Embryonic Stem Cell Proteins and MicroRNAs in the Etiology of Germ Cell Cancer

Ronak Eini

Molecular Medicine
Postgraduate School

The research for this thesis was performed within the framework of the Erasmus Postgraduate School molecular Medicine at the Department of Pathology, Erasmus Medical Center, Rotterdam, The Netherlands.

The research described in this thesis was funded by the Dutch Cancer Society.
Printing of this thesis was financially supported by:
Dutch cancer society
J.E.Jurriaanse Stichting
Novartis Pharma B.V.

Cover design: Nikki Vermeulen, Ridderprint BV, Ridderkerk, the Netherlands
Layout: Nikki Vermeulen, Ridderprint BV, Ridderkerk, the Netherlands
Printing: Ridderprint BV, Ridderkerk, the Netherlands

ISBN: 978-90-5335-693-7

All rights reseved. No parts of this thesis may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by means, without prior written permission of the author.

Embryonic Stem Cell Proteins and MicroRNAs in the Etiology of Germ Cell Cancer

Embryonale stamcel eiwitten en microRNAs in de etiologie van kiemcelkanker

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Wednesday 19 June 2013 at 13:30 hours

by

Ronak Eini

Born in Tehran, Iran

Doctoral committee:

Promotor:	Prof.dr. L.H.J. Looijenga
Other members:	Prof.dr. J.W. Oosterhuis
	Prof.dr. R. Fodde
	Prof.dr. G.W. Jenster

"*Wisdom is not a product of schooling but of the lifelong attempt to acquire it.*"

- Albert Einstein

To my parents

Contents:

List of abbreviations:

AFP	alpha fetoprotein
AR	androgen receptor
CDK-1	cyclin-dependant kinase 1
CIS	carcinoma *in situ*
CNV	copy number variation
COSMIC	catalogue of somatic mutations in cancer
DG	dysgerminoma
DNA	deoxyribonucleic acid
DNMT	DNA methyltrasferase
EC	embryonal carcinoma
ES cells	embryonic stem cells
FACS	fluorescent activated cell sorting
FFPE	formalin-fixed paraffine-embedded
FISH	fluorescent *in situ* hybridization
GB	gonadoblastoma
GCC	germ cell cancer
hCG	human choronic gonadotropine
hMSCs	human mesenchymal stem cells
HRM	high resolution melting
ICM	inner cell mass
IGCNU	intratubular germ cell neoplasia unclassified
iPS cells	induced pluripotent stem cells
LATS2	large tumor suppressor
LIF	Leukemia inhibitory factor
miR	microRNA
MLPA	multiple ligation-dependent probe amplification
NAP-1	nucleasome assembly protein-1
NPM1	nucleophosmin 1
NT	normal testis
PCA	principle component analysis
PGC	primordial germ cell
pRB	phosphorylation of retinoblastoma protein
qRT-PCR	quantitative reverse transcriptase polymerase chain reaction
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
SE	seminoma
siRNA	short interfering RNAs

SOM	self organizing map
SS	spermatocytic seminoma
TGCT	testicular germ cell tumors
TOI	target of interest
TSmiR	targeted serum miR
TSPY	testis-specific-protein Y-encoded
YS	yolk sac tumor
ZGA	zygotic genome activation

CHAPTER 1

General introduction

1.1) Embryonic stem cells:

In the early 1980s, a population of unique cells was isolated from the inner cell mass (ICM) of the mouse pre-implantation embryo named embryonic stem (ES) cells. These cells were generated by removing the ICM from pre-implantation blastocysts. The resulting cells were found to be pluripotent and showed unrestricted proliferative potential [1]. Human ES cell lines have now been derived in many labs from totipotent cells of the early mammalian embryo and are capable of unlimited, undifferentiated proliferation *in vitro* [2-5]. Essential characteristics of ES cells include: a) derivation from the pre-implantation embryo, b) prolonged (undifferentiated) proliferation and c) stable developmental potential to form derivatives of all three embryonic germ layers even after prolonged culture [6]. These unique properties of ES cells make them exceptionally valuable for cell replacement therapy and regenerative medicine, defining their potential clinical applications.

Mouse and human ES cells:

Mouse ES cells express proteins characteristic of the early blastocyst, such as the POU domain transcription factor Oct3/4, the homeobox domain transcription factor Sox2, the zinc finger protein Rex-1, the transcriptional activator Utf-1, as well as carbohydrate epitopes Ssea-1. These findings are based on the use of specific antibodies at the preimplantation embryo stage. Mouse ES cells do not spontaneously differentiate into trophectoderm. They can be propagated in continuous culture on feeder layer of mouse embryonic fibroblasts or without feeders in the presence of leukemia inhibitory factor (Lif), retaining high telomerase levels, with a stable karyotype and retention of their ability to contribute to chimeras and to form teratomas after multiple passages in culture [reviewed in 7].

Recent studies suggest that differences exist between mouse and human ES cell lines. Similar to mouse ES cells expressing Ssea-1 (see above), at least some of the human ES cells lines express specific surface markers (glycolipids and glycoproteins) such as SSEA-1, SSEA-3, SSEA-4, TRA-1-81, TRA-1-60 [8, 9]. On the other hand, human ES cells, unlike their mouse counterparts, do not require LIF for their propagation or for maintenance of pluripotency [6, 10]. Furthermore, in contrast to mouse ES cells, human ES cells are able to differentiate into trophoectoderm [6, 11]

In a study in which a commonly used mouse ES cell line (D3) was compared to the most used human ES cell line (H1), it is shown that several markers show the same profile, independent on the species. However, differences exist, including morphology, patterns of embryonic antigen immunostaining, expression of differentiation markers, as well as expression profiles of cytokines, cell cycle and cell death-regulating genes [7]. Overall, the strategies to regulate self-renewal and differentiation are likely similar among ES cell lines from different species. However, significant differences are present and it will be important to identify and characterize

such differences before a useful set of markers for the differentiated and undifferentiated state can be developed [7].

Some of the most important genes involved in ES cell regulation are shown in the Table indicating their expression in mouse and human ES specifically, or both. Their biology is also indicated.

Table. Overview of mouse and human embryonic stem (ES) cell characteristics. Information on the biology of the markers is provided.

Surface markers	Human ES cells	Mouse ES cells	Classification
SSEA-1	-	+	Carbohydrate associated molecules
SSEA-3	+	-	Carbohydrate associated molecules
SSEA-4	+	-	Carbohydrate associated molecules
TRA-1-60	+	-	Surface antigen
TRA-1-81	+	-	Surface antigen
Frizzled-5	+	+	Seven transmembrane-spanning G-protein-couples receptor
Stem cell factor (SCF or c-Kit ligand)	+	+	Cytokins, exist both as a transmembrane protein and a soluble protein
Cripto (TDGF-1)	+	+	Receptor for the TGF-β signalling pathway
Core nuclear transcription markers			
OCT3/4(Pou5f1)	+	+	POU family transcription factor
SOX2	+	+	POU family binder transcription factor
NANOG	+	+	Transcription factor
REX-1	+	+	Zinc-finger transcription factor
KLF4	+	+	Zinc-finger transcription factor
Signal pathway-related intracellular markers			
SMAD1/5/8	-	+	Smad proteins, BMP signalling pathway
SMAD4	+	+	Smad proteins, TGF-β/Activin/Nodal signalling pathway, BMP signalling pathway
SMAD2/3	+	-	Smad proteins (R-Smad), TGF-β /Activin/Nodal signalling pathway
β- catenin	+	+	Transcription activators,Wnt/ β- catenin signalling pathway

Based on studies on human and mouse ES cells [reviewed in 7, 12].

1.2) Primordial germ cells:

Primordial germ cells (PGCs) are the embryonic precursors of the gametes as found during later life [13]. During gastrulation, PGCs differentiate from pluripotent epiblasts cells by mesodermal induction signals. Although PGCs are in principle unipotent cells that eventually

differentiate into only sperm or oocytes, they can generate pluripotent stem cells known as embryonic germ cells (EGCs) *in vitro* [14]. However, this capacity is limited to a specific window of development. In humans, PGCs can be identified in the wall of the yolk sac by their intrinsic alkaline phosphatase activity, beginning at about day 24 [15]. As PGCs migrate out of the yolk sac into the embryo, they start to proliferate. Dysregulation of this process of migration can result in ectopic germ cells. Persistence of these ectopic embryonic germ cells either during migration to the genital ridge or after arrival at the gonads is one possible mechanism by which germ cell tumors (GCTs) are thought to arise [15, 16].

1.3) Germ cell tumors:

Human GCTs are a heterogeneous group of neoplasm in gonads (both testes and ovaries) and in different extra-gonadal sites along the midline of the body and the midline of the brain. Based on epidemiology, phenotypic characterization, chromosomal constitution and genomic imprinting, GCTs of the testis comprise three main types: Type I, II and III. Type I GCTs are the teratomas and yolk sac tumors of infants and newborns. Type II GCTs, here referred to as Germ Cell Cancer (GCC), are histologically and clinically divided into seminoma (SE)/dysgerminoma (DG) and non-seminomas. The non-seminomas are subdivided into embryonal carcinoma (EC), which is the undifferentiated (ES-like) component and extra-embryonic lineages (yolk sac tumor (YS) and choriocarcinoma), as well as teratoma, which represents the somatic lineage. The origin of these malignancies is either a PGC or a gonocyte blocked in its physiological maturation towards either pre-spermatogonium or oogonium [17 for review]. These malignant GCTs occur mostly in Caucasian males between the ages of 20 to 40 and are the first cancer cause of death in this age group [18, 19]. Since GCC mimic embryogenesis to some extent, they can provide a useful model to study genes involved in normal development as well as oncogenesis [20]. Type III GCT is known as spermatocytic seminoma (SS). They have a different cell of origin called spermatogonia/spermatocyte. Therefore, the Type II and III of GCTs represent two stages of normal germ cell development: PGC/gonocyte and the spermatogonia/spermatocytes [21].
There are various markers for ES cells and (embryonic) germ cells, as well as the different histological elements of GCC. Among which, OCT3/4, NANOG and SOX2 are highly informative and applicable for diagnosis of some types of GCC [22-24].

1.4) microRNAs:

MicroRNAs (miRs) are endogenous small noncoding RNAs ranging from 19-25 nucleotides in length. miRs are phylogenetically conserved and have been shown to be instrumental in a wide variety of key biological processes including cell cycle regulation, apoptosis, control of metabolic pathways, imprinting, differentiation and maintenance of "stemness" among others

[25, 26] suggesting that miRs might be as important as transcription factors in controlling gene regulations [27]. Thereby, ectopic expression of miRs can result in aberrant regulation of cell cycle, differentiation and apoptosis.
miRs are transcribed as long primary transcripts (pri-miRs) [28]. Pri-miRs are cropped by the RNase III Drosha [29] to the hairpin-shaped precursor (pre-miR) which is exported to cytoplasm through the nuclear transporter exportin 5 [30]. Then the cytoplasmic Dicer (another RNase III-like endonuclease) processes the pre-miRs into 20 to 22 nucleotide duplexes [31]. One strand will be loaded onto the RNA-induced silencing complex and acts as the mature miR. This ribonucleoprotein complex promotes the direct inhibition of mRNA translation and/or the destabilization of the target mRNAs [32]. Accordingly, miRs mutations or altered expression of miRs might be related to the development of various cancers, including GCC.

1.5) ES cell miRs and GCC:

So far, 678 miRs within the human genome have been discovered. Specific miRs are expressed in ES cells referred to as ES cell specific miRs. So far, 31 ES cell specific miRs have been identified [33-35] among which the two miR clusters miR-371-3 and miR-302/367 are highly expressed in GCC representing their pluripotent cell type of origin [36]. These miRs are differentially expressed between the different subtypes of GCC. Additionally, in a screen for miRs that act as oncogenes in GCC, mimicking the role of mutated P53, the cluster of miR-371-3 was identified. In fact, increased levels of miR-371-3 in GCC have been shown to substitute for presence of a mutated P53, by inhibiting cellular senescence which is normally induced after oncogenic stress [37, 38].

1.6) Scope of the thesis:

Three master pluripotency factors, OCT3/4, NANOG and SOX2 are introduced in **chapter 2** and their interconnection machinery for the maintenance of pluripotency in ES cells and germ cells is presented. In this review, additionally, ES cell specific miRs and their crucial role in maintenance of self-renewal and differentiation of ES cells through their interaction with core pluripotency factors is discussed. Moreover, properties of the master pluripotency factors and ES cell specific miRs in induction of pluripotent stem cells are presented. Two major ES cell miR which are highly expressed in GCC, miR-371-3 and miR-302/367 are introduced regarding their roles in normal development and in the pathogenesis of GCC and their exceptional sensitivity to DNA damaging agents such as cisplatin.
The transcription factor SOX2 is essential for the maintenance of pluripotency and self-renewal in ES cells [39]. In addition to its normal stem cell function, SOX2 overexpression is associated with cancer development [40-44]. SOX2 is highly expressed in NCCIT (an EC

cell line) and has shown to be amplified in the array comparative genome hybridization (CGH) result in this cell line. In **chapter 3**, results are presented related to the effect of down-regulation of SOX2 in NCCIT. Furthermore, it is assessed whether oncogene dependency of this pluripotency factor in GCC (with SOX2 expression) exist.

Since miR-371-3 and miR-302/367 clusters are highly expressed in GCC, it is investigated whether the copy numbers might explain the high expression level of these miRs in GCC. Therefore, copy number variation and sequence mutation analysis was done on 242 GCC derived DNA samples, of which the results are presented in **chapter 4**.

All miRs are derived from precursor sequences, which are processed by DICER1 to form double stranded RNA duplexes. Recently germline mutations in DICER1 have been indentified in patients with pleuropulmonary blastoma [45] and nonepithelial ovarian tumors. So far, one DICER1 mutation has been reported in a non-seminomatous GCC [46]. In **chapter 5**, results are presented based on an investigation of the frequency of DICER1 mutation in 94 GCC samples using high resolution melting curve analysis.

Currently, two serum markers are predominantly informative (although with some false positive and negative) results for diagnosis of two subtypes of GCC (YS and choriocarcinoma). However, no informative serum markers for other two important subtypes (SE and EC) exist. In **chapter 6**, the result of a study in detection of the two ES miR clusters miR-371-3 and miR-302/367 in serum of GCC patients is shown and the diagnostic and prognostic values of these miR clusters in GCC patients is discussed.

In **chapter 7** an overview of the topics and conclusions obtained from the results in each chapter is described and a general discussion in respect to conclusions is given.

References:

1. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.
2. Lysdahl, H., et al., *Derivation and characterization of four new human embryonic stem cell lines: the Danish experience.* Reprod Biomed Online, 2006. **12**(1): p. 119-26.
3. Carpenter, M.K., E. Rosler, and M.S. Rao, *Characterization and differentiation of human embryonic stem cells.* Cloning Stem Cells, 2003. **5**(1): p. 79-88.
4. Oh, S.K., et al., *Derivation and characterization of new human embryonic stem cell lines: SNUhES1, SNUhES2, and SNUhES3.* Stem Cells, 2005. **23**(2): p. 211-9.
5. Quattrocelli, M., L. Thorrez, and M. Sampaolesi, *Pluripotent stem cell derivation and differentiation toward cardiac muscle: novel techniques and advances in patent literature.* Recent Pat Drug Deliv Formul, 2012.
6. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
7. Ginis, I., et al., *Differences between human and mouse embryonic stem cells.* Dev Biol, 2004. **269**(2): p. 360-80.
8. Draper, J.S., et al., *Surface antigens of human embryonic stem cells: changes upon differentiation in culture.* J Anat, 2002. **200**(Pt 3): p. 249-58.
9. Henderson, J.K., et al., *Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens.* Stem Cells, 2002. **20**(4): p. 329-37.
10. Reubinoff, B.E., et al., *Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro.* Nat Biotechnol, 2000. **18**(4): p. 399-404.
11. Odorico, J.S., D.S. Kaufman, and J.A. Thomson, *Multilineage differentiation from human embryonic stem cell lines.* Stem Cells, 2001. **19**(3): p. 193-204.
12. Zhao, W., et al., *Embryonic stem cell markers.* Molecules, 2012. **17**(6): p. 6196-236.
13. Molyneaux, K. and C. Wylie, *Primordial germ cell migration.* Int J Dev Biol, 2004. **48**(5-6): p. 537-44.
14. Kimura, T. and T. Nakano, *Induction of pluripotency in primordial germ cells.* Histol Histopathol, 2011. **26**(5): p. 643-50.
15. Frazier, A.L., C. Weldon, and J. Amatruda, *Fetal and neonatal germ cell tumors.* Semin Fetal Neonatal Med, 2012. **17**(4): p. 222-30.
16. Horwich, A., J. Shipley, and R. Huddart, *Testicular germ-cell cancer.* Lancet, 2006. **367**(9512): p. 754-65.
17. Oosterhuis, J.W. and L.H. Looijenga, *Testicular germ-cell tumours in a broader perspective.* Nat Rev Cancer, 2005. **5**(3): p. 210-22.
18. Shah, M.N., et al., *Trends in testicular germ cell tumours by ethnic group in the United States.* Int J Androl, 2007. **30**(4): p. 206-13; discussion 213-4.
19. Huyghe, E., P. Plante, and P.F. Thonneau, *Testicular cancer variations in time and space in Europe.* Eur Urol, 2007. **51**(3): p. 621-8.
20. Yang, S., et al., *Tumourigenic characteristics of embryonal carcinoma cells as a model for studying tumour progression of human embryonic stem cells.* Cell Prolif, 2012. **45**(4): p. 299-310.
21. Looijenga, L.H., et al., *Genomic and expression profiling of human spermatocytic seminomas: primary spermatocyte as tumorigenic precursor and DMRT1 as candidate chromosome 9 gene.* Cancer Res, 2006. **66**(1): p. 290-302.
22. de Jong, J., et al., *Diagnostic value of OCT3/4 for pre-invasive and invasive testicular germ cell tumours.* J Pathol, 2005. **206**(2): p. 242-9.
23. de Jong, J., et al., *Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.* J Pathol, 2008. **215**(1): p. 21-30.
24. Looijenga, L.H., et al., *POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.* Cancer Res, 2003. **63**(9): p. 2244-50.
25. Sotillo, E. and A. Thomas-Tikhonenko, *Shielding the messenger (RNA): microRNA-based anticancer therapies.* Pharmacol Ther, 2011. **131**(1): p. 18-32.
26. Ambros, V., *The functions of animal microRNAs.* Nature, 2004. **431**(7006): p. 350-5.

27. Barroso-delJesus, A., et al., *Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter.* Mol Cell Biol, 2008. **28**(21): p. 6609-19.
28. Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* EMBO J, 2002. **21**(17): p. 4663-70.
29. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
30. Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. **303**(5654): p. 95-8.
31. Ketting, R.F., et al., *Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans.* Genes Dev, 2001. **15**(20): p. 2654-9.
32. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
33. Morin, R.D., et al., *Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.* Genome Res, 2008. **18**(4): p. 610-21.
34. Wang, Y., et al., *MicroRNAs in embryonic stem cells.* J Cell Physiol, 2009. **218**(2): p. 251-5.
35. Suh, M.R., et al., *Human embryonic stem cells express a unique set of microRNAs.* Dev Biol, 2004. **270**(2): p. 488-98.
36. Gillis, A.J., et al., *High-throughput microRNAome analysis in human germ cell tumours.* J Pathol, 2007. **213**(3): p. 319-28.
37. Voorhoeve, P.M., et al., *A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.* Cell, 2006. **124**(6): p. 1169-81.
38. Looijenga, L.H., et al., *Relevance of microRNAs in normal and malignant development, including human testicular germ cell tumours.* Int J Androl, 2007. **30**(4): p. 304-14; discussion 314-5.
39. Avilion, A.A., et al., *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. **17**(1): p. 126-40.
40. Wilbertz, T., et al., *SOX2 gene amplification and protein overexpression are associated with better outcome in squamous cell lung cancer.* Mod Pathol, 2011. **24**(7): p. 944-53.
41. Stolzenburg, S., et al., *Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer.* Nucleic Acids Res, 2012. **40**(14): p. 6725-40.
42. Bae, K.M., et al., *Expression of pluripotent stem cell reprogramming factors by prostate tumor initiating cells.* J Urol, 2010. **183**(5): p. 2045-53.
43. Xu, X.L., et al., *The properties of tumor-initiating cells from a hepatocellular carcinoma patient's primary and recurrent tumor.* Carcinogenesis, 2010. **31**(2): p. 167-74.
44. Lengerke, C., et al., *Expression of the embryonic stem cell marker SOX2 in early-stage breast carcinoma.* BMC Cancer, 2011. **11**: p. 42.
45. Hill, D.A., et al., *DICER1 mutations in familial pleuropulmonary blastoma.* Science, 2009. **325**(5943): p. 965.
46. Heravi-Moussavi, A., et al., *Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers.* N Engl J Med, 2012. **366**(3): p. 234-42.

CHAPTER 2

Role of stem cell proteins and microRNAs in embryogenesis and germ cell cancer

Ronak Eini, Lambert C. J. Dorssers, Leendert H. J. Looijenga

Erasmus MC, University Medical Center Rotterdam, Department of Pathology, The Netherlands

IJDB, in press

Abstract:

Embryonic stem (ES) cells are pluripotent cells derived from the inner cell mass of the blastocyst. These cells can proliferate indefinitely and differentiate into all cell lineages. Germ cell cancers (GCC) mimic embryonic development to a certain extent. The origin of GCC trace back to primordial germ cells/gonocytes in the embryo, which determines their specific characteristics such as totipotency and overall (exceptional) sensitivity to DNA damaging agents. Thus GCC provide a useful model system for the study of gene regulation involved in oncogenesis as well as development. Several reports have demonstrated the role of specific proteins and microRNAs (miRs) in the control of pluripotency and thus early development. miRs are small non-coding RNA molecules that post-transcriptionally regulate gene expression by base-paring to protein encoding mRNAs. miRs are predicted to regulate up to 30% of the protein-encoding genes within the human genome. They are expressed in a tissue-specific and developmentally regulated manner. Aberrant miR expression and its correlation with development and progression of cancers is an emerging field. Important evidence has shown that knock-down by synthetic anti-sense oligonucleotides or re-expression of specific miRs by pre-miR can induce drug sensitivity, leading to increased inhibition of cancer cell growth, invasion, and metastasis. In addition, miRs have been found in body fluids of patients with different types of diseases, including cancer. Therefore, investigation of miRs can shed light on the process of pathogenesis, and may provide biomarkers for diagnosis and prognosis. A subset of miRs is specifically expressed in ES cells and GCC, suggesting their critical role in early embryogenesis and development. In this review we discuss the current view of the biology of embryonic stem cell proteins and miRs in GCC, and their potential clinical impact.

Introduction:

During development of a multi-cellular organism processes of growth and differentiation are kept well-controlled, although not always in balance. Post-transcriptional control of gene regulation plays a key role in maintenance of this process by coordinating the functional effect of selected genes at specific moments in time and place. In early mammalian development from maturation of the germ cell lineage to initiation of gastrulation, the role of post-transcriptional regulation is particularly apparent since the rapid growth and the subsequent switch from unspecialized cells into specific cell types requires control on mRNA localization, stability and translation which all provide fundamental ways of gene regulation [1].

The first cell fate decision is initiated by cell polarization at the 8- to 16-cell stage, where inner and outer cells are differentiated from each other. Outer cells adopt the trophoblast lineage, whereas the inner cell population give rise to the inner cell mass (ICM, the collection of cells that eventually will generate the fetus) forming the embryonic tissues and the primitive endoderm. The process of gastrulation converts the epiblast into the three definitive germ layers of the embryo, and generates the extra-embryonic tissues. The trophectoderm represents the extra-embryonic trophoblast lineage which participates in the formation of visceral yolk sac and envelopes the inner cell mass. The early stem cells within the ICM begin to proliferate rapidly, ensuring that stocks of unspecialized cells are established for future differentiation into the three germ layers [2-4].

In fact, from the fully grown oocyte stage until the moment of Zygotic Genome Activation (ZGA), the genome is transcriptionally silent [1]. Therefore, all mRNA functionality regulation must occur post-transcriptionally [2]. One group of post-transcriptional regulators are miRs. These small non-coding RNAs range in size from 18 to 32 nucleotides in length and have emerged in the past decades as major players in posttranscriptional regulation in many multi-cellular organisms [5]. Hundreds of miRs have been identified in worms, flies, fish, frogs, mammals and flowering plants using molecular cloning and bioinformatics prediction strategies [6-8] The first two miRs, lin-4 and let-7, were both initially identified based on their mutant phenotypes in forward genetic screens in the nematode *Caenorhabditis elegans* [9, 10]. Mutations causing loss of function in either lin-4 and let-7 result in defective development but at different developmental stages [11]. After this observation, it has been reported that proper embryonic development, indeed requires a well-organised temporal and spatial expression of miR besides expression of cell-type specific genes [12-14]. In general, it can be concluded that miRs are involved in the regulation of all different processes, including cell division, cell differentiation and cell death. Therefore, based on the fact that miRs interfere with the translation of mRNAs or proteins, it was predicted that aberrant miR expression might be found in cancer [5, 11].

miRNA biogenesis and mechanism:

The miRs are transcribed as long RNA precursors (pri-miR) that contain a stem-loop structure of about 80 bases. These pri-miRs are processed in the nucleus by the RNase III enzyme Drosha and DGCR8/Pasha, which form the microprocessor complex and process long pri-miRs into short hairpins called precursor miRs (pre-miR). The pre-miRs are exported from the nucleus by Exportin-5 [15-17]. In the cytoplasm, another RNase III enzyme, Dicer, cuts the pre-miR to generate the mature miR as part of a short RNA duplex. The RNA is subsequently unwound by a helicase activity and incorporated into a RNA-induced silencing complex (RISC). At this stage, miR can mediate down-regulation of target gene activity by two modes: translational inhibition or target mRNA cleavage. The choice is made based on the degree of complimentarity between the miR and target gene in combination with an Argonaute family protein. Near-perfect complementarity results in cleavage, followed by general RNA degradation of the targets, whereas partial complementarity causes translational inhibition [11].

Embryonic stem/germ cells and pluripotency:

ES cells are capable of indefinite self-renewal as well as generating all cell-types within the body (pluripotency). These capacities of ES cells require unique transcriptional and cell cycle regulations [18]. Gene expression analysis in the early embryo has led to the identification of several factors that may be involved in the regulation of early developmental events. These include genes encoding growth factors, their receptors and numerous transcription factors, among which pluripotency factors seem to play major role [19]. In this review, we will discuss predominantly three master pluripotency factors: OCT3/4, NANOG and SOX2. They function in combinatorial complexes to regulate the pathways involved in ES cell pluripotency and cellular differentiation: OCT3/4 (also known as POU5F1) is a POU-domain, octamer binding transcription factor, expressed in both mouse and human ES cells. OCT3/4 is critical for the maintenance of pluripotency in ES cells and is down-regulated in all differentiated somatic cell types *in vitro* as well as *in vivo* [20]. Oct3/4 expression starts at the morula stage and becomes restricted at first to the inner cell mass during blastocyst formation, subsequently to the epiblast during gastrulation and finally to primordial germ cells (PGC) [21, 22]. PGC lacking Oct3/4 expression has been shown to undergo apoptosis rather than differentiation [23].

However, OCT3/4 is not the only key gene regulating ES cell and germ cell pluripotency. The homeodomain protein Nanog is also required for the maintenance of pluripotency in the inner cell mass *in vivo* [24]. Nanog plays a critical role in regulating the cell fate of the pluripotent epiblast and preventing differentiation to primitive endoderm. In addition, down-regulation of Nanog can induce both mouse and human ES cell differentiation to extra-embryonic lineages [25]. Nanog mRNA is also enriched in pluripotent germ cells. Similarly,

Nanog deficiency in PGC triggers apoptosis but not cell differentiation [26]. Another key pluripotency regulating gene is SOX2. The SOX2 expression factor is a member of the SRY-related HMG box transcription factor family. SOX2 is expressed in pluripotent cells and multi-potent embryonic and extra-embryonic cells. Sox2 cooperates with Oct3/4 to activate the expression of a number of genes that regulate pluripotency including Nanog [27]. This appears essential for the derivation of induced pluripotent stem cells from primary human fibroblasts [28]. Interestingly, priming by Nanog results in improved capacity to generate iPS cells.

We demonstrated that SOX2 is not present in human germ cells, in contrast to mouse (see below), but instead another member of the SOX family, SOX17, is expressed in human germ cells and their malignant counterparts [29]. Knock-out studies and expression data of SOX17 in primitive endoderm, fetal human stem cells, and early germ cells indicate a crucial role for this transcription factor in a number of lineages during embryonic and fetal development [30]. Sox2 has shown to be present in PGC and gonocytes (testis and ovary) of mouse whereas it is absent in human embryonic gonads. [31]. An interesting hypothesis to be tested is that the protein complex containing OCT3/4 and SOX17 is predominantly involved in regulation of cell death and OCT3/4 and SOX2 in differentiation.

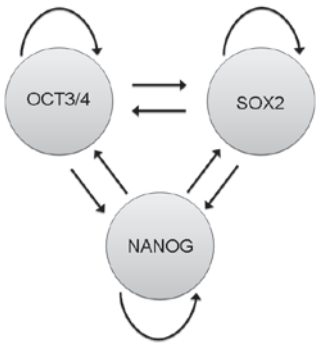


Figure 1. Schematic interconnections between OCT3/4, SOX2 and NANOG.

The core ES cell regulators form intrinsic self-regulatory and feed-back loops at the transcriptional and translational level to insure high expression levels and self-stabilizing system for the maintenance of pluripotency in mouse and human ES cells.

These three transcription factors interact physically with each other and occupy regulatory regions in many target genes towards the regulation of pluripotency. Oct3/4, Sox2 and Nanog co-occupy several hundred genes, often at overlapping genomic sites suggesting that these pluripotency factors generally do not control their target genes independently, but rather act co-ordinately to maintain the transcriptional program required for pluripotency. Furthermore,

Oct3/4, Sox2 and Nanog are also bound to their own promoters forming an inter-connected auto-regulation loop to maintain the ES cell identity [32, 33]. This is schematically illustrated in Figure 1.

Induced pluripotent stem cells:

Regarding the mentioned facts about the critical role of pluripotency factors in ES cells, in 2006, Takahashi and Yamanaka demonstrated that differentiated cells can be converted into so-called induced pluripotent stem cells (iPS cells) by forced expression of four transcription factors Oct3/4, Sox2, Klf4 and c-Myc, which have been termed Yamanaka factors [34]. Recently, in a similar approach towards converting somatic cells into pluripotent cells two independent groups [35]; [36] have reported that mouse and human somatic cells can be reprogrammed to pluripotency by introducing embryonic stem cell miR, miR-302 family and miR-367. It is also reported that introduction of another group of embryonic stem cell specific miR, miR-291-5, enhances production of mouse iPS cells. The miR-291-3p, miR-294 and miR-295, a dominant miR cluster in mouse ES cells with a role in promoting cell proliferation and maintaining pluripotency, increase the efficiency of reprogramming by Oct3/4, Sox2, Nanog and Klf4 [37]. These evidences emphesize on the crucial role of ES cell miR in cell cycle, pluripotency, differentiation and early development of a multicellular organism.

Embryonic stem cell miRs:

So far, a total of 678 miRs within human genome and 472 miRs within the mouse genome have been discovered, but only a subset of these miRs is expressed in ES cells. So far, 31 miRs have been reported to be ES cell specific [1, 38, 39]. In this review, we focus on two ES miR clusters: miR-371-3 cluster (the human counterpart of miR-290-295) and miR302a-d/367 cluster.

The overall roles of miRs in both mouse and human ES cells have been evaluated by analyzing the phenotypes of Dicer and Dgcr8 mouse mutants [40, 41]. *Dicer*-deficient mouse ES cells exhibit defects in differentiation and G_1 cell-cycle arrest [42]. Moreover, deletion of Dicer causes embryonic lethality [43]. Similarly, *Dgcr8*-deficient mouse ES cells show problems in cell-cycle progression and differentiation, evidenced by failing to silence self-renewal genes, such as Oct3/4, Nanog and Sox2, as well as delayed expression of differentiation markers [40]. The *Dgcr8* knockout ES cell phenotype differs from *Dicer1* knockout. Both knockouts are defective in differentiation. *Dicer1* knockout ES cells have a more severe initial proliferation defect that is overcome over time possibly due to additional genetic events, however, *Dgcr8* knock-out ES cells show a stable and more subtle proliferative defect [42, 44]. These overall miR-deficient cells showed a relative increase in the number of cells in the G_1 phase of the cell cycle. This finding suggests that miR suppress inhibitors of the G_1-S transition, allowing the

rapid transition from mitosis to S phase [45]. These data demonstrate the crucial role of miR in ES cell self-renewal and differentiation. How miR regulates gene expression in ES cells was described recently [46], indicating that the key ES cell transcription factors are associated with promoters for miRs that are expressed in ES cells. In fact, miR promoters are occupied by master ES cell transcription factors, including OCT3/4, SOX2 and NANOG. This suggests an important role of ES cell miRs in the maintenance of the pluripotent cell state, as summarized in Figure 2.

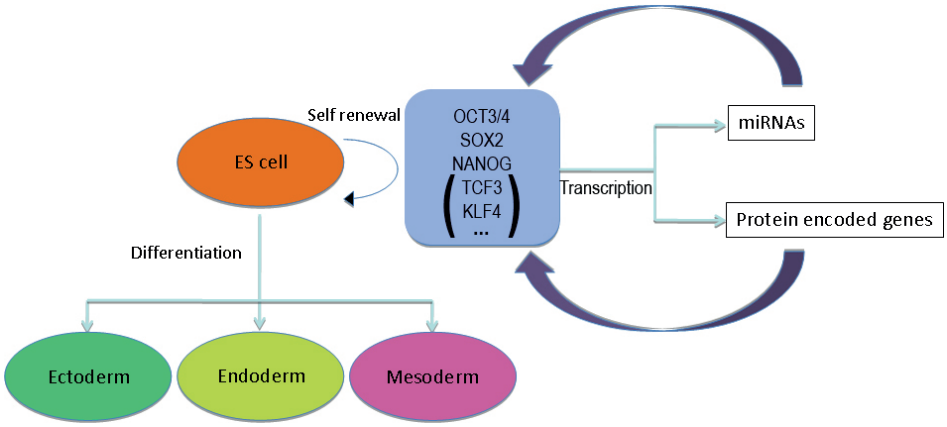


Figure 2. ES cell self-renewal and differentiation.

ES cells have the ability to self-renew or differentiate into cells of all three germ layers (endoderm, ectoderm, and mesoderm). In ES cells, OCT3/4, SOX2, and NANOG form a core transcriptional network influencing the stem cell self-renewal machinery. The interaction of downstream miRs and protein coded genes with these master ES cell transcriptional factors and their opposing functions play major role in the properties of ES cells.

On the other hand, miRs can regulate expression of pluripotency genes by binding to their coding regions. A good example for regulatory circuitry between miRs and pluripotency genes is the one between Lin28 and let-7 miR. Lin-28, a marker of ES cells, also forms a negative feedback loop with the let-7 family miR to control each others levels. Lin-28 regulates expression of let-7 by binding to the precursors and blocking their maturation, whereas in differentiated cells where let-7 levels are increased, let-7 miR targets the Lin-28 mRNA [40].

Germ cell tumor biology in summary:

Malignant germ cell tumors are the most frequent malignancy in Caucasian males between 20 and 40 years of age, and the first cause of cancer death in this age group with increasing incidence [47, 48]. Human germ cell tumors are a heterogeneous group of neoplasm occurring in gonads, both testes and ovaries, and in different extra-gonadal sites along the midline of the body and the midline of the brain. The specific characteristics can either be

due to the process of tumorigenesis or just a reflection of normal embryonal development, which contributes to the complexity of these tumors. Based on epidemiology, phenotypic characterization, chromosomal constitution and genomic imprinting, the group of testicular germ cell tumors comprises three types mainly: Type I, II and III. In the context of the role of ES cell miRs, predominantly the Type I and II will be discussed. Type I germ cell tumors are the teratomas and yolk sac tumors occurring in newborn and infants. The type II germ cell tumors comprise the seminomatous tumors (seminoma (SE)/dysgerminoma (DG) / germinoma) and nonseminomatous tumor, referred to as the malignant germ cell tumors, i.e., germ cell cancers (GCC). The non-seminomas are subdivided into embryonal carcinoma (EC, undifferentiated stem cell component, similar to ES) and the various derivatives: extra-embryonic lineages (yolk sac tumor (YS) and choriocarcinoma) and along the somatic lineage (teratoma) [49, 50]. The precursor cells of GCC are known as carcinoma *in situ* (CIS) or intratubular germ cell neoplasia unclassified (IGCNU) of the testis or gonadoblastoma (GB) of the dysgenetic gonad. They originate from an embryonic germ cell, either PGC or gonocyte, blocked in its physiological maturation process [50, 51]. During normal germ cell development genomic imprinting plays a role. Genomic imprinting is the phenomenon that the paternal and maternal sets of chromosomes have different functionality in mammals, due to parental-specific epigenetic modification. This modification is established somewhere between the stage of a PGC and a spermatozoa (male) and an oocyte (female) [49, 52]. The functional difference between the paternal and maternal haploid set of chromosomes comes from so-called imprinted genes, which are expressed from the paternal or maternal allele only, sometimes in a tissue-specific manner [53]. We demonstrated biallelic expression of imprintined genes in testicular GCC, in line with their erased embryonic germ cell origin [54].

Although the exact mechanistic basis of monoallelic versus biallelic expression of imprinted genes is not solved yet, it is clear that it is related to epigenetic modifications of the genome, including DNA methylation and histone modification [53]. Besides GCC, within the testis there are so-called type III germ cell tumors, known as spermatocytic seminomas (SS). They have shown to have a different cell of origin, i.e., spermatogonia/spermatocyte. These are characterized by a different pattern of chromosomal anomalies, different gene and protein expression, as well as genomic imprinting [55]. Interestingly, these two types of germ cell tumors represent two stages of normal germ cell development: PGC/gonocyte and the spermatogonia/spermatocyte.

OCT3/4 is the most informative diagnostic marker in recognizing CIS/GB, as well as SE/DG and EC [56]. In contrast, SOX2 is expressed in EC, but not the precursor lesions and SE and normal germ cells, while it is more heterogeneous in the differentiated nonseminomatous elements [29]. On the other hand, SOX17 is a suitable marker to distinguish SE (positive) from EC (negative) [29, 57]. Moreover, LIN-28 is expressed in precursor lesions (both

CIS and GB), as well as SE/DG, EC and yolk sac tumor (YS). In addition, elevated level of TSPY has been observed in a variety of tumor tissues, including GCC [58]. Testis-specific-protein Y-encoded (TSPY) is a multi-copy gene mapped to the critical region harbouring the gonadoblastoma locus on the Y chromosome, the only proven oncogenic locus on the male-specific chromosome [58]. The genetic interaction of TSPY and GB has been clinically established [59]. In testicular GCC patients as well as in model cell lines, co-expression of TSPY and androgen receptor (AR) is observed whereas such co-expression was not seen in normal testis. In addition, TSPY is suggested to induce an anti-androgenic microenvironment, possibly leading to a block in maturation of germ cells [60]. This indicates that co-expression of TSPY and OCT3/4 is informative to detect CIS cells from delayed matured germ cells.

p53 in ES cells and GCC:

Since extensive proliferation and differentiation of stem cells can contribute to hyper-proliferative disorders, a coordinated control of stem cell self-renewal and differentiation is fundamental for maintaining tissue and organ homeostasis. p53 appears to contribute to this process by controlling the proliferation, self-renewal and differentiation of both embryonic and adult stem cells. The first observation about a potential role of p53 in ES cells was reported in 1980 [61] in which they observed that p53 was highly expressed in primary cell cultures obtained from 12-14 day old mouse embryos but not in cells from 16 day old embryos. In differentiated cells, p53 is an important regulator of cell proliferation. By controlling expression of the p21 gene, which encodes the inhibitor of cyclin-dependent kinases, p53 induces transition from G_1 into S-phase of the cell cycle. In addition, p53 is able to initiate apoptosis by both the extrinsic and intrinsic pathways. In concordance with these activities in differentiated cells, p53 also controls proliferation and cell death in ES cells. Treatment of human ES cells with nutlin, an inhibitor of p53 degradation, leads to the rapid accumulation of p21 and to cell cycle arrest at the G_1/S boundary [reviewed in 62]. Interestingly, it is shown that knock-down of p53 in human adult fibroblast has dramatically increased the efficiency of iPS cell generation [63].

