Functional Genomics of Germ Cell Tumors

from balls to bytes and back again

Martin A. Rijlaarsdam, MSc. MD

m.a.rijlaarsdam@gmail.com

Erasmus MC
University Medical Center Rotterdam

Molecular Medicine
Postgraduate School

Functional Genomics of Germ Cell Tumors

Het functionele genoom in kiemceltumoren

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam

op gezag van de rector magnificus Prof.dr. H.A.P. Pols

en volgens het besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 26 november 2014 om 15.30 uur

door

Martin Anne Rijlaarsdam

geboren te Leiden

ERASMUS UNIVERSITEIT ROTTERDAM

PROMOTIECOMMISSIE

Promotor: Prof.dr. L.H.J. Looijenga

Overige leden: Prof.dr. E.P.J.G. Cuppen

                     Prof.dr. G.W. Jenster

                     Prof.dr. S. Repping

VISION AS PROJECTION

(Bill Viola, notebook entry, 1986)

voor mijn ouders: Anne en Henriëtte

Table of Contents

PART III: EPIGENETIC CONSTITUTION OF GERM CELL TUMORS

CHAPTER 1

General Introduction

This introduction provides a concise overview of germ cell development followed by a discussion of its relation to human germ cell tumors (GCT). The etiology, pathogenesis and molecular characteristics of the various types of GCTs are extensively reviewed in **chapters 2 and 9.** These chapters also evaluate the onco-fetal relation of GCTs to (early embryonic) germ cell development. The results presented in this thesis describe the (genome-wide) regulation of gene and protein expression, i.e. "functional genomics". This subject is therefore also briefly introduced. In the last section of this chapter, the aims and outlines of this thesis are presented.

Normal germ cell development

During embryogenesis germ cell progenitor cells migrate from the yolk sac to the genital ridge while undergoing massive (epigenetic) changes. Primordial germ cells (PGCs) migrate from the yolk sac via the hindgut. They finally arrive at the progenitor site of the gonad, the so called genital ridge, where they are referred to as gonocytes. The genital ridge normally develops into either a functional testis or an ovary. During migration and most notable after arrival at the genital ridge, PGCs evolve from a totipotent phenotype into sex specific mature germ cells (spermatozoa and oocytes). The developmental program guiding the maturation of PGCs/gonocytes into fully developed germ cells is characterized by changes in RNA and protein expression. Maturing germ cells also undergo an epigenetic "reset" (see also: "functional genomics in a nutshell" and **chapter 9**). Most importantly in the context of this thesis, these cells show characteristic patterns of histone modifications and DNA CpG methylation (see below for an explanation of these terms). Epigenetic reprogramming of germ cells is partly sex specific. It is of particular interest that genomic imprinting differs between males and females while X chromosome reactivation occurs in females only. These features are established during (fetal) germ cell maturation. The characteristic molecular properties of developing germ cells can be assessed experimentally to investigate their developmental status as mirrored in their malignant derivatives (GCTs). It must be kept in mind that most experiments evaluating the various aspects of physiological germ cell development have been performed in mice. While differences between species have to be taken into account, important similarities have been shown as well. For detailed references and an extensive review of normal germ cell development, please see **chapter 2** (core literature: [1-7]).

Germ cell tumors

GCTs are a unique, heterogeneous group of tumors originating from germ cells at different developmental stages and can be classified into five subtypes (I-V). These subtypes are part of the current World Health Organization (WHO) classification (Figure 1). Type I-V GCTs are related to different stages of (embryonic) germ cell development and exhibit distinct characteristics which are frequently directly related to their immature (fetal) non-malignant ancestors. Many of these characteristics can be used as clinical tumor markers in GCT diagnostics and patient follow-up as reviewed in **chapter 2**. GCTs do not arise from epithelial cells (i.e. carcinomas), mesenchymal cells (i.e. sarcomas) or hematopoietic cells (i.e. leukemias or lymphomas). Therefore, they represent a completely separate class of malignant entities [8-11].
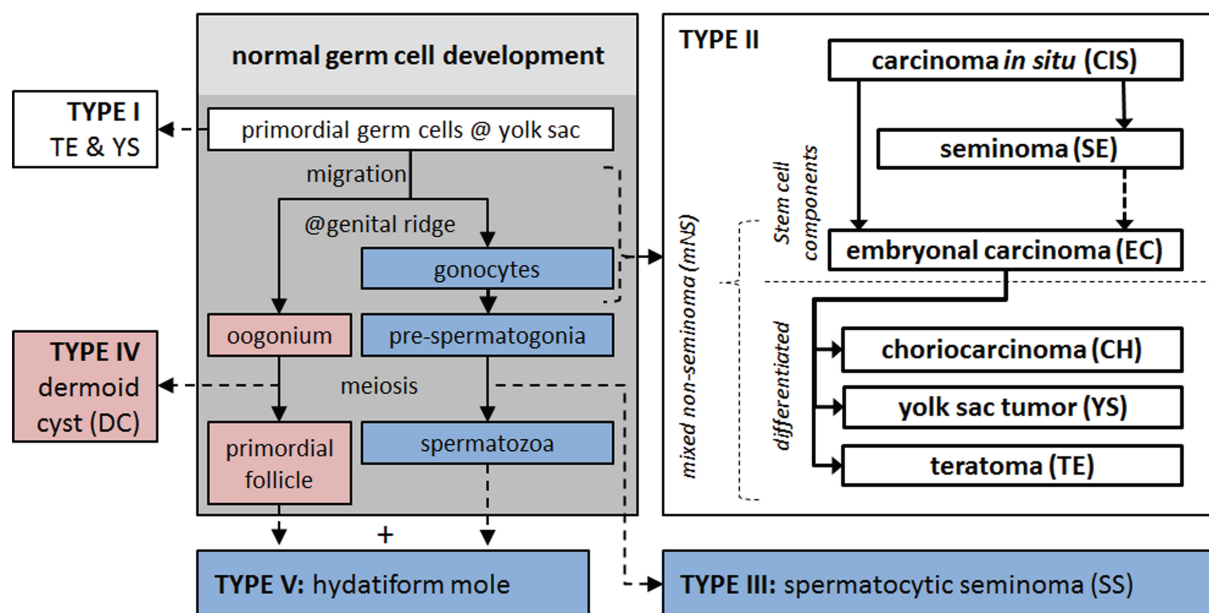
Figure 1. Germ cell tumor subtypes and their relation to normal germ cell development. Migration and maturation of (fetal) germ cells are shown in the grey box. Developmental schemes are indicated in blue (male), red (female) or in white when unrelated to sex. These color schemes are extended to the boxes identifying the various GCT subtypes. This visualization is based on the model proposed by Oosterhuis and Looijenga [8] as used in the current WHO classification [10]. This figure is adapted from chapter 9, Figure 1A. For more details, please see the main text, section "Germ cell tumors".

