterozygous variants

To explore the occurrence of pathogenic changes in GC predisposition genes that follow a recessive inheritance pattern, we explored whether missense and/or truncating variants occurred in a homozygous or compound heterozygous form in our set of patients with an age at diagnosis below 35. Apart from a germline homozygous missense variant in *MYD88* in a patient with GC at age 23 and recurrent candidiasis, which we previously published,³⁸ no other candidate genes were found.

Variant prioritization based on gene function

Because of the large amount of missense and truncating variants, prioritization was performed based on the function and recurrence of the affected genes. To select variants that may be involved in GC predisposition, we created a gene list composed of 1899 genes (for details see Materials and Methods and Supplementary Table 2 Online). Our exome data were then filtered for variants in these genes. After *in silico* prediction, 252 possibly deleterious variants were identified using this approach (on average 5 per patient, range 1–11). The number of variants in the individual pathways and databases can be found in Table 2, variant details are shown in Supplementary Table 4 Online (excluding variants that were not confirmed after validation using Sanger sequencing). Twenty-one variants were identified in known cancer predisposing genes (Table 2; Supplementary Table 4 Online), including four different heterozygous variants (two truncating and two missense) in the *ATM* gene, previously associated with a small increased GC risk (RR = 3.39, 95% CI = 0.86 to 13.4).⁴³ No obvious candidate GC predisposition genes were identified from either the hereditary cancer list, or the other pathways we selected. In addition, for all truncating variants affecting genes not represented in the selected pathways, the putative relation with GC tumorigenesis of the affected gene was evaluated based on the known function of the gene. This did not result in a convincing candidate gene.

Table 2 Number of potential deleterious variant calls in different pathways

Pathway/gene list (number of genes in pathway)	Total number of variants
KEGG Actin cytoskeleton (151) ^a	17
KEGG Adherens junction (73) ^a	20
KEGG Focal adhesion (205) ^a	43
KEGG Helicobacter pylori (68) ^a	4
KEGG JAK-STAT pathway (158) ^a	3
KEGG NFκB signaling (92) ^a	12
KEGG TLR pathway (106) ^a	7
KEGG Pathways in cancer (327) ^a	48
TIGER database (207) ^b	16
Asian Primary ImmunoDeficiencies (247) ^c	38
In-house generated gene list of genes predisposing to immunodeficiencies (271)	37
In-house generated gene list of genes predisposing to hereditary cancer (113)	21
Human TSGene (716) ^d	107

^aSee reference.³⁷

^bSee Liu *et al*.⁴⁰

^cSee Keerthikumar *et al*.³⁹

^dSee Zhao *et al*.³⁶

DISCUSSION

In the current study, whole-exome sequencing was performed on germline DNA from 54 GC patients from 53 families with the aim to identify novel GC susceptibility genes. In order to increase the likelihood of finding a putative causative mutation underlying GC, these patients were selected from families at high risk of a genetic predisposition, who met strict inclusion criteria. No clear novel GC predisposition gene was identified.

Mutations in *CDHI* were not detected in the 21 cases for which no *CDHI* mutation analysis had been performed prior to our study. In two recent studies, germline mutations in *CTNNA1* were identified in families with GC.^{16,21} This gene is in the same pathway as E-cadherin, making it a plausible GC predisposition gene. In our data set of 54 patients no mutations in *CTNNA1* were found, which indicates that mutations in this gene probably do not explain a large proportion of the early onset gastric or familial cancer in patients that tested negative for germline *CDHI* mutations.

Gaston *et al.* reported on variants in the *MAP3K6* gene in familial GC.²² We have also observed variants in this gene, but we do not consider this gene a strong GC candidate gene for two reasons. Firstly, in the study by Gaston *et al.* the gene variant p.P946L was identified in a large family, but the variant does not completely segregate with the disease.²² Secondly, this variant occurs quite frequently in the ExAc database ($n=640/0.5\%$ allele frequency). This argues against its pathogenicity, simply because it is not expected that a variant that occurs so frequently in a database containing exomes of people without suspicion of hereditary cancer would cause GC, a relatively rare form of hereditary cancer.

The observation that frequently occurring variants are reported as candidate genes for GC development stresses the importance to determine the frequency of variants in candidate genes in local and public control datasets in addition to assessment of functional relevance. In the current study, we have used three datasets to compare our exome data with. The first one is an independent in-house database containing 2329 exomes sequenced with high coverage. The second one is the EVS database,⁴¹ which contains sequencing data of ~6500 individuals of European and African descent. The third is the ExAc database,⁴² containing exome data of 60 706 unrelated individuals. These datasets allowed for more stringent filtering of the data.