The presence of functional p53 in ES cells makes them hypersensitive to UV irradiation, whereas the differentiated cells were resistant to UV treatment. Wild-type ES cells undergoing apoptosis expressed functional p53 whereas ES cells lacking p53 showed enhanced proliferation in both undifferentiated and differentiated state and apoptosis was reduced. Undifferentiated ES cells with a functional p53 protein proliferate slower than p53-/- ES cells. This indicates that functional p53 is involved in regulating proliferation of these cells *in vitro* [64]. This is related amongst others, to induction of pre-mature senescence after oncogenic stress. These observations could explain why genes that are involved in the establishment of oncogene-induced senescence such as p53 are amongst the most frequently mutated tumor suppressor genes. However, GCC are an exception to this rule. They have been shown to express wild-

type p53 and mutations in this protein are rare while they proliferate fast and respond to DNA damaging agents, such as cisplatin (in most cases) [65 , 66].

ES cell division:

ES cells have a unique cell cycle characterized by lack of G_0, absence of G_1 checkpoint and subsequently a short G_1/S transition which facilitates quick cell movement through G_1 to S phase allowing the cells to proliferate rapidly leading to rapid growth of the early embryo [67]. This specific characteristic of ES cells makes these cells hypersensitive to DNA damaging agents. In recent studies, miRs have shown to play a central role in achieving this unique cell cycle regulation [38, 68]. In somatic cells, where the G_1/S checkpoint exists, the G_1 phase is a gap period between cytokinesis and DNA replication. During the early G_1 phase, upon stimulation of the mitogenic factors, D-type cyclins are expressed. These cyclins are required during G_1 restriction point to activate the Cdk4/6 and the Cdk2 kinases which are expressed throughout cell cycle. The Cdk4/6-Cyclin D complex then phosphorylates proteins of the retinoblastoma (pRB) family. This leads to the partial inhibition of RB and release of its target, the E2F transcription factor. Upon elevation of E2F, the expression of E-type Cyclin increases resulting in activation of Cdk2 which further phosphorylates RB.

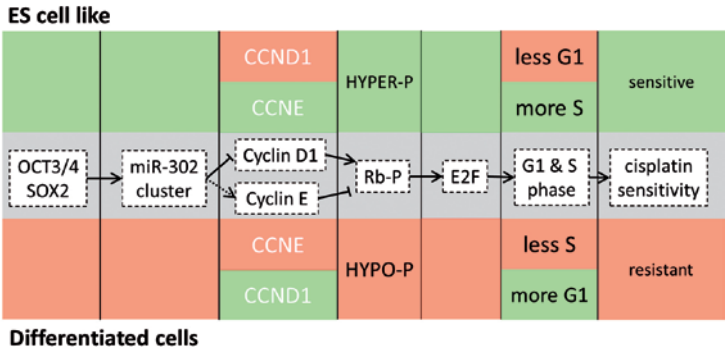


Figure 3. Schematic presentation of G_1-S transition in ES cells and somatic cells.

In ES cells, high expression of OCT3/4 and SOX2 maintains high expression level of miR-302 cluster causing hyper-phosphorylation of Rb via directly suppression of Cyclin D1 directly and activation of Cycin E resulting in high expression of E2F protein. This causes short G_1-S phase by accumulating more cells in S phase and less in G_1. This will lead to sensitivity of ES cells to cisplatin. In differentiated (somatic) cells low expression of main pluripotency factors results in lower expression level of miR-302 cluster leads to hypo-phosphrylation of Rb via up-regulation of cyclin D1 (green) and down-regulation of cyclin E. Rb then inhibits E2F protein. This will cause long G_1-S due to accumulation of more cells in G_1 and less cells in S phase, and consequently result in a more resistant type of differentiated cell to cisplatin. Abbreviations and colors used: Red=low protein/miR; green=high protein/miR; hyper-P = hyper phosphorylation; Hypo-P = hypophosphorylation.

This event leads to the transcription of genes required for progression through the S phase. This will result in the elongation of G_1 phase and slower proliferation of the resulting somatic cells [38, 69] (Figure 3). Aberration in the expression of cell cycle regulatory factors can lead to uncontrolled proliferation which is one of the hallmarks of cancer [70]. In conclusion, ES cells characterized by a very short G_1 phase and lack of a check point at the G_1/S transition, are very similar to many cancer cells [38]. In this context, it is shown that ectopic expression of TSPY increases cell proliferation *in vitro* and tumorigenesis *in vivo* [58]. Expression of TSPY facilitates the transition of the cells through the G_2/M phase of the cell cycle, indirectly up-regulates pro-growth genes and down-regulates apoptosis inducing molecules and growth inhibitory genes, thereby promoting cell proliferation in both *in vitro* and *in vivo*. TSPY harbours a SET/NAP domain represented by the SET oncoprotein and nucleosome assembly protein-1 (NAP-1) respectively. Several evidences have shown that SET/NAP-containing proteins are cell cycle regulators. They regulate the G_2/M transition by modulating cyclin-B-cyclin-dependent kinase 1 (CDK1) activity [71].

Potential role of miR-290 and miR-302 clusters in ES cell division:

In a study in which identification of the specific miR and targets involved in the process of rapid transition from mitosis to S phase was aimed, it was found that introducing specific miR mimics in the *Dgcr8* knockout ES cells, could promote growth and partially rescued the proliferation defect in these cells. In that study, miR-290 cluster which is the most highly expressed miR in mouse ES cells and rapidly down-regulated upon differentiation, was selected [72]. Moreover, in another study, by adding back individual miR into *Dgcr8* knockout ES cells, many miRs, including members of the miR-290 and miR-302 clusters, were found to rescue the prolonged G_1 phenotype by reducing cells accumulation in S phase or G_2-M phase. These data show that ES cell-specific miRs promote ES cell proliferation at least in part by promoting the transition of cells from G_1 to S phase. These miRs are called ES cell-specific cell cycle-regulating miRs or ESCC miRs [45].

It has been reported that expression of miR-302a caused decrease in the number of cells in G_1 phase and increased the population of cells in S-phase and G_2/M-phase in HeLa cells transfected with Pre-miR-302a compared to the cells transfected with negative control pre-miR. Therefore, consistent with the previous reports, in pluripotent human ES cells, the majority of cells were in S-phase and as cells progressively differentiate, the cell cycle profile shifts from a short G_1/high S-phase population to a high G_1-phase cell cycle and this shift is mediated through miR-302 decrease of cell cycle regulator level in pluripotent cells [73]. In the same study, similar increase in the population of G_1-phase of NTera2 (an EC cell line of GCC origin) cells which have been differentiated with retinoic acid is detected. Interestingly, in an experiment done in our lab, we observed that accumulation of differentiated NTera2

cells in S phase was reduced when subjected to OCT3/4 siRNA followed by increasing cell population in G_1-phase analyzed by Fluorescence Activated Cell Sorting (FACS) confirming their ES characteristics (Figure 4).

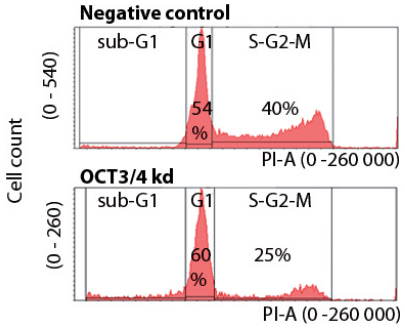


Figure 4. Result of FACS analysis.

NTera-2 cells were subjected to OCT3/4 siRNA and scrambled siRNA (negative control). As it is indicated in OCT3/4 kd situation, more cells are accumulated in G_1 phase and less cells are in S phase compare to negative control.

When ES cells differentiate, G_1 phase lengthens and the rate of cell division slows [67]. We tested the combined down-regulation of OCT3/4 and SOX2 in NT2, and observed massive induction of apoptosis in these cells. This also confirms the major role of these core pluripotency factors in proliferation and cell cycle of EC cells.

miRs in GCC:

We demonstrated that specific miRs are able to mimic the effect of mutated p53 and thereby bypass pre-mature senescence after oncogenic exposure [74]. Using a high throughput unbiased functional screen, a cluster of miR mapped to chromosome 19, known as miR-371-3, was found to be responsible for this phenomenon, subsequently found to be expressed in almost all SE, EC as well as YS. Interestingly, a small number of SE with p53 mutation lacked expression of this set of miRs, and two GCC-derived cell lines with a low or no expression of these miRs showed either a mutated or non-expressed p53 supporting the proposed model. Therefore, miR-371-3 was considered a potential oncogene, able to mimic the effect of mutated p53 by preventing Ras-induced oncogenic senescence in GCC. This process is most likely occurring via suppression of CDK inhibitor LATS2 (Large tumor suppressor) which is the target of miR-371-3 [74].

Based on these findings, a high-throughput screen of 157 miRs using a quantitative RT-PCR based approach in GCC was performed in which only the mature miRs were detected. First, indeed expression of the miR cluster 371-373 was confirmed. Secondly, unsupervised

hierarchical clustering showed different expression of the cluster in different groups of tumors. miR-371-3 was highly expressed in SE, EC and YS, but not expressed in teratoma and normal testicular parenchyma (NT) (Figure 5). In the same study, miR-302a-d cluster were differentially expressed between the clusters of GCC; highly expressed in SE and EC as well as YS and absent in teratoma [75].

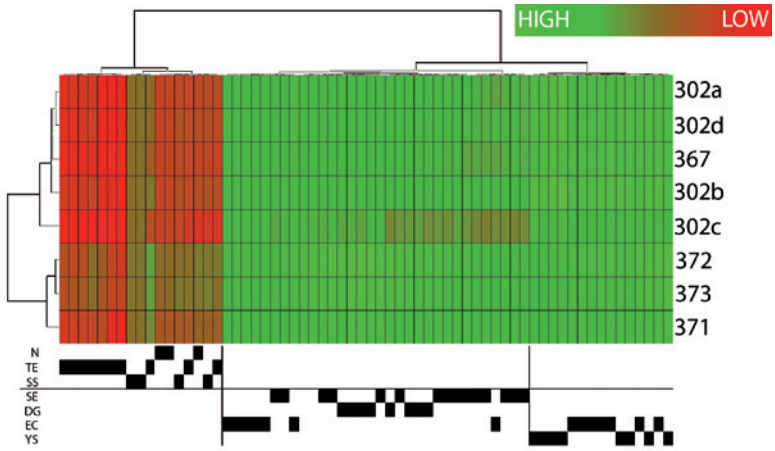


Figure 5. Results of the unsupervised hierarchial clustering.

miR clusters miR-371-3 and miR-302/367 are differentially expressed between different subtypes of GCC; highly expressed in SE/DG and EC and absent in teratoma (TE), SS and normal testis parenchyma (NT). The numbers of each histological subtype are indicated: ($SE_{n=15}$), ($EC_{n=13}$), ($YS_{n=8}$), ($TE_{n=10}$), ($SS_{n=4}$), ($DG_{n=10}$) and ($NT_{n=3}$). For normalization, all miRs on the plate were used using mean expression of all genes (threshold C40) according to [120]. Clustering was performed using Ward's algorithm; both dendrograms were formatted according to euclidean distance. GenEx 5.3.7.332 was used to analyze the data.

Additionally, in an independent study [76] it was demonstrated that the miR-371-3 and miR-302/367 clusters were enriched in CIS samples while they were very low expressed or absent in the normal testis. This indicates that high expression of these miRs is not restricted to the invasive components, but intrinsic to the early pathogenetic stage. In addition, we have shown that no mutation, deletion or amplification was present in the loci of miR-371-3 and miR-302 clusters in the DNA of 242 GCC revealing that no genetic changes affect on the expression levels of these miR clusters (De Boer *et al.* submitted for publication).

Expression levels of miR-371-3 and miR-302 clusters:

We measured expression of the individual members of the miR-371-3 and miR-302 clusters in a series of GCC cell lines and primary GCC (EC, SE and YS, teratoma, normal testis as well as SS). Comparing the patterns of expression, it is of interest that in almost all samples (cell

lines and tumors), within the different clusters, the same variant was showing the highest level of expression: miR-372 and miR-302b and miR-302d, respectively (Figure 6). Since all members of the two clusters are generated from a single primary transcript (see above), this observation can not be explained by differences in transcription, but are likely due to variation in maturation and/or stability of the mature transcript. This needs further investigation.

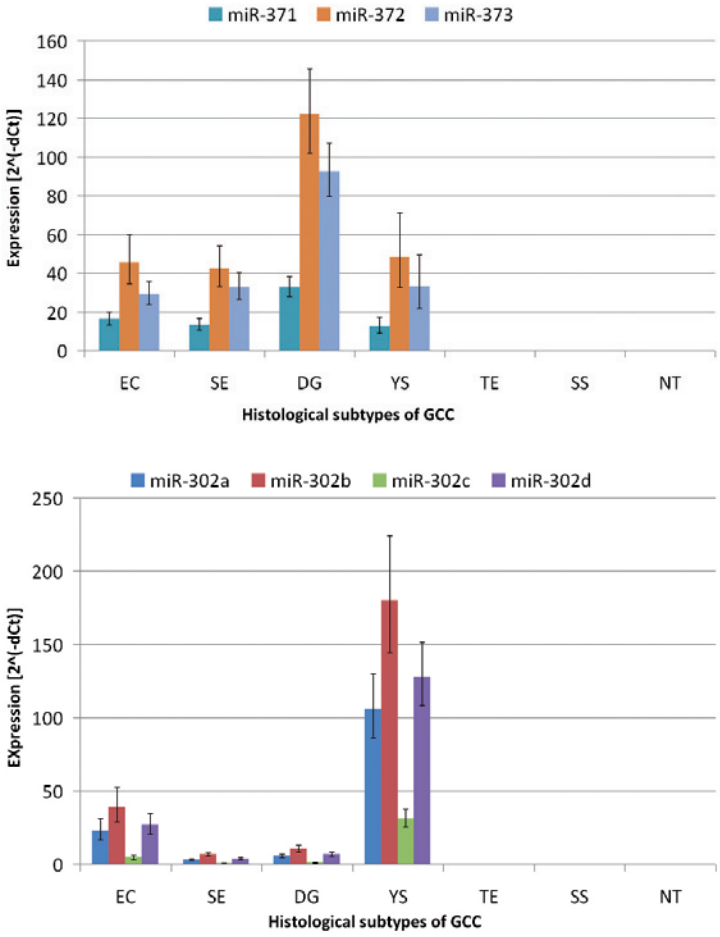


Figure 6. Expression pattern of miR-371-3 and miR-302 clusters in GCC.

These include EC $_{n=14}$, SE $_{n=15,}$ DG $_{n=10}$, YS $_{n=8}$, TE $_{n=10}$, SS $_{n=4}$ and NT $_{n=4}$. A) Within cluster miR-371-3 miR-372 shows the higher level of expression in all samples; B) Within the miR-302 cluster miR-302b and miR-302d show the highest level of expression. (The CT values of miRs are normalized as described before [120].

miR-371-3:

The miR-371–3 cluster and its murine ortholog, the ES cell-specific cell cycle regulating miR-290–5 cluster, are over-expressed in ES cells and down-regulated upon differentiation. These miRs play a crucial role in the maintenance of ES cell renewal [1, 77]. They promote tumor invasion and metastasis in response to hypoxia [74, 78]. Recently, a direct regulation of the miR-371–3 cluster by Myc is reported [79]. This is of interest in the context of iPS cells because c-Myc in combination with other core pluripotency factors was used to generate iPS cells from a differentiated fibroblast [34]. Lack of c-MYC in this set up results in a significantly lower efficiency in the ability to generate iPS cells [80, 81, 82]. Moreover, as recently found for the murine ortholog miR-294 in ES cells [83], it is suggested that Myc is up-regulated by the miR-371–3 cluster. This might be due to the direct binding of the miR to the MYC genomic region or inhibition of an intermediate repressor or direct activation through binding. Moreover, miR-290-295 cluster is reported to be highly enriched in the germ cell population of day 6 testis when compared to the somatic cell population [84]. This cluster is a target of pluripotency factors, and in turn targets the pluripotency factors Oct3/4, Sox2 and Nanog [85]. It was shown that suppression of the miR-290 cluster, besides embryonic lethality, leads to the defective migration of PGC in the gonads. In addition, they are differentially expressed between gonocytes and spermatogonia pointing to their role in the differentiation of gonocytes. These data demonstrate the role of miR-290 cluster in maturation of germ cells, indicating that this miR cluster plays an important role in transition from gonocyte to spermatogonia and in the pre-invasive lesion CIS, which has been disrupted and failed to successfully differentiate into a spermatogonium [86].

miR-302/367:

The cluster of miR miR-302/367 represents the first human miR promoter characterized in human ES cells and their malignant counterpart EC cells [1, 87]. The cluster miR-302-367 is located on chromosome 4 and contains eight different miRs: miR-302a, miR-302a*, miR-302b, miR-302b*, miR-302c, miR-302c*, miR-302d and miR-367. The first seven constitute the miR-302 family with a highly conserved sequence [87]. The major members of the miR-302 family (302a, 302b, 302c and 302d) share a high sequence homology, differing only in the 3'hexanucleotides. All four family members display a cytoplasmatic localization with no significant differences among them, meaning that the 3'hexanucleotides does not promote nuclear localization for any of these miRs.

miR profiling has confirmed cell type specificity for miR-302/367 expression; it is highly expressed in human ES and EC cells but not in adult human mesenchymal stem cells (hMSCs) and normal tissues. The miR-302/367 promoter activity depends on the hierarchical cellular stage. The promoter activity is functional in human and mouse ES cells and human EC cells, but it is turned off later in development [87]. MiR-302/367 can target over 445

human genes and most of these targets are developmental signals involving the initiation and facilitation of lineage-specific cell differentiation during early human embryogenesis [88]. miR-302/367 promoter activity decays upon differentiation of both mouse and human ES cells, demonstrating that its activity is restricted to the ES cell compartment and that the ES cell-specific expression of the miR-302/367 cluster is fully conferred by its core promoter transcriptional activity. Importantly, these data are further supported by the fact that endogenous miR-302/367 expression is also down-regulated during human ES cell differentiation [87]. The important point is that the key ES cell-specific factors such as, OCT3/4, SOX2 and NANOG act as upstream regulators of the miR-302/367 cluster meaning that potential binding sites for these pluripotency factors in the promoter sequence of these miRs are present [73, 87]. It has been demonstrated that in human ES cells the expression of miR-302a under OCT3/4/SOX2-depleted conditions was reduced by more than 50% in comparison with the level in cells transfected with non-targeting siRNAs. These data support the model that the expression of miR-302a is dependent on OCT3/4-SOX2 in human ES cells and suggest that OCT3/4 and SOX2 function as transcriptional activators [73]. Interestingly, siRNAs directed against OCT3/4 and SOX2 in a human EC cell line (NTera2) showed down-regulation of miR-302/367 cluster up to 50% in our experiment confirming the relationship between pluripotency and miR-302/367 cluster (shown in Figure 7). This also confirms that initiation of differentiation reduces the level of miR-302/367 cluster in EC.

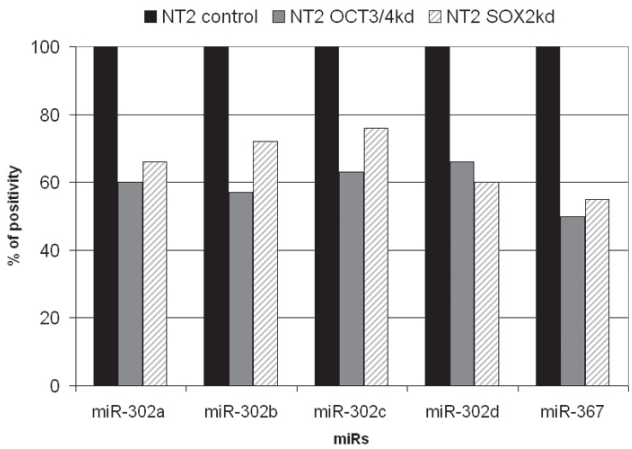


Figure 7. Result of luciferase measurement assay.

It is shown that luciferase activity is down-regulated in both NTera2 cells subjected to LATS2 siRNA and miridian-373. Luciferase activity is also partly down-regulated in cells with a mutant construct of LATS2, which may be due to the partial homology between the siRNA or miridian and LATS2 sequence in the construct. The results are based on six parallel experiments (96 wells) for each condition tested. Standard deviations were: for controls: (0, 015 and 0, 014), LATS2 siRNA (0,079 and 0,058), miridian-373 (0,057 and 0,025) mutant and wild-type respectively.

miR-371-3 and 302/367 in other cancer types than GCC:

miR-371-3 and miR-302/367 have been reported to be involved in other malignancies than GCC. In a study regarding expression of miRs in oesophageal squamous cell carcinoma, it has been shown that miR-373 participates in the carcinogenesis of human oesophageal cancer by directly suppressing LATS2 expression [89]. Additionally, in a study related to human hepatocellular carcinoma (HCC), it is found that miR-373 is up-regulated in HCC tissues as compared with adjacent normal tissues, and promotes the proliferation of HCC cell lines by regulating the G_1-S transition [90]. In a number of studies it is reported that miR-373 can promote tumor invasion and metastasis such as breast and prostate cancer [78, 91, 92]. Additionally, up-regulation of miR-373 in retinoblastoma has been reported [82]. Furthermore, miR-371-3 cluster has shown to be up-regulated in thyroid adenomas [93]. Up-regulation of miR-302/367 has been reported in glioblastoma multiforme. It was shown that this miR cluster is rapidly and strongly induced in glioma-initiating cells [94]. In addition, miR-302 is reported to be enriched in human leukaemia cell lines [95], large B cell lymphoma [96]. miR-367 is also found to be increased in non-small-cell lung cancer [97]. Furthermore, miR-302 has been shown to reprogram human skin cancer cells into a pluripotent ES cell like state [88].

Functional read-out system in GCC cell lines:

Considering all the mentioned facts about miR-371-3, we initiated a study in order to investigate the role of suppression of miR-371-3 in cell cycle and cisplatin sensitivity of GCC cell lines. As a consequence of the poor transfectability of the GCC cell lines, a lentiviral construct was designed to monitor the manipulation of miR levels. In the lentiviral construct, the luciferase reporter is linked to the 3' untranslated region (UTR) of LATS2. In addition, the construct contains a selection marker to allow for stable integration in the GCC cell genome. Its functionality was demonstrated using oligonucleotides for a miR-373 mimic (meridian-373) and a siRNA directed to the 3'UTR. These oligos were introduced into the cells by transient transfection and their consequence on luciferase activity was measured. Both the meridian-373 and the LATS2 siRNA caused strong reduction of the luciferase activity (Figure 8), and thus demonstrate the sensitivity of the reporter system. Recently an alternative reporter system based on a retroviral construct (miR-Sens) was published [98]. In this vector, the Renilla luciferase gene is also linked to the LATS2 3'UTR and the Firefly luciferase is needed for normalization in the absence of a selection marker. Following introduction of a retroviral vector causing expression of miR-373, the authors could demonstrate reduction of the luciferase activity as a result of the targeting of the LATS2 3"UTR [98]. Both approaches are well suited to monitor the consequence of introduction of exogenous miR, but it remains to be established whether these reporters can monitor changes in endogenous levels of miR following cell differentiation or other treatments.

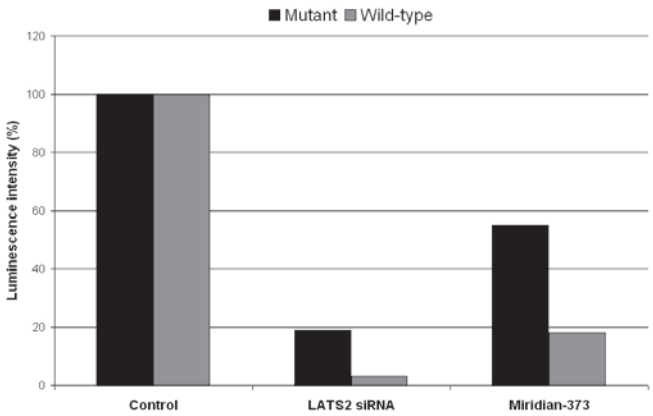


Figure 8. High throughput results show down-regulation of expression of the members of cluster miR-302/367 due to OCT3/4 and SOX2kd in NTera2 cells.

All miR members in the cluster show down-regulation (~ 50%) in cells subjected to OCT3/4 and SOX2 siRNA compare to cells transfected with scrambled siRNA.

Cisplatin sensitivity in GCC and proposed model for involvement of miR clusters 371-3 and 302/367:

Recently, involvement of some miRs (miR-193b and miR-320) in resistancy/sensitivity to platinum agents in ovarian cancer cell lines has been reported [99]. Platinum agents are main chemotherapeutic agents used in the treatment of cancers [100]. As mentioned, ES cell-specific miRs have a central role in G_1-S transition and promotion of cellular proliferation. EC cells, similarly to ES cells, have a short G_1-S transition and are therefore highly responsive to cisplatin-based chemotherapy. Indeed, NTera-2 cells respond to cisplatin as well. The level of expression of the miR-302 cluster in NTera-2 cells is relatively high [75]. This causes sensitivity to DNA-damaging agents such as cisplatin (Figure 3). Our model predicts that suppression of the two ES cell miRs, miR-302 and miR-371-3 cluster in GCC cell lines will cause longer G_1-S transition due to higher level of Cyclin D1 and LATS2 expression respectively. This will leads to suppression of CDK inhibitors and consequently will induce resistance to cisplatin in these cells.

Detection of miR-373, miR-302c and 367 clusters in serum of patients with GCC:

Recently, there has been an enormous interest in investigating circulating miRs in blood plasma/serum as potential biomarkers for early disease, including cancer. In multiple reports, detection of cancer-specific miRs in serum of cancer patients has been reported [101-103]. Currently, serum markers such as alpha feto-protein (AFP) and human Chorionic

Gonadotropine (hCG) are predominantly informative for initial diagnosis as well as follow up of GCC, predominalty related to YS and choriocarcinoma components, respectively [104]. However, no informative serum marker for SE and EC exist, which is highly relevant, because of the distribution of these components in GCC overall, and their stem cell characteristics and behavior. Therefore, the necessity for new serum markers in order to detect and follow-up the other histological component of GCC is apparent. Interestingly, a case report [105] showed the feasibility of measuring serum levels of the miR-371-3 and miR-302 clusters in a single patient. In addition, a decline of these miR level in serum of GCC patients after surgical removal of the (stage I) GCC was shown [106]. A subsequent study [107] demonstrated elevation of these miR clusters in serum of multiple patients at the time of initial diagnosis with different histological subtypes. These findings open new avenues for diagnosis and follow-up the other mentioned cancer patients as well.

The finding of stable extracellular miRs in serum/plasma and other body fluid types suggests the possibility of their involvement in mediating cell-cell communication. This could imply that miRs convey specific information and therefore only some cellular miRs are exported or released from cells in response to biological stimuli. In addition, it is shown that cells release a significant number of RNA-binding proteins into the culture medium and one of them, nucleophosmin 1 (NPM1), can protect miRs from degradation. There is no evidence yet that miR containing complexes are taken up by other cells, although vesicles containing miRs are reported to be taken up by cells. It is possible that miRs outside vesicles, complexed with proteins, could be targeted to specific cell surface receptors [108].

It is controversial whether miRs circulate freely or are encapsulated in micro-vesicles, particularly exosomes; in other words if their transport is active or passive. It is suggested that extracellular miRs are bound to Ago proteins and Ago2 in particular. It is shown that miRs remained stable in the cell lysates for at least 2 month meaning that Ago2/miR complexes are highly nuclease/protease resistant. The miR/Ago2 complexes are released from the cells upon necrosis or apoptosis [109]. This has been confirmed by another study in which it is shown that although some circulating miRs are vesicle-associated, 90% of miRs are present in a non-membrane-bound form consistent with a ribonucleoprotein complex which is identified as Ago2-miR complex [110]. However, it is reported that the majority of miRs detectable in serum and saliva is concentrated in exosomes [111]. In this study it was shown that the concentration of miRs is consistently higher in the exosome pellet compared to the exosome-depleted supernatant. The reason for this discrepancy is unclear but could be due to differences in isolation of exosomes or differences between plasma and serum.

To confirm the findings in GCC patients, we aimed to detect the serum level of the mentioned miR clusters in a series of GCC patients at stage I and higher (i.e., metastatic disease). As a pilot study, we tested the serum level of miRs 373, 302c and 200c (as control) in a limited number of cases. The miR-200c is reported to be involved in metastasis of epithelial cancers

[112] and is not specific for GCC. Serum samples included in our study were high stage (stage II and higher) GCC serum samples, stage I GCC serum samples and inflammatory serum samples as controls (10 sample of each). Indeed a higher level of miRs was detected in serum of GCC patients compared to control samples, correlated with stage of presentation of the disease (Figure 9). However, there are technical limitations in all studies reported, related to low RNA recovery from serum, and issues of normalization (done based on volume of *RNU6B*). To solve these problems, we developed together with Applied Biosystems (part of Life Technologies) a complete pipeline for detection of miRs in body fluids based on a magnetic-bead based purification and qPCR quantitation (Gillis and Rijlaarsdam *et al.,* submitted for publication)

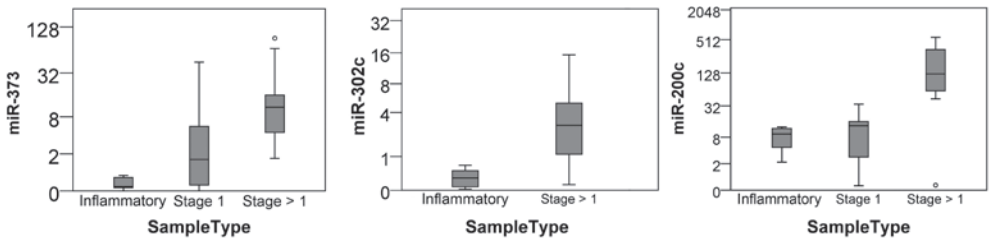


Figure 9. Results of level of miR in serum of patients with stage I and higher stage GCC.

miR-373 and miR-200c show the highest levels in higher stage GCC and less expression level in stage I samples while the expression levels are very low in case of inflammation. The miR-302c was measured in inflammatory samples and high stage tumours showing higher expression level in high stage GCC.

Correlation study between the level of miR-302/367 and miR-371-3 cluster between the tumors and the serum of the same GCC:

To investigate whether the pattern of miRs is serum correlates with the pattern observed in the matched primary tumors, a series of 10 GCC cases (including stage I and higher) were investigated. Statistical analysis showed no correlation between the level of miRs within the tumor and matched serum samples (statistical analysis not shown). This suggests that specific (active) mechanisms are involved in this process.

Methylation status of the promoter of miR-371-3 in GCC:

Gene function in cancer cells can be disrupted either through genetic alteration, directly by mutation, deletion or amplification [113]. However, no such anomalies have been found so far in a series of 242 GCC regarding the miR-371-3 and miR-302/367 clusters (De Boer *et al.* submitted for publication). Alternatively, this can be the result of epigenetic changes, including DNA methylation. This is a major epigenetic modification of the genome that

regulates crucial aspects of its function and allows cells to lock genes in the "off" position. Disruption of DNA methylation associated with specific imprinted genes is a common feature of human cancers [114]. Hypermethylation of promoter regions has been observed for several anti-oncogenes and hypomethylation has been reported for some proto-oncogenes. A genome-wide survey has revealed that the overall degree of methylation in SE is lower than in non-seminoma tissues [115]. Methylation of genomic DNA is performed by DNA methyltransferases that transfer methyl groups to cytosine residues. The five genes known to encode DNA methyltransferases (DNMT) include DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. Here we mainly discuss DNMT1 which maintains methylation of genomic DNA during DNA replication, thus contributing to stability of gene expression from parental to daughter cells. It has been reported that in the DNMT1-/- mice in which loss of imprinting has been transiently induced, the imprint free stem cells formed *in vivo* various neoplasms, including a single tumor mimicking seminoma [116]. Therefore, one hypothesis for the high expression level of miR-371-3 (and possibly miR-302a/367) in GCC, could be that the overall retained demethylation pattern in the precursor cells is involved.

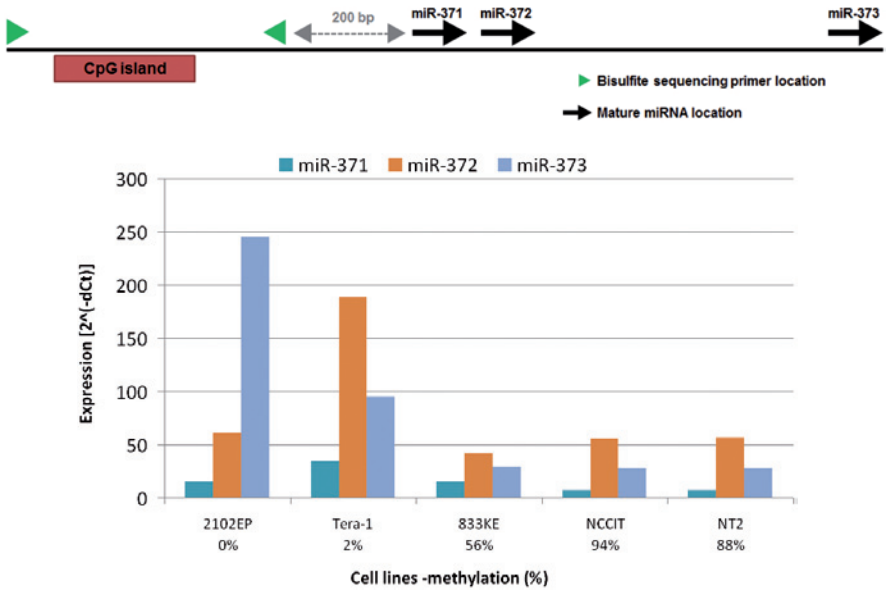


Figure 10. A) miR-371-3 locus. The location of sequencing primers within the promoter of miR-371-3 and the location of mature miRs are shown. B) Expression pattern of miR-371-3 in comparison to methylation. Percentage of the total number of methylated CpG sites is shown for the cell lines. Expression of miR-373 was strongly, significantly negatively correlated with increasing percentages of methylation (Spearman's ρ -0.90, p=0.037). The correlation between expressions of miR-371/2 with methylation was not significant.

Indeed, changes in miR expression can occur through various mechanisms including chromosomal abnormalities, transcription factor binding and epigenetic alterations. Recent studies have shown that, in cancer, expression of some miRs is silenced in association with CpG island hypermethylation [117-119]. In order to determine whether the expression pattern of miR-371-3 in GCC (EC, SE and YS) is due to the methylation status of the supposed promoter region, we examined the methylation status of 12 CpG sites within the CpG island of the promoter of miR-371-3 cluster. The CpG island is located 400 bp upstream miR-371 (Figure 10). Firstly, the methylation status of the promoter of this miR cluster was tested in GCC cell lines using sequencing of bisulfite-treated DNA. A correlation between expression level of miR-371-3 cluster and methylation percentage of the promoter of this cluster was found. The correlation between miR-373 and methylation percentage was significant, however, this correlation for miR-371 and miR-372 was not significant. The reason for this is unclear and needs further investigation. Analysis of micro-dissected GCC samples (EC, SE, YS, and teratoma) showed a different pattern, which needs further investigation (data not shown).

Concluding remarks:

Emerging evidences have revealed the importance of miRs for normal development and maintenance of a multi-cellular organism. It is now proven that formation of ES cells as well as various differentiation lineages is dependent on miRs. The core transcriptional factors and ES cell miRs are linked in a regulatory circuitry that critically regulates both pluripotency and differentiation in human ES cells. These processes are represented in human GCC. In fact the most important reason to study the role of ES cell miRs in GCC is the fact that these cancers mimic embryonic development to a certain extent.

One of the core pluripotency marker in ES cells, OCT3/4, is an informative diagnostic marker for this type of GCC. Detailed investigation on the expression profile of miRs in GCC revealed expression of specific miRs differentially expressed in various histological components. Two of the most informative clusters of miRs were again miR-371-3 and miR-302/367. Interestingly, miR cluster 371-3 is involved in inhibition of cellular senescence after oncogenic stress. This explained the presence of wild type P53 in GCC, in contrast to most other solid cancers. High level of miR-371-3 cluster in GCC leads to short G_1-S transition (via suppression of LATS2) which in turn increases the sensitivity of these tumors to chemotherapy. miR-302 cluster is highly expressed in SE, EC as well as YS. Functional analysis has shown that down-regulation of core pluripotency factors such as OCT3/4 and SOX2, induces reduction of the level of miR-302/367 cluster in an EC cell line NTera-2. Interestingly, OCT3/4 down-regulation in NTera2 induced longer G_1-S transition by accumulating more cells in G_1 phase and less cells is S phase. On the other hand, it is known that differentiated NTera2 cells in which pluripotency factors are down-regulated, show resistance to DNA-damaging agents such as

cisplatin. All these together confirm that the link between ES cell miRs and pluripotency markers plays a major role in the pathogenesis of GCC as well as their exceptional sensitivity to DNA damaging agents. Currently, to improve knowledge about the insight roles of these sets of miRs in pathogenesis of GCC, functional analysis on the effect of suppression of these miRs on cisplatin sensitivity as well as cell cycle regulation of GCC cell lines is in process. These findings will shed novel light on normal and malignant germ cell development leading to diagnostic and prognostic implications.

Acknowledgements:

The authors would like to thank Martin Rijlaarsdam (Department of Pathology, Erasmus MC) for his continuous support in analysis of the data and generating the illustrations.

References:

1. Suh, M.R., et al., *Human embryonic stem cells express a unique set of microRNAs.* Dev Biol, 2004. **270**(2): p. 488-98.
2. Lu, R., et al., *Systems-level dynamic analyses of fate change in murine embryonic stem cells.* Nature, 2009. **462**(7271): p. 358-62.
3. Kuckenberg, P., et al., *Lineage conversion of murine extraembryonic trophoblast stem cells to pluripotent stem cells.* Mol Cell Biol, 2011. **31**(8): p. 1748-56.
4. Tremblay, K.D., N.R. Dunn, and E.J. Robertson, *Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation.* Development, 2001. **128**(18): p. 3609-21.
5. Alvarez-Garcia, I. and E.A. Miska, *MicroRNA functions in animal development and human disease.* Development, 2005. **132**(21): p. 4653-62.
6. Reinhart, B.J., et al., *MicroRNAs in plants.* Genes Dev, 2002. **16**(13): p. 1616-26.
7. Lim, L.P., et al., *The microRNAs of Caenorhabditis elegans.* Genes Dev, 2003. **17**(8): p. 991-1008.
8. Watanabe, T., et al., *Stage-specific expression of microRNAs during Xenopus development.* FEBS Lett, 2005. **579**(2): p. 318-24.
9. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.
10. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.
11. Wienholds, E. and R.H. Plasterk, *MicroRNA function in animal development.* FEBS Lett, 2005. **579**(26): p. 5911-22.
12. Mineno, J., et al., *The expression profile of microRNAs in mouse embryos.* Nucleic Acids Res, 2006. **34**(6): p. 1765-71.
13. Karp, X. and V. Ambros, *Developmental biology. Encountering microRNAs in cell fate signaling.* Science, 2005. **310**(5752): p. 1288-9.
14. Chen, C.Z., *MicroRNAs as oncogenes and tumor suppressors.* N Engl J Med, 2005. **353**(17): p. 1768-71.
15. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
16. Landthaler, M., A. Yalcin, and T. Tuschl, *The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis.* Curr Biol, 2004. **14**(23): p. 2162-7.
17. Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. **303**(5654): p. 95-8.
18. Jaenisch, R. and R. Young, *Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming.* Cell, 2008. **132**(4): p. 567-82.
19. Ovitt, C.E. and H.R. Scholer, *The molecular biology of Oct-4 in the early mouse embryo.* Mol Hum Reprod, 1998. **4**(11): p. 1021-31.
20. Pesce, M. and H.R. Scholer, *Oct-4: gatekeeper in the beginnings of mammalian development.* Stem Cells, 2001. **19**(4): p. 271-8.
21. Rosner, M.H., et al., *A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo.* Nature, 1990. **345**(6277): p. 686-92.
22. Scholer, H.R., et al., *Octamer binding proteins confer transcriptional activity in early mouse embryogenesis.* EMBO J, 1989. **8**(9): p. 2551-7.
23. Boiani, M., J. Kehler, and H.R. Scholer, *Activity of the germline-specific Oct4-GFP transgene in normal and clone mouse embryos.* Methods Mol Biol, 2004. **254**: p. 1-34.
24. Mitsui, K., et al., *The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells.* Cell, 2003. **113**(5): p. 631-42.
25. Pan, G. and J.A. Thomson, *Nanog and transcriptional networks in embryonic stem cell pluripotency.* Cell Res, 2007. **17**(1): p. 42-9.
26. Yamaguchi, S., et al., *Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells.* Development, 2009. **136**(23): p. 4011-20.
27. Masui, S., et al., *Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.* Nat Cell Biol, 2007. **9**(6): p. 625-35.