Type I GCTs, also called infantile or pediatric GCTs, are rare and generally benign. They generally appear before the age of six and occur in 0.12/100,000 individuals. Type I GCTs are composed of teratoma (TE) and/or yolk sac tumor (YS) components and typically require only surgical resection to cure the patient. Interestingly, many primary type I GCTs occur outside of the gonads, along the migration route of the germ cells. They differ from the type II TE and YS in their age of presentation, lack of the precursor lesion *Carcinoma In Situ (CIS)*, absence of the stem cell component embryonal carcinoma and lack of trophoblastic differentiation (choriocarcinoma) [8-10, 12, 13].

Type Ⅱ GCTs include a heterogeneous set of histological subtypes with clearly defined totipotent stem cell components, together accounting for 60% of all cancers in Caucasian males between the ages of 20 and 40. Because of their malignant behavior, type Ⅱ GCTs are also called Germ Cell Cancer (GCC). GCCs are most frequently seen in adolescents and young adults and primarily occur in the testis. They represent ≈1% of all solid cancers in Caucasian males, but account for 60% of all cancers in these males between the ages of 20 and 40 [8, 9]. Incidence has almost doubled in the Netherlands over the last decade (9.43/100,000 ESR 2012, Dutch Caner Registration (IKNL), www.cijfersoverkanker.nl). This corresponds with international observations [14]. The increase is most frequently associated with environmental changes [15]. Indeed, GCC risk has been associated with a complex set of interacting environmental and genetic, i.e. genvironmental, risk factors which are reviewed in **chapter 2** (core literature: [15-19]). Current treatment regiments cure over 95% of the patients. However, GCCs still represent an important cause of (treatment related) mortality in these young men [20, 21]. All GCCs arise from a common precursor lesion (CIS in the testis, gonadoblastoma in the ovary and dysgenetic gonad) [8, 9, 22, 23]. CIS is improperly named *carcinoma* because of the non-epithelial origin of the tumor (see above). The WHO therefore baptized this precursor lesion "intratubular germ cell neoplasia unclassified" (IGCNU) [10]. However, throughout this thesis the term CIS will be used for reasons of consistency with most of the existing literature. CIS cells in many ways represent maturation arrested PGCs/gonocytes and can transform into two invasive entities: seminoma (SE) and embryonal carcinoma (EC). EC has totipotent potential and can differentiate into all embryonic (teratoma, TE) and extra-embryonic (YS and choriocarcinoma) tissue lineages. EC and its differentiated derivatives are collectively called non-seminomas (NS). SE incidence peaks at 35 years of age while NS incidence shows a peak about ten years earlier. GCCs occur only rarely at extra-gonadal sites, in contrast to the type I GCTs [8, 9]. In the testis and anterior mediastinum SEs are called seminomas, in contrast to similar GCCs in the ovary (dysgerminoma) and brain (germinoma). Despite differences in their anatomical localization, these tumors all show very similar molecular profiles [24]. Strong similarities have been shown between early embryonic (germ cell) development and the pluri-/totipotent phenotype of CIS, SE and EC as reviewed in **chapter 2**.

Type Ⅲ, Ⅳ and Ⅴ GCTs are generally benign and arise from more mature germ cells. Type Ⅲ GCTs are called spermatocytic seminomas. They occur after the age of 50 and are exclusively located in the testis. Surgical resection suffices in the majority of the cases. Type Ⅳ GCTs, or dermoid cysts are derivatives of committed female germ cells. Type Ⅴ GCTs (hydatiform moles) are formed by the fertilization of an empty ovum by two mature male germ cells [8-10].

The spectrum of GCTs (type I-V) represents a unique class of malignant entities in young patients with strong and traceable onco-fetal roots that clinical applications can take advantage of. Even though GCTs are relatively rare compared to other neoplasms and their sensitivity to chemo-/radiotherapy is generally extremely good, they still impose a considerable disease burden on

mostly young males. This is relevant both in the short term (disease, treatment) and the long term (fertility, long term side effects of treatment). Moreover, the availability of representative cell lines (see **chapter 2**) and animal models [15] provides research opportunities to evaluate GCC as a model for early embryonic (stem) cells and (germ cell) development. This thesis illustrates that fundamental research into the unique (embryonic) characteristics of these tumors translates to clinical applications, helping to identify individuals who are at risk for GCT and enhancing the quality of care for testicular GCT patients (diagnosis, treatment and follow-up).

Functional genomics in a nutshell

The genome is comprised of DNA which is the core information storage facility in a cell's nucleus and is subsequently transcribed into RNA and in many cases translated into proteins. In humans the genome comprises of ≈3 billion base pairs: adenine (A) pairs with thymine (T) and guanine (G) pairs with cytosine (C). The pairs are organized in a double helix structure and divided over 21 autosomes and 2 sex specific chromosomes (XX: female, XY male). For DNA to exhibit any function, the base-pair sequence needs to be transcribed into RNA. A major subclass of RNA is the messenger RNA (mRNA), which is translated into a polypeptides. These are then modified into a 3-dimensional functional protein. Historically, the term (gene) expression refers to mRNA levels and "gene" refers to protein coding DNA regions. However, the term expression is also frequently used to describe the levels of proteins or other types of RNA in a cell. Proteins are major actors in a cell, for example as building blocks of the cell's physical structures or as transducers in signaling pathways that allow the cell to adjust to internal and external stimuli. The human genome encodes around 20,000 protein coding genes. The large majority of the DNA is therefore non-protein coding, but far from useless as science has only just started to discover [25-29]. Indeed, the "one gene to one protein" process presented above is a major simplification of reality. Within the context of this thesis two additional notions are important to keep in mind: (1) some RNAs are not translated into protein (non-coding RNAs, see below) but still regulate important cell functions and (2) a single gene coding region can produce different alternative transcripts and proteins (iso-forms) [30, 31] (Figure 2).