Even though we identified over 9200 rare variants in these 54 patients, we were unable to unequivocally show that among these are the disease-causing variants and, therefore, GC-predisposing genes. There may be several reasons for the fact that we did not find clear candidate genes in the current study. For example, the large amount of variants in our data set may have influenced our ability to recognize candidate genes as such. Additionally, a gene similar to the previously described *CTNNA1* mutations (which account for only a small portion of GC families) may well have remained undetected in our data set. Furthermore, our strictly selected patient cohort might be too small to identify candidate genes, especially to determine enrichment of genes or pathways compared to controls. Another reason is that, even though we used strict criteria for the selection of families, the patient group that we included in this study is still quite heterogeneous. We included patients with both DGC and IGC, who were either diagnosed at young age or had a family history of GC. For a number of cases the histological subtype was unknown, underscoring the importance of extensive pathological review and reporting of GC according to current guidelines.²⁰ It may be very well possible that performing whole-exome sequencing in a more homogeneous patient cohort may allow for improved detection of candidate genes, although other

studies also did not identify promising new GC predisposition genes.^{16,44,45} Also, we performed exome sequencing, whereas predisposing variants may also be in non-protein coding parts of the genome, currently not analyzed. Since we collected only one family member for each family and affected family members are often deceased due to the cancer, we were not able to follow-up on potential candidate genes. Finally, it could be possible that some of the patients we included developed cancer because of chance and occasional familial clustering or complex inheritance involving multiple genomic alterations.

Future perspectives

Taken together, we performed exome sequencing in 54 *CDHI* mutation-negative patients from 53 families. In this study we did not identify obvious candidate genes for GC predisposition. Future studies should be performed in larger, homogeneous patient groups and we would suggest that data from different research groups should be combined to identify candidate genes in these families. If candidate genes are identified this way, it will enable better preventive care in carriers of these mutations.

CONFLICT OF INTEREST

JRL has stock ownership in 23andMe and Lasergen, is a paid consultant for Regeneron, and a coinventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from the clinical exome sequencing offered in the Baylor Miraca Genetics Laboratory (BMGL; <http://www.bmgl.com/BMGL/>). The remaining authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Lilian Vreede, Hanna Feunekes, Eveline Kamping and Neeltje Arts from the Radboud university medical center for excellent technical assistance. This work was supported by the Radboud university medical center for Oncology, granted in 2010, the Dutch Cancer Society (KUN 2013-5876, RSvdP), ZONMW (917-10-358), the Deutsche Krebshilfe Grant 70-2371, FEDER - *Fundo Europeu de Desenvolvimento Regional* funds through the COMPETE 2020 - Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, by Portuguese funds through the Portuguese Foundation for Science and Technology (FCT)/Ministério da Ciência, Tecnologia e Inovação, in the framework of the projects: (1) 'Institute for Research and Innovation in Health Sciences' (POCI-01-0145-FEDER-007274); (2) FCOMP-01-0124-FEDER-015779 (ref. FCT PTDC/SAU-GMG/110785/2009), a Post-doc grant to HP (ref. SFRH/BPD/79499/2011), by the US National Human Genome Research Institute (NHGRI) National Heart Lung and Blood Institute (NHBLI) grant U54HG006542 to the Baylor-Hopkins Center for Mendelian Genomics, NINDS grant RO1 NS058529 to JRL and NHGRI 5U54HG003273 to RAG.

- 1 Ferlay J, Soerjomataram I, Ervik M *et al*: GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC CancerBase No. 11 (Internet). *Int Agency Res Cancer* 2013; **42**: 124-140.
- 2 Webb PM, Law M, Varghese C *et al*: Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut* 2001; **49**: 347-353.
- 3 Infection with *Helicobacter pylori*. *IARC Monogr Eval Carcinog Risks Hum* 1994; **61**: 177-240.
- 4 Suerbaum S, Michetti P: *Helicobacter pylori* infection. *N Engl J Med* 2002; **347**: 1175-1186.
- 5 Lauren P: The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* 1965; **64**: 31-49.