28. Huangfu, D., et al., *Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds.* Nat Biotechnol, 2008. **26**(7): p. 795-7.
29. de Jong, J., et al., *Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.* J Pathol, 2008. **215**(1): p. 21-30.
30. Kanai-Azuma, M., et al., *Depletion of definitive gut endoderm in Sox17-null mutant mice.* Development, 2002. **129**(10): p. 2367-79.
31. Perrett, R.M., et al., *The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture.* Biol Reprod, 2008. **78**(5): p. 852-8.
32. Loh, Y.H., et al., *The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.* Nat Genet, 2006. **38**(4): p. 431-40.
33. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. **122**(6): p. 947-56.
34. 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.
35. Miyoshi, N., et al., *Reprogramming of mouse and human cells to pluripotency using mature microRNAs.* Cell Stem Cell, 2011. **8**(6): p. 633-8.
36. Anokye-Danso, F., et al., *Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency.* Cell Stem Cell, 2011. **8**(4): p. 376-88.
37. Judson, R.L., et al., *Embryonic stem cell-specific microRNAs promote induced pluripotency.* Nat Biotechnol, 2009. **27**(5): p. 459-61.
38. Wang, Y., et al., *MicroRNAs in embryonic stem cells.* J Cell Physiol, 2009. **218**(2): p. 251-5.
39. Morin, R.D., et al., *Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.* Genome Res, 2008. **18**(4): p. 610-21.
40. Liu, C., et al., *The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44.* Nat Med, 2011. **17**(2): p. 211-5.
41. Liu, C. and D.G. Tang, *MicroRNA regulation of cancer stem cells.* Cancer Res, 2011. **71**(18): p. 5950-4.
42. Kanellopoulou, C., et al., *Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing.* Genes Dev, 2005. **19**(4): p. 489-501.
43. Bernstein, E., et al., *Dicer is essential for mouse development.* Nat Genet, 2003. **35**(3): p. 215-7.
44. Wang, Y., et al., *DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal.* Nat Genet, 2007. **39**(3): p. 380-5.
45. Wang, Y., et al., *Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation.* Nat Genet, 2008. **40**(12): p. 1478-83.
46. Marson, A., et al., *Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.* Cell, 2008. **134**(3): p. 521-33.
47. Huyghe, E., P. Plante, and P.F. Thonneau, *Testicular cancer variations in time and space in Europe.* Eur Urol, 2007. **51**(3): p. 621-8.
48. Shah, M.N., et al., *Trends in testicular germ cell tumours by ethnic group in the United States.* Int J Androl, 2007. **30**(4): p. 206-13; discussion 213-4.
49. Oosterhuis, J.W. and L.H. Looijenga, *Testicular germ-cell tumours in a broader perspective.* Nat Rev Cancer, 2005. **5**(3): p. 210-22.
50. Looijenga, L.H., et al., *Relevance of microRNAs in normal and malignant development, including human testicular germ cell tumours.* Int J Androl, 2007. **30**(4): p. 304-14; discussion 314-5.
51. Skakkebaek, N.E., *Possible carcinoma-in-situ of the testis.* Lancet, 1972. **2**(7776): p. 516-7.
52. McLaren, A., *Primordial germ cells in the mouse.* Dev Biol, 2003. **262**(1): p. 1-15.
53. Hajkova, P., et al., *Epigenetic reprogramming in mouse primordial germ cells.* Mech Dev, 2002. **117**(1-2): p. 15-23.
54. van Gurp, R.J., et al., *Biallelic expression of the H19 and IGF2 genes in human testicular germ cell tumors.* J Natl Cancer Inst, 1994. **86**(14): p. 1070-5.
55. Looijenga, L.H., et al., *Genomic and expression profiling of human spermatocytic seminomas: primary spermatocyte as tumorigenic precursor and DMRT1 as candidate chromosome 9 gene.* Cancer Res, 2006. **66**(1): p. 290-302.

56. Looijenga, L.H., et al., *POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.* Cancer Res, 2003. **63**(9): p. 2244-50.
57. Korkola, J.E., et al., *Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors.* Cancer Res, 2006. **66**(2): p. 820-7.
58. Oram, S.W., et al., *TSPY potentiates cell proliferation and tumorigenesis by promoting cell cycle progression in HeLa and NIH3T3 cells.* BMC Cancer, 2006. **6**: p. 154.
59. Page, D.C., *Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads.* Development, 1987. **101 Suppl**: p. 151-5.
60. Akimoto, C., et al., *Testis-specific protein on Y chromosome (TSPY) represses the activity of the androgen receptor in androgen-dependent testicular germ-cell tumors.* Proc Natl Acad Sci U S A, 2010. **107**(46): p. 19891-6.
61. Mora, P.T., K. Chandrasekaran, and V.W. McFarland, *An embryo protein induced by SV40 virus transformation of mouse cells.* Nature, 1980. **288**(5792): p. 722-4.
62. Solozobova, V. and C. Blattner, *p53 in stem cells.* World J Biol Chem, 2011. **2**(9): p. 202-14.
63. Zhao, Y., et al., *Two supporting factors greatly improve the efficiency of human iPSC generation.* Cell Stem Cell, 2008. **3**(5): p. 475-9.
64. Sabapathy, K., et al., *Regulation of ES cell differentiation by functional and conformational modulation of p53.* EMBO J, 1997. **16**(20): p. 6217-29.
65. Kersemaekers, A.M., et al., *Role of P53 and MDM2 in treatment response of human germ cell tumors.* J Clin Oncol, 2002. **20**(6): p. 1551-61.
66. Masters, J.R. and B. Koberle, *Curing metastatic cancer: lessons from testicular germ-cell tumours.* Nat Rev Cancer, 2003. **3**(7): p. 517-25.
67. He, S., D. Nakada, and S.J. Morrison, *Mechanisms of stem cell self-renewal.* Annu Rev Cell Dev Biol, 2009. **25**: p. 377-406.
68. Hong, Y. and P.J. Stambrook, *Restoration of an absent G1 arrest and protection from apoptosis in embryonic stem cells after ionizing radiation.* Proc Natl Acad Sci U S A, 2004. **101**(40): p. 14443-8.
69. White, J., et al., *Developmental activation of the Rb-E2F pathway and establishment of cell cycle-regulated cyclin-dependent kinase activity during embryonic stem cell differentiation.* Mol Biol Cell, 2005. **16**(4): p. 2018-27.
70. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
71. Canela, N., et al., *The SET protein regulates G2/M transition by modulating cyclin B-cyclin-dependent kinase 1 activity.* J Biol Chem, 2003. **278**(2): p. 1158-64.
72. Calabrese, J.M., et al., *RNA sequence analysis defines Dicer's role in mouse embryonic stem cells.* Proc Natl Acad Sci U S A, 2007. **104**(46): p. 18097-102.
73. Card, D.A., et al., *Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells.* Mol Cell Biol, 2008. **28**(20): p. 6426-38.
74. Voorhoeve, P.M., et al., *A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.* Cell, 2006. **124**(6): p. 1169-81.
75. Gillis, A.J., et al., *High-throughput microRNAome analysis in human germ cell tumours.* J Pathol, 2007. **213**(3): p. 319-28.
76. Novotny, G.W., et al., *MicroRNA expression profiling of carcinoma in situ cells of the testis.* Endocr Relat Cancer, 2012. **19**(3): p. 365-79.
77. Stadler, B., et al., *Characterization of microRNAs involved in embryonic stem cell states.* Stem Cells Dev, 2010. **19**(7): p. 935-50.
78. Huang, Q., et al., *The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis.* Nat Cell Biol, 2008. **10**(2): p. 202-10.
79. Cairo, S., et al., *Stem cell-like micro-RNA signature driven by Myc in aggressive liver cancer.* Proc Natl Acad Sci U S A, 2010. **107**(47): p. 20471-6.
80. Nakagawa, M., et al., *Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.* Nat Biotechnol, 2008. **26**(1): p. 101-6.
81. Wernig, M., et al., *c-Myc is dispensable for direct reprogramming of mouse fibroblasts.* Cell Stem Cell, 2008. **2**(1): p. 10-2.

82. Zhao, J.J., et al., *Identification of miRNAs associated with tumorigenesis of retinoblastoma by miRNA microarray analysis.* Childs Nerv Syst, 2009. **25**(1): p. 13-20.
83. Melton, C., R.L. Judson, and R. Blelloch, *Opposing microRNA families regulate self-renewal in mouse embryonic stem cells.* Nature, 2010. **463**(7281): p. 621-6.
84. Medeiros, L.A., et al., *Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects.* Proc Natl Acad Sci U S A, 2011. **108**(34): p. 14163-8.
85. Zheng, G.X., et al., *A latent pro-survival function for the mir-290-295 cluster in mouse embryonic stem cells.* PLoS Genet, 2011. **7**(5): p. e1002054.
86. McIver, S.C., et al., *A unique combination of male germ cell miRNAs coordinates gonocyte differentiation.* PLoS One, 2012. **7**(4): p. e35553.
87. Barroso-delJesus, A., et al., *Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter.* Mol Cell Biol, 2008. **28**(21): p. 6609-19.
88. Lin, S.L., et al., *Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state.* RNA, 2008. **14**(10): p. 2115-24.
89. Lee, K.H., et al., *MicroRNA-373 (miR-373) post-transcriptionally regulates large tumor suppressor, homolog 2 (LATS2) and stimulates proliferation in human esophageal cancer.* Exp Cell Res, 2009. **315**(15): p. 2529-38.
90. Wu, N., et al., *MicroRNA-373, a new regulator of protein phosphatase 6, functions as an oncogene in hepatocellular carcinoma.* FEBS J, 2011. **278**(12): p. 2044-54.
91. Negrini, M. and G.A. Calin, *Breast cancer metastasis: a microRNA story.* Breast Cancer Res, 2008. **10**(2): p. 203.
92. Yang, K., A.M. Handorean, and K.A. Iczkowski, *MicroRNAs 373 and 520c are downregulated in prostate cancer, suppress CD44 translation and enhance invasion of prostate cancer cells in vitro.* Int J Clin Exp Pathol, 2009. **2**(4): p. 361-9.
93. Rippe, V., et al., *The two stem cell microRNA gene clusters C19MC and miR-371-3 are activated by specific chromosomal rearrangements in a subgroup of thyroid adenomas.* PLoS One, 2010. **5**(3): p. e9485.
94. Fareh, M., et al., *The miR 302-367 cluster drastically affects self-renewal and infiltration properties of glioma-initiating cells through CXCR4 repression and consequent disruption of the SHH-GLI-NANOG network.* Cell Death Differ, 2012. **19**(2): p. 232-44.
95. Yu, J., et al., *Human microRNA clusters: genomic organization and expression profile in leukemia cell lines.* Biochem Biophys Res Commun, 2006. **349**(1): p. 59-68.
96. Lawrie, C.H., et al., *Expression of microRNAs in diffuse large B cell lymphoma is associated with immunophenotype, survival and transformation from follicular lymphoma.* J Cell Mol Med, 2009. **13**(7): p. 1248-60.
97. Campayo, M., et al., *Low miR-145 and high miR-367 are associated with unfavorable prognosis in resected NSCLC.* Eur Respir J, 2012.
98. Beillard, E., et al., *miR-Sens--a retroviral dual-luciferase reporter to detect microRNA activity in primary cells.* RNA, 2012. **18**(5): p. 1091-100.
99. Ziliak, D., et al., *Genetic variation that predicts platinum sensitivity reveals the role of miR-193b* in chemotherapeutic susceptibility.* Mol Cancer Ther, 2012. **11**(9): p. 2054-61.
100. Williams, C.J. and J.M. Whitehouse, *Cis-platinum: a new anticancer agent.* Br Med J, 1979. **1**(6179): p. 1689-91.
101. Morimura, R., et al., *Novel diagnostic value of circulating miR-18a in plasma of patients with pancreatic cancer.* Br J Cancer, 2011. **105**(11): p. 1733-40.
102. Chen, X., et al., *Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases.* Cell Res, 2008. **18**(10): p. 997-1006.
103. Zahm, A.M., et al., *Circulating microRNA is a biomarker of pediatric Crohn disease.* J Pediatr Gastroenterol Nutr, 2011. **53**(1): p. 26-33.
104. Horwich, A., J. Shipley, and R. Huddart, *Testicular germ-cell cancer.* Lancet, 2006. **367**(9512): p. 754-65.
105. Murray, M.J., et al., *Identification of microRNAs From the miR-371~373 and miR-302 clusters as potential serum biomarkers of malignant germ cell tumors.* Am J Clin Pathol, 2011. **135**(1): p. 119-25.
106. Belge, G., et al., *Serum levels of microRNAs miR-371-3: a novel class of serum biomarkers for testicular germ cell tumors?* Eur Urol, 2012. **61**(5): p. 1068-9.
107. Murray, M.J. and N. Coleman, *Testicular cancer: a new generation of biomarkers for malignant germ cell tumours.* Nat Rev Urol, 2012. **9**(6): p. 298-300.

108. Wang, K., et al., *Export of microRNAs and microRNA-protective protein by mammalian cells.* Nucleic Acids Res, 2010. **38**(20): p. 7248-59.

109. Turchinovich, A., et al., *Characterization of extracellular circulating microRNA.* Nucleic Acids Res, 2011. **39**(16): p. 7223-33.

110. Arroyo, J.D., et al., *Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma.* Proc Natl Acad Sci U S A, 2011. **108**(12): p. 5003-8.

111. Gallo, A., et al., *The majority of microRNAs detectable in serum and saliva is concentrated in exosomes.* PLoS One, 2012. **7**(3): p. e30679.

112. Hurteau, G.J., et al., *Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin.* Cancer Res, 2007. **67**(17): p. 7972-6.

113. Baylin, S.B., et al., *Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer.* Hum Mol Genet, 2001. **10**(7): p. 687-92.

114. Plass, C. and P.D. Soloway, *DNA methylation, imprinting and cancer.* Eur J Hum Genet, 2002. **10**(1): p. 6-16.

115. Smiraglia, D.J., et al., *Distinct epigenetic phenotypes in seminomatous and nonseminomatous testicular germ cell tumors.* Oncogene, 2002. **21**(24): p. 3909-16.

116. Holm, T.M., et al., *Global loss of imprinting leads to widespread tumorigenesis in adult mice.* Cancer Cell, 2005. **8**(4): p. 275-85.

117. Toyota, M., et al., *Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer.* Cancer Res, 2008. **68**(11): p. 4123-32.

118. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer.* Carcinogenesis, 2010. **31**(1): p. 27-36.

119. Lujambio, A., et al., *A microRNA DNA methylation signature for human cancer metastasis.* Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13556-61.

120. Mestdagh, P., et al., *A novel and universal method for microRNA RT-qPCR data normalization.* Genome Biol, 2009. **10**(6): p. R64.

CHAPTER 3

Role of SOX2 in the etiology of germ cell cancer

Ronak Eini, Hans Stoop, Ad J.M.Gillis, Katharina Biermann,
Lambert C.J. Dorssers, Leendert H.J. Looijenga

Erasmus MC, University Medical Center Rotterdam, Department of Pathology,
The Netherlands

Submitted for publication

Abstract:

The transcription factor SOX2, associated with amongst others OCT3/4, is essential for maintenance of pluripotency and self-renewal of embryonic stem cells. In addition, SOX2 is highly expressed in embryonal carcinoma (EC), the stem cell component of malignant nonseminomatous germ cell tumors, referred to as germ cell cancer (GCC). In fact, OCT3/4 together with SOX2 is an informative diagnostic tool for EC in a clinical setting. Several studies support the hypothesis that SOX2 is a relevant oncogenic factor in various cancers and recently, SOX2 has been suggested as a putative therapeutic target for early stage EC. We demonstrate the presence of genomic amplification of SOX2 in an EC cell line, NCCIT, using array comparative genome hybridization and fluorescence *in situ* hybridization. Down-regulation of SOX2 by targeted siRNA provokes NCCIT cells towards apoptosis, while inhibition of OCT3/4 expression induced differentiation, with retained SOX2 levels. Mice pluripotent xenografts from NCCIT (N-NCCIT and N2-NCCIT) show a consistent SOX2 expression, in spite of loss of the expression of OCT3/4, and differentiation, with retained presence of genomic amplification. No SOX2 amplification has been identified in primary pure and mixed EC *in vivo* patient samples so far. These data support the model of SOX2-oncogene dependency of EC, which however, does not exclude induction of differentiation. These findings shed novel light on understanding the role of stem cell pluripotency factors in the etiology of GCC.

Introduction:

SOX2 (sex-determining region Y-box2) is a 317 amino-acid transcription factor containing an HMG domain located at 3q26, being a critical transcription factor of normal embryonic stem (ES) cell development and maintenance, as well as neural stem cells [1, 2]. Sox2 is required for epiblast maintenance, and formation of multipotent cell lineages in early mouse development depends on Sox2 function [3]. Moreover, Sox2 is one of the four transcription factors successfully used to induce pluripotent stem cell (iPS) from mouse and human fibroblast cells [4, 5]. In particular, SOX2 physically interacts with OCT3/4 and NANOG forming an interconnection machinery that binds to promoters of numerous but defined stem cell genes, regulating their expressions [1]. This seems essential since generating iPS cells from primary human fibroblast has become possible with the single use of OCT3/4 and SOX2 [5]. Relative hyper- or hypo-expression of these pluripotency factors may result in aberrant self-renewal of ES cells and can possibly even promote oncogenesis [6]. Recent studies have shown that SOX2 over-expression leads to aberrant stem cell self-renewal signaling in breast cancer cells [7, 8]. Moreover, several studies have shown over-expression of SOX2 in various cancers including glioblastoma [9], non-small cell lung cancer [10, 11], prostate cancer [12] and hepatocellular carcinomas [13] supporting SOX2 as a relevant oncogene in these malignancies. Specifically, SOX2 is reported as a lineage-survival oncogene in squamous cell carcinoma of the lung [14-16] and its over-expression is associated with tumor progression and poor clinical outcome in breast cancer [7, 17]. These reports suggest that SOX2 could activate important gene cascades involved in initiation and progression of tumors and maintenance of a poorly differentiated state [18].

Besides in these epithelial cancers, SOX2 has also been proven to be of diagnostic value in the context of human germ cell cancers (GCC) [19]. Testicular GCC originate from either a primordial germ cell (PGC) or gonocyte during early development [20-22]. Histologically and clinically, GCC are classified into seminoma (SE) and non-seminoma (NS). They both originate from the same precursor known as carcinoma *in situ* (CIS), also referred to as intratubular germ cell neoplasia unclassified (IGCNU) [20-23] . NS can contain both embryonal and extra-embryonal lineages, including embryonal carcinoma (EC), somatic differentiation (teratoma) and extra-embryonal differentiation (choriocarcinoma (CH) and yolk sac tumor (YS)). EC is the malignant ES cell counterpart, in principle able to differentiate into virtually all tissue lineages [24-26]. EC cells show a gene expression profile similar to that of ES cells, including high expressing of the core pluripotency transcription factors OCT3/4 and SOX2. These transcription factors act in concert to control stem cell self-renewal and pluripotency [27, 28]. OCT3/4 is expressed in CIS, SE and EC. In contrast, SOX2 is expressed in EC but not the precursor lesions and SE and normal germ cells. Currently, the expression of OCT3/4 and SOX2 are used for the diagnosis of EC while combination of the presence of OCT3/4 and SOX17 is used for the diagnosis of SE [19]. Additionally, a single study [29]

using an EC cell line NEC8 model reported SOX2-siRNA induced apoptotic cell death *in vitro* and growth suppression *in vivo*. In view of the similarity between EC and human ES cells, disruption of the orchestrated activity of these transcription factors could possibly induce lethal effects in EC cells [29].

There are various cell lines representing EC tumors i.e. NTera2 (NT2) [30], NCCIT [31], and 2102Ep [32] in which OCT3/4 and SOX2 are highly expressed [33]. As a result of array comparative genomic hybridization (CGH) on multiple EC cell lines, NCCIT cell line showed amplification at the long arm of chromosome 3, band q23 including the SOX2 gene locus [34]. In this study we explored whether oncogene dependency of the pluripotency gene SOX2 in EC tumors exist, which might explain the biological and clinical difference(s) between different histological subtypes of GCC. Previously we have shown that reduction of OCT3/4 and SOX2 in NT2 cause induction of differentiation [33]. Here, we investigate the effect of reduction of OCT3/4 and SOX2 by means of targeted siRNA in NCCIT cells. In addition, in order to decipher whether specific amplification of pluripotency genes associates with the undifferentiated state of EC tumors, we investigate whether primary GCC including pure EC and mixed NS including EC components, have amplification of OCT3/4 and SOX2 using Fluorescence *In Situ* hybridization (FISH).

Materials and methods:

Cell culture and manipulation:

NCCIT cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum supplemented with penicillin and streptomycin at 37°C under 5% CO2. Cells were passaged at 80% confluence. Briefly, cells were seeded at the density of 2 x 10^6 in 75ml flasks (Greiner bio-one/Germany). Incubated cells were harvested with trypsin-EDTA, washed with PBS, centrifuged and resuspended in DMEM medium. Approximately 8000 cells were used to make cytospins.

siRNA transfection for SOX2 and OCT3/4:

NCCIT cells were transfected with siRNA based SOX2 (s13294; Ambion/Invitrogen, Breda, the Netherlands), siRNA against OCT3/4 (Qiagen) [35] and Silencer Negative Control siRNA (Ambion, 4611) using Lipofectamine 2000 (Invitrogen) in 24 well-plates (Greiner bio-one/Germany) according to the manufacturers' protocol.

Protein Isolation and Western blotting analysis:

Isolation of protein and Western blotting analysis were essentially performed as previously described [36]. The antibodies are described in Immunohistochemistry (see below). In

addition, mouse monoclonal beta-actin (clone AC-15; Sigma Aldrich, St Louis, MO, USA) was used. Binding of the primary antibodies was visualized by using IRDYe donkey anti-mouse or donkey anti-goat secondary antibodies and the blots were scanned on the Odyssey infrared imaging system (from LI COR Biosciences, NE, USA).

Immunohistochemistry (IHC):

Staining was performed on cell lines as well as FFPE tissues. Unfixed cells were incubated for 1 hour at room temperature with the primary antibodies. Formalin-fixed paraffin-embedded (FFPE) tissues (4µm thick sections) were incubated overnight at 45˚C. After antigen retrieval treatment with the antibodies SOX2 (1:250, AF2018; R&D System, USA). OCT3/4 (sc-5279; 1:350; sc-Santa Cruz Biotechnology, Santa Cruz, CA, USA). Ki67 (1:50, A047; Dako, CA, USA) and Caspase 3 (1:500, 9579; Cell Signaling). Visualization was performed by using horseradish peroxidase avidin-biotin complex (Vectastain, PK6100; Vector) and DAB/H202 as substrate. For negative controls, primary antibody was omitted, resulting in complete absence of signals.

Percentage of knock-down:

Cells were harvested by trypsinization and analyzed for knock-down of the proteins under investigation using IHC on cytospins (see below). Percentage of positive cells for each gene was calculated by counting five different regions on the cytospin in which each region contained 100 cells, therefore, the total number of counted cells were 500 cells.

Cell viability test:

Cell death induced by siRNA transfection was evaluated by trypan blue exclusion. Briefly, NCCIT cells were harvested after 48h, 54h and 60h after siRNA transfections. Cells were washed with PBS and resuspended in 100 µl PBS. After mixing with 100 µl 0.8% trypan blue cells were calculated as the number of blue cells/total number of cells.

Flow cytometry:

Propidium Iodide: In order to measure the percentage of live and dead cells, Propidium Iodide (PI) assay was used. Briefly, cell pellets were resuspended in 500µl of PBS. Cells were fixed by adding 5 ml of 70% cold ethanol. Then the cell suspension was centrifuged at 400g for 5 min and the supernatant was removed. Cell pellet was washed with PBS and resuspended in 1 ml of PI staining solution (50µg/ml) according to the standard protocol. Cells were incubated for at least 30 min at room temperature in the dark to be analyzed by flow cytometry using FACSAria III machine using 488-nm laser line for excitation.

AnnexinV: Percentage of apoptotic cells was measured using BD Pharmingen FITC AnnexinV apoptosis detection kit according to the manufacturers' protocol. Briefly, harvested cells were washed with PBS, centrifuged and resuspended in Annexin buffer provided with the kit. The final concentration used for Annexin was 0.5µg/µl. Cells were analyzed by FACSAria III machine.

In Situ hybridization:

Fluorescent *in situ* hybridization (FISH) was performed with probe mapped to chromosome 3 band 26.33 (RP11-43F17) for SOX2 detection. In addition, an OCT3/4 specific probe was used (RP11-1058J10). In addition to this specific region, a probe mapped to centromeric region of chromosome 12 was assessed as control [37]. Briefly, FFPE tissue section of 4µm thickness were deparaffinized, pretreated in a 0.01M Sodium citrate solution under high pressure in pressure cooker , subsequently with pepsin (4.000U) at 37°C followed by washing and dehydration. Probes were labeled by nick-translation, according to standard procedures, either with dioxigenin-11-dUTP (Roche) or biotin-16-dUTP (Roche, Manheim, Germany), and applied in 10-15 µl hybridization mixture on the tissue slides. The probes were denatured together with the target by placing the slide for 10 min on the 80°C oven. After hybridization overnight at 37°C, the slides were washed stringently and the hybrids were detected by FITC-conjugated sheep-anti-digoxigenin (Roche) and CYE3-conjugated avidine (Jackson Immuno research laboratories, Cambridgeshure, UK) Results were studied with a fluorescent microscope LSM700 Zeiss. Cells were pretreated in -20°C methanol/acetone for 20 min and the rest of the procedure for FISH on NCCIT cells was the same as above mentioned with omitting of the high pressure treatment.

Xenografts generation:

The NMRI/Nu-Nu strain of immune-compromised female mice (Mus musculus), aged 6-8 weeks (Charles River Laboratories, Wilmington, Massachusetts, USA) was used in this study. The animals were housed in the individually ventilated cages with sterile bedding, water, rodent chew feed and air. Experiments were carried out in accordance with the ARRIVE guidelines [38] and the animal research protocol was approved by the Institutional Animal Ethical Committee (DEC), Erasmus University Medical Center, Rotterdam, the Netherlands. Six mice were used. The mice were anesthetized by the exposure to 2% isoflurane (Pharmachemie BV, Haarlem, The Netherlands) and placed ventral side up on a pre-warmed injecting pad. Approximately 10-12 million cells from each cell line suspended in 200 µl fresh culture medium with 20% serum were implanted under the skin with a 1-cc syringe and 24-gauge needle. Two week after implantation, xenograft tumor growth was checked by palpation and the size of xenografts was measured using a vernier calliper [xeno-tumor size (mm^2) = length x width] twice a week and the mice were followed for a total period of maximal 12 months until they were sacrificed. For ethical reasons, primary xenograft tumors

reaching the size of 225 mm^2 were taken from study and sacrificed. The primary tumors were preserved by direct freezing in liquid nitrogen as well as by fixation in 10% formalin solution (Sigma-Aldrich, St. Louis, Missouri, USA) for subsequent histological analyses.

Results:

OCT3/4 and SOX2 in Embryonal Carcinoma (EC) cell lines:

SOX2 amplification confirmed in NCCIT cells:

Our previous genome wide copy number investigation of multiple EC cell lines showed a specific amplification at the long arm of chromosome 3, band q23, including the SOX2 gene locus only in the NCCIT cell line [34]. The borders were defined as 177.604.206 bp (detected by RP11-71G7 probe) until 184.060.761 bp (detected by RP11-553E4 probe), encompassing a region of about 6.4Mb. Various genes are mapped to this genomic fragment including SOX2 (Figure 1A & B).

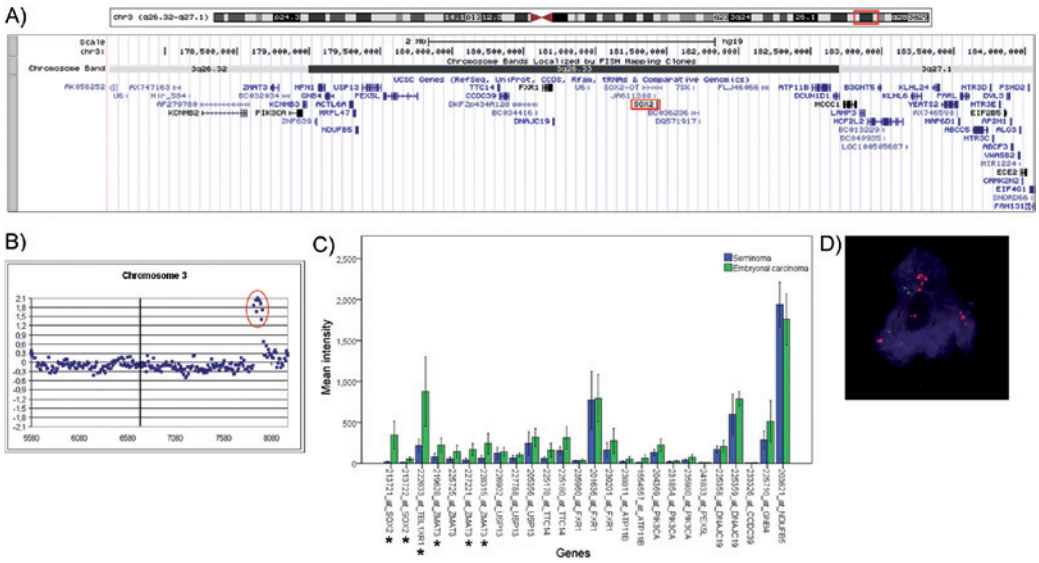


Figure 1. A) UCSC genome browser (version hg19) representation of the genomic region of amplification at the long arm of chromosome 3, band q23.33, in NCCIT cells. The borders are 177.604.260 bp and 184.060.761 bp (encompassing a region of about 6.4Mb). The genes mapped to this region are shown including SOX2 locus; B) Array CGH result, the region of amplification in chromosome 3q is indicated in red circle, the borders are defined between the probes RP11-71G7 and RP11-553E4, respectively; C) Expression histogram indicating the relative expression of genes mapped within the amplified region in a series of SE and EC. The stars indicate genes significantly differentially expressed between SE and EC (according to Mann Whitney U test). Most of the genes are represented by multiple specific probes; D) DNA-FISH result for SOX2 in NCCIT cells. A centromere 12 (C12) specific probe is used as control. Red dye (Cye3) shows SOX2 probe. For C12 probe green dye (FITC) is used. Multiple red spots for SOX2 are detectable in each cell containing two green spots for C12, indicating SOX2 amplification.

To investigate the pattern of expression of all genes within the amplified region in a series in primary GCC, including multiple pure EC, high throughput Affymetrix expression data analysis was performed, and compared to seminomas (SE). The results showed a significant difference between the expression of SOX2 in EC and SE, being high versus low in expression respectively, which is in line with previous findings [19]. In total, 13 genes were analyzed in this region in which three genes, including SOX2, showed a significant difference between EC and SE (Figure 1C). Although there are other candidate genes within the amplified region, due to critical role of SOX2 in early development in close connection with OCT3/4 and its diagnostic value in the diagnosis of GCC, SOX2 was selected for further investigations. To verify the presence of SOX2 amplification in NCCIT, DNA FISH was performed using a verified probe for SOX2 gene (labeled with biotin). A centromere 12 (C12) specific probe (labeled with digoxigenin) was used in a double FISH experiment as control for copy number changes. The frequency of the signals obtained for SOX2 and C12 confirmed the amplification of SOX2, suggested to be at a single chromosome (Figure 1D).

Silencing OCT3/4 and SOX2 in NT2:

To investigate the effect of reduced levels of OCT3/4 and SOX2 in the NT2 cell line, representative for pluripotent EC, siRNA-based OCT3/4 and SOX2 inhibition was performed. Two independent siRNAs against OCT3/4 and three independent siRNA against SOX2 were tested for specificity by means of Western blotting at different time points after transfection (48, 72, 96 and 120 hours) (Supplementary figure). These included previously reported sequences, referred to as "Matin" and "Hay" for OCT3/4 [35, 39] and "13294", "13295", "13296" for SOX2 [33]. Based on the results obtained, specific siRNAs were chosen for further experiments (boxed in supplementary figure), being "Hay" for OCT3/4 and "13294" for SOX2. Because of the time effects, cells obtained 72h after transfection were chosen for subsequent analysis. These conditions showed the most profound down-regulation of expression at the protein level (over 90%). Both OCT3/4 and SOX2 RNA- knock-down in the NT2 cell line under these conditions resulted in defined induction of differentiation as described before [33].

Silencing OCT3/4 and SOX2 in NCCIT:

To investigate the effect of reduced levels of OCT3/4 and SOX2 in the NCCIT cell line, as done in the NT2 cell line, the selected siRNAs for OCT3/4 and SOX2 (see above) were transfected, and the cells were investigated at three different time points (48, 48+6h and 48+12h). Use of each siRNA specifically led to a down-regulation of OCT3/4 and SOX2 protein expression in NCCIT. The percentage of positive cells for these proteins was measured in negative control cells and cells transfected with siRNAs, by immunohistochemistry, a method shown to be informative in such a setting [33]. The results indicate that the expression level of OCT3/4 and SOX2 protein expression were reduced significantly in time (Figure 2A).

OCT3/4-down-regulation in NCCIT results in differentiation:

To investigate the effect of OCT3/4 down-regulation on the identity of the NCCIT cells, expression level of a selected panel of genes representative for pluripotency and differentiation (for all embryonic germ layers: mesoderm, endoderm and ectoderm) was measured using q-RT-PCR, as reported before [33]. As in the NT2 cells, OCT3/4 down-regulation resulted in differentiation, demonstrated by loss of the pluripotency factors (*OCT3/4*, *NANOG* and *LIN28*) and up-regulation of some differentiation genes, including *OTX1* (ectoderm), *brachyury* and *HAND1* (mesoderm), and *LAMB1* (endoderm) (Figure 2B).

SOX2-siRNA caused apoptosis in NCCIT:

As reported for OCT3/4, reduction of SOX2 expression in NT2 resulted in induction of differentiation. In contrast, in the NCCIT cell line SOX2 reduction did not result in differentiation, but instead, it caused a progressive loss of cells in time (48, 48+6, 48+12 and 72 hours after transfection). Cell death was obvious after 54 hours and prominent at 60 hours. Because at the last time point, almost no viable cells were present, no further analysis could be done including expression profiling. Using Trypan blue staining, the presence of live cells was measured at each time point (Figure 2C). The results demonstrate at the latest time point after transfection a progressive and significant effect of SOX2 and OCT3/4 reduction on the amount of living cells in time, indicating only 20% living cells in the SOX2 reduced cells and 50% living cells in the OCT3/4 reduced cells. To investigate the effect of SOX2 and OCT3/4 down-regulation on proliferation, immunohistochemical staining was performed for Ki67 on cytospin slides at the latest time point (60h). The results (Figure 2D) demonstrate that OCT3/4, inducing differentiation (see above), also resulted in a decrease in proliferation status compared to the negative siRNA control and untreated cells for about 50%. The effect of SOX2 reduction was even more severe, resulting in about 10% positive cells ($p<0.01$). Subsequently, we investigated whether loss of cells was also the result of increased apoptosis, for which various approaches were used. These include immunohistochemistry for the apoptosis marker Caspase 3 (Figure 2E). Positive staining was found predominantly and significantly in the cells with SOX2 inhibition, especially at the 60 hours time point ($p = 0.04$). In contrast, inhibition for OCT3/4 resulted in less that 4% apoptotic cells. This pattern was confirmed by FACs analysis using Propidium Iodide (PI) exclusion test and Annexin V staining (Figure 2F, G). These data demonstrate that the major decrease in the amount of living NCCIT cells due to reduced SOX2 expression is explained by induction of apoptosis.

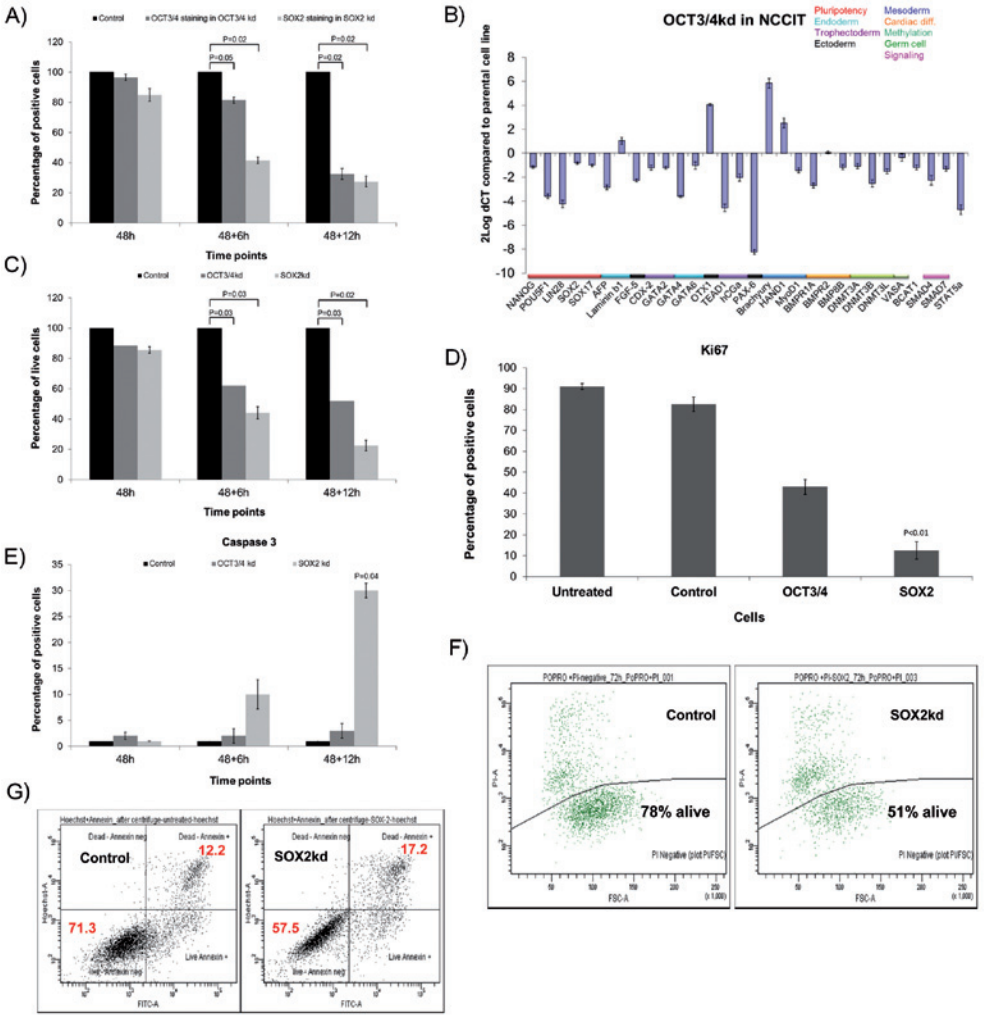


Figure 2. A) Percentage of positive cells for OCT3/4 and SOX2 in cells with reduced OCT3/4 and SOX2 levels compared to cells transfected with control siRNA in the NCCIT cells in three different time-points post transfection based on immunohistochemistry; The controls are set to 100 in all cases. B) Relative expression pattern of 32 genes representing targets for pluripotency and differentiation (ectoderm, mesoderm and endoderm). The expression levels are normalized based on the housekeeping gene *HPRT*; C) Percentage of living NCCIT cells with reduced level of OCT3/4 and SOX2 compared to cells transfected with control siRNA at each time point based on Trypan blue measurement; D) Percentage of positive NCCIT cells for Ki67 in untreated cells, cells transfected with control siRNA and cells with reduced levels of OCT3/4 and SOX2 at 48+12h after the transfection; E) Percentage of positive cells for Caspase 3 in untreated NCCIT cells, Cells transfected with control siRNA show a reduced level of OCT3/4 and SOX2 at all three time points after the transfection; F) FACS analysis with Propidium Iodide staining in cells transfected with control siRNA and cells transfected with SOX2 siRNA at 60h after transfection. NCCIT cells transfected with control siRNA show the presence of 78% living cells, while cells transfected with SOX2 siRNA show 51% living cells; G) Annexin V assay results in cells with siRNA control and cells with reduced SOX2 at 60h after transfection. In the control, percentages of living-annexin negative cells are 89.6%, while dead-annexin positive cells are 12.2 %. In SOX2kd cells, the amount of living-annexin negative cells is 57.5 % while the amount of dead-annexin positive cells is 17.2%.

N-NCCIT and N2-NCCIT xenografts show a consistent positive expression for SOX2:

It is known that NCCIT has the capacity to differentiate [31, 40], in line with the results obtained from the OCT3/4 inhibition experiments (see above). SOX2 is found to be an absolute marker for EC in combination with OCT3/4, although it can also be found more heterogeneously in differentiated components, especially teratoma [19]. In the context of the induction of apoptosis by means of SOX2 suppression in NCCIT, it is interesting to investigate whether NCCIT cells grown as a xenograft *in vivo* remains SOX2 positive throughout, in spite of possible differentiation. To answer this question, multiple mice xenografts were generated from the parental NCCIT cell line (referred to N- and N2-NCCIT). The established tumors were characterized, and a sub-line was subsequently cultured continuously *in vitro.* The *in vivo* tumors as well as the derived cell lines were again tested for SOX2 and OCT3/4. The overall pattern (representative examples shown in Figure 3) indicated that in spite of morphological induction of differentiation, supported by loss of OCT3/4 by immunohistochemistry, this was not accompanied by loss of the expression of SOX2. In other words, SOX2 remained positive in all components of the tumors.