The functional genome controls which part the genome is used. A cell has many complex interacting tools at its disposure to regulate (gene) expression. This is necessary to ensure that a cell only generates the molecular machinery needed for its specific function. In many diseases, including (germ cell) cancer, this process is disturbed [32, 33]. Regulation starts even before transcription. In a non-dividing cell, the DNA is tightly wound up around coils (histones). The term histone modification refers to chemical additions to these coils that alter the accessibility of the DNA for transcription [34-36]. The DNA itself can also be modified to make it less accessible, most notably by CpG methylation. This constitutes the chemical addition of a methyl group to a CG

dinucleotide in the DNA [35, 37]. Modification of the DNA's availability for transcription is called epigenetics. The epigenomic blueprint organizes the DNA archive according to the needs of a specific cell type in certain environmental conditions [38]. Epigenetics should not be confused with "genetics." Genetics is defined as to the genetic code or base-pair sequence of the DNA which is not changed by epigenetic modifications. On top of epigenetic regulation, transcription depends on the activation of a gene's transcription regulatory regions. These typically consist of a region directly before/upstream of the gene (promoter) and can include more distant sites as well (enhancers). Promoters and enhancers provide binding sites for proteins (transcription factors) which can, for example, stimulate the expression of a group of genes with a similar function [30, 31, 39]. If a gene is transcribed, post-transcriptional regulation takes over, targeting the mRNA. In this thesis, emphasis is placed on a specific type of post-transcriptional regulation by micro-RNAs (miRs). miRs are 22 nucleotide long strings of RNA that bind to the tail (3'UTR) of a mRNA and usually inhibit translation / promote degradation of that mRNA [29, 40]. The term post-translational modification is used to indicate chemical modification of proteins, e.g. to activate them or prime them for degradation (Figure 2).

Cross sections of a cell at specific functional or regulatory levels are referred to as "omics" or "ome." They are used to describe the simultaneous quantification of all members of a certain functional class in a cell. Examples of these classes are: mRNA, miRs, DNA, etc. [41, 42]. The textbook example of an "-ome" is the genome: the complete DNA base-pair sequence of a cell. Epigenomics is the investigation of epigenetic modifications as present across the whole genome. This includes investigation of the methylome, the methylation status of all CG dinucleotides in the genome. The term transcriptome is used to refer to all RNAs. The proteome refers to all proteins in a cell. Many more "omes" have been identified [42]. These are not directly relevant to this thesis and are therefore not discussed here. With the exception of the genome in normal diploid cells from the same individual, "omes" are generally different for each cell type and environmental condition. The terms "omics" and "ome" are also used when only a representative sample of all members in a class is investigated instead of al members. For example, the methylation array used in this thesis inquires the methylation status of ≈480,000 CG sites on the genome. This number does not even come close to the total number of CG dinucleotides in the total genome, but provides a useful indication of the total methylome. Such a representative sample is also referred to as a "genome-wide profile." Omic datasets are basically big buckets of numbers representing cross sections of the state of a cell at certain functional or regulatory levels. Such datasets contain technical artifacts and difficulties inherent to the complex methods used to produce them, requiring normalization / pre-processing and validation. Normalization and pre-processing refer to all steps necessary to remove technical artifacts, noise and other non-biological effects from the data (relevant for this thesis: [43, 44]). Validation refers to the targeted experiments required to confirm the omics findings (Figure 2).
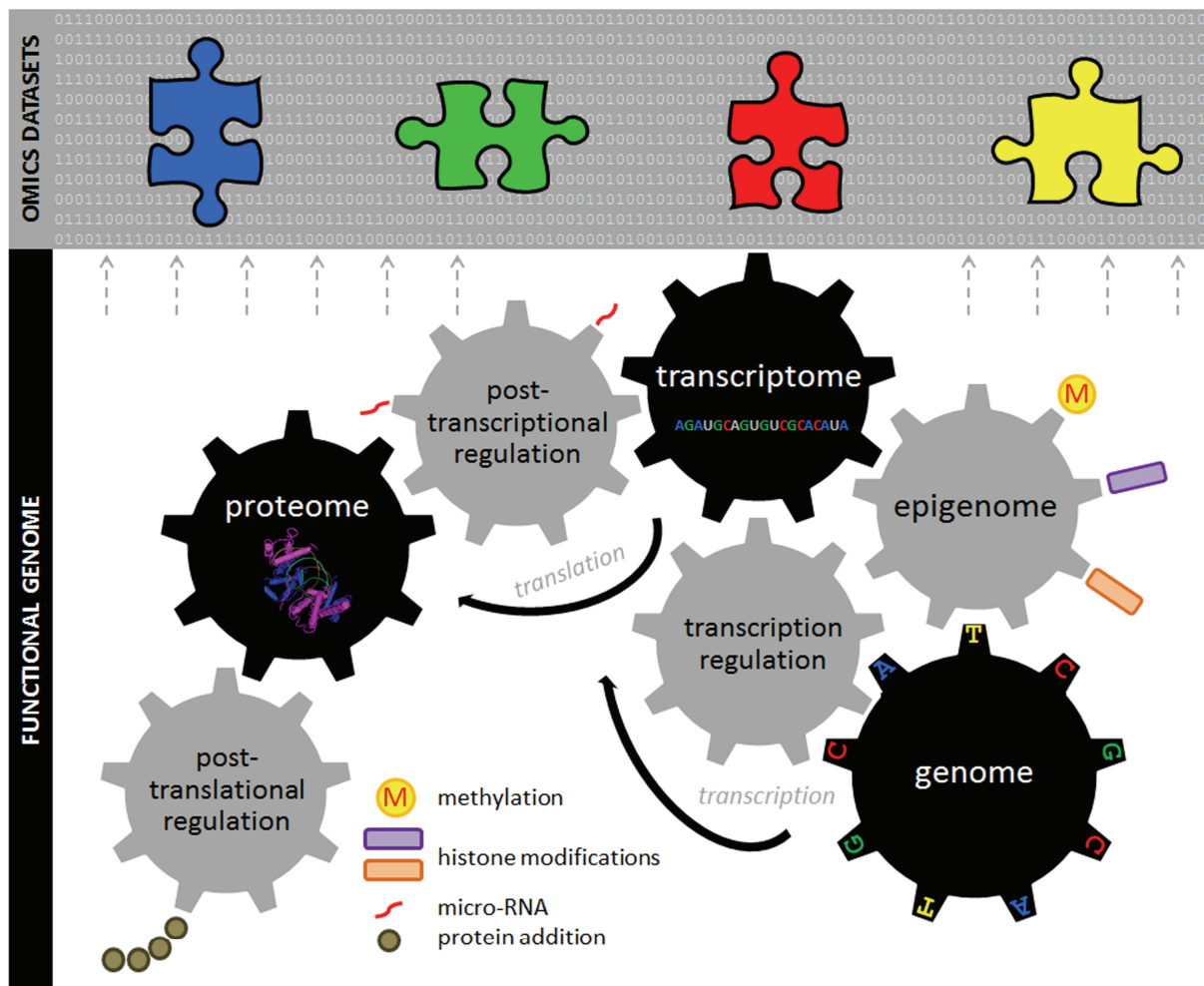
Figure 2. Schematic represtnation of the functional genome and its quantification into "omics" datasets. This highly simplified schematic illustration depicts the transcription of the DNA (genome) into RNA under the control of epitenetic mechanisms and other actors like transcription factors. Epigenetic mechanisms include DNA CpG methylation and various histone modifications. Translation of the RNA (transcriptome) is further influenced by post-transcriptional regulation, including micro-RNAs. The resulting set of proteins is called the proteome. In the proteome cogwheel, the 3-dimensional structure[a] of the OCT3/4 protein is shown as an example. Proteins can be modified by the addition of various biochemical groups, i.e. post-translational modification. The dashed grey arrows illustrate the genome-wide experimental evaluation of the factors represented by the cogwheels. This results in large omics dataset. The aim of integrated functional genomics is to evaluate and if possible integrate these datasets. These datasets are quite literally the pieces of the puzzle that need to be put together to decipher the functional genome. For more details, see the main text.