- 6 Bosman FT, Carneiro F, Hruban RH, Theise ND: *WHO Classification of Tumours of the Digestive System, 4th edition*. IARC: Lyon, France, 2010.
- 7 Han MA, Oh MG, Choi IJ *et al*: Association of family history with cancer recurrence and survival in patients with gastric cancer. *J Clin Oncol* 2012; **30**: 701–708.
- 8 Bernini M, Barbi S, Roviello F *et al*: Family history of gastric cancer: a correlation between epidemiologic findings and clinical data. *Gastric Cancer* 2006; **9**: 9–13.
- 9 Kawasaki K, Kanemitsu K, Yasuda T, Kamigaki T, Kuroda D, Kuroda Y: Family history of cancer in Japanese gastric cancer patients. *Gastric Cancer* 2007; **10**: 173–175.
- 10 La Vecchia C, Negri E, Franceschi S, Gentile A: Family history and the risk of stomach and colorectal cancer. *Cancer* 1992; **70**: 50–55.
- 11 Roviello F, Corso G, Pedrazzani C *et al*: High incidence of familial gastric cancer in Tuscany, a region in Italy. *Oncology* 2007; **72**: 243–247.
- 12 Guilford P, Hopkins J, Harraway J *et al*: E-cadherin germline mutations in familial gastric cancer. *Nature* 1998; **392**: 402–405.
- 13 Oliveira C, Seruca R, Carneiro F: Genetics, pathology, and clinics of familial gastric cancer. *Int J Surg Pathol* 2006; **14**: 21–33.
- 14 Kaurah P, MacMillan A, Boyd N *et al*: Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. *JAMA* 2007; **297**: 2360–2372.
- 15 Oliveira C, Senz J, Kaurah P *et al*: Germline CDH1 deletions in hereditary diffuse gastric cancer families. *Hum Mol Genet* 2009; **18**: 1545–1555.
- 16 Hansford S, Kaurah P, Li-Chang H *et al*: Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond. *JAMA Oncol* 2015; **1**: 23–32.
- 17 Benusiglio PR, Malka D, Rouleau E *et al*: CDH1 germline mutations and the hereditary diffuse gastric and lobular breast cancer syndrome: a multicentre study. *J Med Genet* 2013; **50**: 486–489.
- 18 Molinaro V, Pensotti V, Marabelli M *et al*: Complementary molecular approaches reveal heterogeneous CDH1 germline defects in Italian patients with hereditary diffuse gastric cancer (HDGC) syndrome. *Genes Chromosomes Cancer* 2014; **53**: 432–445.
- 19 van der Post RS, Vogelaar IP, Manders P *et al*: Accuracy of hereditary diffuse gastric cancer testing criteria and outcomes in patients with a germline mutation in CDH1. *Gastroenterology* 2015; **149**: 897–906.
- 20 van der Post RS, Vogelaar IP, Carneiro F *et al*: Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J Med Genet* 2015; **52**: 361–374.
- 21 Majewski IJ, Kluijft I, Cats A *et al*: An alpha-E-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer. *J Pathol* 2013; **229**: 621–629.
- 22 Gaston D, Hansford S, Oliveira C *et al*: Germline mutations in MAP3K6 are associated with familial gastric cancer. *PLoS Genet* 2014; **10**: e1004669.
- 23 Vissers LE, de Ligt J, Gilissen C *et al*: A de novo paradigm for mental retardation. *Nat Genet* 2010; **42**: 1109–1112.
- 24 de Voer RM, Geurts van Kessel A, Weren RD *et al*: Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. *Gastroenterology* 2013; **145**: 544–547.
- 25 Lupski JR, Gonzaga-Jauregui C, Yang Y *et al*: Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. *Genome Med* 2013; **5**: 57.
- 26 Bosch DG, Boonstra FN, Gonzaga-Jauregui C *et al*: NR2F1 mutations cause optic atrophy with intellectual disability. *Am J Hum Genet* 2014; **94**: 303–309.
- 27 Bainbridge MN, Wiszniewski W, Murdock DR *et al*: Whole-genome sequencing for optimized patient management. *Sci Transl Med* 2011; **3**: 87re83.
- 28 Shen Y, Wan Z, Coarfa C *et al*: A SNP discovery method to assess variant allele probability from next-generation resequencing data. *Genome Res* 2010; **20**: 273–280.
- 29 Li H, Handsaker B, Wysoker A *et al*: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
- 30 Gonzaga-Jauregui C, Lotze T, Jamal L *et al*: Mutations in VRK1 associated with complex motor and sensory axonal neuropathy plus microcephaly. *JAMA Neurol* 2013; **70**: 1491–1498.
- 31 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164.
- 32 Tarone RE: A modified Bonferroni method for discrete data. *Biometrics* 1990; **46**: 515–522.
- 33 Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073–1081.
- 34 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
- 35 Tavtigian SV, Deffenbaugh AM, Yin L *et al*: Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 2006; **43**: 295–305.
- 36 Zhao M, Sun J, Zhao Z: TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res* 2013; **41**: D970–D976.
- 37 KEGG: Kyoto Encyclopedia of Genes and Genomes.
- 38 Vogelaar IP, Ligtenberg MJ, van der Post RS *et al*: Recurrent candidiasis and early-onset gastric cancer in a patient with a genetically defined partial MYD88 defect. *Fam Cancer* 2016; **15**: 289–296.
- 39 Keerthikumar S, Raju R, Kandasamy K *et al*: RAPID: Resource of Asian Primary Immunodeficiency Diseases. *Nucleic Acids Res* 2009; **37**: D863–D867.
- 40 Liu X, Yu X, Zack DJ, Zhu H, Qian J: TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* 2008; **9**: 271.
- 41 Exome Variant Server NESPE, Seattle, WA: *Exome Variant Server, NHLBI Exome Sequencing Project (ESP)*.
- 42 Lek M, Karczewski KJ, Minikel EV *et al*: Analysis of protein-coding genetic variation in 60 706 humans. *Nature* 2016; **536**: 285–291.
- 43 Thompson D, Duedal S, Kirner J *et al*: Cancer risks and mortality in heterozygous ATM mutation carriers. *J Natl Cancer Inst* 2005; **97**: 813–822.
- 44 Thutkawkorapin J, Picelli S, Kontham V, Liu T, Nilsson D, Lindblom A: Exome sequencing in one family with gastric- and rectal cancer. *BMC Genetics* 2016; **17**: 41.
- 45 Donner I, Kiviluoto T, Ristimaki A, Aaltonen LA, Vahteristo P: Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. *Fam Cancer* 2015; **14**: 241–246.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2017

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)