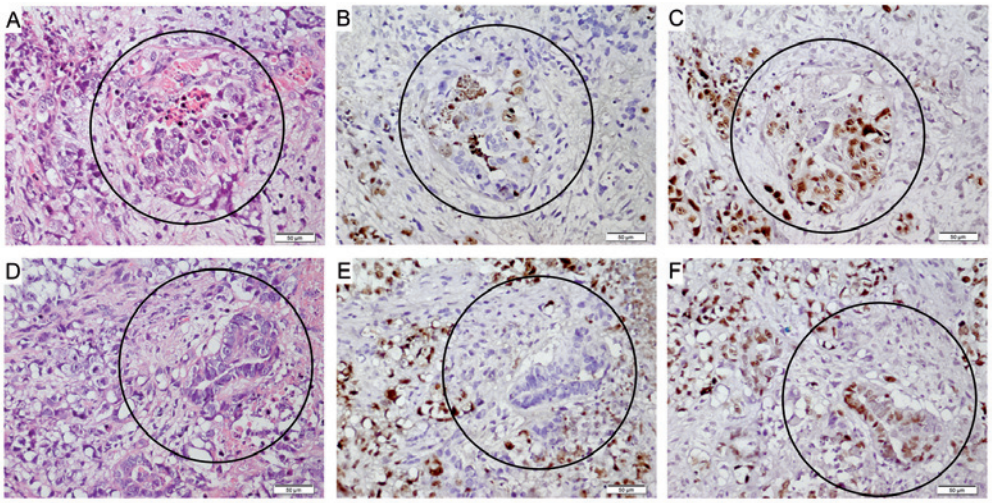


Figure 3. Immunohistochemistry for SOX2 and OCT3/4 on nude mice xenografts of NCCIT.

Images A, B and C belong to N-NCCIT; D, E and F belong to N2-NCCIT. A & D) H & E staining demonstrates the histological composition, showing regions with differentiation (indicated with a circle); B & E) Staining for OCT3/4 showing the presence of heterogeneity, showing undifferentiated (positive) and differentiated (negative) cells ; C & F) Staining for SOX2 shows that the malignant cells are consistently positive, in spite of the presence or absence of OCT3/4.

Detection of amplification for SOX2 and OCT3/4 in EC:

Because *SOX2* is amplified in the EC cell line NCCIT, with a functional effect on survival, it is of interest to check whether primary GCC, have *SOX2* amplification. Particularly, the so-called nullipotent EC might be of interest to be investigated, although this component might also be present in mixed nonseminomas, i.e., tumors with a histologically mixed composition. A series of 12 pure EC and 34 mixed GCCs including an EC component were studied using the double FISH method (see above). No amplification for *SOX2* or *OCT3/4* (i.e. more than 6 copies) was found in the cases investigated.

Discussion:

Pluripotent stem cells have been isolated from a variety of human and mouse sources as models to investigate processes involved in early embryonal development [41, 42]. Two of the well-studied cell types are ES cells derived from the inner cell mass of blastocyst-stage embryos and EC cells, the nonseminomatous stem cells of GCC [42, 43]. By definition, pluripotent stem cells have extensive self-renewal capacity and the ability to differentiate into wide variety of cell types [41, 42, 44]. In fact EC cells derived from the progenitor of the germ line are a malignant equivalent of ES cells [45], thus they provide a good model to study early embryonal development as well as tumorigenesis [42].

OCT3/4 and SOX2 are transcription factors essential to the pluripotent and self-renewing phenotypes of ES cells. These master ES cell pluripotency factors are highly expressed in EC [19, 46]. Representative cell lines of EC which are capable of differentiation (NT2, NCCIT, 2102Ep) have been generated and extensively used for studies [30, 31, 40, 43]. Based on our molecular data derived from array CGH data, the NCCIT cell line, derived from an extra-gonadal GCC, shows a restricted amplification of the 3q23 region. This region contains *SOX2*, amongst many other genes. Here we demonstrate that indeed *SOX2* maps within the minimal region of overlap of the amplification. This triggered a more detailed analysis of the role of SOX2 in this cell line compared to an EC cell line without such a genomic amplification. In fact, *SOX2* amplification and its pathogenic role association with oncogenesis have been reported in human lung squamous cell carcinoma [47]. In addition, targeting SOX2 in breast cancer cell lines have shown that siRNA-mediated knock-down of SOX2 resulted in cell cycle arrest by down-regulation of Cyclin D1 and this arrest in cell cycle was accompanied by an inhibition of tumor cell proliferation in xenograft models [18]. Most recently, it has been reported that inhibition of SOX2 might be of therapeutic potential for EC [29]. In that particular study, SOX2 down-regulation in the EC cell line NEC8, when established *in vivo*, induced tumor growth suppression in case of a limited tumor size. In contrast, down-regulation showed no effect on progression in case of a tumor of larger size. This is likely related to loss of SOX2 expression to some extent and differentiation. This study indicates that suppression of the expression of SOX2 expression might be useful to block cell proliferation

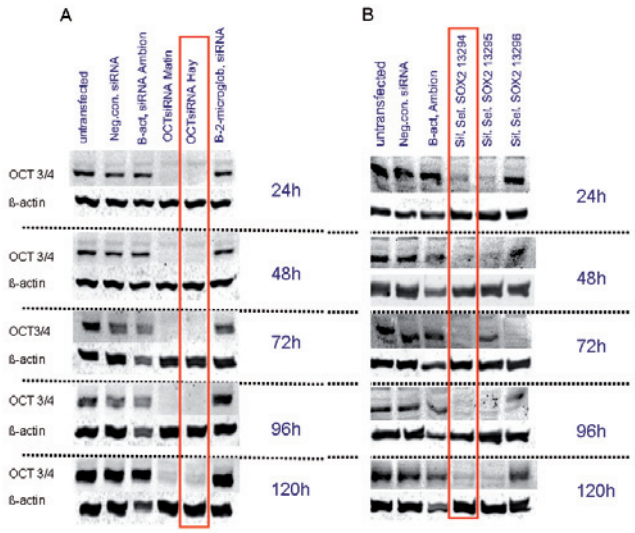
in early stage EC. Interestingly, as a result of our array CGH study on multiple EC cell lines, NCCIT cell line showed amplification of the long arm of chromosome 3, band q23 including the SOX2 gene locus [34].

Unlike the majority of GCC cell lines which are derived from testicular cancers, NCCIT is derived from an extragonadal mediastinal GCC. In 1988, NCCIT was established as an *in vitro* cell line, composed of developmentally pluripotent cells capable of somatic and extra-embryonic differentiation. Nude mouse-xenografts of NCCIT contained foci of EC, yolk sac tumor, immature somatic tissues, and trophoblastic giant cells indicating that this cell line is indeed developmentally pluripotent [40]. In 1993, retinoic-induced differentiation of NCCIT into all three embryonic germ layers and extra-embryonic cell lineages was reported, and in fact, it was suggested that the parental NCCIT cells show characteristics intermediate between SE and EC [31]. We investigated the effect of OCT3/4 and SOX2 down-regulation in NCCIT. The results demonstrate that inhibition of OCT3/4 resulted in differentiation, while inhibition of SOX2 led to cell death. This suggests that survival of NCCIT is dependent on the presence of SOX2 expression, referred to as oncogene-dependence [48]. This was supported by various methods and read out systems. As a control, inhibition of OCT3/4 and SOX2 was done similarly in the NT2 cell line, which showed induction of differentiation under the experimental conditions applied [33]. The next step in our study was to investigate the potential of the NCCIT cells to undergo differentiation *in vivo*. Therefore multiple xenografts were generated. Interestingly, in spite of induction of differentiation, supported by loss of OCT3/4, all tumor cells remained positive for SOX2. This is in line with the hypothesis that differentiation is possible, as found in the *in vitro* experiments, even in spite of SOX2 amplification. The lineages formed are selected by continuous expression of SOX2 [49]. In this context the potential use of inhibition of SOX2 in a clinical setting, should be considered carefully because down-regulation of SOX2 might result in the induction of differentiation, leading to potentially highly metastatic clones.

The absence of SOX2 amplification in a series of *in vivo* pure EC and EC containing mixed nonseminomas, suggest that SOX2 oncogene dependence, at least due to gene amplification, is not a frequent mechanism in GCC, which questions indeed the approach of targeted therapy in a clinical setting. In this context it might be of interest to check whether the NEC8 cell line contains SOX2 amplification [29].

Most recently it has been shown that site-specific phosphorylation of OCT3/4 regulated by AKT, promotes stemness of EC cells (i.e., NCCIT) compared to ES cells [50]. Presence of this site-specific phosphorylation promotes release of the OCT3/4 protein from the *AKT1* promoter, resulting in induced expression, which will lead to suppression of apoptosis, and simultaneously, enhances capacity of OCT3/4 to form a complex with NANOG and SOX2, promoting pluripotency. This additional effect promotes tumorigenic capacity of EC compared to ES, which might be of interest for further investigation.

In conclusion, these data shed novel light on the role of ES cell pluripotency factors, particularly SOX2 and its dependence in relation to differentiation, in the etiology of GCC and regulation of stem cell differentiation.



Supplementary figure. Western blot analysis of down-regulation of OCT3/4 and SOX2 in NT2 cells at various time points (24, 48, 72, 96 and 120 hours).

A) NT2 cells are transfected with two independent OCT3/4 siRNAs ("Matin" and "Hay"), two independent β-actin siRNAs and negative control siRNA. OCT3/4 "Hay" is selected for further experiments. B) NT2 cells are transfected with three independent SOX2 siRNAs (13294, 13295 and 13296), one β-actin siRNA and negative control siRNA. SOX2-13294 siRNA is selected for further experiments. (In 72h incubation, SOX2- siRNA 13295 and 13296 have been switched)

References:

1. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. **122**(6): p. 947-56.
2. Episkopou, V., *SOX2 functions in adult neural stem cells.* Trends Neurosci, 2005. **28**(5): p. 219-21.
3. Avilion, A.A., et al., *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. **17**(1): p. 126-40.
4. 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.
5. Huangfu, D., et al., *Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds.* Nat Biotechnol, 2008. **26**(7): p. 795-7.
6. Zhang, Y., et al., *Estrogen Receptor alpha Signaling Regulates Breast Tumor-initiating Cells by Downregulating miR-140 which Targets the Transcription Factor SOX2.* J Biol Chem, 2012.
7. Leis, O., et al., *Sox2 expression in breast tumours and activation in breast cancer stem cells.* Oncogene, 2012. **31**(11): p. 1354-65.
8. Lengerke, C., et al., *Expression of the embryonic stem cell marker SOX2 in early-stage breast carcinoma.* BMC Cancer, 2011. **11**: p. 42.
9. Gangemi, R.M., et al., *SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity.* Stem Cells, 2009. **27**(1): p. 40-8.
10. Lu, Y., et al., *Evidence that SOX2 overexpression is oncogenic in the lung.* PLoS One, 2010. **5**(6): p. e11022.
11. Leung, E.L., et al., *Non-small cell lung cancer cells expressing CD44 are enriched for stem cell-like properties.* PLoS One, 2010. **5**(11): p. e14062.
12. Bae, K.M., et al., *Expression of pluripotent stem cell reprogramming factors by prostate tumor initiating cells.* J Urol, 2010. **183**(5): p. 2045-53.
13. Xu, X.L., et al., *The properties of tumor-initiating cells from a hepatocellular carcinoma patient's primary and recurrent tumor.* Carcinogenesis, 2010. **31**(2): p. 167-74.
14. Bass, A.J., et al., *SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas.* Nat Genet, 2009. **41**(11): p. 1238-42.
15. Hussenet, T. and S. du Manoir, *SOX2 in squamous cell carcinoma: amplifying a pleiotropic oncogene along carcinogenesis.* Cell Cycle, 2010. **9**(8): p. 1480-6.
16. Yuan, P., et al., *Sex determining region Y-Box 2 (SOX2) is a potential cell-lineage gene highly expressed in the pathogenesis of squamous cell carcinomas of the lung.* PLoS One, 2010. **5**(2): p. e9112.
17. Chen, Y., et al., *The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer.* J Biol Chem, 2008. **283**(26): p. 17969-78.
18. Stolzenburg, S., et al., *Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer.* Nucleic Acids Res, 2012. **40**(14): p. 6725-40.
19. de Jong, J., et al., *Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.* J Pathol, 2008. **215**(1): p. 21-30.
20. Horwich, A., J. Shipley, and R. Huddart, *Testicular germ-cell cancer.* Lancet, 2006. **367**(9512): p. 754-65.
21. Oosterhuis, J.W. and L.H. Looijenga, *Testicular germ-cell tumours in a broader perspective.* Nat Rev Cancer, 2005. **5**(3): p. 210-22.
22. Skakkebaek, N.E., *Possible carcinoma-in-situ of the testis.* Lancet, 1972. **2**(7776): p. 516-7.
23. Woodward, P.J. (2004). *world health organization classification of tumours pathology and genetics of the urinary system and male genital organs*. IARC Press, Lyon, p. 17-278.
24. Ulbright, T.M., *Germ cell neoplasms of the testis.* Am J Surg Pathol, 1993. **17**(11): p. 1075-91.
25. Masters, J.R., et al., *Cancer stem cells.* BJU Int, 2003. **92**(7): p. 661-2.
26. Honecker, F., et al., *Germ cell lineage differentiation in non-seminomatous germ cell tumours.* J Pathol, 2006. **208**(3): p. 395-400.
27. Andrews, P.W., et al., *Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin.* Biochem Soc Trans, 2005. **33**(Pt 6): p. 1526-30.

28. Andrews, P.W., *From teratocarcinomas to embryonic stem cells.* Philos Trans R Soc Lond B Biol Sci, 2002. **357**(1420): p. 405-17.
29. Ushida, H., et al., *Therapeutic potential of SOX2 inhibition for embryonal carcinoma.* J Urol, 2012. **187**(5): p. 1876-81.
30. Andrews, P.W., et al., *Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours.* Int J Cancer, 1996. **66**(6): p. 806-16.
31. Damjanov, I., B. Horvat, and Z. Gibas, *Retinoic acid-induced differentiation of the developmentally pluripotent human germ cell tumor-derived cell line, NCCIT.* Lab Invest, 1993. **68**(2): p. 220-32.
32. Damjanov, I. and P.W. Andrews, *Ultrastructural differentiation of a clonal human embryonal carcinoma cell line in vitro.* Cancer Res, 1983. **43**(5): p. 2190-8.
33. Gillis, A.J., et al., *Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis.* Int J Androl, 2011. **34**(4 Pt 2): p. e160-74.
34. de Jong, J., et al., *Further characterization of the first seminoma cell line TCam-2.* Genes Chromosomes Cancer, 2008. **47**(3): p. 185-96.
35. Hay, D.C., et al., *Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells.* Stem Cells, 2004. **22**(2): p. 225-35.
36. de Jong, J., et al., *Diagnostic value of OCT3/4 for pre-invasive and invasive testicular germ cell tumours.* J Pathol, 2005. **206**(2): p. 242-9.
37. Looijenga, L.H., et al., *Localization and polymorphism of a chromosome 12-specific alpha satellite DNA sequence.* Cytogenet Cell Genet, 1990. **53**(4): p. 216-8.
38. Kilkenny, C., et al., *Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research.* PLoS Biol, 2010. **8**(6): p. e1000412.
39. Matin, M.M., et al., *Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells.* Stem Cells, 2004. **22**(5): p. 659-68.
40. Teshima, S., et al., *Four new human germ cell tumor cell lines.* Lab Invest, 1988. **59**(3): p. 328-36.
41. Gearhart, J., *New potential for human embryonic stem cells.* Science, 1998. **282**(5391): p. 1061-2.
42. Yang, S., et al., *Tumourigenic characteristics of embryonal carcinoma cells as a model for studying tumour progression of human embryonic stem cells.* Cell Prolif, 2012. **45**(4): p. 299-310.
43. Andrews, P.W., et al., *Inhibition of proliferation and induction of differentiation of pluripotent human embryonal carcinoma cells by osteogenic protein-1 (or bone morphogenetic protein-7).* Lab Invest, 1994. **71**(2): p. 243-51.
44. Preston, S.L., et al., *The new stem cell biology: something for everyone.* Mol Pathol, 2003. **56**(2): p. 86-96.
45. Andrews, P.W., B. Fenderson, and S. Hakomori, *Human embryonal carcinoma cells and their differentiation in culture.* Int J Androl, 1987. **10**(1): p. 95-104.
46. Looijenga, L.H., et al., *POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.* Cancer Res, 2003. **63**(9): p. 2244-50.
47. 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.
48. Sharma, S.V. and J. Settleman, *Oncogene addiction: setting the stage for molecularly targeted cancer therapy.* Genes Dev, 2007. **21**(24): p. 3214-31.
49. Lim, J.H., et al., *Isolation and characterization of neural progenitor cells from adult canine brains.* Am J Vet Res, 2012. **73**(12): p. 1963-8.
50. Lin, Y., et al., *Reciprocal regulation of akt and oct4 promotes the self-renewal and survival of embryonal carcinoma cells.* Mol Cell, 2012. **48**(4): p. 627-40.

CHAPTER 4

DICER1 RNase IIIb domain mutations are infrequent in testicular germ cell tumours

Carmela M. de Boer, Ronak Eini, Ad M.J. Gillis, Hans Stoop,
Leendert H.J. Looijenga, Stefan J. White

Center for Reproduction and Development, Monash Institute for Medical Research, Melbourne, Australia

Department of Pathology, Erasmus MC, University Medical Center Rotterdam, Josephine Nefkens Institute, The Netherlands

BMC, research notes; 5:569

Abstract

Background:
Testicular Germ Cell Tumours (TGCT) are the most frequently occurring malignancy in males from 15-45 years of age. They are derived from germ cells unable to undergo physiological maturation, although the genetic basis for this is poorly understood.
A recent report showed that mutations in the RNase IIIb domain of DICER1, a micro-RNA (miRNA) processing enzyme, are common in non-epithelial ovarian cancers. *DICER1* mutations were found in 60% of Sertoli-Leydig cell tumours, clustering in four codons encoding metal-binding sites. Additional analysis of 14 TGCT DNA samples identified one case that also contained a mutation at one of these sites.

Findings:
A number of previous studies have shown that *DICER1* mutations are found in <1% of most cancers. To provide a more accurate estimate of the frequency of such mutations in TGCTs, we have analysed 96 TGCT samples using high resolution melting curve analysis for sequence variants in these four codons. Although we did not detect any mutations in any of these sites, we did identify a novel mutation (c.1725 R>Q) within the RNase IIIb domain in one TGCT sample, which was predicted to disturb DICER1 function.

Conclusion:
Overall our findings suggest a mutation frequency in TGCTs of ~1%. We conclude therefore that hot-spot mutations, frequently seen in Sertoli-Leydig cell tumours, are not common in TGCTs.

Keywords: miRNA, DICER1, cancer, Testicular Germ Cell Tumours, mutation detection

Introduction

Testicular germ cell tumours (TGCTs), also referred to as type II TGCTs, are the most common cancer affecting adolescent and young men, and are typically diagnosed around the age of 30. TGCTs are classified based on a number of features, including clinical characteristics and histological markers [1]. TGCTs are derived from immature germ cells that did not mature during embryonic development [2, 3], and can be subdivided into seminomas and non-seminomas. Seminomas consist of transformed germ cells that do not exhibit overt pluripotency, although expressing SOX17, OCT3/4 and NANOG [4, 5]. In contrast, non-seminomas are transformed germ cells that have reactivated pluripotency, as evidenced by the expression of SOX2, OCT3/4 and NANOG [6], three main regulators of pluripotency.

Micro-RNAs (miRNAs) play key roles in regulating mRNA levels [7]. All miRNAs are derived from precursor sequences, which are processed by DICER1 to form double stranded RNA duplexes. These duplexes consist of a principle miRNA strand and the (imperfectly) complementary miRNA strand, referred to as miRNA* [8].

Germline mutations in *DICER1* have been identified in patients with pleuropulmonary blastoma [9], often associated with goiter and Sertoli-Leydig cell tumours. A recent report described the identification of recurrent, somatic mutations in the *DICER1* gene in nonepithelial ovarian cancers [10].

Table 1. DNA constructs created to simulate DICER1 hot-spot mutations.

Variant*	Oligonucleotide sequences^
1705_1709	
Reference sequence	5'-GGTGCTTGGTTATGAGGTAGTCCaaaatcgcatctcccaggaattctaagCGCTGGTAACAATCTGAGGG-3' 3'-CCACGAACCAATACTCCATCAGGttttagcgtagagggtccttaagattgGCGACCATTGTTAGACTCCC-5'
c.5113G>A p.E1705K	5'-GGTGCTTGGTTATGAGGTAGTCCaaaatcgcatctcccaggaatt**T**taagCGCTGGTAACAATCTGAGGG-3' 3'-CCACGAACCAATACTCCATCAGGttttagcgtagagggtccttaa**A**attgGCGACCATTGTTAGACTCCC-5'
c.5125G>A p.D1709N	5'-GGTGCTTGGTTATGAGGTAGTCCaaaatcgcat**T**tcccaggaattctaagCGCTGGTAACAATCTGAGGG-3' 3'-CCACGAACCAATACTCCATCAGGttttagcgta**A**agggtccttaagattgGCGACCATTGTTAGACTCCC-5'
c.5126A>G p.D1709G	5'-GGTGCTTGGTTATGAGGTAGTCCaaaatcgca**C**ctcccaggaattctaagCGCTGGTAACAATCTGAGGG-3' 3'-CCACGAACCAATACTCCATCAGGttttagcgt**G**gagggtccttaagattgGCGACCATTGTTAGACTCCC-5'
c.5127T>A p.D1709E	5'-GGTGCTTGGTTATGAGGTAGTCCaaaatcgc**T**tctcccaggaattctaagCGCTGGTAACAATCTGAGGG-3' 3'-CCACGAACCAATACTCCATCAGGttttagcg**A**agagggtccttaagattgGCGACCATTGTTAGACTCCC-5'
1810_1813	
Reference sequence	5'-CATGTAAATGGCACCAGCAAgcgactcaaaaatatcccccatggCCTTTGGAACTTCAATATCCTCTT-3' 3'-GTACATTTACCGTGGTCGTTcgctgagtttttatagggggtaccGGAAACCTTGAAGTTATAGGAGAA-5'
c.5428G>T p.D1810Y	5'-CATGTAAATGGCACCAGCAAgcgactcaaaaatat**A**ccccatggCCTTTGGAACTTCAATATCCTCTT-3' 3'-GTACATTTACCGTGGTCGTTcgctgagtttttata**T**ggggtaccGGAAACCTTGAAGTTATAGGAGAA-5'
c.5437G>C p.E1813Q	5'-CATGTAAATGGCACCAGCAAgcgact**G**aaaaatatcccccatggCCTTTGGAACTTCAATATCCTCTT-3' 3'-GTACATTTACCGTGGTCGTTcgctga**C**tttttatagggggtaccGGAAACCTTGAAGTTATAGGAGAA-5'

* Nucleotide and amino acid numbering are based on DICER1 reference sequence [GenBank:NM_177438].
^ Priming sequences for PCR amplification (using primers listed in Table 2) are in capital letters. The nucleotides representing the mutations are in bold.

The highest frequency of *DICER1* mutations were found in Sertoli-Leydig cell tumours, where 26 of 43 (60%) contained a somatic variant within one of four hot-spot codons. All four codons encode for acidic amino acids acting as metal binding sites within the RNase IIIb domain of DICER1. Mutations affecting any of these residues resulted in reduced RNase IIIb activity.

Additional analysis of other tumour types identified a somatic *DICER1* hotspot mutation in one of 14 TGCT samples, raising the possibility of mutations within this domain of DICER1 also playing a role in TGCT development. To better estimate the frequency of somatic variants within these regions in TGCTs, we have analysed 96 TGCT samples using High Resolution Melting Curve analysis, a powerful method widely used for identifying variants in genomic DNA [11, 12].

Table 2. PCR amplification primers used in this study.

Genomic region covered*	Forward primer (5'-3')	Reverse Primer (5'-3')	Product size	Comment^
chr14:95560431-95560500	GGTGCTTGGTTATGAG-GTAGTCC	CCCTCAGATTGTTACC-AGCG	70 bp	Product includes codons 1705-1709 of DICER1; primer sequences from this study
chr14:95557604-95557671	CATGTAAATGGCACCA-GCAA	AAGAGGATATTGAAGT-TCCAAAGG	68 bp	Product includes codons 1810-1813 of DICER1; primer sequences from this study
chr14:95560345-95560533	CTTCTGCACAAGCTTAC-GGTTCCA	CAGCGATGCAAAGATG-GTGTTGT	188 bp	Product includes codons 1705-1709 of DICER1; primer sequences from [10].
chr14:95557565-95557759	TGGACTGCCTGTA-AAAGTGG	ACACACCTGCCAGACT-GTCTCC	194 bp	Product includes codons 1810-1813 of DICER1; primer sequences from [10].

* Genomic location is based on reference sequence hg19.
^ Amino acid numbering is based on DICER1 reference sequence [GenBank:NM_177438].

Results

We have used HRM analysis to screen 96 TGCT samples for sequence variants in the four mutation hot-spots codons identified in the RNase IIIb domain of DICER1. To first demonstrate that HRM was able to detect the specific variants previously identified in DICER1, we created six different DNA templates (Table 1), each containing one of the DICER1 mutations described in [10] that were identified in more than one sample. Combined, these six mutations cover 79% (26/33) of all cases where a mutation within one of the hot-spot codons was identified. For each variant, a heterozygous mutation was simulated by combining the variant template

with an equimolar amount of a DNA template containing the reference sequence. As shown in Figure 1, all six variants could clearly be identified using HRM analysis.

We then used PCR primers (Table 2) to amplify the corresponding genomic regions in 96 TGCT samples to screen for these six variants. No samples showed an aberrant melting curve. As this initial analysis only covered a small amount of genomic sequence (70 bp and 68 bp for the 1705/1709 and 1810/1813 codons respectively), we rescreened the same 96 TGCT samples using primers described in [10] (Table 2). These reactions generated products of 188 bp and 194 bp, covering more of the RNase IIIb domain. In these expanded assays, only one sample (a seminoma) showed an aberrant curve with either primer pair (Figure 2A). Sanger sequencing revealed a G>A transition (Figure 2B), predicted to change an Arginine to a Glutamine at position 1725 (Figure 1C). This variant is not listed in the 1000 Genome Project data [13], nor is it present in Catalogue of Somatic Mutations in Cancer (COSMIC), a database that curates mutations from a range of different cancers [14].

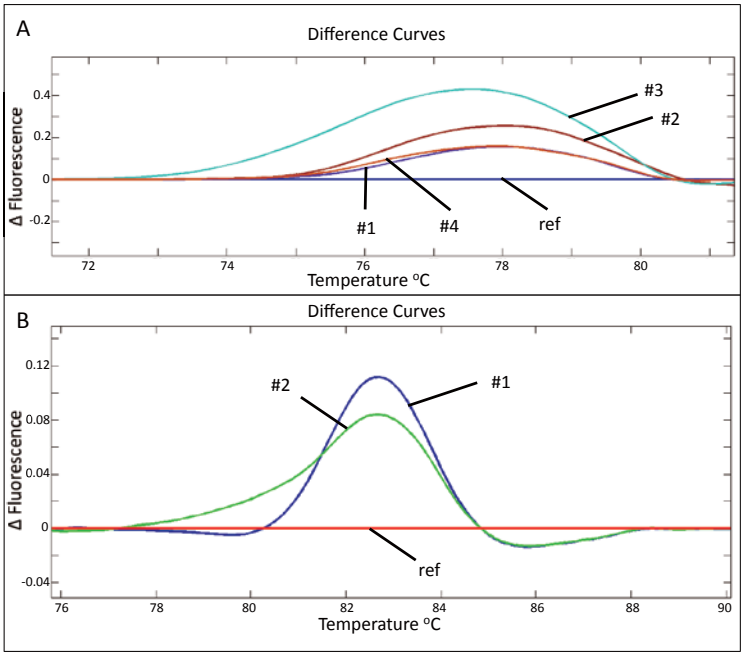


Figure 1. HRM analysis detection of previously identified mutations within the DICER1 RNase IIIb domain. Nucleotide and amino acid numbering are based on DICER1 reference sequence [NCBI:NM_177438].

1A. Aberrant HRM curves resulting from four different heterozygous mutations, affecting amino acids 1705 and 1709 of DICER1. #1 = c.5113G>A (p.E1705K); #2 = c.5125G>A (p.D1709N); #3 = c.5126A>G(p.D1709G); #4 = c.5127T>A (p.D1709E); Ref = reference sequence.

1B. Aberrant HRM curves resulting from two different heterozygous mutations, affecting amino acids 1810 and 1813 of DICER1. #1 = c.5428G>T (p.D1810Y); #2 = c.5437G>C (E1813Q); Ref = reference sequence.

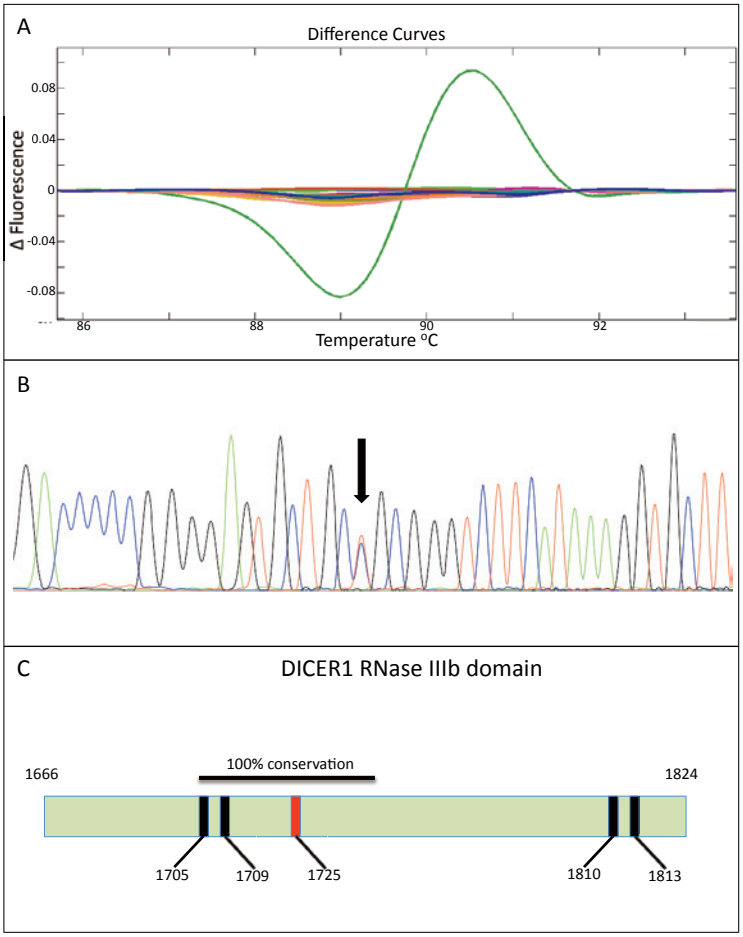


Figure 2. A novel mutation identified within the DICER1 RNase IIIb domain in a single seminoma sample. Amino acid numbering is based on DICER1 reference sequence [GenBank:NM_177438].

2A. The aberrant HRM curve corresponding to the *DICER1* mutation. The curves on the baseline represent samples without a sequence variant (this was confirmed with Sanger sequencing in one sample).

2B. A Sanger sequencing trace from the sample that showed the aberrant HRM curve in 2A. The heterozygous sequence variant is indicated by an arrow. This mutation is predicted to change an Arginine to a Glutamine at position 1725 of DICER1.

2C. The RNase IIIb domain in DICER1, containing amino acids 1666-1824. The location of the mutation (R1725Q) identified in this study is shown by the red bar. The locations of the four metal-binding residues (all acidic amino acids) frequently mutated in Sertoli-Leydig cell tumours are shown by black bars. The region of 100% conservation across at least 42 species at the amino acid level (residues 1705-1741) is indicated by the horizontal black bar.

Although the affected amino acid is not acidic, and not predicted to directly function as a metal-binding site, it is within a contiguous sequence of 36 amino acids that show 100% conservation across at least 42 species. This supports the hypothesis that this region within the RNase IIIb domain is critical for normal DICER1 function. Indeed, analysis using PolyPhen2

[15] predicts the impact of this variant to be "probably damaging" (score 1.0, sensitivity 0.0, specificity 1.0).

Discussion

Somatic sequence variants are rare in TGCTs. Analysis of 518 kinase genes in seven seminoma and six non-seminoma samples identified a single somatic point mutation, with an estimated mutation frequency of 0.12 per Mb [16]. A small number of genes are recurrently mutated in TGCT, including *KIT*, *KRAS2* and *BRAF* [17, 18], but in each case these represent <10% of tumours analysed.

Although Heravi-Moussavi *et al.* found *DICER1* mutations in non-epithelial tumours, including a TGCT, >85% of these mutations were restricted to Sertoli-Leydig cell tumours of the ovary. Sertoli-Leydig tumours are composed of both Sertoli and Leydig cells, which are cell types normally found in the testis [19]. They are derived from the sex cords, which originate from the gonadal ridge prior to sex determination. In contrast, seminomas and non-seminomas are derived from germ cells that have not undergone appropriate maturation. As such, these two tumour types represent distinct cell lineages.

The mutations in the RNaseIIIb domain were shown to alter rather than abolish DICER1 activity, and it was proposed that a specific miRNA expression profile would be derived as a result of changes in RNase IIIb activity [10]. The RNase IIIb domain is known to cut the miRNA strand, with the miRNA* strand being cut by the RNase IIIa domain [20]. Impaired RNase IIIb activity would therefore be expected to result in a relative increase in miRNA* production.

Several studies have shown increased expression of certain miRNAs in TGCTs, primarily the miR302 and miR37-373 clusters [21, 22]. Indeed, expression of these miRNAs is a hallmark of these cancers. These miRNAs are also upregulated in embryonic stem cells and other pluripotent cell types [23], and their high expression levels in TGCTs are thought to represent the pluripotent cell type of origin rather than specific genetic mutations affecting expression levels [24]. We have recently performed copy number variation and sequence analysis of these miRNA loci in a large cohort of TGCTs, and did not identify any mutations likely to be responsible for altering miRNA levels (unpublished observations, de Boer *et al.*).

Conclusion

In conclusion, we show that previously described hot-spot mutations within the RNase IIIb domain of DICER1 are not frequent in TGCTs. It is likely that the high rate of recurrent mutations observed in Sertoli-Leydig cell tumours is due to specific characteristics of this tumour type.

Materials and Methods

TGCT Samples

All TGCT DNA samples were extracted at Erasmus University Medical Center, Rotterdam, the Netherlands. The 96 samples consisted of 32 primary seminomas and 64 primary non-seminomas. None had been treated by either chemotherapy or irradiation. Use of tissue samples for scientific reasons was approved by an institutional review board (MEC 02.981 and CCR2041). Samples were used according to the "Code for Proper Secondary Use of Human Tissue in The Netherlands" as developed by the Dutch Federation of Medical Scientific Societies (FMWV (Version 2002, update 2011). Genomic DNA was isolated from peripheral blood lymphocytes following standard protocols.
This project was approved by the Monash University Human Research Ethics Committee, #CF11/1841.

Mutation analysis

All oligonucleotides were ordered from Sigma-Aldrich (Castle Hill, Australia). Double-stranded DNA sequences containing either the reference sequence, or one of the six most commonly identified variants in [10], were generated by annealing two complementary single-stranded oligonucleotides of the appropriate sequences (Table 1).
Primers for PCR amplification are listed in table 2. Amplification reactions were performed in 10 µl reaction volumes, consisting of HRM Master Mix (Idaho Technologies, USA), 5 µM each of forward and reverse primer, and either 25 ng genomic DNA (TGCT samples) or 1 fmol of variant DNA template.
PCR reactions were carried out under the following conditions; an initial hold at 95 °C for 2 min, followed by 45 cycles of 94 °C for 30 sec and 58 °C for 30 sec. PCR products were analysed in a 96 well plate in the LightScanner (Idaho Technologies, USA). The HRM settings for the LightScanner were as follows; start temperature of 70 °C, end temperature at 96 °C, with a hold temperature at 67 °C. HRM curves were normalized using GeneMelt software supplied with the instrument, and samples that showed an aberrant melting curve were analysed by Sanger sequencing at the Gandel Charitable Trust Sequencing Centre at the Monash Health Translation Precinct, Melbourne, Australia. Sequence traces were visually assessed using 4Peaks software (Mek&Tosj, Amsterdam, the Netherlands).

Bioinformatic analysis.

The predicted effect of missense mutations was determined using PolyPhen2 [15].

Competing interests.

The authors declare that they have no competing interests.

Authors' contributions.

SdB performed the genetic studies and helped draft the manuscript. RE, AG, and HS performed the molecular analysis and preparation of the samples. LL conceived of the study, participated in its design and coordination and helped draft the manuscript. SW conceived of the study, participated in its design and coordination and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by funding from Monash University, the Marian and E.H. Flack Trust, and the Victorian Government's Operational Infrastructure Support Program. R.E. is supported by Dutch Cancer Society Grant 2007-3081.

References

1. Oosterhuis, J.W. and L.H. Looijenga, *Testicular germ-cell tumours in a broader perspective.* Nat Rev Cancer, 2005. **5**(3): p. 210-22.
2. Looijenga, L.H., et al., *Dissecting the molecular pathways of (testicular) germ cell tumour pathogenesis; from initiation to treatment-resistance.* Int J Androl, 2011. **34**(4 Pt 2): p. e234-51.
3. Honecker, F., et al., *Pathobiological implications of the expression of markers of testicular carcinoma in situ by fetal germ cells.* J Pathol, 2004. **203**(3): p. 849-57.
4. de Jong, J., et al., *Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.* J Pathol, 2008. **215**(1): p. 21-30.
5. Hart, A.H., et al., *The pluripotency homeobox gene NANOG is expressed in human germ cell tumors.* Cancer, 2005. **104**(10): p. 2092-8.
6. van de Geijn, G.J., R. Hersmus, and L.H. Looijenga, *Recent developments in testicular germ cell tumor research.* Birth Defects Res C Embryo Today, 2009. **87**(1): p. 96-113.
7. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
8. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
9. Hill, D.A., et al., *DICER1 mutations in familial pleuropulmonary blastoma.* Science, 2009. **325**(5943): p. 965.
10. Heravi-Moussavi, A., et al., *Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers.* N Engl J Med, 2012. **366**(3): p. 234-42.
11. Aten, E., et al., *Methods to detect CNVs in the human genome.* Cytogenet Genome Res, 2008. **123**(1-4): p. 313-21.
12. Vossen, R.H., et al., *High-resolution melting analysis (HRMA): more than just sequence variant screening.* Hum Mutat, 2009. **30**(6): p. 860-6.
13. *1000 Genomes*. Available from: http://www.1000genomes.org/.
14. *Catalogue of Somatic Mutations in Cancer*. Available from: http://www.sanger.ac.uk/genetics/CGP/cosmic/.
15. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
16. Bignell, G., et al., *Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults.* Genes Chromosomes Cancer, 2006. **45**(1): p. 42-6.
17. Sakuma, Y., et al., *Alterations of the c-kit gene in testicular germ cell tumors.* Cancer Sci, 2003. **94**(6): p. 486-91.
18. Sommerer, F., et al., *Mutations of BRAF and RAS are rare events in germ cell tumours.* Int J Cancer, 2005. **113**(2): p. 329-35.
19. Young, R.H., *Sex cord-stromal tumors of the ovary and testis: their similarities and differences with consideration of selected problems.* Mod Pathol, 2005. **18 Suppl 2**: p. S81-98.
20. Zhang, H., et al., *Single processing center models for human Dicer and bacterial RNase III.* Cell, 2004. **118**(1): p. 57-68.
21. Palmer, R.D., et al., *Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets.* Cancer Res, 2010. **70**(7): p. 2911-23.
22. Gillis, A.J., et al., *High-throughput microRNAome analysis in human germ cell tumours.* J Pathol, 2007. **213**(3): p. 319-28.
23. Suh, M.R., et al., *Human embryonic stem cells express a unique set of microRNAs.* Dev Biol, 2004. **270**(2): p. 488-98.
24. Looijenga, L.H., et al., *Relevance of microRNAs in normal and malignant development, including human testicular germ cell tumours.* Int J Androl, 2007. **30**(4): p. 304-14; discussion 314-5.