Functional genomics aims to integrate multiple (omics) datasets to better understand the functional interactions that determine a cells identity and behavior, requiring a multidisciplinary approach and a cultural change towards sharing and disclosing data & tools. Functional genomics is closely related to the field of systems biology which aims to model biological systems based on observations. Analogous, an important ambition of integrated genomics is to combine various (omics) datasets and create a complete map of the functional genome to better understand the functional interactions that determine a cells identity and behavior [45] (Figure 2). A great number

---

[a] http://www.ncbi.nlm.nih.gov/Structure/

of global initiatives strive to disclose complex (integrated) omics datasets. Examples of these initiatives include mainstream genome browsers/databases (Ensembl[b], UCSC[c], NCBI[d]) which include numerous data tracks concerning functional genomics. Specialized databases are also freely available online. Examples are: genomics (NHGRI GWAS catalog [e], dbSNP [f]), epigenomics (ENCODE[g], Blueprint project[h], WADIMEX[i], geneimprint.com[j], Cancer Genome Atlas[k]), miR target prediction (microRNA.org[l], TargetScan[m]) and disease/pathway relations (OMIM[n], KEGG[o]). GEO[p] and ArrayExpress[q] are online repositories where researchers can submit their data, which is often mandatory for publication. On the bioinformatics side, platforms like Bioconductor[r] and CRAN[s] are available, facilitating the exchange, development and the standardization of bioinformatic tools. Making data and tools available is not enough however. A multidisciplinary approach including, among others, mathematicians, biologists and clinicians is crucial to successfully retrieve biologically and clinically relevant results from integrated functional genomic analysis [46]. Ultimately such results contribute to better understanding of (patho)biology and optimization of health care.

---

[b] http://www.ensembl.org/index.html
[c] https://genome.ucsc.edu/
[d] http://www.ncbi.nlm.nih.gov/
[e] http://www.genome.gov/gwastudies/
[f] http://www.ncbi.nlm.nih.gov/projects/SNP/
[g] http://genome.ucsc.edu/ENCODE/
[h] http://www.blueprint-epigenome.eu/
[i] https://atlas.genetics.kcl.ac.uk/atlas.php
[j] http://geneimprint.com/
[k] http://www.broadinstitute.org/cancer/cga/
[l] http://www.microrna.org/microrna/home.do
[m] http://www.targetscan.org/
[n] http://www.ncbi.nlm.nih.gov/omim
[o] http://www.genome.jp/kegg/
[p] http://www.ncbi.nlm.nih.gov/geo/
[q] http://www.ebi.ac.uk/arrayexpress/
[r] http://www.bioconductor.org/
[s] http://cran.r-project.org/

Aims & outline of the thesis

The work discussed in the thesis aims to elucidate the role of the functional genome in GCT (pathogenesis) by applying newly developed and existing computation methods to (genome-wide) datasets. Epigenetic and (post-)transcriptional regulation in GCTs is studied to gain deeper understanding about disease pathogenesis, also aiming at clinical application of the findings. The exploratory approach of the studies also contributes to the emergence of new hypothesis that can be experimentally and clinically validated in the future.

Chapter 2 / part I of this thesis starts with an in depth review of the etiology, developmental biology and onco-fetal roots of GCC.

Part II discusses the application of oncofetal proteins and miRs in GCC diagnostics and follow-up as well as a tool to identify functional miR-mRNA interactions in genome-wide datasets. **Chapter 3** verifies the specificity of a distinct OCT3/4 isoform as a marker for GCCs by comparing mRNA and protein levels of three OCT3/4 isoforms in GCCs, non-GCC malignancies and representative cell lines. This chapter highlights the importance of alternative splicing in gene function and clinical marker specificity. **Chapter 4** describes miMsg, a bioinformatic tool to effectively reduce a high number of predicted miR-mRNA interactions to a small, high confidence set directly applicable to research. In this chapter, miMsg is also applied to a number of in-house GCT and public non-GCT datasets to illustrate its validity. This kind of tools is relevant to identify new disease specific targets and molecular networks that can be further studied in targeted experiments. **Chapter 5** describes the application of serum embryonic miR levels as a new biomarker for GCCs, which is sensitive for the stem cell components SE and EC. The identification of these miRs as GCC markers could be applied to optimize diagnosis and follow-up of GCC patients. **Chapter 6** compares the serum levels of ≈750 miRs in GCC cases and controls, validating known miR markers and identifying novel GCC specific targets / potential biomarkers.

Part III focusses on the epigenetic constitution of GCTs and their cells or origin. **Chapter 7** outlines DMRforPairs, a tool developed to detect differentially methylated genomic regions between unique samples using array based methylation data. This tool is applied in **chapters 8 and 9** which elaborate on the epigenetic constitution of GCTs and their representative cell lines. **Chapter 8** sketches the functional genomic footprints of SE versus EC cell lines by integrating genome-wide data describing histone modifications, methylation status and RNA expression. **Chapter 9** provides insight into the developmental timing and underlying biology of the various subtypes of GCTs and their (embryonic) cells of origin by identifying specific and global methylation differences between GCT subtypes.

Overall, the findings presented in this thesis outline interacting levels of the functional genomic landscape in GCTs and demonstrate (potential) clinical applications. **Chapter 10** integrates the results. This last chapter provides a concise set of concluding remarks and future prospects regarding three main topics. Firstly, the unique onco-fetal hallmarks of GCC and the related clinical applications are discussed. Secondly, the insights gained into the developmental biology of GCT through the analysis of characteristic epigenetic properties of GCTs are revisited. Lastly, challenges faced when exploring (integrated) functional genomics underlying disease are examined from a more general point of view.

References

1.   Matsui Y, Mochizuki K: **A current view of the epigenome in mouse primordial germ cells.** *Mol Reprod Dev* 2014, 81:160-170.
2.   Wermann H, Stoop H, Gillis AJ, Honecker F, van Gurp RJ, Ammerpohl O, Richter J, Oosterhuis JW, Bokemeyer C, Looijenga LH: **Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance.** *J Pathol* 2010, 221:433-442.
3.   Almstrup K, Nielsen JE, Mlynarska O, Jansen MT, Jorgensen A, Skakkebaek NE, Rajpert-De Meyts E: **Carcinoma in situ testis displays permissive chromatin modifications similar to immature foetal germ cells.** *Br J Cancer* 2010, 103:1269-1276.
4.   Kristensen DG, Skakkebaek NE, Rajpert-De Meyts E, Almstrup K: **Epigenetic features of testicular germ cell tumours in relation to epigenetic characteristics of foetal germ cells.** *The International journal of developmental biology* 2013, 57:309-317.
5.   Seisenberger S, Peat JR, Reik W: **Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells.** *Curr Opin Cell Biol* 2013, 25:281-288.
6.   Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, Eggan K, Meissner A: **DNA methylation dynamics of the human preimplantation embryo.** *Nature* 2014, 511:611-615.
7.   Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J, et al: **The DNA methylation landscape of human early embryos.** *Nature* 2014, 511:606-610.
8.   Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5:210-222.
9.   Looijenga LH: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J Pathol* 2009, 218:146-162.
10.   Woodward PJ, Heidenreich A, Looijenga LHJ, al. e: **Testicular germ cell tumors.** In *World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs.* Edited by Eble JN, Sauter G, Epstein JI, Sesterhann IA, eds. Lyon: IARC Press; 2004: 17-278
11.   Boublikova L, Buchler T, Stary J, Abrahamova J, Trka J: **Molecular biology of testicular germ cell tumors: unique features awaiting clinical application.** *Critical reviews in oncology/hematology* 2014, 89:366-385.
12.   Gobel U, Schneider DT, Calaminus G, Haas RJ, Schmidt P, Harms D: **Germ-cell tumors in childhood and adolescence. GPOH MAKEI and the MAHO study groups.** *Ann Oncol* 2000, 11:263-271.
13.   Rescorla FJ: **Pediatric germ cell tumors.** *Seminars in surgical oncology* 1999, 16:144-158.
14.   Horwich A, Shipley J, Huddart R: **Testicular germ-cell cancer.** *Lancet* 2006, 367:754-765.
15.   Looijenga LH, Van Agthoven T, Biermann K: **Development of malignant germ cells - the genvironmental hypothesis.** *The International journal of developmental biology* 2013, 57:241-253.
16.   Dieckmann KP, Pichlmeier U: **Clinical epidemiology of testicular germ cell tumors.** *World journal of urology* 2004, 22:2-14.
17.   McGlynn KA, Cook MB: **Etiologic factors in testicular germ-cell tumors.** *Future Oncol* 2009, 5:1389-1402.
18.   Rajpert-De Meyts E: **Developmental model for the pathogenesis of testicular carcinoma in situ: genetic and environmental aspects.** *Human reproduction update* 2006, 12:303-323.
19.   van der Zwan YG, Biermann K, Wolffenbuttel KP, Cools M, Looijenga LHJ: **Gonadal Maldevelopment as Risk Factor for Germ Cell Cancer: Towards a Clinical Decision Model.** *European Eurology* 2014, in press.
20.   Garner MJ, Turner MC, Ghadirian P, Krewski D: **Epidemiology of testicular cancer: an overview.** *Int J Cancer* 2005, 116:331-339.
21.   Alanee SR, Feldman DR, Russo P, Konety B: **Long-term mortality in patients with germ cell tumors: effect of primary cancer site on cause of death.** *Urol Oncol* 2014, 32:26 e29-15.
22.   Oosterhuis JW, Stoop H, Honecker F, Looijenga LH: **Why human extragonadal germ cell tumours occur in the midline of the body: old concepts, new perspectives.** *International journal of andrology* 2007, 30:256-263; discussion 263-254.
23.   Fan R, Ulbright TM: **Does intratubular germ cell neoplasia, unclassified type exist in prepubertal, cryptorchid testes?** *Fetal and pediatric pathology* 2012, 31:21-24.