CHAPTER 5

Mutation screening of the miR-371-3 and miR-302 clusters in testicular germ cell tumours

C.M. de Boer, A.J. Notini, R. Eini, A.J.M. Gillis, J. Jongenotter, L.C.J. Dorssers, L.H.J. Looijenga, S.J. White

Center for Reproduction and Development, Monash Institute for Medical Research, Melbourne, Australia

Department of Pathology, Erasmus MC, University Medical Center Rotterdam, Josephine Nefkens Institute, The Netherlands

In preparation for submission

Abstract

MicroRNAs (miRNAs) are endogenous small RNAs, which regulate basic cellular processes by interfering with the process of translation. There is considerable evidence that expression of miRNAs can be deregulated in human cancer, and previous studies have shown that the miR302(a-e)/367 and miR371~373 clusters are highly expressed in human germ cell tumours, i.e., seminomas and nonseminomas. We have used multiplex ligation-dependent probe amplification (MLPA) and high resolution melting (HRM) curve analysis to screen 242 testis germ cell tumour-derived DNA samples for copy number variants (CNVs) and sequence mutations affecting these two miRNA clusters, encompassing eight miRNAs in total. A 4 bp deletion within mir302c was found in ~6% of the cases, disturbing MLPA probe ligation. This was found to be represented in the general population in the same frequency. No other CNVs were identified. Two previously unreported sequence variants were identified in miR367 with HRM analysis, although neither was likely to be pathogenic.

Our findings demonstrate that the increased expression levels of these miRNAs are unrelated to changes in copy number of the corresponding chromosomal loci. Instead, the high expression levels are intrinsic to the embryonic origin of the cancer.

Introduction

Testicular germ cell tumours of adolescents and young adults (TGCTs) are the most common malignancy affecting males between the ages of 15 and 45 years, and are associated with significant morbidity [1]. The incidence of TGCTs has increased in the past 40 years in Caucasian populations [2], although the reason for this is unknown. It is most likely to be an interaction between genetic and environmental factors, which we refer to as Genvironment. The two main types of TGCTs are seminomas (SE) and non-seminomas (NS), which are classified as Type II TGCTs [1]. Both SE and NS originate from a common precursor, referred to as carcinoma *in situ* (CIS), representing embryonic germ cells that have failed to mature and undergo mitotic arrest [3, 4]. SE are transformed embryonic germ cells that express NANOG, OCT3/4, as well as SOX17. The stem cell component of NS, known as embryonal carcinoma, expresses SOX2 in addition to NANOG and OCT3/4, and can give rise to diverse histological components, i.e., yolk sac tumour, choriocarcinoma and teratoma [5, 6].

MicroRNAs (miRNAs) are short noncoding RNAs involved in the regulation of gene expression[7]. The initial pri-miRNA is cleaved in the nucleus by an RNase III endonuclease, Drosha, to produce a short, pre-miRNA of 60-70 nt in length [8]. This pre-miRNAs fold in a stemloop structure that can contain multiple mismatches. The pre-miRNA is transported to the cytoplasm, where it is processed by Dicer to produce a mature miRNA of ~22nt in length. This mature form can interact with multiple (partially) complementary sequences at the 3' untranslated region of mRNAs, and in association with a protein complex it can either inhibit translation or promote degradation of the mRNA molecule [9, 10].

A number of miRNAs are highly expressed in pluripotent cells [11]. Of these, the miR-371~373 and miR302/367 clusters have been shown to be expressed at a high level in TGCTs [12, 13], and elevated levels of these miRNAs have been identifed in the serum of patients with malignant germ cell tumours [14, 15]. Expression levels of these miRNAs in TGCTs have been used to specifically segregate malignant from non-malignant tumours, suggesting that they may have diagnostic utility as biomarkers [13].

Increased levels of miR371-373 in TGCTs have been shown to substitute for mutated p53, by inhibiting cellular senescence normally induced by oncogenic stimuli [16, 17]. Overexpressing the miR302 cluster has been shown to reprogram differentiated cells into induced pluripotent stem cells, with all the characteristics of pluripotency [18]. The high level of expression of these miRNAs in TGCTs may therefore represent the cause (reactivating stem cell-like potential), or the consequence of the pluripotent phenotype of the cell of origin.

Increased expression levels can be due to changes in copy number of the miRNA locus, or sequence changes that influence the transcription process or the transcript stability. Array comparative genomic hybridization analysis of TGCT cell lines found little correlation between copy number and miRNA expression levels [12], and a CNV study of seminoma samples did not reveal consistent amplification of the genomic regions containing either of the

miR302/367 or miR371~373 clusters [19]. This contrasts with findings in epithelial cancers (e.g. ovarian cancer, breast cancer and melanoma) in which copy number alteration affecting miRNA loci were highly prevalent, and in many cases specific gains or losses correlated with changes in miRNA expression [20].

The miR-371~373 and miR-302/367 clusters are located on chromosomes 19 and 4 respectively, and each cluster is <2 kb (Figure 1). The relatively low resolution of the microarrays used in previous studies of TGCTs (~100 kb), compared to the size of the miRNA clusters, imply that microduplications affecting these loci could easily have been missed. Several methods for identifying copy number changes at high resolution have been described [21]. One powerful approach is multiplex ligation-dependent probe amplification (MLPA), which allows analysis of up to 40 loci in a single PCR amplification [22]. A particularly attractive feature of this approach is the ability to differentiate highly similar sequences [23], which is important when studying miRNAs sharing significant sequence homology.

A sensitive, cost effective and high throughput method for identifying variants in DNA is high resolution melting (HRM) curve analysis [24]. It has been applied to the analysis of many genes, and has been shown to be capable of detecting variants in as few as 10% of cells [21]. We have used both MLPA and HRM analysis to screen for copy number changes and sequence variants that may be causative for the high miRNA levels in TGCTs.

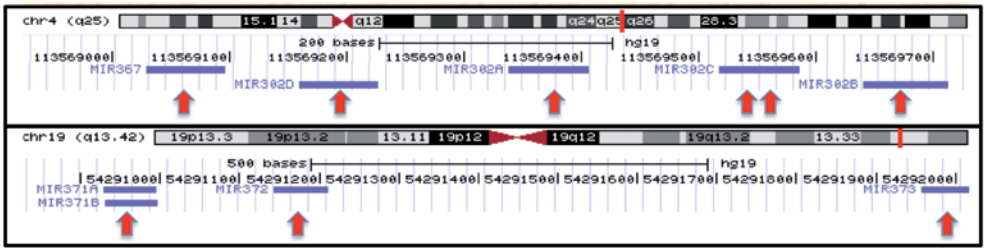


Figure 1. The two miRNA clusters. The locations of the MLPA probes are indicated by arrows.

Methods

Samples

All TGCT DNA samples were extracted at Erasmus University Medical Center, Rotterdam, the Netherlands. Use of tissue samples for scientific reasons was approved by an institutional review board (MEC 02.981 and CCR2041). Samples were used according to the "Code for Proper Secondary Use of Human Tissue in The Netherlands" as developed by the Dutch Federation of Medical Scientific Societies (FMWV (Version 2002, update 2011). This project was approved by the Monash University Human Research Ethics Committee, #CF11/1841.

MLPA

MLPA probes were designed for each of the eight miRNA loci, with each probe producing a product of a different length (Table 1). Oligonucleotides were purchased from Sigma Aldrich (Australia). Two control probes on different chromosomes were included for normalization. These probes have previously been shown to give a low standard deviation in MLPA analysis [25], and are in loci that do not normally show evidence for CNV in TGCT samples.
The MLPA reaction was performed essentially as described [26], and fragment analysis was performed on the Applied Biosystems 3130xl Genetic Analyzer at the Gandel Charitable Trust Sequencing Centre, Clayton, Australia. Data analysis was carried out as described [27], by comparing the peak height of each miRNA-specific probe with the two control probes. Thresholds for deletions and amplifications were set at 0.75 and 1.25 respectively. Any sample that showed evidence of a copy number change at one or more loci was tested at least twice.

Table 1. The MLPA probe sequences for the eight miRNAs analysed. The left-hand (LH) oligonucleotides all had the following sequence added to the 5' end; 5'-GGGTTCCCTAAGGGTTGGA-3'. The right-hand (RH) oligonucleotides all had the following sequence at the 3' end; 5'-TCTAGATTGGATCTTGCTGGC-3'. The RH oligos had a 5' phosphorylation modification to facilitate ligation.

miRNA	LH oligo (5'-3')	RH oligo (5'-3')
mir302a	GAAGCACTTACTTCTTTAGTTTCAAAGCA AGTACATCCACG	TTTAAGTGGTGGGGAGCCCAGTCTTGG
mir302b	CACAGAAAGCACTTCCATGTTAAAGTTGA AGG	GAGCCCACCCAACATACAACTTCTTTGGA CTT
mir302c #1	GCTGGCTTGGAGACACCTCCACTGAAACA TGGAAGCACTTACTT	TTGTTTCACACAGCAGGTACCCCCATGTT AAAGCAAAGGGGATC
mir302c #2	CGCACACAGCAGGTACCCCCATGTTAAAG CAAAG	GGGATCCCTTCAAATGAGGTTAGCGTGTT CTATTTTGGAG
mir302d	GGTCATGTCACAGCAAGTGCCTCCATGTT AAAG	TAGAGGGGGCCCCTTAACAGATGTAAA
mir367	CATATTAGCAACAGTAATGGCCTGTA	GCCAAGAACTGCACACAGTGTGGGCG
mir371	GGGCACTTTCTGCTCTCTGGTGAAAGTGC CGCCATCTTTTGA	GTGTTACCGCTTGAGAAGACTCAACCTGC GGAGAAGATACCA
mir372	GTGGAGCACTATTCTGATGTCCAAGTG	
mir373	CTGTACTGGGAAGTGCTTCGATTTTGGGG TGTCCCTGTTT	GAGTAGGGCATCACGAACCATCCTGCTTC AAGGGAGCCTG

Sequence variant analysis.

We used HRM Master Mix (TrendBio) for the PCR followed by HRM analysis performed on the LightScanner System (Idaho Technology Inc.). Primer pairs were designed for the eight miRNA loci, and were designed to generate products <200 bp (Table 2).
PCR amplifications were performed in 10 µl reaction volumes, consisting of HRM Master Mix (Idaho Technologies), 5 µM each of forward and reverse primer, and 25 ng genomic DNA.

The PCR reaction was carried out under the following conditions; an initial hold at 95°C for 2 min, followed by 45 cycles of 94 °C for 30 sec and 58 °C for 30 sec. PCR products were analysed in a 96 well plate in the LightScanner. The HRM settings for the LightScanner were as follows; start temperature of 70 °C, end temperature at 96 °C, with a hold temperature at 67 °C. HRM curves were normalized using GeneMelt software supplied with the instrument, and samples that showed an aberrant melting curve were analysed by Sanger sequencing at the Gandel Charitable Trust Sequencing Centre at the Monash Health Translation Precinct, Melbourne, Australia.

Specific screening for the 28 bp tandem duplication identified in miR367 was performed by PCR analysis using two primers (Table 2), followed by agarose gel electrophoresis.

Table 2. The PCR primer sequences used for sequence variant identification. All primers listed were used for HRM analysis, with the exception on miR367#2, which was used for agarose gel screening.

miRNA	Forward	Reverse
miR302a	TGCGGTCAATACAATAAAGTTATTTTCTA	TTCCAAGACTGGGCTCC
miR302b	ATGAGGTTAGCGTGTTCTAT	TCTGAAGTCCAAAGAAGTTGTATG
miR302c	AAAGGTGTGCTGGCTTG	AGAACACGCTAACCTCATTT
miR302d	TAACGCAATTGCTGATTAGGT	GCAGCTCATATATTTAAGCTTTATTTTGTA
miR367	GGTTTAAATTCTGTCATTGGCTT	ACACTGTGTGCAGTTCTT
miR371	GTGCTTCCACTTGCGAT	TTCTCCGCAGGTTGAGTC
miR372	TGATATAAATTTCTTGGCCGGG	ATCCGTTGATATGGGCG
miR373	GTCACAGTGATGGCAGAT	CCTTGAAGCAGGATGGTT
miR367#2	GCCCCATGTCTTAACGGAGAGCT	GGTTCCTACCTAATCAGC

Results

We have used MLPA and HRM to look for genetic variants affecting eight miRNA sequences in 242 tumour-derived DNA samples from TGCT patients. This cohort consisted of 121 seminoma and 121 non-seminoma samples.

The only locus that showed any evidence for specific gain or loss compared to control loci was mir302c, where a putative deletion was seen in 14/242 TGCT samples. An independent MLPA analysis using a different miR302c probe ~100 bp further away showed no evidence of this CNV. Sanger sequence analysis that covered the ligation site of the first miR302c MLPA probe showed the same 4 bp deletion (NR_029858; n.40_43del) in all samples with the apparent deletion. This deletion was located within 4 bp of the ligation site, and disrupted probe ligation and hybridisation sufficiently for an apparent loss of a single allele to be observed.

This variant is not listed in the most recent version of dbSNP (dbSNP 135). Analysis of data from 445 blood-derived DNA samples that had been analysed by whole exome sequencing showed that ~5% were heterozygous for this variant.

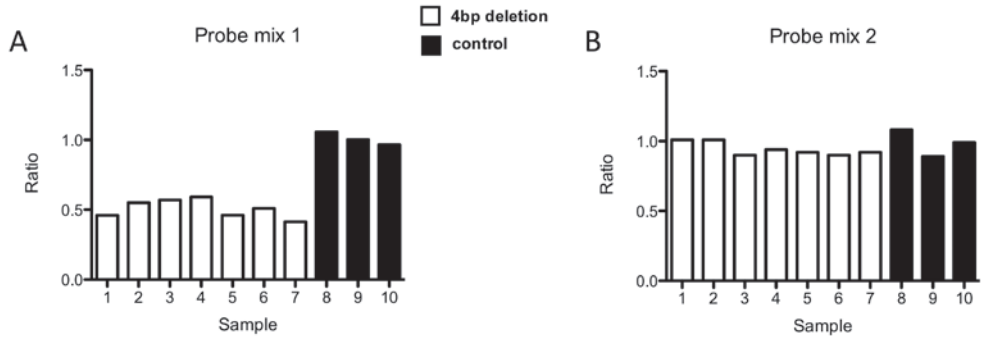


Figure 2. MLPA results for miR302c.

Analysis with a MLPA probe ligating at either A) the site of the 4 bp deletion found in ~6% of samples, or B) 100 bp outside of this region.

These samples were chosen for sequencing due to congenital conditions such as mental retardation, and are not be expected to be enriched for TGCT susceptibility alleles. All variants identified with HRM analysis are listed in Table 3. One variant in dbSNP was seen, with rs114318553 in miR302b being found in four samples. The 4 bp deletion in miR302c seen with MLPA analysis was also readily detectable with HRM (Figure 3). No other variants in miR302c were seen. Two other previously undescribed variants were identified, both in miR367. The first was shown by Sanger sequencing to be a 28 bp tandem duplication (NR_029860; n.57_+16dup). PCR and agarose gel analysis of a further 166 extra TGCT DNA samples did not identify any with the 28 bp duplication. A novel single nucleotide variant was found in miR 367 in a separate sample, at position 11 of the mature transcript (NR_029860; n.54T>C). This nucleotide is highly conserved, but is not predicted to affect the stem loop structure as it does not normally pair with another base.

Table 3. All sequence variants affecting pre-miRNA sequences that were identified in this study. Sequence changes are described relative to the following reference sequences: miR302b, NR_029857; miR302c, NR_029858; miR367, NR_029860.

miRNA	SNP ID	Variant	Number of samples
miR302b	rs114318553	n.19A>C	4
miR302c	n.a.	n.40_43del	14
miR367	n.a.	n.54T>C	1
miR367	n.a.	n.57_+16dup	1

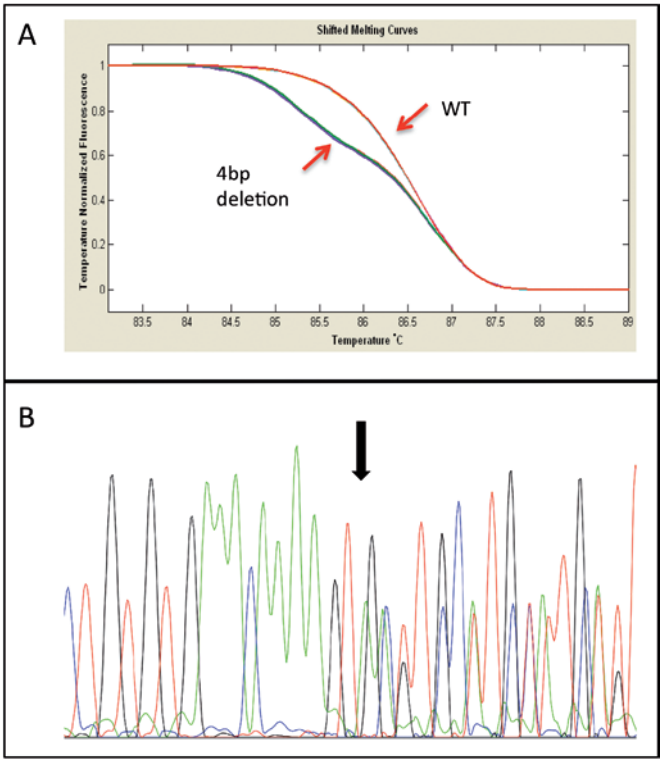


Figure 3. Sequence analysis of miR302c.

A) HRM curves for samples with or without the 4 bp deletion. B) A Sanger sequence trace of a PCR product containing the 4 bp deletion. The arrow indicates the start of the deletion.

Discussion

The miRNAs studied in this report are relatively highly expressed in cells of pluripotent origin such as ES cells, as well as in the non-differentiated components of TGCTs. The expression level is determined by the lineage of differentiation, although differences have been reported [12, 13, 16]. It remains to be answered whether the high level of expression of these miRNAs in TGCTs is due to persistence of expression from the pluripotent cell of origin, or reactivation of expression as a consequence of malignant transformation. Genetic variants that lead to increased levels of these miRNAs, such as duplications or sequence variants affecting regulation or stability, could therefore play a causative role in TGCT development and/or progression.

Germ line variants affecting miRNA loci are rare. An analysis of well validated CNV regions showed that miRNA loci are underrepresented [28], and SNPs are also significantly less common [29]. Sequence analysis of miRNAs in cancer samples has also identified few

mutations. A study of 15 miRNAs in 91 cancer-derived cell lines identified several variants affecting either the precursor or primary miRNA transcripts, but no variants were seen in the mature sequences [30]. A similar pattern was seen in a study of 10 miRNAs in ovarian cancer, where HRM analysis did not identify any somatic variants affecting the mature miRNA sequences [31]. Conversely, copy number changes affecting miRNA loci in cancers are known to occur frequently, and often correlate with changes in miRNA expression [20, 32].

To identify any genetic changes affecting miRNA expression levels we have used MLPA and HRM analysis. Applying these approaches we have identified three previously undescribed sequence variants affecting the miRNA sequence. The first was a 4 bp deletion in mir-302c in 14 out of 242 TGCT samples (6%), seen with both MLPA and HRM. The MLPA result was consistent with the loss of one allele i.e. a deletion, but a second MLPA probe designed within 100 bp of the first mir302c MLPA probe did not show any evidence for a larger deletion.

Inspection of data from whole exome sequencing analysis shows that this variant is present in ~5% of the general population. Given this population frequency it is highly unlikely that this variant has a pathological effect. This finding does however highlight the need to confirm apparent deletions identified with a single MLPA probe. Previous reports have described how sequence variants at or near the MLPA probe ligation site can decrease the peak height sufficiently to appear as a mosaic or complete deletion [33, 34].

The second variant that we identified was a tandem duplication of 28 bp in mir367 in a single case. This variant was not identified in DNA from 407 other TGCT cases, nor is it present in any sequence variant databases. Any effect is likely to decrease the amount of mature miR367 transcript, as it is probable that processing would be disturbed.

The third variant was a single base pair variant in miR367, located at position 11 in the mature miRNA transcript. Although this variant shows a high degree of conservation across species, it is located outside the 2-8 seed sequence, which is considered to be the most critical region for recognising a complementary sequence in the 3'UTR of target mRNAs [35]. In addition it is not predicted to base pair with another nucleotide in the stem loop. This variant is therefore unlikely to have any pathogenic effect.

In conclusion, we show that genetic variants affecting the miR-302 and miR-371~373 clusters are not major contributors to the increased miRNA expression levels observed in TGCTs. Rather, these findings are consistent with TGCTs being composed of a pluripotent cell type, reflecting the germ cell origin of these tumours.

Acknowledgements

We thank Joris Veltman and Christiaan Gilissen (Department of Human Genetics, University Medical Center Nijmegen The Netherlands) for their support in analyzing the presence of the different variants in the general population.

References

1. Oosterhuis, J.W. and L.H. Looijenga, *Testicular germ-cell tumours in a broader perspective.* Nat Rev Cancer, 2005. **5**(3): p. 210-22.
2. Walsh, T.J., et al., *Incidence of testicular germ cell cancers in U.S. children: SEER program experience 1973 to 2000.* Urology, 2006. **68**(2): p. 402-5; discussion 405.
3. Looijenga, L.H., et al., *Dissecting the molecular pathways of (testicular) germ cell tumour pathogenesis; from initiation to treatment-resistance.* Int J Androl, 2011. **34**(4 Pt 2): p. e234-51.
4. Honecker, F., et al., *Pathobiological implications of the expression of markers of testicular carcinoma in situ by fetal germ cells.* J Pathol, 2004. **203**(3): p. 849-57.
5. van de Geijn, G.J., R. Hersmus, and L.H. Looijenga, *Recent developments in testicular germ cell tumor research.* Birth Defects Res C Embryo Today, 2009. **87**(1): p. 96-113.
6. de Jong, J., et al., *Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.* J Pathol, 2008. **215**(1): p. 21-30.
7. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
8. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
9. Guo, H., et al., *Mammalian microRNAs predominantly act to decrease target mRNA levels.* Nature, 2010. **466**(7308): p. 835-40.
10. Pillai, R.S., et al., *Inhibition of translational initiation by Let-7 MicroRNA in human cells.* Science, 2005. **309**(5740): p. 1573-6.
11. Suh, M.R., et al., *Human embryonic stem cells express a unique set of microRNAs.* Dev Biol, 2004. **270**(2): p. 488-98.
12. Gillis, A.J., et al., *High-throughput microRNAome analysis in human germ cell tumours.* J Pathol, 2007. **213**(3): p. 319-28.
13. Palmer, R.D., et al., *Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets.* Cancer Res, 2010. **70**(7): p. 2911-23.
14. Murray, M.J., et al., *Identification of microRNAs From the miR-371~373 and miR-302 clusters as potential serum biomarkers of malignant germ cell tumors.* Am J Clin Pathol, 2011. **135**(1): p. 119-25.
15. Belge, G., et al., *Serum levels of microRNAs miR-371-3: a novel class of serum biomarkers for testicular germ cell tumors?* Eur Urol, 2012. **61**(5): p. 1068-9.
16. Voorhoeve, P.M., et al., *A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.* Cell, 2006. **124**(6): p. 1169-81.
17. Looijenga, L.H., et al., *Relevance of microRNAs in normal and malignant development, including human testicular germ cell tumours.* Int J Androl, 2007. **30**(4): p. 304-14; discussion 314-5.
18. Anokye-Danso, F., et al., *Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency.* Cell Stem Cell, 2011. **8**(4): p. 376-88.
19. LeBron, C., et al., *Genome-wide analysis of genetic alterations in testicular primary seminoma using high resolution single nucleotide polymorphism arrays.* Genomics, 2011. **97**(6): p. 341-9.
20. Zhang, L., et al., *microRNAs exhibit high frequency genomic alterations in human cancer.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 9136-41.
21. Aten, E., et al., *Methods to detect CNVs in the human genome.* Cytogenet Genome Res, 2008. **123**(1-4): p. 313-21.
22. Schouten, J.P., et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Res, 2002. **30**(12): p. e57.
23. White, S.J., et al., *Variation of CNV distribution in five different ethnic populations.* Cytogenet Genome Res, 2007. **118**(1): p. 19-30.
24. Vossen, R.H., et al., *High-resolution melting analysis (HRMA): more than just sequence variant screening.* Hum Mutat, 2009. **30**(6): p. 860-6.
25. Roelfsema, J.H., et al., *Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease.* Am J Hum Genet, 2005. **76**(4): p. 572-80.
26. White, S.J., et al., *Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses.* Hum Mutat, 2004. **24**(1): p. 86-92.

27. White, S.J., M.H. Breuning, and J.T. den Dunnen, *Detecting copy number changes in genomic DNA: MAPH and MLPA.* Methods Cell Biol, 2004. **75**: p. 751-68.
28. Marcinkowska, M., et al., *Copy number variation of microRNA genes in the human genome.* BMC Genomics, 2011. **12**: p. 183.
29. Saunders, M.A., H. Liang, and W.H. Li, *Human polymorphism at microRNAs and microRNA target sites.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3300-5.
30. Diederichs, S. and D.A. Haber, *Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing.* Cancer Res, 2006. **66**(12): p. 6097-104.
31. Bearfoot, J.L., et al., *Genetic analysis of cancer-implicated MicroRNA in ovarian cancer.* Clin Cancer Res, 2008. **14**(22): p. 7246-50.
32. Calin, G.A. and C.M. Croce, *MicroRNAs and chromosomal abnormalities in cancer cells.* Oncogene, 2006. **25**(46): p. 6202-10.
33. 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.
34. Janssen, B., et al., *MLPA analysis for the detection of deletions, duplications and complex rearrangements in the dystrophin gene: potential and pitfalls.* Neurogenetics, 2005. **6**(1): p. 29-35.
35. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.

CHAPTER 6

Targeted Serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients

Ad J.M. Gillis[1*], Martin A. Rijlaarsdam[1*], Ronak Eini[1], Lambert C.J. Dorssers[1], Katharina Biermann[1], Matthew J.Murray[2], James C. Nicholson[2], Nicholas Coleman[3], Klaus-Peter Dieckmann[4], Gazanfer Belge[5], Jörn Bullerdiek[5], Tom Xu[6], Nathalie Bernard[6] and Leendert H.J. Looijenga[1]

[*] Both authors contributed equally to the work

[1]Department of Pathology, Erasmus MC - University Medical Center Rotterdam, Rotterdam, The Netherlands [2]Department of Paediatric Haematology and Oncology, Addenbrooke's Hospital, and [3]Department of Pathology, Medical Research Council Cancer Cell Unit, University of Cambridge, Cambridge, United Kingdom, [4]Department of Urology, Albertinen-Krankenhaus, Hamburg, Germany, [5]Department of Human Genetics, University of Bremen, Germany, [6]Applied Biosystems, Europe and USA, part of Life Technologies Group, Foster City, United States of America

Erasmus MC, University Medical Center Rotterdam, Department of Pathology, The Netherlands

Submitted for publication

Abstract

Germ cell cancers (GCC) are the most frequent malignancy in young Caucasian males. GCC can consist of seminomas (SE) and non-seminomas (malignant NS: embryonal carcinoma (EC), yolk sac tumor (YS), choriocarcinoma (CH) and teratoma (TE)). Current serum-markers used for diagnosis and follow-up (AFP, hCG) are predominantly related to YS and CH and marker positivity can vary during the course of the disease. Therefore, stable markers consistently identifying more GCC components, specifically the stem cell components SE and EC, are of interest. Current reports indicate specific expression of the embryonic stem cell miR-371-3 and miR-302/367 clusters in SE/EC/YS, suggesting possible application of these micro-RNAs as GCC tumor-markers. The TSmiR protocol described in the manuscript constitutes a complete, quality-controlled pipeline for the detection of miRs in serum, based on magnetic bead-based purification and qPCR quantification. TSmiR was applied to five independent serum sample series including 80 GCC samples, 47 controls, 11 matched pre- and post orchidectomy samples and 12 no-GCC testicular masses. TSmiR demonstrated a consistent, significant ($p<0.0064$) increase of miR-371/2/3 and miR-367 levels in the GCC serum samples when compared to controls. Analogous, in patients treated for local disease, the serum levels of these miRs returned to baseline after orchidectomy. Moreover, there was a trend toward higher miR levels in patients with metastasis. These results imply suitability for follow-up. TSmiR showed an overall sensitivity of 98%, clearly outperforming the traditional serum markers AFP/hCG (36%/57%, $Sensitivity_{AFP}$= 3% and 45%; $sensitivity_{hCG}$= 62% and 66%, in SE and malignant NS respectively). The combined serum AFP/hCG and TSmiR test identified all T samples correctly while only applying TSmiR lead to one misclassification of a tumor sample as a control. In conclusion, TSmiR constitutes a highly sensitive and reproducible serum test for GCC patients, suitable to be prospectively tested for diagnostic and follow-up purposes.

Keywords: testicular cancer; germ cell cancer; serum markers; AFP; hCG; miRNA.

Introduction

Human germ cell tumors are a heterogeneous group of neoplasms, derived from the germ cell lineage and originating at various stages of development with different cells of origin and pathogenesis [1]. The proposed classification system has been adopted by the World Health Organization (WHO) [2], and specialized pathologists [3,4]. Malignant germ cell tumors, referred to as germ cell cancer (GCC), include seminoma (SE) and nonseminoma (NS) and are predominantly found in adolescents and young adults, although also diagnosed in neonates and infants [5]. It is the most frequent malignancy in young Caucasian males and incidence is increasing. GCC originate from a pluripotent embryonic germ cell (primordial germ cell/gonocyte) blocked in its maturation. This is referred to as carcinoma *in situ* (CIS) [6] or intratubular germ cell neoplasia unclassified (IGCNU) [2] in the context of testicular GCC of adolescents and adults. Their pluripotency is reflected in the capacity to form the germ cell lineage (CIS/IGCNU, SE, *de novo* germ cells [7]), embryonic stem cell components (embryonal carcinoma, EC) and all differentiation lineages (teratoma (TE), yolk sac tumor (YS) and choriocarcinoma (CH)) as found in adolescents and adults. In neonates and infants, TE and YS are distinguished, in which YS is the malignant component. Pluripotent GCC (including their precursor lesions) exhibited expression of various embryonic pluripotency markers of significant diagnostic value (OCT3/4 (POU5F1) [8], NANOG [9], SOX2/SOX17 [10] and LIN28 [11,12]). In addition, CIS/SE/EC/YS components, diagnosed at pediatric and adult age, express a specific set of embryonic stem cell related microRNAs (miRs), including the miR-371-372-373 (miR-371-3) and miR302a,b,c,d/367 (miR-302/367) clusters [13-16] (Figure 1, Figure S1).

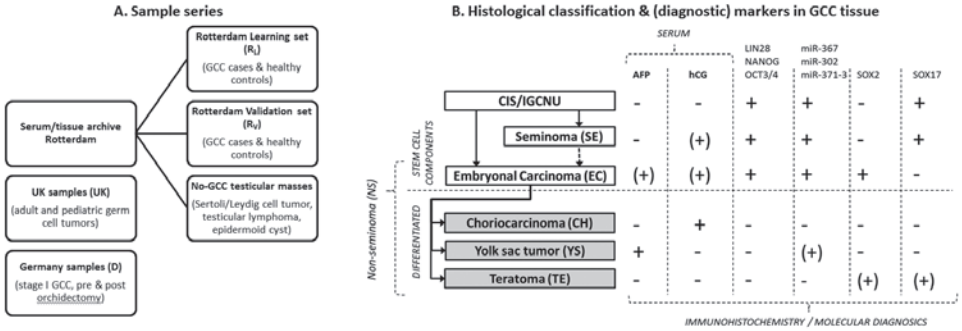


Figure 1. Sample series & histological classification / markers.

(A) Sample series investigated. Five independent series were investigated: the Rotterdam-Learning (R_L) and Validation (R_V) sets including GCC cases and controls, a series of serum samples of patients with a testicular tumor other than GCC (no-GCC) and two previously published series: UK, including GCC samples (pediatric and adult) and controls [21,23] and Germany (D), composed of pre- and post orchidectomy samples (local disease only, stage I) [22]. (B) Histological classification & overview of immunohistochemical, molecular and serum GCC markers [8-16]. GCC originates from a common progenitor, further developing into the stem cell components SE and EC. EC can further differentiate into the various types of differentiated NS. Markers are summarized as described throughout the main text. + = consistently positive; - = consistently negative; (+) = sporadic positivity (inconsistent).

Depending on tumor stage and histology, pediatric and adult GCC (SE/EC/YS/CH) are highly treatable by surgery, possibly followed by either irradiation and/or chemotherapy [17]. However, significant long term effects of these treatment protocols have become evident [18,19]. Therefore, early and accurate diagnosis as well as detailed follow-up of GCC patients is of crucial relevance for optimal treatment preventing possible under- or overtreatment. In clinical management of GCC, evaluation of a set of serum markers is informative, both for diagnosis, risk assessment, as well as follow-up: alpha feto-protein (AFP; predominantly informative for YS, sporadic positivity in EC) and human Chorionic Gonadotrophine (hCG; predominantly informative for CH, sporadic positivity in SE/EC) [17]. LDH-1 has also been reported as a (less informative) serum marker [20]. So far, no consistent markers for SE and EC are available, which limits the use for diagnosis/follow-up in a large proportion of GCC patients.

Recently, three studies reported that a higher expression of members of the embryonic miR clusters miR-371-3 and miR-302/367 can be detected in serum of GCC patients (adult and pediatric) as compared to controls [21-23]. One study showed a decline to normal levels after surgical removal of stage I GCC [22]. These findings could be a significant step forwards in clinical management (diagnosis and follow-up) of GCC patients, especially for the high number of "marker-negative" cases, i.e., those without elevated AFP or hCG serum levels.

Here, we describe a stringently controlled assay (TSmiR) for the detection of these miRs in serum, including magnetic bead-based purification for miR recovery. A number of quality control steps was implemented to ensure reproducible, comparable, high quality results. TSmiR shows a high sensitivity and specificity to distinguish serum of GCC patients from controls, indicating that it can become an informative additional tool for clinical application.

Materials and Methods

Patient and control serum samples

Detailed information on the composition of the various sample series (R_V/R_L/UK/D/no-GCC) is presented in Figure 1 & Table 1. Samples in the R_V/R_L/no-GCC series were selected to be distributed over different clinical stages and histological subtypes. R_V/R_L samples were extracted at the Department of Pathology, Erasmus MC, Rotterdam, the Netherlands. Use of tissue samples for scientific reasons was approved by an institutional review board (MEC 02.981 and CCR2041). Samples were used according to the "Code for Proper Secondary Use of Human Tissue in The Netherlands" developed by the Dutch Federation of Medical Scientific Societies (FMWV (Version 2002, update 2011). The UK and D samples were sent frozen to Rotterdam, handled similarly, and used in accordance to regulations set by the local institutional review boards.

Anti-miRNA Magnetic Beads

The target-specific anti-miRNA magnetic beads were supplied by Applied Biosystems, Foster City, California, USA (Part Number 4473087, TaqMan® miRNA ABC Purification Kit - Human Panel A). The TaqMan® miRNA ABC (Anti-miRNA Bead Capture) Purification Kit is designed for rapid purification of miRNA from small inputs of all human sample types including body fluids, tissues, and cell cultures. The kit contains buffers and reagents for single-tube isolation of human miRNA molecules. The A panel consists of superparamagnetic Dynabeads® covalently bound to a unique set of ~380 anti-miRNA oligonucleotides. The miRNA isolation relies on hybridization of miRNA molecules to the corresponding anti-miRNA oligonucleotide probes attached to the beads. A common set of control anti-miRNA probes that can be used to capture both exogenously added and endogenous miRNAs (including the ones used in this study), are also bound to the beads. The complete lists of human miRNAs and control miRNAs that on the panel can be found in Appendix C, "Human Panel Beads" (page 29 in the User Guide of TaqMan® miRNA ABC Purification Kit, Publication Part Number 4473439 Rev. B).

miR analysis of primary GCC and normal testis

Small RNA was prepared from 47 primary, independent GCC and 3 normal testis samples. Expression of 156 miRs was analyzed, quantified and normalized as described [14,24].

miRNA purification and detection

Detailed protocol was provided by the supplier (Applied Biosystems/Invitrogen). miRs were purified from 50µl serum using the Taqman miRNA ABC Purification Buffer Kit (PN 4473084, Applied Biosystems). 100 µl of lysis buffer was added as provided by the supplier as well as 2 µl of a 1nM Spike in Controls (SiC: 6 different non-human control miR) solution and and $80x10^6$ washed human panel A magnetic beads in 1.5ml LoBind tubes (Eppendorf). Of note, hsa-miR-302d is not represented in the panel A bead set and was therefore not included in the study. After shaking at 12,000 rpm at 30°C for 40 minutes, unbound RNA was removed by three wash steps using a magnetic bead separator. Bound miRs were eluted from the beads with 100 µl of elution buffer (Applied Biosystems) at 12,000 rpm at 70°C for three minutes and stored at -80°C. DNA synthesis was performed using the TaqMan MicroRNA Reverse Transcription kit (Applied Biosystems, PN: 4366597) according to the instructions of the supplier. Quantification of miR levels was performed in duplicate (except for the R_L series) using the various TaqMan assays (Applied Biosystems catalogue ID: hsa-miR-371-3p (002124); hsa-miR-372 (000560); hsa-miR-373 (000561); hsa-miR-302a (000529); hsa-miR-302b (000531); hsa-miR-302c (000533); hsa-miR-367 (000555); hsa-miR-200c (002300); RNU6b (001093); RNU44 (001094); RNU48 (001006); U6snRNA (001973); ath-miR-159a (000338); cel-miR-39 (000200); hsa-miR-103 (000439); hsa-miR-16 (000391);

hsa-miR-192 (000491); hsa-miR-20a (000580); hsa-miR-451 (001105); hsa-miR-93 (000432), hsa-let-7a (000377). Expression levels were measured on the TaqMan 7900HT using standard procedures.

Quality control and missing data

miR-200c, implicated in carcinomas (e.g. breast cancer) but not in GCC was included as a control [26]. This miR is of additional interest because it is mapped to the short arm of chromosome 12 which is found to be consistently gained in invasive GCC [27]. Amplification curves were inspected individually. Only in the R_L set nine samples showed irregular curves with no semi-linear phase. These were set to UNDETERMINED, and replaced with the top Ct value measured per assay in the R_V set +2 (Ct_{max}). Ct_{max} indicates a miR level just outside the known sensitivity of the assay. This way the risk of underestimating the miR-level by choosing a fixed Ct_{max} (for example the total number of runs) is reduced. Assay-specific cDNA was generated separately from TCam-2 cells (SE cell line [25]) diluted in serum from a healthy donor to be used for inter-plate calibration (plate controls (PlC)) and positive control (PC). Four tubes (50,000 TCam-2 cells in 50 µl normal serum) were pooled for lysis, followed by RNA pull down as described above. The resulting RNA was used as positive control (PC) for cDNA synthesis when analyzing patient samples. cDNA was generated separately as well to be used for inter-plate calibration (plate controls (PlC)). Reference measurements for the PlCs were calculated as the average of triplicate PCRs for each assay. PlC and PC were included in duplicate (per assay) on each plate with patient samples. Inter-plate variation and variation in assay efficiency was investigated in most samples and found to be constant (below 0.3 Ct) with the exception of the miR-93 and miR-302b assays (0.6-0.8 Ct, consistent between plates), without affecting suitability for usage. Non-human Spike In Controls (SICs) (ath-miR-159a and cel-miR-39) were included to monitor RNA recovery. One sample (N-UK) showed 22-fold less recovery and was discarded. For negative control (No Template Control: NTC) elution buffer was added instead of the bead eluate. Ct values >NTC were treated as UNDETERMINED (R_V set, n=1).

Calibration & Normalization

Stable target recovery using the magnetic bead method was investigated and demonstrated using SiCs in all subgroups with occasionally different distributions between subgroups ($p<0.01$, Kruskal-Wallis test), warranting calibration for recovery (Figure S2) before determination of the most stable reference miRs for normalization. Some samples in the UK set were corrected for <50ul input volume. A panel of possible targets for normalization (RefN) was selected based on published data: RNU44, U6snRNA, hsa-miR-16, -93, -103, -192, -451 (haemolysis effects), miR-20a, let-7a [28]. Based on established algorithms for stability analysis (Normfinder [29] and geNorm [30]), miR-20a and miR-93 were the most

stable combination after calibration for normalization in the R_L set, which was validated in the R_V set (Figure S2).

Statistical analysis & Software

Data pre-processing was performed in Microsoft Excel 2010 & GenEx 5.3.7.332. Comparison of mean miR levels (dCt) was performed using a Mann-Whitney U test, two-sided in GenEx. If not otherwise specified, significant indicates $p \leq 0.000639$, yielding a risk of type I error of 5% after correction for multiple testing (Bonferroni). Uncorrected comparisons are indicated separately ($p<0.05$ = significant), as putative additional interesting results might be lost because of the stringent rule set. Principal Component Analysis (PCA), Kohonen Self Organizing Map (SOM) and Heatmap analysis were performed in GenEx. ROC analysis and SiC stability testing were performed in SPSS 18 & 20. Visualizations were generated using Microsoft Powerpoint & Excel 2010 and Photoshop CS4 & 5.5. Analyses were performed on a 64-bit Windows 7 system.