24.   Looijenga LH, Hersmus R, Gillis AJ, Pfundt R, Stoop HJ, van Gurp RJ, Veltman J, Beverloo HB, van Drunen E, van Kessel AG, 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**:**290-302.

25.   Palazzo AF, Gregory TR: **The case for junk DNA.** *PLoS Genet* 2014, 10:e1004351.

26.   Ledford H: **Circular RNAs throw genetics for a loop.** *Nature* 2013, 494:415.

27.   Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J: **Natural RNA circles function as efficient microRNA sponges.** *Nature* 2013, 495:384-388.

28.   Cech TR, Steitz JA: **The noncoding RNA revolution-trashing old rules to forge new ones.** *Cell* 2014, 157**:**77-94.

29.   He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation.** *Nature reviews* 2004, 5**:**522-531.

30.   Weinberg RA: **The biology of cancer**. New York, USA: Garland Science, Taylow & Francis Group; 2007.

31.   Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: **Molecular Biology of the cell**. New York, USA: Garland Science, Taylow & Francis Group; 2002.

32.   Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, 144**:**646-674.

33.   Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, 100**:**57-70.

34.   Zhou VW, Goren A, Bernstein BE: **Charting histone modifications and the functional organization of mammalian genomes.** *Nature reviews* 2011, 12**:**7-18.

35.   Cedar H, Bergman Y: **Linking DNA methylation and histone modification: patterns and paradigms.** *Nature reviews* 2009, 10**:**295-304.

36.   Tee WW, Reinberg D: **Chromatin features and the epigenetic regulation of pluripotency states in ESCs.** *Development* 2014, 141**:**2376-2390.

37.   Smith ZD, Meissner A: **DNA methylation: roles in mammalian development.** *Nature reviews* 2013, 14**:**204-220.

38.   Mirbahai L, Chipman JK: **Epigenetic memory of environmental organisms: a reflection of lifetime stressor exposures.** *Mutation research Genetic toxicology and environmental mutagenesis* 2014, 764-765**:**10-17.

39.   Lee E, de Ridder J, Kool J, Wessels LF, Bussemaker HJ: **Identifying regulatory mechanisms underlying tumorigenesis using locus expression signature analysis.** *Proceedings of the National Academy of Sciences of the United States of America* 2014, 111**:**5747-5752.

40.   Nana-Sinkam SP, Croce CM: **Clinical applications for microRNAs in cancer.** *Clinical pharmacology and therapeutics* 2013, 93**:**98-104.

41.   Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, Miriami E, Karczewski KJ, Hariharan M, Dewey FE, et al: **Personal omics profiling reveals dynamic molecular and medical phenotypes.** *Cell* 2012, 148**:**1293-1307.

42.   Baker M: **Big biology: The 'omes puzzle.** *Nature* 2013, 494**:**416-419.

43.   Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerstrom-Billai F, Jagodic M, Sundberg CJ, Ekstrom TJ, Teschendorff AE, Tegner J, Gomez-Cabrero D: **An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform.** *Epigenetics : official journal of the DNA Methylation Society* 2013, 8**:**333-346.

44.   Schmid R, Baum P, Ittrich C, Fundel-Clemens K, Huber W, Brors B, Eils R, Weith A, Mennerich D, Quast K: **Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3.** *BMC Genomics* 2010, 11**:**349.

45.   Kristensen VN, Lingjaerde OC, Russnes HG, Vollan HK, Frigessi A, Borresen-Dale AL: **Principles and methods of integrative genomic analyses in cancer.** *Nat Rev Cancer* 2014, 14**:**299-313.

46.   Teiten MH, Grigorakaki C, Orlikova B, Diederich M: **OMICS, a multidisciplinary friendship.** *Cell Death Dis* 2012, 3**:**e267.

CHAPTER **2**

An onco-fetal and developmental perspective on testicular germ cell cancer

Martin A. Rijlaarsdam[1], Leendert H.J. Looijenga[1]

Department of Pathology, Erasmus MC - University Medical Center, Rotterdam, The Netherlands [1]

Abstract

Germ cell tumors (GCTs) represent a diverse group of tumors presumably originating from (early fetal) developing germ cells. Most frequent are the testicular germ cell cancers (TGCC). Overall, TGCC is the most frequent malignancy in Caucasian males (20-40yr) and remains an important cause of (treatment related) mortality in these young men. The strong association between the phenotype of TGCC stem cell components and their totipotent ancestor (fetal primordial germ cell or gonocyte) makes these tumors highly relevant from an onco-fetal point of view. This review subsequently discusses the evidence for the early embryonic origin of TGCCs, followed by an overview of the crucial association between TGCC pathogenesis, genetics, environmental exposure and the (fetal) testicular micro-environment (genvironment). This culminates in an evaluation of three genvironmentally modulated hallmarks of TGCC directly related to the oncofetal pathogenesis of TGCC: (1) maintenance of pluripotency, (2) cell cycle control/cisplatin sensitivity and (3) regulation of proliferation/migration/apoptosis by KIT-KITL mediated receptor tyrosine kinase signaling. Briefly, TGCC exhibit identifiable stem cell components (seminoma and embryonal carcinoma) and progenitors that show large and consistent similarities to primordial/embryonic germ cells, their presumed totipotent cells of origin. TGCC pathogenesis depends crucially on a complex interaction of genetic and (micro-)environmental, i.e. genvironmental risk factors that have only been partly elucidated despite significant effort. TGCC stem cell components also show a high degree of similarity with embryonic stem/germ cells (ES) in the regulation of pluripotency and cell cycle control, directly related to their exquisite sensitivity to DNA damaging agents (e.g. cisplatin). Of note, (ES specific) micro-RNAs play a pivotal role in the crossover between cell cycle control, pluripotency and chemosensitivity. Moreover, multiple consistent observations reported TGCC to be associated with KIT-KITL mediated receptor tyrosine kinase signaling, a pathway crucially implicated in proliferation, migration and survival during embryogenesis including germ cell development. In conclusion, TGCCs are a fascinating model for onco-fetal developmental processes especially with regard to studying cell cycle control, pluripotency maintenance and KIT-KITL signaling. The knowledge presented here contributes to better understanding of the molecular characteristics of TGCC pathogenesis, translating to identification of at risk individuals and enhanced quality of care for TGCC patients (diagnosis, treatment and follow-up).

Introduction

Germ cell tumors (GCTs) represent a diverse group of tumors presumably originating from (early fetal) developing germ cells. Five GCT subtypes can be defined based on different stages in germ cell maturation. Most frequent are the (testicular) germ cell cancers ((T)GCC, type II GCTs). TGCC accounts for ≈1% of all solid cancers in Caucasian males. In contrast to other solid cancers, TGCC occurs in adolescents and young adults, accounting for 60% of all malignancies diagnosed in men between 20 and 40 years of age [1-3]. In the Netherlands, TGCC incidence in 2012 was 9.43/100,000 (European standardized rate), an increase of 45% since 2002 (Dutch Caner Registration (IKNL), www.cijfersoverkanker.nl) which is consistent with increasing incidence outside The Netherlands (+3-6%/year) [4]. 5-year survival rates under current treatment regiments exceed 96% (IKNL). Prognosis depends strongly on the composition / location of the tumor and patient characteristics [1, 2]. In spite of the overall success of treatment, TGCC remains an important cause of (treatment related) mortality in these young men [5, 6].

A common precursor lesion called carcinoma-in-situ (CIS) or intratubular germ cell neoplasia unclassified (IGCNU, WHO definition [7]) precedes TGCC [1, 2, 8, 9]. Because of the non-epithelial origin of TGCC, CIS is technically not a proper term but will be used throughout this review in the interest of consistency with existing literature. At least 70% of all CIS progresses into TGCC within 7 years and the prevalence of CIS in autopsies was comparable to the lifetime risk of TGCC. This suggests that all patients with CIS eventually develop TGCC [10, 11], which has however not yet been proven conclusively. Invasive TGCCs are divided into seminoma (SE) and non-seminoma (NS, stem cell component embryonal carcinoma (EC)). ECs can differentiate into all somatic (teratoma) and extra-embryonic (yolk sac tumor and choriocarcinoma) lineages including germ cells, hence exhibiting the totipotent potential of TGCC [1, 2, 7, 12] (Figure 1). TGCC subtypes have distinct histopathological and molecular marker profiles used in research and diagnosis [1, 2, 13-27]. For follow-up, serum markers Alpha Feto-Protein (AFP), human Chorionic Gonadotrophine (hCG) and in a limited fashion Lactate DeHydrogenase 1 (LDH-1) are currently used, although they have limited specificity and are not sensitive for detecting TGCC stem cell components (SE or EC) or their precursor lesion CIS [4, 28-30]. Novel diagnostic strategies include immunohistochemical analysis of semen in search of CIS which has shown high specificity, but low sensitivity, warranting further investigation and optimization [31-34]. Quantifying the methylation status of specific regions (XIST promoter) of serum cell free DNA has been suggested as follow-up marker, but has so far not been validated [35, 36]. Very recently, a number of independent studies showed that serum levels of embryonic micro-RNA (miR) clusters 371-3 and 302abc/367 are predictive for the presence of TGCC. This tool proved especially sensitive for identifying the SE and EC components which are indeed known to express these miRs [13, 14, 37-41]. Many (suggested) TGCC markers are functionally related to the fetal origin and pluripotent biology of TGCC (see below, Figure 1).
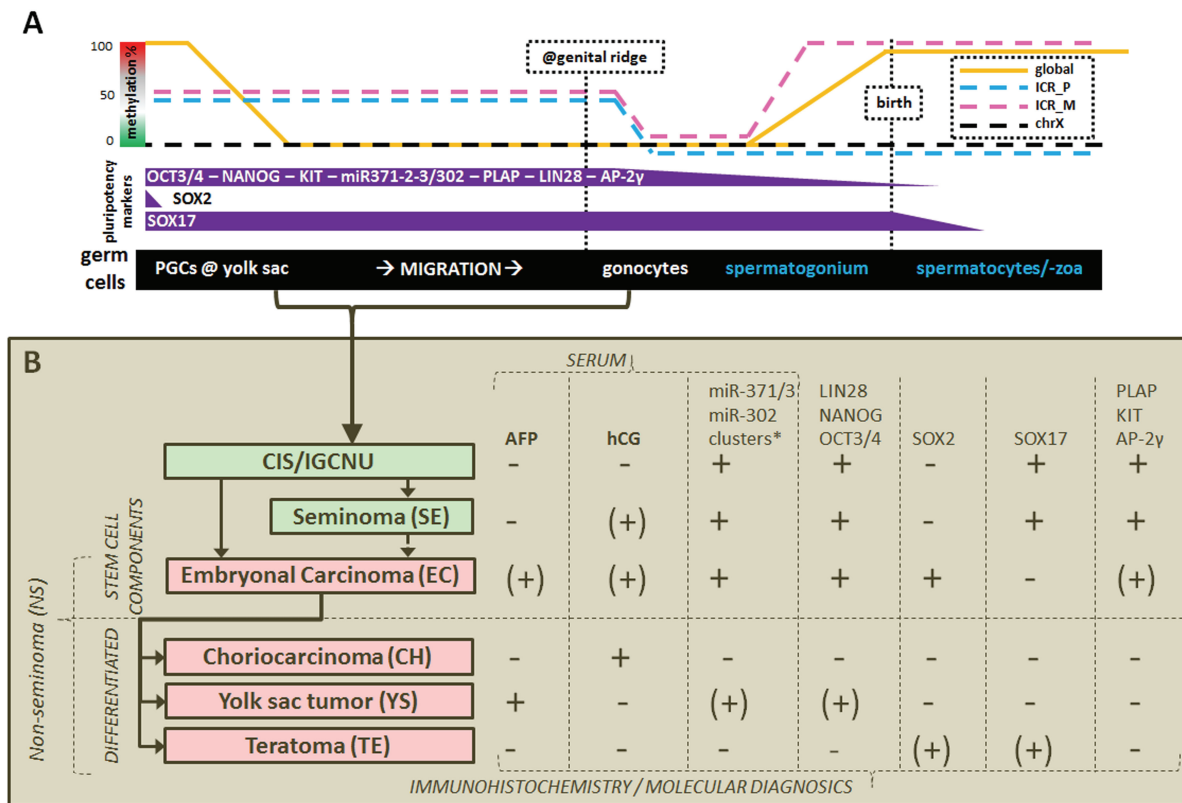
Figure 1. Placing the TGCC cell of origin in an onco-fetal perspective as related to early embryonic germ cell development. The figure summarizes the observations reviewed in the main text. Panel A illustrates the physiological epigenetic reset and expression of core pluripotency markers during fetal germ cell development and the process of migration and maturation. Based on the reviewed evidence and anatomical localizations, the cell of origin of (T)GCC is positioned between the early PGC (extragonadal seminomas/germinomas) and gonocyte stage (testicular seminoma in the case of a male patient). ICR_P/M: imprinting control regions regulating expression of the paternal/maternal allele. Panel B summarizes the interrelation between the various types of TGCC and illustrates their totipotent potential. The methylation status and profile of mainstream diagnostic/onco-fetal markers is displayed as reviewed in the main text. * = miR-371/372/373 cluster & miR-302abcd/367 cluster.