Results

A novel and complete pipeline (TSmiR-test) was developed for miR purification, recovery, as well as qPCR (including stringent quality controls for assay stability, inter-plate variability, reproducibility of cDNA synthesis, RNA recovery efficiency, and data processing) (Figure 2). The TSmiR test allows for standardized comparison between various sample series. ToIs included seven miRs from the miR-371-3 and miR-302/367 clusters, known to be expressed in SE, EC and YS (expression patterns in tumors is shown in Figure S1) [13-15].

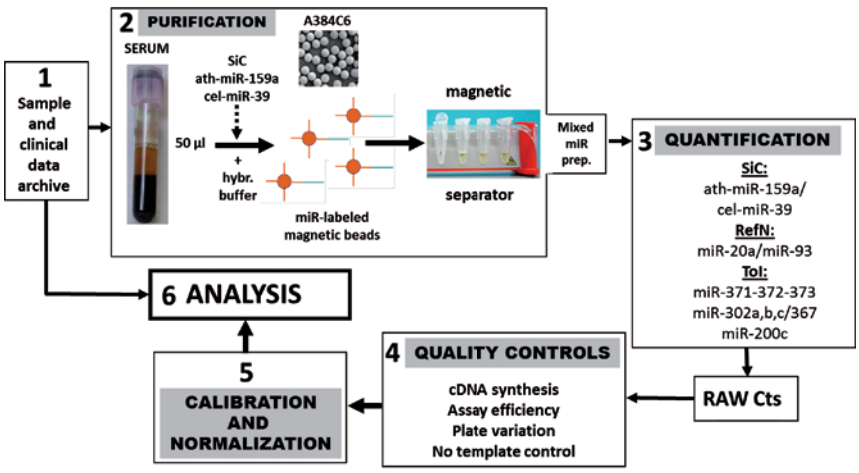


Figure 2. Complete pipeline of the TSmiR test. See text for explanation.

Table 1. Information of included series of patient samples.

miR	Series	N vs T	N vs SE	N vs NS	T meta vs no meta	SE meta vs no meta	NS meta vs no meta	pre- vs post-orchidectomy	N vs post-orchidectomy/ N vs no-GCC
				difference in miR levels in serum					
371	**RL**	68.7 (-31.3; +57.6)	41.9 (-21.4; +43.8)	112.4 (-62.5; +141)	7.7 (-4.1; +8.7)	2.2 (-1.4; +3.9)	24.7 (-16.8; +52.5)		
	RV	69.8 (-22.8; +33.8)	33.7 (-13.7; +23)	113.4 (-37.7; +56.4)	7.7 (-3.6; +6.8)	11.5 (-7.8; +24.3)	4.4 (-2.4; +5)		
	UK	80.1 (-33.2; +56.8)							
	D							46.5 (-16.5; +25.5)	0.7 (-0.2; +0.2)
	no-GCC								0.2 (-0; +0.1)
372	**RL**	19.7 (-5.9; +8.4)	16 (-5.5; +8.3)	24.2 (-9.2; +14.9)	4 (-1.4; +2.3)	5.8 (-2.4; +4.1)	2.6 (-1.4; +2.8)		
	RV	31 (-6; +7.4)	19.1 (-4.3; +5.6)	42.9 (-8.8; +11)	3.6 (-0.8; +1.1)	4 (-1.3; +2)	2.8 (-0.8; +1)		
	UK	4.1 (-1.7; +3)							
	D							4.1 (-1; +1.2)	1.3 (-0.2; +0.3)
	no-GCC								0.8 (-0.2; +0.2)
373	**RL**	44.9 (-14.1; +20.4)	34.7 (-12.4; +19.2)	58.2 (-23; +37.9)	6.7 (-2.3; +3.4)	9.9 (-3.6; +5.6)	4.3 (-2.2; +4.3)		
	RV	575.6 (-149.5; +202)	426 (-132.1; +191.5)	703.4 (-197.2; +274.1)	3.9 (-1; +1.3)	3.6 (-1.6; +2.7)	3.8 (-1.2; +1.7)		
	UK	5.9 (-2.8; +5.2)							
	D							6.1 (-1.7; +2.4)	3.2 (-0.8; +1)
	no-GCC								1.3 (-0.4; +0.5)
367	**RL**	57 (-22; +35.7)	62 (-28.1; +51.4)	52.4 (-24.2; +45)	4.4 (-1.3; +1.9)		2.7 (-1.2; +2.1)		
	RV	206.4 (-51.8; +69.2)	168.4 (-38.5; +49.9)	236.4 (-73; +105.7)	3.8 (-1.2; +1.9)	2.6 (-0.7; +1)	4.8 (-2.2; +4.2)		
	UK	32.9 (-14; +24.3)							
	D							6.7 (-2; +2.9)	1.1 (-0.3; +0.4)
	no-GCC								0.9 (-0.3; +0.4)
302a	**RL**	5.7 (-1.6; +2.1)	5 (-1.6; +2.4)	6.5 (-2.2; +3.3)	4.4 (-1.4; +2)	7.8 (-2.8; +4.4)	2.4 (-1.1; +2)		
	RV	18 (-4.8; +6.5)	5.4 (-1.8; +2.7)	40.1 (-9.2; +12)	5.1 (-1.9; +2.9)	7 (-4; +9.3)	2.4 (-0.6; +0.9)		
	UK	3.5 (-0.7; +0.9)							
	D							3.4 (-1; +1.3)	4.6 (-1.1; +1.4)
	no-GCC								1 (-0.2; +0.3)

302b	**RL**	5.2 (-2.3; +4)	3.5 (-1.6; +3.1)	7.6 (-4.2; +9.3)	7.2 (-3.6; +7)	6.2 (-3.3; +6.9)	7.7 (-5.4; +18.1)		
	RV	1.1 (-0.4; +0.6)	0.6 (-0.2; +0.4)	1.6 (-0.6; +1)	2.8 (-1.3; +2.3)	0.7 (-0.4; +1.4)	6.6 (-3.3; +6.7)		
	UK	11.2 (-4.8; +8.4)							
	D							2.3 (-0.6; +0.7)	8.3 (-2.7; +3.9)
	no-GCC								0.6 (-0.2; +0.3)
302c	**RL**	3.4 (-1; +1.4)	4.3 (-1.4; +2.1)	2.8 (-1; +1.5)	3.4 (-1.2; +1.8)	7.2 (-2.7; +4.3)	1.7 (-0.9; +1.7)		
	RV	9.6 (-2.1; +2.7)	6.6 (-1.6; +2.2)	12.5 (-3; +3.9)	3.5 (-0.9; +1.3)	2.4 (-1; +1.7)	3.9 (-1.3; +1.9)		
	UK	1.7 (-0.4; +0.5)							
	D							2.4 (-0.5; +0.6)	2.9 (-0.6; +0.8)
	no-GCC								0.6 (-0.1; +0.2)
200c	**RL**	0.9 (-0.1; +0.2)	1 (-0.2; +0.2)	0.8 (-0.2; +0.2)	1.7 (-0.3; +0.4)	1.7 (-0.4; +0.4)	1.7 (-0.5; +0.7)		
	RV	2.5 (-0.5; +0.6)	2 (-0.4; +0.6)	2.9 (-0.6; +0.8)	1.6 (-0.4; +0.5)	1.4 (-0.5; +0.7)	1.7 (-0.5; +0.6)		
	UK	1.8 (-0.6; +0.8)							
	D							1.3 (-0.1; +0.1)	1.1 (-0.2; +0.2)
	no-GCC								0.8 (-0.1; +0.2)
nr. of samples									
nr. samples	**RL**	12 / 32	12 / 16	12 / 16	16 / 16	8 / 8	8 / 8		
	RV	30 / 30	30 / 12	30 / 18	12 / 18	6 / 6	6 / 12		47
	UK	6 / 7							
	D							6/5	11
	no-GCC								12

Numbers indicate times (x) difference between two group identified as "A vs B" in column headers, for example N vs T). Positive numbers indicate relative higer levels in group B. Numbers between brackets indicate SEM (Standard Error of Mean), asymetric because of exponential transformation of dCT values. Black background indicates significant difference between mean dCt values of both groups after Bonferroni correction; grey without correction for multiple testing. Abbreviations used: R_L: Rotterdam Learning set; R_V: Rotterdam Validation set; UK: United Kingdom set; D: Germany set; N = Normal (control); T = Tumor; SE = seminoma; NS = Nonseminoma; meta: metastasis; The T samples in the R_L and R_V were classified as stage I ($n_{RL,SE}$=8, $n_{RV,SE}$=6, $n_{RL,NS}$=7, $n_{RV,NS}$=6), stage II ($n_{RL,SE}$=6, $n_{RV,SE}$=6, $n_{RL,NS}$=0, $n_{RV,NS}$=6) and stage IV ($n_{RL,SE}$=2, $n_{RV,SE}$=0, $n_{RL,NS}$=9, $n_{RV,NS}$=6). The N series constituted of serum of healthy individuals (n_{RL}=2, n_{RV}=10), patients with testicular torsion (n_{RL}=10, n_{RV}=10) and patients with epididymitis (n_{RL}=0, n_{RV}=10). The UK dataset consisted of N (n=6) and T (n_{adult}=4, $n_{pediatric\ YS}$=3) samples. In the D dataset 11 matched pre and post orchidectomy samples were included of patients with stage I (local) disease (n_{SE}=2x6, n_{NS}=2x5). The no-GCC series consistent of serum of patients with a Sertoli/Leydig cell tumor (n=3/3), testicular B-cell lymphoma (n=3) or epidermoid cyst (n=3).

The TSmiR test was applied to five independent (patient) sample series (R_V/R_L/UK/D/no-GCC, Figure 1, Table 1). Results of quality control and normalization according to the TSmiR pipeline (Figure 1) are described extensively in the Material & Methods section. Results of the different (patient) series will be presented subsequently (actual dCt values presented in Table S1, available upon request).

Rotterdam Learning and Validation series

Comparison of subgroups (fold difference & significance with and without Bonferroni correction) in R_L and R_V is presented in Figures 3, 4 and 5 and Table 1 (Table S2 for p-values). miR-371-3 and miR-367 levels were significantly higher in both series (T/SE/NS vs N). miR-302a was significantly higher in T/NS vs N. miR-302c was significantly higher in T/SE/NS vs N in the R_V set only. A more heterogeneous pattern was observed for the other targets. Without Bonferonni correction, additionally, miR-302a showed a consistent significantly higher expression in T/SE/NS vs N. Distinction between absence and presence of metastases based on miR expression level was possible for miR-372 (R_V only), miR-373 (R_L and R_V) and miR-367 (R_L only) (after Bonferroni correction). In addition, uncorrected analysis showed significantly higher levels of miR-371-3 and miR-302 cluster members, except for miR302b in R_V. The histology-specific analysis (SE and NS) revealed a more heterogeneous pattern with no significant differences between histological subgroups,

Correlation between the different miRs tested showed that the highest correlation exists between levels of members of the miR-371-3 cluster, followed by miR-367, miR-302a, and miR-302c (Spearman's rho, both for the R_L and R_V, Table S3, available upon request). Low correlation or non-significant correlation was found for the others.

Based on the findings presented above, only the most discriminating miRs (miR-371-3 and miR-367) were used for further analysis. A Kohonen Self Organizing Map (SOM) was generated using the R_L data (Figure 4A). The R_L set (Figure 3A) indicated that one group only contains T samples, while the other group contains all N as well as some T samples. The SOM learned using the R_L data was highly discriminative in the R_V series (Figure 4B), separating a group composed of 29 N and one T sample and a group composed of 30 T samples and one N sample. In addition, Principal Component Analysis (PCA) on the R_L series (Figure 5A) also demonstrated that most T samples cluster separately from the N samples, although some T samples positioned in the N group. The R_V series showed this separation even more stringent (Figure 5B): all N samples cluster separately from the T samples. The SOM and PCA results were in line with conventional heatmap analysis (Figure S4).

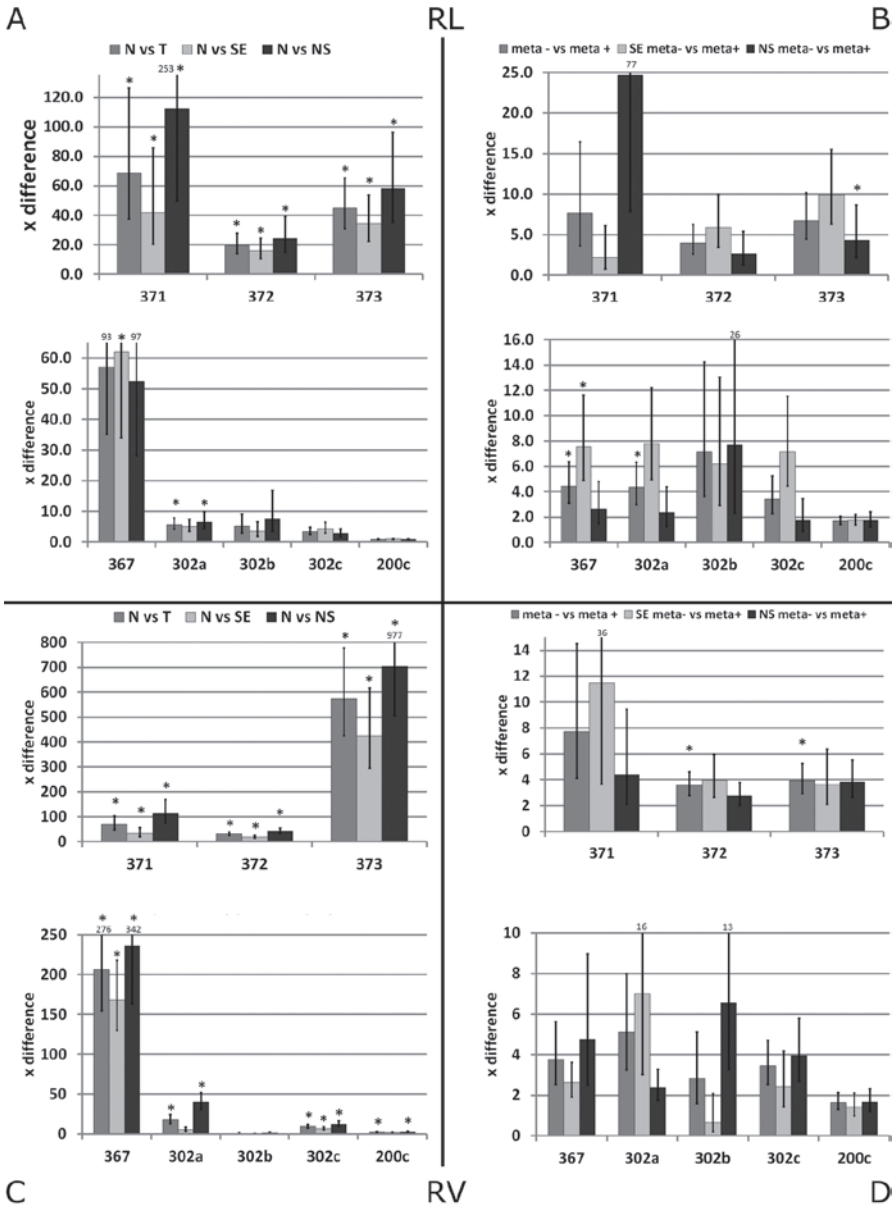


Figure 3. Bars represent times (x) difference between groups, defined as difference between the mean of the dCt of group A vs B (fold change=2^(dCtA-dCtB)). High values represent relatively higher levels in group B. Error bars depict standard error of the mean. * = group A and B show significantly different levels based on the dCt values (Bonferroni corrected)). R_L/R_V series: A/C) Comparison of N versus histological subtype (SE or malignant NS) and all tumor samples (T=SE+NS); B/D) Comparison of miR levels in serum samples from patients with localized and metastatic disease (stage I vs pooled stage II, III and IV) for all tumor samples together (meta+/meta-) and SE/NS separately.

UK-series

The UK series consisted of a more heterogeneous set of serum samples (Table 1), including serum from either pediatric or adult patients. Inter-plate variation and assay efficiency was found to be comparable to the $R_{V/L}$ series (data not shown). Again a significant difference was found for miR-371-3p between N and T (low versus high) after Bonferonni correction, and additionally for miR-367, miR-302a,b without correction. Application of the R_L based SOM (see above) and PCA analysis demonstrated that N and T samples could be separated for the majority of cases. PCA also showed a distinction between the pediatric and adult T samples (Figure 4C and 5C).

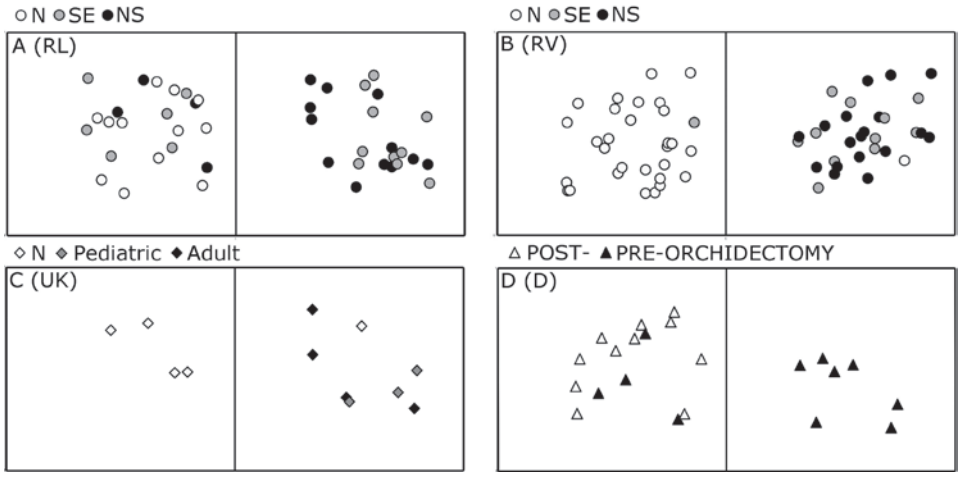


Figure 4. A Kohonen Self-Organizing-Map (SOM) was generated using the R_L series as training data. The SOM was set to identify two groups (alpha=0.40, 500 iterations). Only miR-371-3/367 data were included because these showed the highest, most significant and most consistent difference between serum samples of patients with a GCC and controls. The generated SOM was then applied to the R_V, D and UK series: A) In the R_L series, the right group only contained tumor samples, however the left group contained a mixed set of samples; B) The same SOM identified a left group with almost only control samples (N) and one SE in the R_V series. The right group contained all other tumor samples and one control sample; C) The same SOM proved to also be able to separate all but one N samples from the T cases in the UK samples; D) All post-orchidectomy samples from the D set clustered together, together with some pre-orchidectomy samples.

D-series

The D series consisted of paired stage I SE and NS before and after orchidectomy (Table 1). miR-371-3 and miR-367 showed significantly higher serum levels in the pre-orchidectomy samples compared to post-orchidectomy samples after correction, as well as miR-302ac without. The post-orchidectomy samples showed miR levels that were not significantly different from those of all pooled N samples for miR-371-3 and miR-367 after correction.

miR-302b remained overexpressed after orchidectomy (Bonferroni corrected) as did miR-373/302a,c (not corrected) (Figure 6, Table 1). The same SOM (see above) was applied, showing one group comprised only of pre-orchidectomy samples while the other group was composed of both post-orchidectomy and pre-orchidectomy samples (Figure 4D), in line with the PCA results (Figure 5D).

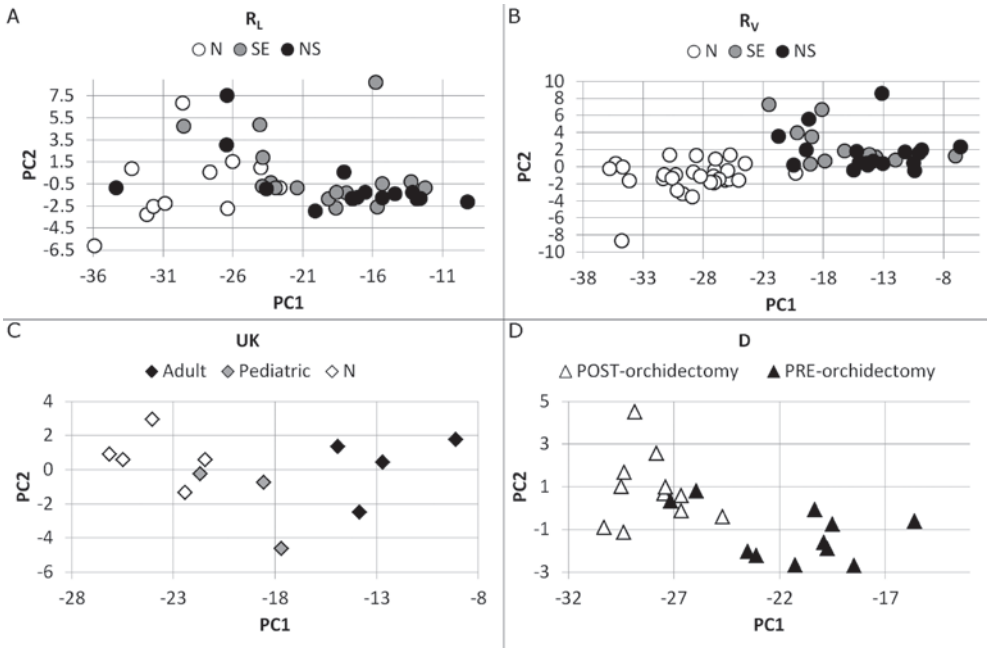


Figure 5. Principal component analyses for all individual datasets. In all cases the first and second PC based on miR371-3 explained > 99% of the variance. A) R_L: although most of the tumor samples are clearly clustered and separated from the controls, there are some T (n=7) intermixed with N; B) R_V: All T and N samples cluster separately. One N (testicular torsion) is situated on the edge of the T/N border; C) UK: all N samples and T samples cluster separately. The pediatric YS are situated between the N samples and the adult GCC clusters; D) D: all but two pre-orchidectomy samples cluster together. All post-orchidectomy samples cluster together. This cluster also included two pre-orchidectomy samples.

No-GCC series

This series consisted of serum samples of patients with Sertoli cell tumors, Leydig cell tumors, testicular lymphomas, or epidermoid cysts (n=3/3/3/3, Table S1). In the tissue, these masses showed no expression of miR-371-3/302abc/367 (unpublished data). In this series none of the miRs sowed significant elevation (miR-371-3p even significantly lower) compared to all pooled controls (N) samples, confirming the sensitivity as well as specificity of the investigated serum miR markers for GCC in the context of differentially diagnosing testicular masses (Table 1, Figure S5).

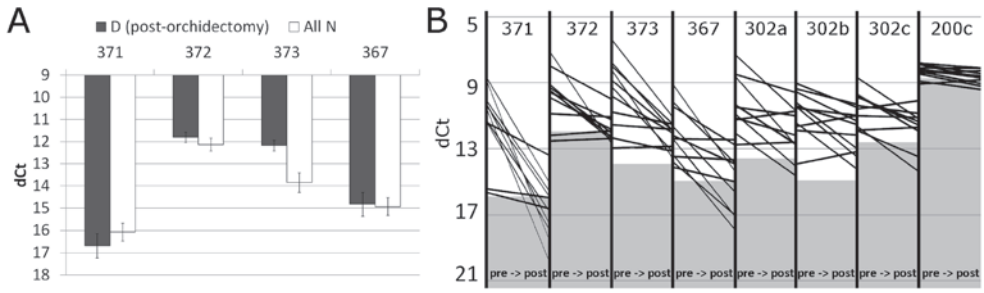


Figure 6. A) When all N from the R_L, R_V and UK series are pooled and compared to the post-orchidectomy D samples, there is no significant difference between the serum-miR levels in both groups for the most differentiating miRs (371-3/367) (all miRs summarized in Table 1). X-axis: most consistently differentially present miR (between N and T). Y-axis: dCt; B) Plot of all pre (left) and post (right) orchidectomy dCt levels per miR per case. Grey area indicates the average baseline level in all pooled N. This does not necessarily constitute a valid cutoff (see Figure 7).

Sensitivity and specificity of TSmiR and comparison to the current clinical gold standard (serum AFP and hCG)

ROC analysis was done for the pooled miR-371-3 and miR-367 data (Figure 7AB; subsets and other miRs in Figure S6). The discriminating power was very high (Area Under the Curves (AUC): miR-371-3p=0.88; 372-3/367>0.91), with only marginal differences when SE or NS subgroups were analyzed separately. Inclusion of the post-orchidectomy D samples as "N" did not significantly influence the AUC. The results were compared to matched AFP and hCG measurements for all T samples (not available for N). This demonstrated that TSmiR is more informative than AFP and hCG, particularly for SE (Figure 7C-F). In the total series investigated, sensitivity of AFP and hCG for all GCC patients is 36% and 56%, respectively (3% and 45% for the SE and 62% and 66% for the NS group). Even when the YS subgroup is investigated specifically, TSmiR shows higher sensitivity than AFP. Cutoffs were set for a sensitivity of 90% , preserving a specificity of 60-91% (Figure 7F). Combined interpretation of the established serum markers AFP/hCG and the TSmiR test led to the correct classification of all investigated serum samples of GCC patients, i.e., those with a malignancy. In contrast, application of the serum markers AFP and hCG alone, resulted in ≈50% false-negative cases. Application of only TSmiR resulted in one false negative NS serum sample.

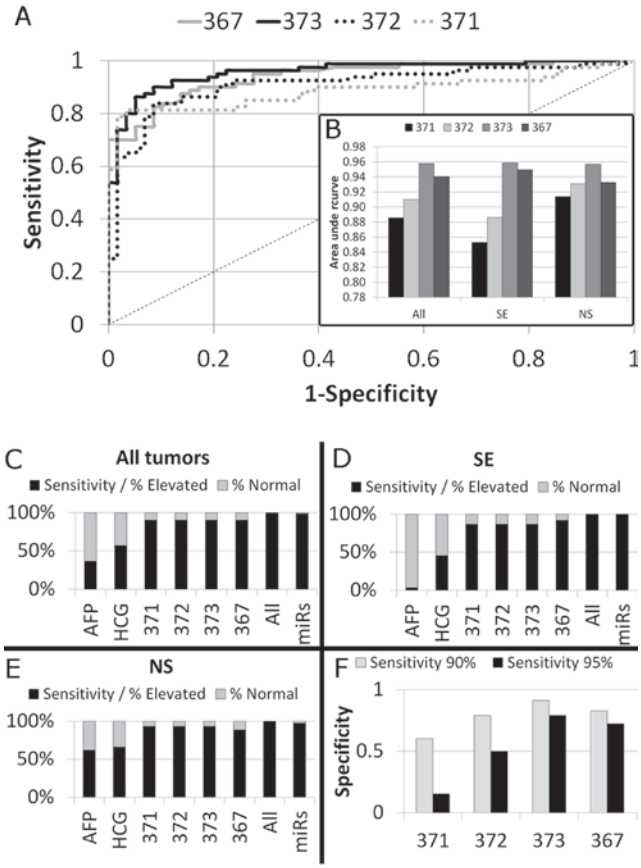


Figure 7. A) ROC curves were generated for the four most differentiating miRs (371-3/367). Samples from the R_L/R_V/UK/D series were pooled for analysis.; B) The AUC varied from 0.89 to 0.96 for all ROCs. Separate ROCs generated for only the SE or the NS showed minor differences between the quality of these miRs as marker. C, D, E) Stacked bar diagrams of the sensitivity and false negative rate for the conventional "gold standard" in current clinical use (AFP, hCG) and the four miRs: miR-371-3/372/373/367. Sensitivity was calculated when all miRs were assessed in combination, and again in combination with AFP/hCG. Missing values were discarded in the calculation of the frequencies; C) All tumors; D) SE; E) NS; F) A sensitivity of 90% (grey) in the ROC analysis lead to an acceptable remaining specificity (y-axis). A sensitivity of 95% (black) resulted in a big loss of specificity, especially for miR-371-3p/372. Cutoffs for dCt values indicating T were identified using the ROC including all pooled samples: miR-371-3p≤15.62, miR372≤11.02, miR-373≤10.45, miR-367≤12.48.

Discussion

GCC patients require intensive follow-up after primary diagnosis and treatment [17]. For diagnosis and follow-up, various approaches are currently used, including the established diagnostic/follow-up serum-markers AFP and hCG (Figure 1). AFP is primarily related to YS

and hCG to CH although they can be positive in sporadic cases of SE and SE/EC respectively as well. Therefore, a significant number of some patients with a germ cell malignancy without YS and CH, will be negative for these serum markers. In addition positivity for either marker can change during disease progression. Because of this, there is need for additional serum markers, particularly ones that are consistently positive in SE and EC. Several possible candidates have been suggested, including demethylated promoter regions of the *XIST* gene [31] (sensitivity of 64% in serum of GCC patients, small single series). More recently, a selected number of miRs are reported to be highly expressed in SE, EC and YS (both pediatric and adult patients, figure S1). They belong to the embryonic stem cell miR-371-3 and miR-302/367 clusters and were validated in multiple independent GCC series [13-15]. Although exosomes have been suggested as a miR-transporters / -repository, the mechanism behind miR release into serum remains to be elucidated [32,33], levels of these miRs have proven to be detectable in the serum in the serum of GCC patients [21-23]. The protocols applied in these studies lacked the relevant quality controls, calibration and normalization steps required for clinical implementation. Although amplification steps were performed, no specific miR purification/recovery step was included, resulting in possible underscoring of miR levels in serum by qPCR.

In this study, a robust and informative pipeline is presented to detect the levels of these miRs in serum of GCC patients and controls (Figure 2)., Quality control steps for miR purification and recovery artifacts were implemented (Figure S2) and suitable targets for normalization of Ct values in serum identified (Figure S3). Moreover, inter-plate and inter-assay variations were evaluated and corrected for. The described pipeline is potentially suitable for clinical serum analysis in the context of other malignancies. miRs identified as candidate markers need to show a specific and consistent high level of expression in serum samples from patients with cancer compared to those from normal, healthy individuals.

The value of GCC specific miR-detection in serum using the TSmiR pipeline was demonstrated in learning and validation set of independent serum samples from controls and GCC patients. miR-371-3 and miR-367 were most informative (Figure 3). At time of diagnosis, these specific serum miR levels allowed almost complete separation of control- and patient samples (Figures 4, 5, S4). When the miR-371-3 and miR-367 results were combined, only one tumor sample was misclassified as control, independent of the histology of the tumor (no consistent difference between SE and NS). (sensitivity 98%). In contrast, the sensitivity of AFP/hCG was much worse (max ≈60%), especially in the SE (max ≈40%) samples (Figure 7,S6). This proof of concept demonstrates the additional value of TSmiR as a "liquid biopsy" in primary GCC diagnosis, indicating superiority over the currently applied AFP/hCG test for the detection of the stem cell components (SE/EC). This is further confirmed by investigating a series of patients with a testicular mass of other origin, showing a pattern similar to the control group (Figure S5). Although the primary aim of this study was to investigate the additional value of

quality controlled detection of GCC specific miRs in primary diagnosis of GCC, differences between cases with and without metastasis (Figure 3, Table S2) and complete normalization of levels after complete resection of a local tumor (Figure 6) suggest an important role for TSmiR in follow-up as well (currently investigated in a prospective setup). This proof of concept indicates that TSmiR has the potential to be of significant importance in the context of additional diagnostic ("liquid biopsy") and follow-up tools for GCC patients, warranting further investigation.

Acknowledgements

The authors would like to thank all the patients and clinicians who contributed to the study. Ronak Eini is financially supported by a Grant from the Dutch Cancer Society (NKI 2007-3081) and Martin Rijlaarsdam by a Translational Grant, Erasmus MC.

List of author contributions

AJMG, MAR, LD, RE and LHJL designed the experiments. AJMG carried out the experiments. MAR analyzed/visualized the data. KB was responsible for the clinical (pathological) diagnosis. MJM, JCN, NC, KPD, GB and JB supplied additional samples. TX and NB assisted in technical issues concerning the magnetic beads. LHJL was responsible for the overall process. All authors were involved in writing the paper and had final approval of the submitted and published versions.

List of online Supporting Information

- Figure S1. Expression analyses of the ToI (miR-371-372-373, miR-302, miR-367, 200c) in a series of GCC tumor samples and controls.
- Figure S2. Spike in levels in the various sample series.
- Figure S3. Selection of stable normalizers in serum.
- Figure S4. Heatmaps of the four main data sets.
- Figure S5. Analysis of no-GCC samples in comparison to the GCC series.
- Figure S6. Extended ROC analysis.
- Table S1. dCts all (excel sheet)
- Table S2. Mean comparison of subgroups, p-values
- Table S3. Correlation between miRs (RL & RV)
- Table 1, Table S2 and Supplementary Figures

Supplementary Figures:

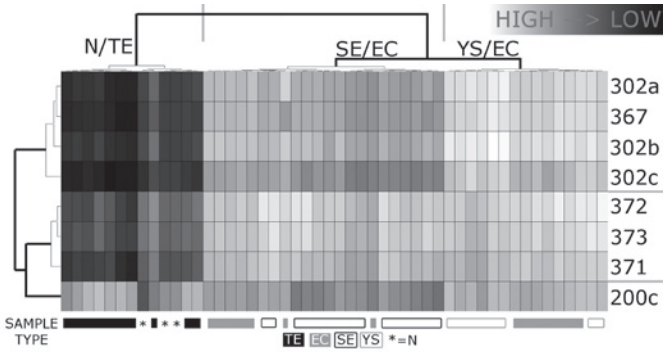


Figure S1. Expression analyses of the ToI (miR-371-372-373, miR-302, miR-367, 200c) in a series of GCC tumor samples and controls.

All miRs (n=156) on the plate were used to normalize using the mean expression of all genes (threshold Ct 40). Samples included are N (n=3), TE (n=10), SE (n=15), EC (n=14), YS (yolk sac tumor) (n=8). Clustering was performed using Ward's algorithm; both dendrograms were formatted according to Euclidean distance. The clustering clearly distinguished between N+TE (differentiated tumor types with expected low level) and SE/EC/YS (known high level). The vertical dendrogram correctly separated the miR clusters (miR-371-3 vs miR302abc/367). miR-200c clustered separately and did not contribute significantly to the clustering. Mean expression levels were significantly higher in the pooled SE/EC/YS or SE/EC groups when compared to N for all miRs shown, except 200c (as expected.)

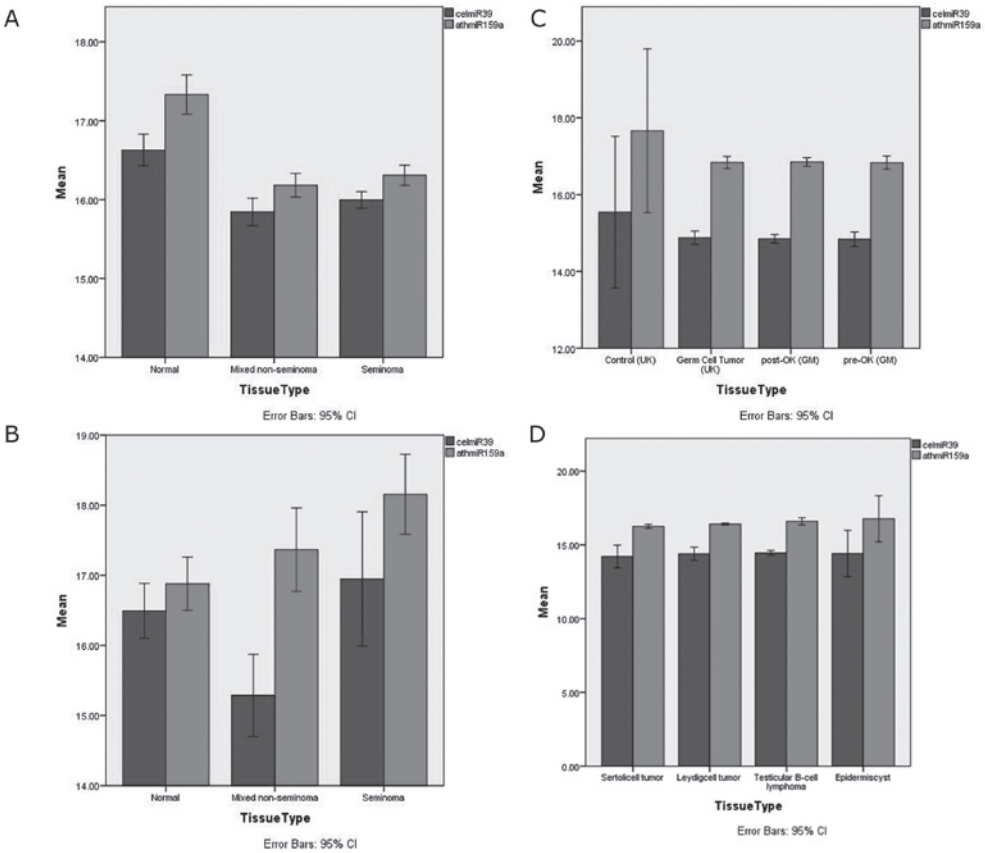
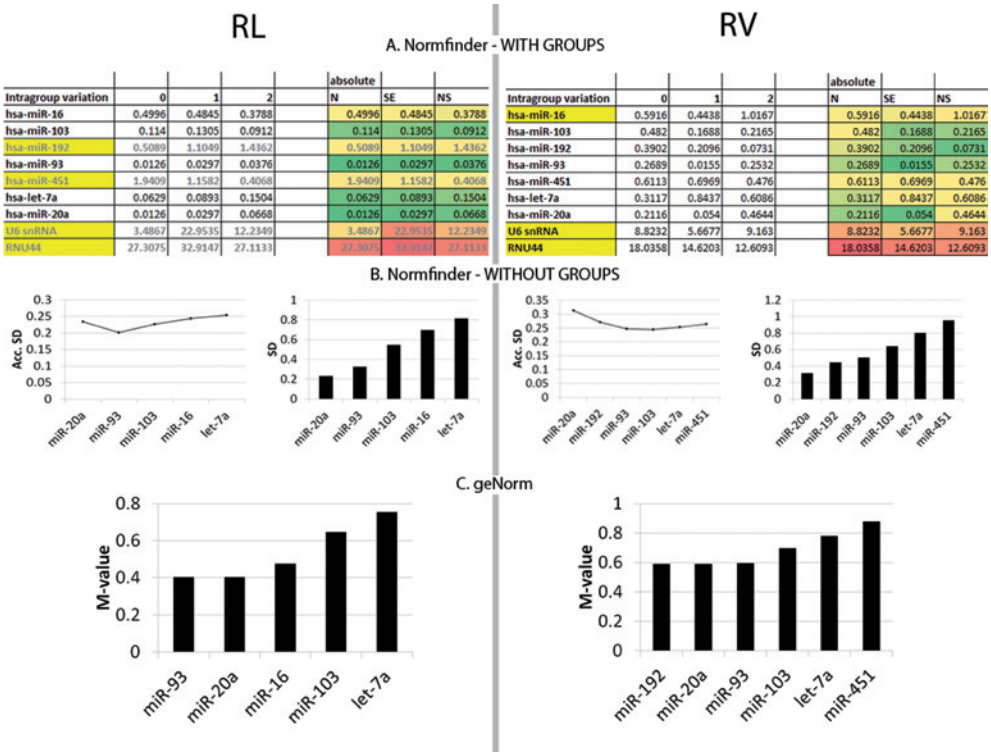


Figure S2. Spike in levels in the various sample series.

A, B) Stable spike in control (SiC) expression was observed in both the R_L and the R_V datasets. However, the distribution of the values was different between the N/SE/malignant NS groups ($p<0.01$, Kruskal-Wallis test). This warranted calibration for recovery efficiency as applied in all cases before analysis; C) SiC expression in the D and UK series; no significant difference between the groups (UK and GM compared separately) ($p>0.05$, Mann Whitney, two-sided); D) Spike-in levels in the no-GCC samples were highly stable, although statistical testing revealed non-similar distributions for ath-miR-159a based on n=3 per group ($p=0.04$, Kruskal-Wallis test, cel-miR-39 $p=0.79$). Calibration was applied.



Intragroup variation	0	1	2	absolute N	SE	NS
hsa-miR-16	0.4996	0.4845	0.3788	0.4996	0.4845	0.3788
hsa-miR-103	0.114	0.1305	0.0912	0.114	0.1305	0.0912
hsa-miR-192	0.5089	1.1049	1.4362	0.5089	1.1049	1.4362
hsa-miR-93	0.0126	0.0297	0.0376	0.0126	0.0297	0.0376
hsa-miR-451	1.9409	1.1582	0.4068	1.9409	1.1582	0.4068
hsa-let-7a	0.0629	0.0893	0.1504	0.0629	0.0893	0.1504
hsa-miR-20a	0.0126	0.0297	0.0668	0.0126	0.0297	0.0668
U6 snRNA	3.4867	22.9535	12.2349	3.4867	22.9535	12.2349
RNU44	27.3075	32.9147	27.1133	27.3075	32.9147	27.1133

Intragroup variation	0	1	2	absolute N	SE	NS
hsa-miR-16	0.5916	0.4438	1.0167	0.5916	0.4438	1.0167
hsa-miR-103	0.482	0.1688	0.2165	0.482	0.1688	0.2165
hsa-miR-192	0.3902	0.2096	0.0731	0.3902	0.2096	0.0731
hsa-miR-93	0.2689	0.0155	0.2532	0.2689	0.0155	0.2532
hsa-miR-451	0.6113	0.6969	0.476	0.6113	0.6969	0.476
hsa-let-7a	0.3117	0.8437	0.6086	0.3117	0.8437	0.6086
hsa-miR-20a	0.2116	0.054	0.4644	0.2116	0.054	0.4644
U6 snRNA	8.8232	5.6677	9.163	8.8232	5.6677	9.163
RNU44	18.0358	14.6203	12.6093	18.0358	14.6203	12.6093

Figure S3. Selection of stable normalizers in serum.

In both sets: miR-93 and miR-20a were top-scoring with respect to (inter-/intragroup) stability. Ct values (series R_v: average of technical replicates) were normalized against the average expression of these RefNs using GenEx. A) Normfinder was used to assess intra-group variability (N/SE/malignant NS). RefNs with variation > 1 for one or more groups were discarded; B) Normfinder was run again on the remaining RefNs using all samples WITHOUT group classification. From this analysis the optimal number of RefNs was assessed; C) geNorm was run separately using all RefNs to obtain an independent ranking of the stability of the RefNs. geNorm does not take group classification into concern; R_L set: A) Normfinder WITH groups: miR-192/451, U6 snRNA & RNU44 were discarded; B) Normfinder WITHOUT groups: Combining 2 RefNs showed the lowest acc. SD (variation) = miR-20a/93 (sorted desc.); C) geNorm: 3 RefNs reached the arbitrary threshold of <0.5 (M-value) = miR-93/20a/16 (sorted descending); R_V set: Normfinder WITH groups: miR-16, U6 snRNA & RNU44 were discarded; B) Normfinder WITHOUT groups: Combining 4 RefNs showed the lowest acc. SD (variation) = miR-20a/192/93/103 (sorted desc.); C) geNorm: 0 RefNs reached the arbitrary threshold of <0.5 (M-value).

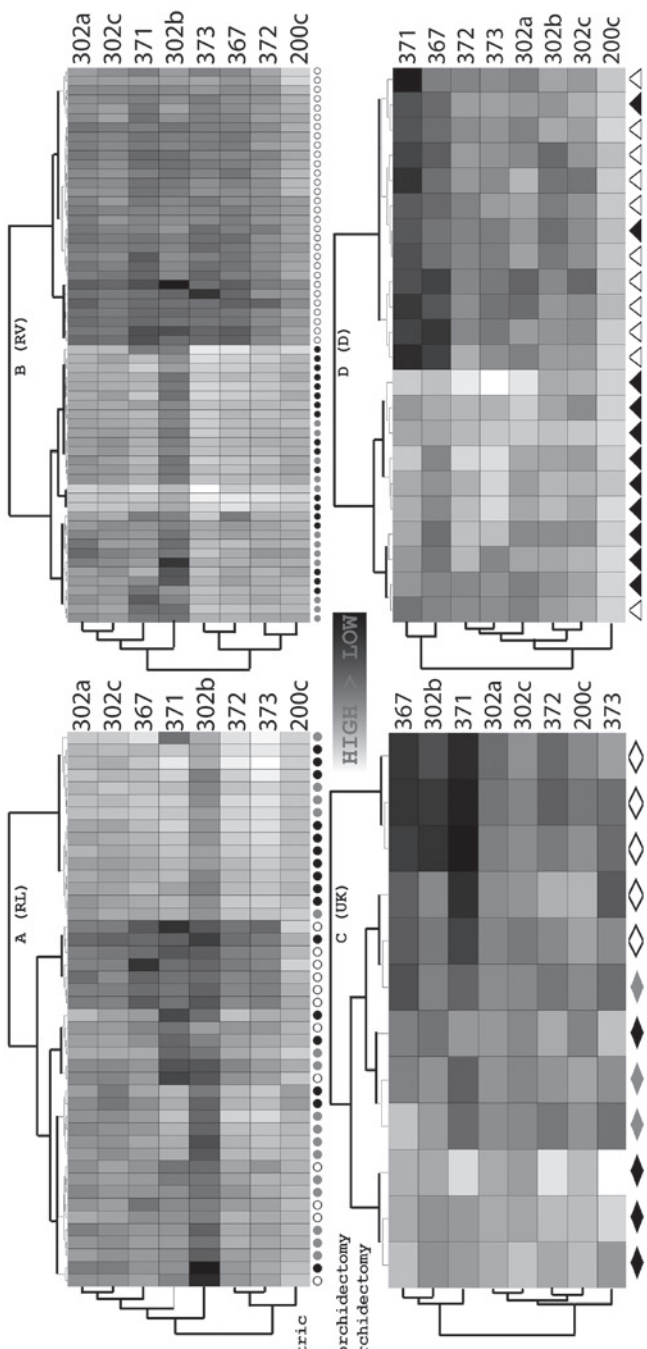


Figure S4. Heatmaps of the four main data sets.

A) R_L dataset. Clustering showed one group with only tumor samples (SE/NS) and a second group containing all N samples and 11 SE and 6 NS; B) R_V dataset. Clustering completely separated the SE/NS and the N samples; C) UK dataset. The N and T samples cluster separately, but two controls and one tumor were placed under one branch; D) D dataset. The clustering separated the pre- and post-orchidectomy samples quite well, although one post-orchidectomy sample was placed in the pre-orchidectomy branch and two pre-orchidectomy samples were situated in the post-orchidectomy branch. Clustering was performed using Ward's algorithm; both dendrograms were formatted according to euclidean distance. These results matched the patterns observed in the SOMs and PCA.

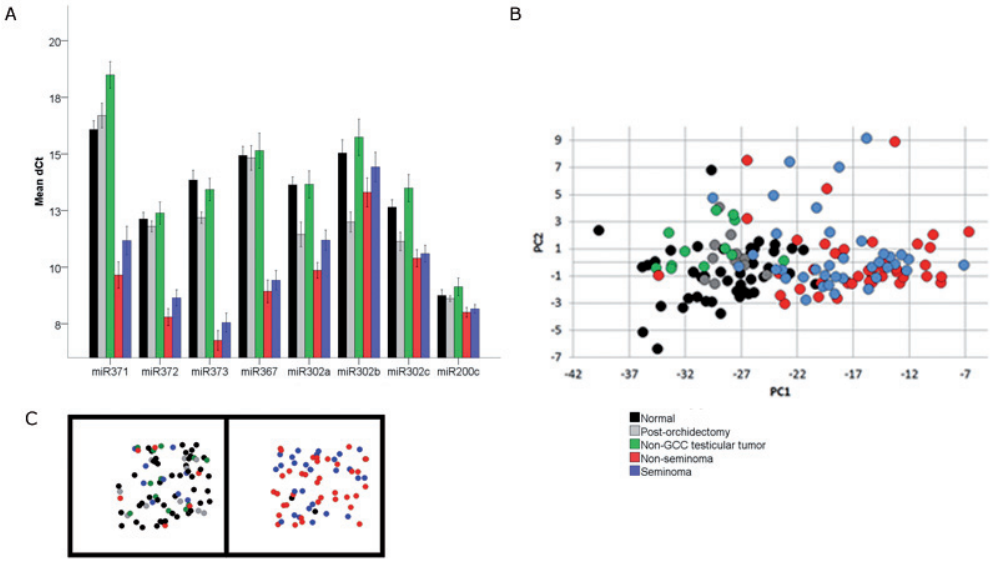


Figure S5. Analysis of no-GCC samples in comparison to the GCC series.

Comparison of control (N), Post-orchidectomy, Non-GCC testicular tumor (no-GCC), Non-seminoma (NS) and Seminoma (SE) groups. Samples from all series were pooled. A) Bar graph of dCt values (high values = low expression), indicating similar (miR371-3p/373/367/302abc/200c, p>0.05) or even significantly lower expression (miR-371-3p, p=0.002) in the no-GCC cases as compared to all N (two sided Mann-Whitney test, black vs. green bar). The difference between the top-miRs (371-3/367) for the N/post-orchidectomy/no-GCC groups as opposed to the NS/SE groups is again much more apparent than in the miR-302abc/200c cluster; B) PCA applied to all pooled samples. PCA separates the N/post-orchidectomy/no-GCC samples with some SE/NS samples intermixed in the area of the controls, especially on the border of both groups. Only one N sample is clearly clustering with the SE/NS samples; C) The SOM learned from the R_1 series was applied to all pooled samples. The right group contains almost exclusively tumor samples: only two N samples cluster with the SE/NS samples. The left group contains all other N samples, all post-orchidectomy samples and all no-GCC samples together with some SE/NS samples.

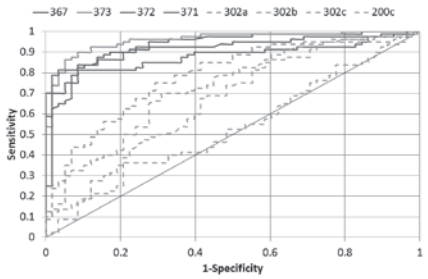
Figure S6. Extended ROC analysis. →

A) ROC curves were generated for all TOIs using different datasets & subsets of samples (columns). ALL indicates the analysis of the aggregated samples of the RL, RV, UK and D datasets. AUC = Area Under Curve (0-1) of each ROC. AUCs > 0.8 are displayed using increasing shades of grey to indicate better discriminative power. The inclusion of the D dataset in ALL does not significantly change the AUCs of any of the miRs investigated (ALL vs. no D column): B) ROC curves were generated for all TOIs using the aggregated data of all datasets (ALL). Top-miRs (371-3p/372/373/367) are indicated in solid lines while the rest are plotted using a dashed pattern; C, left) A sensitivity of 90% (grey) in the ROC analysis lead to an acceptable remaining specificity (y-axis). A sensitivity of 95% (black) lead to a big loss of specificity, especially for miR-371-3p/372/302ab. Cutoffs for dCt values indicating T were identified using the ROC (sensitivity 90%) including all samples: miR-371-3p≤15.6175, miR372≤11.0175, miR-373≤10.45, miR-367≤12.4825, miR-302a≤13.9075, miR-302b≤19.025, miR-302c≤13.255, miR-200c≤9.4175; C, right) Using the chosen sensitivity of 90%, the array of miRs including miR-302abc shows a much lower specificity than the array of the top-miRs; D) Stacked bar diagrams of the sensitivity and false negative rate for the golden standard (AFP, hCG) and miR-371-3p/372/373/367/302abc/200c. Sensitivity was calculated when all combining all or only the top-miRs with and without the established GCC serum markers AFP and bHCG (coded as elevated (1) or normal (0) as indicated in medical information or based on the cutoffs used in the Erasmus MC Rotterdam (AFP <=9, hCG<=1.9). The minimal detected concentration of <2 (hCG, UK) was interpreted as normal. miR expression data was also dichotomized (0 normal, 1 elevated) per miR based on the ROC data, retaining a sensitivity of 90%. Missing values were discarded in the calculation of the frequencies. The observations for the tumor markers are in line with the known hGC positivity of some SE and a higher sensitivity of AFP/hCG in NS (especially YS) as opposed to the SE subgroup. (left top) All tumors; (right top) SE; (left bottom) NS; (right bottom) YS.

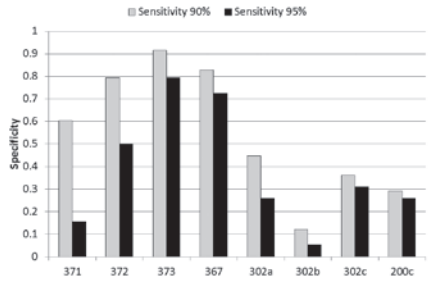
Dataset		ALL	RL/RV/UK	RL	RL	RL	RL	RV	RV	RV	RV	UK	UK	UK	D	D	D
Subset			*no D*		*SE+N*	*NS+N*	*YS+N*		*SE+N*	*NS+N*	*YS+N*		*pediatric+N*	*adult+N*		*SE+N*	*EC+N*
Known	positive (n)	80	69	32	16	16	14	30	12	18	11	7	3	4	11	6	5
	negative (n)	58	47	12	12	12	12	30	30	30	30	5	5	5	11	11	11
AUC	371	89%	87%	84%	83%	85%	83%	90%	84%	94%	93%	100%	100%	100%	97%	94%	100%
	372	91%	92%	90%	89%	91%	92%	99%	98%	100%	100%	74%	53%	90%	86%	83%	89%
	373	96%	97%	94%	94%	94%	93%	100%	100%	100%	100%	80%	60%	95%	85%	88%	82%
	367	94%	95%	91%	92%	91%	93%	98%	100%	97%	100%	94%	87%	100%	90%	85%	96%
	302a	79%	82%	80%	79%	81%	88%	89%	74%	99%	100%	97%	93%	100%	61%	65%	56%
	302b	54%	54%	70%	63%	77%	84%	47%	37%	54%	58%	91%	93%	90%	77%	73%	82%
	302c	72%	76%	72%	74%	69%	79%	89%	85%	92%	94%	66%	40%	85%	47%	48%	45%
	200c	65%	65%	49%	51%	48%	51%	73%	69%	76%	79%	71%	60%	80%	70%	61%	82%

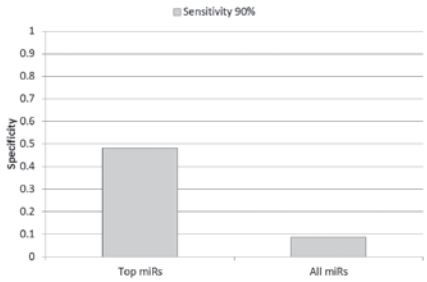
A

B

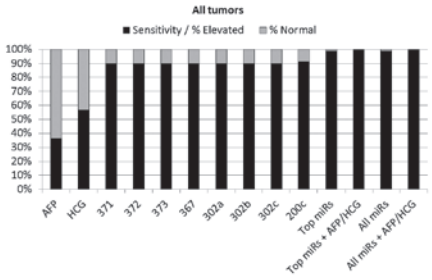


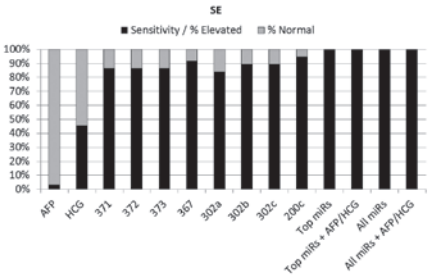
C

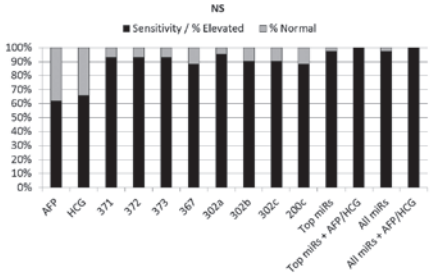




D







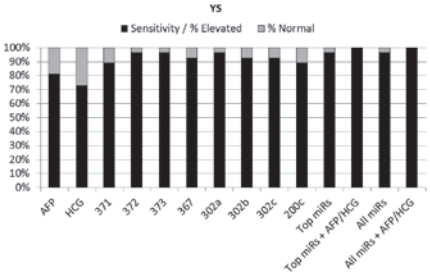


Table S2. Mean comparison of subgroups, p-values

Groups	N vs T	N vs SE	N vs NS	SE stage 1 vs >1	NS stage 1 vs >1	meta vs no meta	SE vs NS		pre- vs post-orchidectomy	N vs post-orchidectomy		N vs T		N vs no-GCC
bonferroni corrected threshold forp-value	6.390E-03	6.390E-03	6.390E-03	6.390E-03	6.390E-03	6.390E-03	6.390E-03		6.390E-03	6.390E-03		6.390E-03		6.390E-03
R_L									D	D		UK		no-GCC
	P-Value	P-Value	P-Value	P-Value	P-Value	P-Value	P-value		P-value			P-value		
hsa-miR-200c	9.685E-01	9.815E-01	9.076E-01	5.203E-02	2.443E-01	1.922E-02	8.211E-01		5.687E-02	4.510E-01		2.556E-01		6.244E-01
hsa-miR-302a	2.337E-03	9.980E-03	5.741E-03	3.132E-02	1.688E-01	5.839E-03	3.758E-01		3.024E-02	6.781E-03		9.366E-03		8.728E-01
hsa-miR-302b	4.664E-02	2.554E-01	1.906E-02	8.312E-02	1.688E-01	2.347E-02	1.468E-01		5.687E-02	5.489E-03		2.299E-02		3.228E-01
hsa-miR-302c	2.778E-02	3.087E-02	9.018E-02	5.203E-02	2.443E-01	2.125E-02	9.850E-01		2.557E-02	2.314E-02		4.168E-01		4.919E-01
hsa-miR-367	2.956E-05	1.861E-04	3.209E-04	3.876E-03	5.674E-02	5.124E-04	8.358E-01		1.622E-03	9.447E-01		1.485E-02		7.704E-01
hsa-miR-371-3p	5.296E-04	3.199E-03	1.727E-03	3.720E-01	1.107E-02	6.550E-03	1.689E-01		2.358E-04	1.806E-01		5.766E-03		1.889E-03
hsa-miR-372	6.541E-05	6.445E-04	3.209E-04	6.608E-02	9.034E-02	9.171E-03	3.365E-01		4.749E-03	4.691E-01		1.939E-01		5.594E-01
hsa-miR-373	8.981E-06	8.752E-05	1.060E-04	1.008E-02	3.426E-02	6.772E-04	2.206E-01		5.817E-03	2.769E-02		1.044E-01		6.178E-01
R_V														
hsa-miR-200c	1.857E-03	5.649E-02	2.584E-03	6.889E-01	6.065E-01	4.335E-01	3.198E-01							
hsa-miR-302a	2.473E-07	1.542E-02	2.564E-08	2.298E-01	1.012E-01	2.924E-02	3.259E-03							
hsa-miR-302b	7.062E-01	2.052E-01	6.779E-01	9.362E-01	2.061E-01	1.824E-01	1.223E-01							
hsa-miR-302c	2.377E-07	5.275E-04	1.745E-06	4.233E-01	4.405E-02	2.225E-02	2.276E-01							
hsa-miR-367	1.000E-07	5.802E-07	6.720E-08	9.270E-02	5.486E-02	8.149E-03	4.091E-01							
hsa-miR-371-3p	1.254E-07	7.174E-04	5.298E-07	1.282E-01	2.061E-01	2.625E-02	1.442E-01							
hsa-miR-372	1.000E-07	1.565E-06	1.078E-08	4.533E-02	6.780E-02	3.259E-03	5.959E-02							
hsa-miR-373	1.000E-07	5.802E-07	1.000E-09	1.735E-01	5.486E-02	4.875E-03	2.444E-01							

significant after correction for multiple testing

significant without correction for multiple testing (p<0.05)

10E-8 = <1*10^-8

References

1. Oosterhuis JW, Looijenga LH. Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer* 2005; 5: 210-222.
2. Woodward PJ, Heidenreich A, Looijenga LHJ, *et al.* Testicular germ cell tumors. In: World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs. Eble JN, Sauter G, Epstein JI, *et al.*, (ed)^(eds). IARC Press: Lyon, 2004; 217-278.
3. Reuter VE. Origins and molecular biology of testicular germ cell tumors. *Mod Pathol* 2005; 18 Suppl 2: S51-60.
4. Ye H, Ulbright TM. Difficult differential diagnoses in testicular pathology. *Arch Pathol Lab Med* 2012; 136: 435-446.
5. Bray F, Richiardi L, Ekbom A, *et al.* Trends in testicular cancer incidence and mortality in 22 European countries: Continuing increases in incidence and declines in mortality. *Int J Cancer* 2006; 118: 3099-3111.
6. Skakkebæk NE. Possible carcinoma-in-situ of the testis. *Lancet* 1972: 516-517.
7. Honecker F, Stoop H, Mayer F, *et al.* Germ cell lineage differentiation in nonseminomatous germ cell tumors. *J Pathol* 2006; 208: 395-400.
8. Looijenga LHJ, Stoop H, De Leeuw PJC, *et al.* POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. *Cancer Res* 2003; 63: 2244-2250.
9. Hart AH, Hartley L, Parker K, *et al.* The pluripotency homeobox gene NANOG is expressed in human germ cell tumors. *Cancer* 2005; 104: 2092-2098.
10. Looijenga LH. Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights. *J Pathol* 2009; 218: 146-162.
11. Gillis AJ, Stoop H, Biermann K, *et al.* Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis. *Int J Androl* 2011; 34: e160-174.
12. Looijenga LHJ. Human testicular (non)seminomatous germ cell tumors:

Patho-biology and optimal diagnostics. *J Pathol* 2008; 218: 146-162.

13. Voorhoeve PM, le Sage C, Schrier M, *et al.* A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* 2006; 124: 1169-1181.
14. Gillis AJ, Stoop HJ, Hersmus R, *et al.* High-throughput microRNAome analysis in human germ cell tumours. *J Pathol* 2007; 213: 319-328.
15. Palmer RD, Murray MJ, Saini HK, *et al.* Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets. *Cancer Res* 2010; 70: 2911-2923.
16. Novotny GW, Belling KC, Bramsen JB, *et al.* MicroRNA expression profiling of carcinoma in situ cells of the testis. *Endocr Relat Cancer* 2012; 19: 365-379.
17. Horwich A, Shipley J, Huddart R. Testicular germ-cell cancer. *Lancet* 2006; 367: 754-765.
18. Van Leeuwen FE, Stiggelbout AM, Van den Belt-Dusebout AW, *et al.* Second cancer risk following testicular cancer: A follow-up study of 1,909 patients. *JournalofClinicalOncology* 1993; 11: 415-424.
19. Haugnes HS, Wethal T, Aass N, *et al.* Cardiovascular risk factors and morbidity in long-term survivors of testicular cancer: a 20-year follow-up study. *J Clin Oncol* 2010; 28: 4649-4657.
20. von Eyben FE, Liu FJ, Amato RJ, *et al.* Lactate dehydrogenase isoenzyme 1 is the most important LD isoenzyme in patients with testicular germ cell tumor. *Acta Oncol* 2000; 39: 509-517.
21. Murray MJ, Halsall DJ, Hook CE, *et al.* Identification of microRNAs From the miR-371~373 and miR-302 clusters as potential serum biomarkers of malignant germ cell tumors. *Am J Clin Pathol* 2011; 135: 119-125.
22. Belge G, Dieckmann KP, Spiekermann M, *et al.* Serum Levels of MicroRNAs miR-371-3: A Novel Class of Serum Biomarkers for Testicular Germ Cell Tumors? *Eur Urol* 2012.
23. Murray MJ, Coleman N. Testicular cancer: A new generation of biomarkers for malignant germ cell tumours. *Nat Rev Urol* 2012.
24. Mestdagh P, Van Vlierberghe P, De Weer A, *et al.* A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 2009; 10: R64.
25. de Jong J, Stoop H, Gillis AJ, *et al.* Further characterization of the first seminoma cell line TCam-2. *Genes Chromosomes Cancer* 2008; 47: 185-196.

26. Shimono Y, Zabala M, Cho RW, *et al.* Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* 2009; 138: 592-603.
27. Van Echten-Arends J, Oosterhuis JW, Looijenga LHJ, *et al.* No recurrent structural abnormalities in germ cell tumors of the adult testis apart from i(12p). *Genes Chromosom & Cancer* 1995; 14: 133-144.
28. Zubakov D, Boersma AW, Choi Y, *et al.* MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *Int J Legal Med* 2010; 124: 217-226.
29. Vandesompele J, De Preter K, Pattyn F, *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002; 3: RESEARCH0034.
30. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 2004; 64: 5245-5250.
31. Kawakami T, Okamoto K, Ogawa O, *et al.* XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer. *Lancet* 2004; 363: 40-42.
32. Chen X, Liang H, Zhang J, *et al.* Horizontal transfer of microRNAs: molecular mechanisms and clinical applications. *Protein Cell* 2012; 3: 28-37.
33. Somasundaram R, Herlyn M. Melanoma exosomes: messengers of metastasis. *Nat Med* 2012; 18: 853-854.

CHAPTER 7

General discussion

7.1) Introduction:

Totipotent stem cells, present in the early stages of the development of multicellular organisms, can generate all cell types (somatic and extra-embryonic differentiation) and are capable of self-renewal [1]. Embryonic stem (ES) cells were first derived from the inner cell mass of mouse blastocysts in the early 1980s [2, 3]. Pluripotent stem cells, primarily ES cells, have been used extensively in studies of embryogenesis, development and gene function [4]. Some factors and signalling pathways that are involved in cell proliferation, differentiation and self-renewal of ES cells, are associated with oncogenesis [5]. In many respects, germ cell tumorigenesis resembles early embryogenesis because the cells of origin are embryonic germ cells, either primordial germ cells (PGCs) or gonocytes [6]. Thus they provide an interesting model to study early development as well as cell transformation in tumorigenesis [5].

Malignant testicular germ cell tumors of adolescents and young adults, here referred to as germ cell cancer (GCC), are the most common malignancy in males between the ages of 20 and 40 years. GCC is thought to be derived from embryonic cells committed to the germ cell lineage that are blocked in their physiological process of maturation. This will give rise in the testis to carcinoma *in situ* (CIS), which is the precursor of GCC [7]. The two main clinical and pathological subtypes of GCC are seminomas (SE) and non-seminomas (NS). SE consists of transformed embryonic germ cells that closely resemble the PGCs/gonocytes. These malignant cells are blocked in their normal maturation/differentiation process and cannot undergo normal spermatogenesis, but accumulate instead [8]. NS, resulted from reprogramming of CIS or SE cells to pluripotent ES-like, so-called embryonal carcinoma (EC) cells, are the stem cells of NS and can be composed of different histological elements. In fact, EC is an accumulation of the undifferentiated stem cells, in principle able to give rise to extra-embryonic lineages known as yolk sac tumor (YS) and choriocarcinoma (CH), as well as somatic lineage (teratoma) [6, 9]. Down-regulation of BLIMP1, which suppresses activation of pluripotency in the normal germ lineage, is mechanistically involved in reprogramming of CIS/SE cells to pluripotency [10]. Several transcription factors play a key role in the maintenance of undifferentiated state of ES cells, including OCT3/4, SOX2/SOX17 and NANOG, discussed in this thesis. These pluripotency factors interact with each other and form an integrated network in ES cells [11, 12] . In a very recent study, by analysing Sox17/Oct3/4 co-binding sites, it was shown that these two transcription factors co-operate and bind to a specific compressed Sox/Oct motif. A few hundred specific enhancer sites bound by both Oct3/4 and Sox17 were identified. In the same study, it was shown that this specific partnership targets defined loci in order to regulate cell fate determination. Cooperation between these transcription factors is therefore important for its function, i.e., binding to specific enhancer sites [13]. Furthermore, they co-occupy promoters of many target genes [11, 12], and control their own expression, for which, a good example is the negative feedback loop between Oct3/4 and Nanog, as shown in the mouse. Oct3/4 maintains Nanog expression by directly binding to

its promoter; Oct3/4 activates the Nanog promoter when expressed below a normal level, yet represses it at or above normal concentration in ES cells. This means that when the expression level of Oct3/4 rises above a steady level, it represses its own promoter as well as Nanog thus maintaining the ES cell properties [14].
Since GCC is thought to be derived from embryonic cells committed to the germ cell lineage that are blocked in their physiological maturation, it is expected and in fact also proven that ES cell markers such as OCT3/4 an SOX2 can be used as highly informative diagnostic markers for GCC [15, 16]. Besides specific proteins, also non-coding RNAs play a role in establishment and maintenance of (embryonic) stem cells. These include the so-called microRNAs (miRs). Over the past decade, the role of miRs in controlling gene regulation has been extensively discussed [17-19]. miRs represent a class of small regulatory RNAs that have been implicated as essential regulators in development and various aspects of stem cell biology [20].
MiRs are endogenous small noncoding RNAs ranging from 19-25 nucleotides in length, encoded by the genome, and transcribed as long primary transcripts (pri-miRs)[21]. Pri-miRs are cropped by the RNase III Drosha [22] to the hairpin-shaped precursor (pre-miR) which is exported to cytoplasm through the nuclear transporter exportin 5 [23]. Then the cytoplasmic Dicer (another RNase III-like endonuclease) processes the pre-miRs into 20 to 22 nucleotide duplexes [24]. The RNA is subsequently unwound by a helicase activity and incorporated into a RNA-induced silencing complex (RISC). At this stage, miRs can mediate their target gene activity either by target mRNA cleavage or translational inhibition. This choice is made based on the degree of complimentarily between the miR and target gene. Near-perfect complimentarily results in cleavage, followed by RNA degradation whereas partial complementarily results in translational inhibition [25].
miRs are phylogenetically conserved and have been shown to be instrumental in a wide variety of key biological processes including cell cycle regulation, apoptosis, control of metabolic pathways, imprinting, differentiation and maintenance of "stemness" among others [17].

7.2) ES pluripotency proteins and miRs in GCC:

ES cells are known to have a unique cell cycle characterized by a short G_1-S transition which facilitates quick cell movement through G_1 to S phase allowing the cells to proliferate rapidly [26]. This specific characteristic of stem cells makes them hypersensitive to DNA damaging agents in which miRs are shown to play a central role [27, 28]. The role of interaction between core ES pluripotency factors and ES miRs, in particular miR-371-3 and miR-302/367 in the maintenance of the ES/germ cell identity has been described (Chapter 2). Tight connection between core ES cell pluripotency proteins and ES miRs seems to play a major role in the pathogenesis of GCC as well as their exceptional sensitivity to chemotherapy. The two ES cell miR clusters, miR-371-3 and miR-302/367 are differentially expressed between various

subtypes of GCC being highly expressed in EC, SE and YS and absent in teratoma [29]. High expressions of miR-371-3 and miR-302/367 clusters in GCC, like-wise ES cells, cause short G_1-S transition via suppression of LATS2 [30] and Cyclin D1 [31] respectively. This in turn leads to sensitivity of these tumors to DNA damaging agents such as cisplatin.

A good example of the connection between ES cell proteins and ES miRs became apparent with down-regulation of ES cell genes experiments. Targeted down-regulation of OCT3/4 and SOX2 in NTera2 caused initiation of differentiation [32] and resulted in down-regulation of miR-302/367 cluster up to 50%. This has also been demonstrated in human ES cells that the expression of miR-302a under OCT4/SOX2-depleted conditions was reduced by more than 50% in comparison with the level in cells transfected with non-targeting siRNAs [31]. These reveal that ES specific proteins act as upstream regulators of the miR302/367 cluster confirming the presence of potential binding sites for key pluripotency factors in the promoter sequence of these miRs. In this context it is of interest to test the effect of down-regulation of OCT3/4 and SOX2 on the ES cell miRs expression profile including miR-371-3 and miR-302/367 in other EC cell lines such as NCCIT and 2102Ep as well. This is one of the aims of our research regarding these miRs.

Since miR-371-3 is involved in inhibition of cellular senescence after oncogenic stress [30], presence of wild type P53 in GCC, in contrast to most solid cancers, is likely explained. We have shown that miR-371-3 is able to mimic the effect of mutated p53 by preventing RAS-induced oncogenic senescence in GCC by inhibition of its downstream target LATS2 in the P53 pathway. In the context of the effect of miR-371-3 and miR-302/367 on G_1-S transition and cisplatin sensitivity in GCC, we performed a functional study using luciferase reporter construct including the 3'UTR of LATS2. We showed that indeed miR-371-3 could reduce the luciferase activity of LATS2 within the construct efficiency in NTera2 cells (Chapter 2). It remains to be investigated what the effect of suppression of miR-371-3 on the cisplatin sensitivity of these cells will be. Since suppression of miR-371-3 enhances the level of LATS2, longer G_1-S transition will be induced which can possibly result in resistance to cisplatin. This hypothesis can also be tested for miR-302/367 cluster by using a similar luciferase reporter system with replacing the target gene with Cyclin D1. It is also reported that LATS2 is the downstream target of miR-302/367 cluster [20], therefore, one interesting hypothesis to be tested is whether suppression of total miR-302/367 cluster by means of sponge, increases the level of LATS2 expression and consequently causes longer G_1-S transition which in turn will lead to resistance of the cells to DNA damaging agents such as cisplatin.

Furthermore, it is of interest to investigate the impact of these miRs on the metastatic ability of GCC cell lines. The results of such tests can further contribute to the understanding of the significance in biological and clinical impact of these miRs in GCC.

SOX2 and EC:

Most recently, it has been reported that inhibition of SOX2 might be of therapeutic potential for EC [33]. In that particular study, SOX2 down-regulation in the EC cell line NEC8, when established *in vivo*, induced tumor growth suppression in the case of a limited tumor size. In contrast, down-regulation had no effect on progression in the case of a larger tumor size, related to loss of SOX2 expression to some extent and possibly differentiation. This study indicates that suppression of SOX2 expression might be useful to block cell proliferation of early stage EC. Interestingly, as a result of our array CGH study on multiple EC cell lines, NCCIT cell line showed amplification at the long arm of chromosome 3, band q23 including the SOX2 gene locus [34].

The effect of down-regulation of OCT3/4 and SOX2 in NCCIT is investigated (Chapter 3). Inhibition of OCT3/4 resulted in induction of differentiation, while inhibition of SOX2 led to cell death. This suggests that survival of NCCIT is dependent on the presence of SOX2, referred to as oncogene-dependence [35]. This is consistent with the anti-apoptotic property of SOX2 in prostate cancer [36]. The potential of NCCIT cells to show differentiation *in vivo* has also been investigated by generating multiple xenografts. In spite of induction of differentiation, supported by loss of OCT3/4, the lineages formed were characterized by continuous expression of SOX2. In line with this study, it is shown that undifferentiated neural progenitor cells, which are capable of producing multiple cell lineages, as found in the *in vitro* study, were SOX2 positive and after differentiation, they were migrating out of the neurospheres [37]. This emphasizes the critical role of SOX2 in formation of different lineages. In addition, its down-regulation results in induction of differentiation leading to formation of metastatic clones. Therefore, the pros and cons of SOX2 down-regulation in the clinical setting requires careful consideration.

Because OCT3/4 and SOX2 are highly expressed in EC, in the same study, we investigated whether their high expression is due to their genomic amplification. Therefore a series of 12 pure EC and 34 mixed NS including EC component were selected for the analysis using DNA-FISH technique. No amplification for OCT3/4 and SOX2 was found in the cases investigated. This suggests that high expression of OCT3/4 and SOX2 in EC is due to their pluripotent origin rather than genomic amplification affecting their expression. In addition, lack of SOX2 amplification in EC components suggests that SOX2 oncogene dependence is not a frequent mechanism in GCC, which indeed questions the approach of targeted therapy in a clinical setting. In this context it might be of interest to check whether the NEC8 cell line contains SOX2 amplification [33].

7.3) MiR-371-3 and miR-302/367 in GCC:

As mentioned above, cytoplasmic Dicer plays a crucial role in the development of miRs by processing the pre-miRs into 20 to 22 nucleotide duplexes [24]. These duplexes consist of a principle miR strand and the complimentary miR stand, referred to as miR* [22]. *DICER1* encodes a 218kd ribonuclease (RNase) III endonuclease that participates in the generation of small RNAs [38] (schematically presented in the Figure). The RNase IIIb domain of DICER1 is known to cut the miR strand while RNase IIIa domain is responsible to cut miR* strand [39]. In this context, germline mutations in DICER1 have been identified in cancer patients with pleuropulmonary blastoma [40] and non-epithelial ovarian cancers [41]. The highest frequency of DICER1 mutation was found in the Sertoli-Leydig cells of the latter where 60% contained a somatic variant within one of four host-spot codons. All variants affected amino acids acting as metal binding sites within the RNase IIIb domain of DICER1 which are critical for miR interaction and cleavage, leading to reduced RNase IIIb activity. Mutations in the RNase IIIb domain are shown to alter DICER1 activity rather than abolish its activity. It is proposed that a specific miR expression profile would be derived as a result of changes in RNase IIIb activity [41]. A number of studies have reported the increased expression of miR-371-3 and miR-302/367 clusters in GCC [29, 42]. These are indeed ES cell miRs and are up-regulated in pluripotent cell types [43] (see above). Regarding the mentioned facts, it is possible to hypothesize that an oncogenic miR profile is derived from altered DICER1 activity in the RNases IIIb domain. One hypothesis is that such activity may result in a positive bias toward the processing or selection of the RNase IIIa and subsequently miR* strand [41]. In this context, recent studies have suggested that miR* species may be important in gene regulation rather than simply being an inert strand [44, 45].

On the other hand, it seems to be controversial whether the hot-spot mutations in DICER1 protein have any effect on miR processing as it is reported that the ability of DICER1 to load miR into RISC does not depend on the integrity of its RNaseIII domains [46]. Regarding the impact of RNase IIIb mutations on miR biogenesis, including miR processing or RISC loading in the context of cancer development, it is interesting that mutation in the metal binding residues in the RNase IIIb domain of DICER1 causes inhibition of cleavage of the 3'end of the 5p-strand of the miRs. This is especially of interest because it is related to normal maturation of only 3p-strand of miR and a large strand bias in the mature miR population. This will enhance the overall mature miR level and consequently targeting a restricted set of mRNAs through the RISC process [47]. Therefore, we have analyzed copy number variation and sequence analysis of miR-371-3 and miR-302/367 clusters loci in order to identify whether any mutations are likely to be responsible for altering these miR levels. Analysis of initially 14 GCC samples including both SE and NS resulted in identification of one somatic *DICER1 (*hotspot) mutation which raised the possibility of existing mutations within this domain of DICER1 in GCC. Therefore, we used high resolution melting curve (HRM), a

powerful method for identifying variants in the four mutation hot-spots codons indentified in the RNase IIIb domain of *DICER1* which was found previously [41]. In our set-up, 96 GCC samples including 32 SE and 64 NS were analysed (Chapter 4). To first demonstrate that HRM was able to detect the specific variants previously identified in DICER1, six different DNA templates were created, each containing one of the *DICER1* mutations described before [41]. These six mutations covered 79% of all cases where a mutation within one of the hot-spot codons was identified. All six variants could clearly be identified using HRM analysis. Then PCR primers were used to amplify the corresponding genomic regions in 96 GCC samples to screen for these six variants. No sample showed an aberrant melting curve. As the initial analysis only covered a small number of genomic sequence (70bp and 68bp), samples were rescreened by using PCR primers described before [41] to amplify the corresponding genomic regions. No mutations within those sites were found, although a novel mutation within the RNase IIIb domain in one SE sample was detected. The affected amino acid was not predicted to directly function as a metal-binding site; however, it is within a contiguous sequence of 37 amino acids that show 100% conservation across at least 42 species, thus supports the hypothesis that this region within the RNase IIIb domain is critical for normal *DICER1* function. Thus, it is predicted that this mutation disturbs DICER1 function. As a result of this study, we conclude that mutations within the RNase IIIb domain of DICER1 are not frequent in GCC. It is likely that the high rate of mutations observed in Sertoli-Leydig cells in non-epithelial ovarian cancer [41] is due to their different origins and specific characteristics and development.

After excluding DICER1 mutations, it was hypothesized that the high expression level of miR-302/367 and miR-371-3 clusters in GCC might be due to variations in copy number of the miR locus, or sequence changes that influence the transcription process or the transcript stability. We used multiplex ligation-dependant probe amplification (MLPA) and HRM curve analysis to screen 242 GCC derived DNA samples for copy number variations and sequence mutation affecting the two ES miR cluster, miR-371-3 and miR-302/367 loci in a large cohort of GCC (242 samples) (Chapter 5). An apparent deletion of miR-302c in 6% of cases was found to be due to a 4bp deletion that disturbed MLPA probe ligation. This was found to be equivalent to the frequency of this mutation in the general population. Two other variants were found for miR-367. The first variant was not identified in the DNA of other GCC cases, nor is present in any sequence variant databases. The latter variant showed a high degree of conservation across species, however, it is located outside the seed sequence, which is considered as the most critical region for recognising a complementary sequence in the 3'UTR of target mRNAs. Therefore, these variants are unlikely to have any pathogenic role.

Thus, mutations in the loci of these miRs are unlikely to be responsible for altering the miR levels. Therefore, the high expression levels of miR-371-3 and miR-302/367 in GCC are likely to be related to the pluripotent cell type of their origin rather than specific genetic mutations

affecting expression levels [9]. This resembles high expression level of OCT3/4, NANOG and SOX2 in EC while to our knowledge no genomic amplification for OCT3/4 and NANOG has been reported, whereas, SOX2 is amplified in NCCIT cell line (see above) and is reported to be amplified in squamous cell carcinoma as well [48]. Moreover, its over-expression is reported in a number of cancers [49-52] supporting the role of SOX2 as relevant oncogene in these malignancies. These reports suggest that SOX2 can activate important gene cascades which are involved in initiation and progression of tumors and maintenance of a poorly differentiated state.

Another mechanism that can cause aberrant miR expression levels is epigenetic alterations. Epigenetic modifications, such as aberrant CpG methylation, contribute to deregulated gene expression in cancer cells. Methyl-groups are transferred to the carbon atom at position 5 of cytosines by DNA methyltransferases (DNMT1 or DNMT3A and DNMT3B). Methylation of promoter-associated CpG dinucleotides (especially in CpG islands) usually correlates with reduced transcription levels of the respective gene [53]. The association between methylation in CpG islands of the promoter of some miRs and cancer has been reported in several studies [54-56]. We hypothesized whether the expression pattern of miR-371-3 in EC, SE and YS, is due to the methylation status of the supposed promoter region. Therefore we assessed the methylation status of 12 CpG sites within the CpG island of the promoter of miR-371-3. A significant correlation between expression level of miR-373 and methylation percentage of the promoter in the cell lines of GCC, containing seven cell lines, was found. Furthermore, correlation between expression of miR-371-3 cluster and methylation status of their promoter in micro-dissected EC and SE tumor samples (in total 10) in comparison to their matched normal sample were also analyzed, in which the expression of miR-371-3 cluster did not show correlation with the DNA methylation pattern of the promoter of this miR cluster. However, other epigenetic mechanisms such as histone modification of the region are also known to have profound effect on gene expression. Unlike methylation, histone modification can lead to gene activation depending on the type of modification present [54]. Therefore, studying other epigenetic mechanisms that could contribute to the regulation of gene expression could give more insight in this matter.

7.4) Clinical application of miR-371-3 and miR-302/367 in GCC:

GCC are highly treatable by surgery, possibly followed by irradiation and/or chemotherapy, depending on stage and histology [57]. Currently, the serum markers such as Alpha Feto-Protein (AFP) and human Chorionic Gonatrophine (hCG) are used, both for diagnosis, risk assessment as well as follow-up of YS, CH, and pure teratoma, respectively. However, for a significant number of GCC patients, i.e. those with pure SE and EC and/or pure teratoma, no informative serum markers are available. Thus there is need for new serum markers for diagnosis and follow-up of these patients. Recently, four studies reported on detection of miR-

371-3 and miR-302/367 clusters in serum of GCC patients [58-61]. However, all these studies have some technical limitations due to the lack of relevant quality controls and normalization steps. A robust and informative pipeline for determining miR levels in serum was developed (Chapter 6). For this purpose, stringent quality controls were included in a routine serum miR protocol, sensitive enough to detect low serum levels and minimal frequency of false negative results. Therefore, the protocol included a magnetic bead-based purification step for detection of specific miRs, referred to as the Targeted Serum miR or TSmiR test. This approach allowed standardized comparison between various data sets. The target of interest (TOI) included seven miRs known to be expressed in SE, EC and YS, belonging to the previously mentioned two distinct clusters: miR-371-3 and miR-302/367. The TSmiR test was performed on four independent patient series including Rotterdam-Learning (R_L) and validation (R_V) set as well as two previously published series (UK, including GCC samples and controls [58, 60] and Germany [59]) composed of pre- and post surgery samples. Suitable targets for normalization were identified, allowing sample by sample comparison. In addition, inter-plate and inter-assay variations were evaluated. A panel of nine possible targets for normalization was studied among which miR-20a and miR-93 were the most stable combination for normalization based on established algorithms for stability analysis (Normfinder [62] and geNorm [63]).

Statistical analyses indicated that the miR-371-3 and miR-367 were found to be most informative at the time of diagnosis. In fact, these specific serum miR levels allowed stringent separation between control and patient samples. Combined interpretation of the established serum markers AFP/hCG and the TSmiR test led to the correct classification of all investigated serum samples of GCC patients. Application of the serum markers AFP and hCG alone, resulted in about 50% false-negative cases while application of only the TSmiR test resulted in a single malignant NS serum sample as false negative. This demonstrates the superiority of the TSmiR test over the currently applied AFP/hCG test or the primary diagnosis of GCC. Application of this protocol might lead to a significant improvement in clinical management of GCC patients, especially for marker-negative cases (i.e., no elevated AFP or hCG serum levels).

In addition, analysis of pre- and post- surgery serum samples of paired primary stage I SE and NS before and after orchidectomy showed a significantly higher serum level of miR-371-3 and miR-367 in the pre-surgery samples compared to post-surgery samples. In fact, there was no serum miR level difference in post-surgery samples compared to normal (non-GCC) serum samples. This reveals the putative value of serum level of miR-371-3 and miR-367 as biomarkers for diagnosis and follow-up of GCC patients.

It is controversial whether these miRs enter into the serum actively or passively via exosomes [64, 65] which needs further investigation. If it appears that miRs are encapsulated in microvesicles, particularly exosomes, it will shed novel light on the biology behind the circulation of miRs in the bloodstream. Furthermore, it might simplify the purification step and therefore will have a great impact on the technical aspect of this approach.

In a separate study, in order to investigate whether the pattern of miR in serum correlates with the pattern observed in the matched primary tumors, we tested a series of 10 GCC samples including stage I and higher stage. Statistical analysis showed no correlation between the levels of miRs investigated within the cancer and matched serum samples. This suggests that specific (active) mechanisms are likely involved in this process. In addition, it is of interest to investigate further whether there are differences between serum miR levels in GCC patients without and with metastasis in a large series. This might have additional impact of the clinical utility of the assay as developed.
In conclusion, the work presented in this thesis shed new light on the role of two clusters of ES miRs, miR-371-3 and miR-302/367 (in close collaboration with ES pluripotency genes) in the pathogenesis of GCC (in connection with embryogenesis) and their (exceptional) sensitivity to chemotherapy as well as on their clinical application in diagnosis and follow-up of GCC patients (presented in the Figure).

7.5) Future perspective:

Most recently it has been shown that specific-site phosphorylation of OCT3/4 regulated by AKT, promotes stemness of EC cells (i.e., the NCCIT cell line) compared ES cells [66]. Presence of this site-specific phosphorylation promotes release of the OCT3/4 protein from the *AKT1* promoter, resulting in induced expression and consequently suppression of apoptosis. Simultaneously, this will enhance the capacity of OCT3/4 to form a complex with NANOG and SOX2, promoting pluripotency. This additional effect promotes tumorigenic capacity of EC compared to ES, which might be of interest to be further investigated in the context of therapeutic effect. In fact, this finding suggests that early stages of GCC can be detected by analysis of the phosphorylated or non-phosphorylated OCT3/4 status, which can have great impact on understanding the earliest pathobiological changes, which significantly affect on the possibilities of early diagnosis of these cancers.
In the context of understanding the biology underlying the sensitivity of GCC to chemotherapy, functional studies involving manipulation of the two ES miR clusters, miR-371-3 and miR-302/367 would be of interest. Moreover, finding possible new targets for these miRs and mediating the key pluripotency factors (which are extensively discussed in this thesis), would decipher new aspects regarding the mechanism in which these ES miRs in collaborations with ES pluripotency factors control the cell cycle. Moreover, the specific property of EC cells, being the malignant counterparts of ES cells, facilitates a unique opportunity for parallel experiments in both EC and ES cell lines regarding ES pluripotency factors and ES miR function.

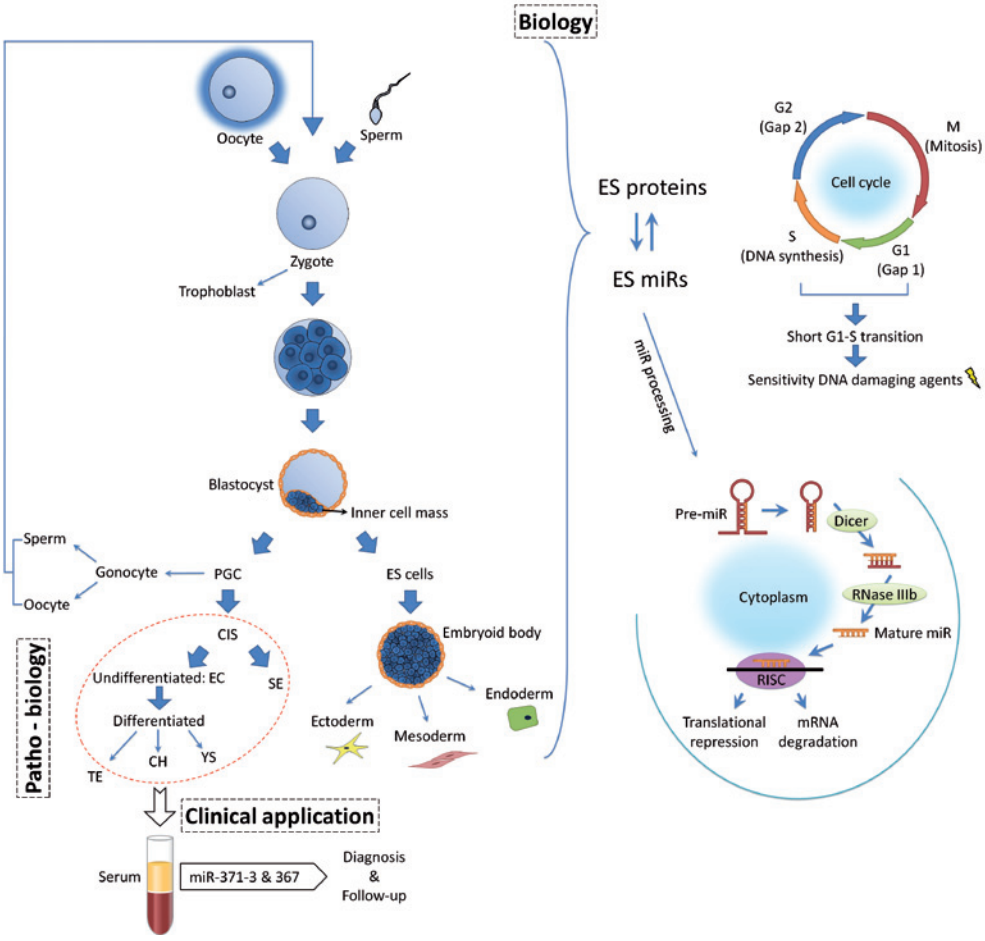


Figure. Schematic representation of the interrelationship between biology, patho- and clinical biology of ES miRs and proteins.

Origin and establishment of pluripotent ES cells and PGC from inner cell mass of the blastocyst. ES cells form embryoid body which gives rise to all three germ layers; ectoderm, mesoderm and endoderm and consequently all the organs of the body. PGC can give rise to malignant counterpart of embryonic germ cell. CIS is the precursor of GCC and gives rise to either SE or NS. The latter comprise of undifferentiated cells EC. EC can differentiate and give rise to YS, CH and TE. ES pluripoteny proteins and miRs work in an orchestrated machinery to regulate all these processes including ES cells regulation, GCC development and sensitivity of these to DNA damaging agents by keeping short G_1-S transition . Mature miR is produced in the cytoplasm by Dicer, which cuts the pre-miR to a short RNA duplex and then RNase IIIb domain of Dicer generates the mature miR which then incorporates into RISC. At this stage miR can mediate regulation of its target gene by translational inhibition or RNA degradation. Serum expression level of two ES miR cluster, miR-371-3 and miR-367 can be used for diagnosis and follow-up of GCC patients. These are expressed in EC, SE and YS and are absent in TE. Abbreviations used: ES cells: embryonic stem cells, PGC: primordial germ cell, CIS: carcinoma in situ, SE: seminoma, EC: embryonal carcinoma, YS: yolk sac tumor, CH: choriocarcinoma, TE: teratoma. miR: microRNA, Pre-miR: precursor miR, RISC: RNA silencing induced complex.

References:

1. Gearhart, J., *New potential for human embryonic stem cells.* Science, 1998. 282(5391): p. 1061-2.
2. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. 292(5819): p. 154-6.
3. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. 78(12): p. 7634-8.
4. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. 282(5391): p. 1145-7.
5. Yang, S., et al., *Tumourigenic characteristics of embryonal carcinoma cells as a model for studying tumour progression of human embryonic stem cells.* Cell Prolif, 2012. 45(4): p. 299-310.
6. Oosterhuis, J.W. and L.H. Looijenga, *Testicular germ-cell tumours in a broader perspective.* Nat Rev Cancer, 2005. 5(3): p. 210-22.
7. Skakkebaek, N.E., *Possible carcinoma-in-situ of the testis.* Lancet, 1972. 2(7776): p. 516-7.
8. van de Geijn, G.J., R. Hersmus, and L.H. Looijenga, *Recent developments in testicular germ cell tumor research.* Birth Defects Res C Embryo Today, 2009. 87(1): p. 96-113.
9. Looijenga, L.H., et al., *Relevance of microRNAs in normal and malignant development, including human testicular germ cell tumours.* Int J Androl, 2007. 30(4): p. 304-14; discussion 314-5.
10. Eckert, D., et al., *Expression of BLIMP1/PRMT5 and concurrent histone H2A/H4 arginine 3 dimethylation in fetal germ cells, CIS/IGCNU and germ cell tumors.* BMC Dev Biol, 2008. 8: p. 106.
11. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. 122(6): p. 947-56.
12. Wu, Q., et al., *Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells.* J Biol Chem, 2006. 281(34): p. 24090-4.
13. Aksoy, I., et al., *Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm.* EMBO J, 2013. 32(7): p. 938-53.
14. Pan, G., et al., *A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal.* FASEB J, 2006. 20(10): p. 1730-2.
15. Looijenga, L.H., et al., *POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.* Cancer Res, 2003. 63(9): p. 2244-50.
16. de Jong, J., et al., *Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.* J Pathol, 2008. 215(1): p. 21-30.
17. Ambros, V., *The functions of animal microRNAs.* Nature, 2004. 431(7006): p. 350-5.
18. Barroso-delJesus, A., et al., *Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter.* Mol Cell Biol, 2008. 28(21): p. 6609-19.
19. Al-Ali, B.M., et al., *MicroRNAs in renal cell carcinoma: implications for pathogenesis, diagnosis, prognosis and therapy.* Anticancer Res, 2012. 32(9): p. 3727-32.
20. Lipchina, I., L. Studer, and D. Betel, *The expanding role of miR-302-367 in pluripotency and reprogramming.* Cell Cycle, 2012. 11(8): p. 1517-23.
21. Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* EMBO J, 2002. 21(17): p. 4663-70.
22. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. 425(6956): p. 415-9.
23. Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. 303(5654): p. 95-8.
24. Ketting, R.F., et al., *Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans.* Genes Dev, 2001. 15(20): p. 2654-9.
25. Wienholds, E. and R.H. Plasterk, *MicroRNA function in animal development.* FEBS Lett, 2005. 579(26): p. 5911-22.
26. He, S., D. Nakada, and S.J. Morrison, *Mechanisms of stem cell self-renewal.* Annu Rev Cell Dev Biol, 2009. 25: p. 377-406.
27. Hong, Y. and P.J. Stambrook, *Restoration of an absent G1 arrest and protection from apoptosis in embryonic stem cells after ionizing radiation.* Proc Natl Acad Sci U S A, 2004. 101(40): p. 14443-8.

28. Wang, Y., et al., *MicroRNAs in embryonic stem cells.* J Cell Physiol, 2009. 218(2): p. 251-5.
29. Gillis, A.J., et al., *High-throughput microRNAome analysis in human germ cell tumours.* J Pathol, 2007. 213(3): p. 319-28.
30. Voorhoeve, P.M., et al., *A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.* Cell, 2006. 124(6): p. 1169-81.
31. Card, D.A., et al., *Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells.* Mol Cell Biol, 2008. 28(20): p. 6426-38.
32. Gillis, A.J., et al., *Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis.* Int J Androl, 2011. 34(4 Pt 2): p. e160-74.
33. Ushida, H., et al., *Therapeutic potential of SOX2 inhibition for embryonal carcinoma.* J Urol, 2012. 187(5): p. 1876-81.
34. de Jong, J., et al., *Further characterization of the first seminoma cell line TCam-2.* Genes Chromosomes Cancer, 2008. 47(3): p. 185-96.
35. Sharma, S.V. and J. Settleman, *Oncogene addiction: setting the stage for molecularly targeted cancer therapy.* Genes Dev, 2007. 21(24): p. 3214-31.
36. Jia, X., et al., *SOX2 promotes tumorigenesis and increases the anti-apoptotic property of human prostate cancer cell.* J Mol Cell Biol, 2011. 3(4): p. 230-8.
37. Lim, J.H., et al., *Isolation and characterization of neural progenitor cells from adult canine brains.* Am J Vet Res, 2012. 73(12): p. 1963-8.
38. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. 116(2): p. 281-97.
39. Zhang, H., et al., *Single processing center models for human Dicer and bacterial RNase III.* Cell, 2004. 118(1): p. 57-68.
40. Hill, D.A., et al., *DICER1 mutations in familial pleuropulmonary blastoma.* Science, 2009. 325(5943): p. 965.
41. Heravi-Moussavi, A., et al., *Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers.* N Engl J Med, 2012. 366(3): p. 234-42.
42. Palmer, R.D., et al., *Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets.* Cancer Res, 2010. 70(7): p. 2911-23.
43. Suh, M.R., et al., *Human embryonic stem cells express a unique set of microRNAs.* Dev Biol, 2004. 270(2): p. 488-98.
44. Guo, L. and Z. Lu, *The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule?* PLoS One, 2010. 5(6): p. e11387.
45. Yang, J.S., et al., *Widespread regulatory activity of vertebrate microRNA* species.* RNA, 2011. 17(2): p. 312-26.
46. Noland, C.L., E. Ma, and J.A. Doudna, *siRNA repositioning for guide strand selection by human Dicer complexes.* Mol Cell, 2011. 43(1): p. 110-21.
47. Anglesio, M., et al., *Cancer-associated somatic DICER1 hotspot mutations cause defective miRNA processing and reverse strand expression bias to predominantly mature 3p strands through loss of 5p strand cleavage.* J Pathol, 2012.
48. Brcic, L., et al., *Morphologic and clinicopathologic features of lung squamous cell carcinomas expressing Sox2.* Am J Clin Pathol, 2012. 138(5): p. 712-8.
49. Gangemi, R.M., et al., *SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity.* Stem Cells, 2009. 27(1): p. 40-8.
50. Bass, A.J., et al., *SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas.* Nat Genet, 2009. 41(11): p. 1238-42.
51. Leis, O., et al., *Sox2 expression in breast tumours and activation in breast cancer stem cells.* Oncogene, 2012. 31(11): p. 1354-65.
52. Bae, K.M., et al., *Expression of pluripotent stem cell reprogramming factors by prostate tumor initiating cells.* J Urol, 2010. 183(5): p. 2045-53.
53. Baer, C., R. Claus, and C. Plass, *Genome-Wide Epigenetic Regulation of miRNAs in Cancer.* Cancer Res, 2013. 73(2): p. 473-7.
54. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer.* Carcinogenesis, 2010. 31(1): p. 27-36.

55. Toyota, M., et al., *Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer.* Cancer Res, 2008. 68(11): p. 4123-32.
56. Lujambio, A., et al., *A microRNA DNA methylation signature for human cancer metastasis.* Proc Natl Acad Sci U S A, 2008. 105(36): p. 13556-61.
57. Horwich, A., J. Shipley, and R. Huddart, *Testicular germ-cell cancer.* Lancet, 2006. 367(9512): p. 754-65.
58. Murray, M.J., et al., *Identification of microRNAs From the miR-371~373 and miR-302 clusters as potential serum biomarkers of malignant germ cell tumors.* Am J Clin Pathol, 2011. 135(1): p. 119-25.
59. Belge, G., et al., *Serum levels of microRNAs miR-371-3: a novel class of serum biomarkers for testicular germ cell tumors?* Eur Urol, 2012. 61(5): p. 1068-9.
60. Murray, M.J. and N. Coleman, *Testicular cancer: a new generation of biomarkers for malignant germ cell tumours.* Nat Rev Urol, 2012. 9(6): p. 298-300.
61. Dieckmann, K.P., et al., *MicroRNAs miR-371-3 in serum as diagnostic tools in the management of testicular germ cell tumours.* Br J Cancer, 2012. 107(10): p. 1754-60.
62. Vandesompele, J., et al., *Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.* Genome Biol, 2002. 3(7): p. RESEARCH0034.
63. Andersen, C.L., J.L. Jensen, and T.F. Orntoft, *Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets.* Cancer Res, 2004. 64(15): p. 5245-50.
64. Chen, X., et al., *Horizontal transfer of microRNAs: molecular mechanisms and clinical applications.* Protein Cell, 2012. 3(1): p. 28-37.
65. Somasundaram, R. and M. Herlyn, *Melanoma exosomes: messengers of metastasis.* Nat Med, 2012. 18(6): p. 853-4.
66. Lin, Y., et al., *Reciprocal regulation of akt and oct4 promotes the self-renewal and survival of embryonal carcinoma cells.* Mol Cell, 2012. 48(4): p. 627-40.

CHAPTER 8

Summary

Samenvatting

Summary

In mammals, one totipotent stem cell, i.e., the fertilized oocyte generates a complex organism containing about 200 types of specialized somatic cells. This developmental potential is due to the capacity of embryonic stem (ES) cells to self-renew and to differentiate into all somatic cells of the body including the germ line. Numerous genes, transcriptional factors and signaling pathways work in concert to regulate cell proliferation, differentiation and self-renewal of ES cells. If this orchestrated regulation gets disrupted, these factors might become associated with oncogenesis. In many aspects, germ cell tumorigenesis resembles early development because the cells of origin are embryonic germ cells. Thus, they provide a useful model to study early development as well as tumorigenesis.

Malignant testicular germ cell tumors of adolescents and young adults, here referred to as germ cell cancer (GCC), are the most common malignancy in Caucasian males between the ages of 20 and 40 years. GCC is thought to be derived from embryonic cells in the germ cell lineage that are blocked in their physiological maturation. This will give rise to carcinoma *in situ* (CIS) which is the precursor of GCC. Clinically and pathologically, the two main types of GCC are seminomas (SE) and non-seminomas (NS). NS are subdivided into embryonal carcinoma (EC) which is the undifferentiated stem cell component, able to give rise to extra-embryonic lineages known as yolk sac tumor (YS) and choriocarcinoma as well as somatic lineage (teratoma). In EC, SOX2, which is one of the main stem cell pluripotency factors, is expressed together with OCT3/4 and NANOG, whereas, in SE, SOX17 instead of SOX2 is expressed.

MicroRNAs (miRs) are endogenous small noncoding RNAs ranging from 19-25 nucleotides. They are shown to be instrumental in a wide variety of key biological processes including cell cycle regulation, apoptosis, metabolic pathways, imprinting, differentiation and maintenance of "stemness".

In chapter 2 of this thesis, the current view on the role of interaction between core ES pluripotency factors and ES miRs, in particular miR-371-3 and miR-302/367 in the maintenance of the ES cell identity and normal development as well as its role in the development of GCC is described. Moreover, the effect of these miRs on the cell cycle of GCC and subsequently their (unique) response to chemotherapy is discussed. The potential of clinical application of miR-371-3 and miR-302/367 in GCC in respect to diagnosis and prognostic implications is discussed as well.

In **chapter 3**, we investigate whether specific amplification of pluripotency genes associates with the undifferentiated state of EC tumors. No amplification for OCT3/4 and SOX2 genes was detected in a number of pure EC and mixed NS including EC tumors. In addition, we investigated the effect of SOX2 down-regulation in NCCIT cell line (with SOX2 amplification) by means of targeted siRNA and it resulted in cell death. The role of core pluripotency factors was also investigated in vivo by generating xenografts from NCCIT.

Using immunohistochemistry for OCT3/4 and SOX2 in differentiated parts of the tumors, we observed that in spite of induction of differentiation, SOX2 continued to be expressed in the lineages formed. This study reveals new aspects of role of ES cell factor SOX2 in relation to differentiation in the development of GCC and regulation of stem cell differentiation.

A number of previous studies have reported that *DICER1* mutations are found in less than 1% of most cancers. To provide a more accurate estimate of the frequency of such mutations in GCC, we analyzed 94 GCC samples using high resolution melting curve analysis in **chapter 4**. Our findings suggest a true mutation frequency of around 1% in GCC. This questions a significant role of this gene in the pathogenesis of GCC.

Several studies have shown increased expression of certain miRs in GCC, primarily the miR-302/367 and miR-371-3. Indeed, expression of these miRs is a hallmark of these cancers. In **chapter 5**, we describe a copy number variation and sequence analysis of these miRs loci in a large cohort of GCC and an apparent deletion of miR-302c in 6% of cases is found. However, this is equivalent to the frequency of this variation in the general population. Two other variants were found for miR-367 which is unlikely to have any pathogenic role. So, mutations in the loci of these miRs are unlikely to be responsible for the miR levels as observed. Therefore, the expression pattern of these miRs in GCC is most likely due to their intrinsic pluripotent capacity of the cell of origin, rather than specific genetic mutations affecting their expression level. This is an important conclusion in the context of using these miRs as potential biomarkers for GCC patients.

GCC are highly treatable by surgery, possibly followed by either irradiation and/or chemotherapy, dependant on stage and histology. Currently, serum markers such as alpha feto-protein (AFP) and human Chorionic Gonadotropine (hCG) are predominantly informative, both for diagnosis and follow-up of YS tumor and choriocarcinoma, respectively. However, for some GCC patients, predominantly having a pure SE and EC, no informative serum markers are available. This increases the need for new serum markers for diagnosis and follow-up of different sub types of GCC. In chapter 6 a robust and informative pipeline for underscoring miR levels in serum is presented. The protocol includes all relevant control steps required for implementation of the approach in a clinical setting. Based on multiple statistical approaches, miR-371-3 and miR-367 are found to be most informative at the time of diagnosis. In addition, analysis of pre- and post- surgery serum samples show a putative value for follow-up of GCC patients as all post-surgery samples show no miR level difference compared to normal serum samples. This is a unique starting point for further clinical evaluation of the value of the TSmiR test.

Overall, this thesis deciphers the role of two clusters of ES miRs, miR-371-3 and miR-302/367 together with the key pluripotency genes in the pathogenesis of GCC and their (exceptional) sensitivity to chemotherapy as well as clinical application of these miRs regarding the diagnosis and follow-up of GCC patients.

Samenvatting

In zoogdieren, een totipotente stamcel, oftewel de bevruchte eicel genereert een complex organisme bestaande uit ongeveer 200 soorten gespecialiseerde lichaamscellen. Dit ontwikkelingspotentieel is te danken aan het vermogen van embryonale stam (ES) cellen om zichzelf te vernieuwen en te differentiëren in alle typen somatische cellen van het lichaam, inclusief de kiemlijn. Tal van genen, transcriptie factoren en signaaltransductie routes werken samen om celproliferatie, differentiatie en zelfvernieuwing van ES cellen te reguleren. Als deze georkestreerde regulatie verstoord wordt, kunnen deze factoren betrokken zijn bij het ontstaan van oncogenese. In veel opzichten lijken kiem cel tumoren op vroege ontwikkeling, omdat de cellen van herkomst embryonale kiemcellen zijn. Daarom zijn zij informatief als model voor zowel vroege ontwikkeling als tumorigenese.

Testiculaire kiemceltumoren van adolescenten en jonge volwassenen, hier aangeduid als kiemcel kanker (KCK), zijn de meest voorkomende maligniteit bij blanke mannen in de leeftijd tussen 20 en 40 jaar. KCK wordt verondersteld te ontstaan uit embryonale stamcellen uit de kiemcellijn die geblokkeerd zijn in hun fysiologische uitrijping. Dit zal leiden tot carcinoom *in situ* (CIS), de voorloper van KCK. Klinisch en pathologisch zijn twee vormen van KCK te onderscheiden, en wel seminomen (SE) en niet-seminomen (NS). NS zijn onder te verdelen in embryonaal carcinoom (EC), die de ongedifferentieerde stamcellen component representeren, en aanleiding kunnen geven tot extra-embryonale lineages, zoals dooierzak tumor (YS) en choriocarcinoom, maar ook de somatische component (teratoom). In EC komt SOX2, een van de belangrijkste stamcellen pluripotentie factoren, tot expressie in combinatie met OCT3/4 en NANOG, terwijl in SE, SOX17 in plaats van SOX2 tot expressie komt. MicroRNA (miR) zijn endogene kleine niet-coderende RNAs, variërend 19 tot 25 nucleotiden. Hety is aangetoond dat zij instrumenteel zijn in diverse belangrijke biologische processen zoals celcyclus, apoptose, metabole wegen, imprinting, differentiatie en onderhoud van "stemness".

In **hoofdstuk** 2 van dit proefschrift wordt de huidige visie op de rol van de interactie tussen structurele ES pluripotentie factoren en ES miRs, in het bijzonder miR-371-3 en miR-302/367, in het onderhouden van de ES-cel identiteit en normale ontwikkeling, alsmede de rol ervan bij de ontwikkeling van KCK beschreven. Bovendien wordt het effect van deze miRs op de celcyclus van KCK en vervolgens de (unieke) respons op chemotherapie besproken. Het mogelijk klinische toepasbaar maken van miR-371-3 en miR-302/367 voor KCK met betrekking tot diagnose en prognostische implicaties wordt besproken.

In **hoofdstuk** 3 hebben we onderzocht of specifieke amplificatie van pluripotentie genen geassocieerd is met de ongedifferentieerdheid van EC. Geen amplificatie voor OCT3/4 en SOX2 genen werd aangetoond in een aantal op pure en gemengde NS, waarin EC vertegenwoordigd was. Bovendien hebben we het effect van SOX2 downregulatie in NCCIT cellijn (met SOX2 amplificatie) met behulp van gerichte siRNA geanalyseerd, wat resulteerde

in celdood. De rol van de structurele pluripotentie factoren werd ook *in vivo* onderzocht door het genereren van zogenaamde xenotransplantaten van NCCIT. Met gebruikmaking van immunohistochemie voor OCT3/4 en SOX2 is gebleken dat ondanks differentiatie steeds hoge expressie van SOX2 aanwezig bleef. Deze studie illustreert nieuwe aspecten betreffende de rol van de ES cel factor SOX2 ten aanzien van differentiatie in de ontwikkeling van KCK en regulering van stamceldifferentiatie.

Een aantal eerdere studies hebben aangetoond dat DICER1 mutaties in minder dan 1% van meeste kankers voorkomen. Om een meer nauwkeurige schatting van de frequentie van dergelijke mutaties in KCK te verkrijgen, hebben we 94 KCK geanalyseerd met behulp van hoge resolutie smeltcurve analyse, zoals in **hoofdstuk** 4 beschreven. Onze bevindingen geven aan dat de mutatie frequentie van ongeveer 1% in KCK is. Hierdoor wordt verondersteld dat afwijkingen van dit geen geen grote rol speelt in de pathogenese van KCK.

Verschillende studies hebben aangetoond dat een verhoogde expressie van bepaalde miRs in KCK aanwezig is, met name betreffende miR-302/367 en miR-371-3. Inderdaad, de relatief hoge expressie van deze miRs is kenmerkend voor deze kanker. In **hoofdstuk** 5 beschrijven we een copie aantal variatie en sequentie-analyse van deze miRs in een relatief grote serie van KCK en een mogelijk verlies van een deel van miR-302c in ongeveer 6% van de gevallen. Dit percentage is echter gelijk aan de frequentie van deze variant in de totale populatie.

Twee andere varianten werden gevonden voor miR-367, die waarschijnlijk ook geen pathogene rol spelen. Dus mutaties in en van deze miRs hebben zeer waarschijnlijk geen rol in de regulatie van expressie van deze miRs. Daarom wordt aangenomen dat het expressiepatroon van deze miRs in KCK het resultaat is van hun intrinsieke karakteristieken, gerelateerd aan de cel van oorsprong. Dit is een belangrijke conclusie in de context van het gebruik van deze miRs als potentiële biomarkers voor KCK patiënten.

KCK zijn goed te behandelen door middel van chirurgie, eventueel gevolgd door bestraling en/of chemotherapie, afhankelijk van het stadium en histologie. Momenteel zijn de serum markers alfa-feto proteine (AFP) en humaan Chorionic Gonadotropine (hCG) informatief voor het stellen van de diagnose en opvolging van KCK patiënten, voornamelijk informatief in aanwezigheid van YS en respectievelijk choriocarcinoma. Voor sommige KCK patiënten, met name die met een puur SE en EC, zijn geen informatieve serum markers beschikbaar. Dit illustreert de waarde van nieuwe serum markers voor diagnose en opvolging van patiënten met bepaalde subtypen van KCK. In **hoofdstuk** 6 wordt een robuuste en informatieve methode voor het aantonen van miRs in serum gepresenteerd. Het protocol bevat alle relevante controle stappen die nodig zijn voor het bepalen van de test in een klinische setting. Op basis van meerdere statistische benaderingen, blijken miR-371-3 en miR-367 het meest informatief te zijn op het moment van diagnose. Daarnaast geeft analyse van pre- en post-operatie sera aam dat deze test van mogelijke waarde is voor de follow-up van KCK patiënten. Het blijkt dat de post-operatie sera dezelfde niveau's van miRs te zien geven als controle sera. Dit is een

uniek uitgangspunt voor verdere klinische evaluatie van de waarde van de TSmiR test.
Concluderend kan gesteld worden dat dit proefschrift een bijdrage geleverd heeft aan het ontrafelen van de rol van twee clusters van ES gerelateerde miRs, miR-371-3 en miR-302/367. Dit in het kader van het samenspel met de sleutel pluripotentie genen in de pathogenese van KCK en hun (bijzondere) gevoeligheid voor chemotherapie, waarbij mogelijke klinische toepassing van deze miRs met betrekking tot diagnose en follow-up van KCT patiënten geïllustreerd is.

Appendices

Acknowledgments:

Getting a PhD is a journey that without the help of others is not possible. The success of any project depends largely on the encouragement and guidance of many people. I take this opportunity to thank all the people who have been beside me in this adventure.

First of all, I am grateful to my supervisor and promoter Prof. Looijenga for giving me the opportunity to work on this PhD project. Without his persistent guidance and help, finishing this thesis would not have been possible. His scientific attitude and continuous effort in regard to research has been inspiring along the way.

I would like to thank my reading committee Prof. Oosterhuis, Prof. Fodde and Prof. Jenster for the time they spent for my thesis and the defense ceremony.

Many thanks to my colleagues in the Lepo lab, particularly Ad and Hans for all the technical supports. Without their help all the lab works for my projects would not have succeeded till the end. Special thanks to Hans for all the useful discussions. I learnt so much from him. His energy in the lab is amazing!

I am thankful to Lambert for his persistent support in the lab especially in complicated experiments and all the thoughtful discussions and efforts to answer all the difficult questions! (which he thinks are always referred to him) I learnt a lot from him and his scientific spirit has been inspiring to me.

I am grateful to my current officemates Will, Martine, Remko, Berdine and Ali for being so patient during my writings and for all the pleasant discussions in our room and for all the supports for finishing my thesis.

Special thanks to Will for her kindness and small chats that always brings in good atmosphere. Thanks to Martine and Berdine for helping me to improve my Dutch before my exam! I also enjoyed our discussions about different topics.

Thanks to Remko for his useful suggestions regarding lab works and also my thesis. His scene of humor is a gifted talent! I am very thankful to Martin for all his help and support regarding statistical analysis, pictures and computer stuff. I really appreciate it. Thank so much to Joun for all his help for my figures and photoshops!

I thank all the other members of Lepo group, Ton, Yvonne, Katharina for all the supports and discussions. I also had a good time with them during Lepo dinners!

I also acknowledge all the students who temporary joined Lepo for short or long periods, Jana, Bestari, Mandana, Mike, Briya, Vezna, Cristina. Mandana was a good student and it was a nice period of five month working with her at Lepo lab.

I thank former Postdocs at Lepo, Gert Jan and Berti whom I only got to know for a couple of months at the beginning of my PhD. Thank you for your support and help in the lab.

I thank all the PhD students including recently graduated ones in the JNI; Marcia, Muhammad, Marijn, Carolien, Katharina, Lale, Tommaso, Matthias, Hamed, Diana for all the nice discussions about our projects and common issues and thoughts during our PhD journey and all the nice PhD drinks and dinners.

Thanks to the Iranian community within the Erasmus MC for all their kindness and friendship during these years, particularly Alireza, Yaser, Farshid and Sara. I would like to thank all my friends, with whom I shared my time and my life outside work.

At the end I am grateful to my relatives and especially my parents and my brother who have always encouraged me with their never-ending support.

List of publications:

- Carmela M. de Beor, **Ronak Eini**, Ad M.J. Gillis, Hans Stoop, Leendert H.J. Looijenga, Stefan J.White. DICER1 RNase IIIb domain mutations are infrequent in testicular germ cell tumours. [BMC. 2012 Oct 15; 5(1):569]

- **Ronak Eini**, Leendert H.J. Looijenga, Lambert .C.J. Dorssers. Role of embryonic stem cell miRNAs in embryogenesis and germ cell cancer. [IJDB, in press]

- Ad J.M. Gillis*, Martin A. Rijlaarsdam*, **Ronak Eini**, Lambert C.J. Dorssers, Matthew J.Murray,James C. Nicholson, Nicholas Coleman, Gazanfer Belge, Klaus-Peter Dieckmann, Jörn Bullerdiek, Tom Xu, Nathalie Bernard and Leendert H.J. Looijenga. A microRNA serum test for diagnosis and follow-up for (testicular) germ cell cancer patients. [Submitted for publication]

- **Ronak Eini**, Hans Stoop, Ad J.M.Gillis, Katharina Biermann, Lambert C.J. Dorssers, L.H.J.Looijenga. Role of SOX2 in the etiology of germ cell cancer. [Submitted for publication]

- Remko Hersmus , Hans Stoop, Gert Jan van de Geijn, **Ronak Eini** , Katharina Biermann, Wolter J. Oosterhuis, Catharina DHooge, Dominik T.Schneider, Isabelle C. Meijssen, Winand N.M. Dinjens, Hendrikus Jan Dubbink, Stenvert L.S. Drop, Leendert H.J. Looijenga. Prevalence of c-KIT mutations in gonadoblastoma and dysgerminomas of patients with disorders of sex development (DSD) and ovarian dysgerminomas. [PLos One.2012; 7(8):e43952]

- Carmela M. de Beor, Amanda J. Notini, **Ronak Eini**, Ad M.J. Gillis, Jochem Jogenotter, Lambert C.J. Dorssers, Leendert H.J. Looijenga, Stephan J. White. Mutation screening of the miRNA-371-3 and miRNA-302 clusters in testicular germ cell tumours. [In preparation for submission]

PhD Portfolio

Ronak Eini
Department of Pathology, Erasmus MC
Molecular Medicine Postgraduate School
PhD period: 1st January 2009 to 31st June 2013
Promoter: Prof.dr.L.H.J.Looijenga
Supervisor: Prof.dr.L.H.J.Looijenga

Courses and workshops:	**year**	**ECTS**
Molecular Medicine	2009	1
Molecular diagnosis	2009	1.5
Basic and translational oncology	2009	1.5
Workshop on basic data analysis on gene expression arrays	2009	1
English biomedical writing and communication	2010	4
Phylogeny workshop	2009	0.5
Introduction into clinical and fundamental Oncology 2011	2011	3
Presentations:		
Annual scientific meeting presentation at the JNI	2009-2013	3
Weekly Lepo workgroup meetings	2009-2013	3
Seminars and conferences:		
Weekly JNI scientific meetings	2009-2013	2.5
Monthly JNI Oncology symposium	2009-2013	1
Annual Molecular Medicine day (poster presentation)	2010-2013	2.5
Annual meeting of the KWF-society tumor cell biology	2010	1
International seminars:		
7th Copenhagen workshop on CIS Testis and Germ cell Cancer Copenhagen, Denmark (poster presentation)	2010	2
From Pluripotency to Senescence: Molecular Mechanisms of Development Disease and Aging Spetses, Greece (poster presentation)	2012	3
Teaching experience:		
Supervision of master student (March-April)	2009	2
Supervision of student internship (Feb-July)	2012	8

Curriculum Vitae

Ronak Eini was born on 4 august 1981 in Tehran, Iran. In 2001 she finished her studies in natural science at Niyayesh high school, Tehran, Iran, and started in September the same year her studies at Tehran University. She studied Agricultural engineering as bachelor at the University of Tehran. Because she was interested in the molecular biology, she decided to continue her studies in this field in the Netherlands. She obtained her Masters degree in medical biotechnology at the University of Wageningen. In January 2009, she started her PhD project in the Laboratoy of Experimental Patho-Oncology in the department of pathology, Josephine Nefkens institute, Erasmus MC University. During her PhD, she has been active in elucidation of the mechanisms behind the role of specific embryonic proteins and microRNAs in the development of human germ cell cancer, the frequent malignant tumor in adolescent and young Caucasian males. Using a multidisciplinary approach she identified factors that shed novel light on the mechanisms behind this type of cancer which is so sensitive to DNA damaging agents such as cisplatin.