The strong association between TGCC and their totipotent ancestor (fetal primordial germ cell or gonocyte) makes these tumors highly relevant from an onco-fetal point of view, especially since a number of animal and cell line models are available for functional studies (supplementary data). Here, the fetal origin and intrinsic pluripotent oncofetal properties of TGCC cells are reviewed, especially with regard to their stem cell components SE and EC. Risk factors that might trigger malignant transformation of embryonic germ cells and disease progression are also discussed.

From fetal germ cells to CIS and beyond

Normal germ cell development

Embryonic germ cells or primordial germ cells (PGC) are detected in the proximal epiblast at week 5-6 gestagional age in humans (E6.5 in mice). During fetal development these cells migrate along the midline of the body, where GCTs are also located: from the yolk sac, via the hindgut to the genital ridge. These cells are characterized by positive immunohistochemical staining for a number of pluripotency/germ cell markers: e.g. AP+/VASA+/OCT3/4+, [2, 42-44]. Part of the maturation processes is an epigenetic "reset". This epigenetic reprograming includes early global demethylation followed by *de novo* methylation, imprinting reset (erasement → uniparental) and X chromosome reactivation (females only) after arrival at the genital ridge [45-56]. Most of the epigenetic studies showing this were done in mice, but even though germ cell development in mice and humans differs [57] similar methylation results were shown in human (fetal) germ cells [58, 59] (Figure 1A). PGCs at genital ridge are called gonocytes and further mature into oogonia and spermatogonia [2, 42-44] under strict regulation of the genomic constitution (XX vs XY) and the (gonadal) micro-environment. In an XX constitution the transcription factor *SRY* will be expressed leading to *SOX9* expression and the formation of Sertoli cells: the supporting cells in the testis that guide gonocytes into maturation along the male germ cell lineage [60-63]. If SRY is not (sufficiently) expressed (physiological in XY condition), granulosa cells will be formed instead which, in the presence of e.g. *FOXL2*, support the formation of the female germ cell lineage [64, 65]. During maturation, germ cells lose expression of embryonic/pluripotency markers, although expression has been shown in the first years after birth. Delayed extinguishing of pluripotency markers does not necessarily indicate (pre-)malignant transformation, but might just indicate delayed maturation [66-70]; a distinction for which additional markers have been investigated [71, 72].

Malignant transformation of PGCs/gonocytes

TGCC share many of the characteristics of PGCs/gonocytes and the standing hypothesis is that all TGCC arises from PGCs via a common precursor CIS (Figure 1). This hypothesis is based on various lines of evidence. Firstly, a number of targeted studies reviewed the marker profile of CIS and different developmental stages of (fetal) germ cells, identifying strong overlap between PGCs and CIS regarding pluripotency and germ cell markers, most notably KIT, POU5F1 NANOG and TCFAP2C (AP-2γ) [68, 70, 73-78] (Figure 1B). Gene expression analysis of CIS and SE also showed overlapping expression patterns related to early embryogenesis and fetal germ cells [13, 20, 79-88]. Despite their common origin, significant differences in gene expression were observed between the histological subtypes of invasive TGCC [14, 82, 85, 89-91]. Chromosomal constitution, as well as cell cycle control and DNA repair have also been shown to work analogous to pre-

meiotic germ cells in TGCC (the latter two are reviewed below and in [73, 74]). With regard to their epigenetic status there are strong similarities between TGCC and PGCs [57, 92-95]. Specifically, SE mirror the methylation profile of their hypomethylated progenitor CIS [96], which in turn reflects the generally hypomethylated state of PGCs [49, 51]. EC show higher levels of methylation [58, 97]. This presumably reflects a process of *de novo* methylation during tumor progression, potentially indicating a reversal to a more hypermethylated embryonic stem cell (ES) like phenotype. Because all CIS is strongly hypomethylated [96], this observation argues against different types of CIS leading to SE or EC as suggested in some studies [98, 99]. In a recent investigation of whole genome methylation profiling, we confirmed these global findings and relate TGCC to PGCs and early gonocytes based on methylation of specific functional regions (e.g. imprinting control regions) (Rijlaarsdam et al, 2014, submitted for publication) (Figure 1).

The progression of CIS to invasive TGCC is not yet completely understood, although the loss of PTEN has been shown to be a common event [100]. Even though a role of the immune system in TGCC pathogenesis has been suggested [101, 102], studies regarding the specificity and the effect of immune response upon invasion contradict each other [103, 104]. With regard to their genomic constitution upon malignant transformation, polyploidization to a tetraploid stage is thought to be an early event, occurring before the formation of CIS [105, 106]. During progression, TGCCs are thought to acquire heterogeneous chromosomal gains and losses ending up in a near triploid state. A consistent causative relation of these copy number variations to disease (progression) has not been determined so far [107-116]. Most markedly, gain of predominantly the whole short arm of chromosome 12 (12p) is specific to invasive GCCs. This aberration might be functionally relevant in the pathogenesis of these tumors [79, 117-121], a hypothesis further inspired by the fact that induced pluripotent stem cells (iPS) with gain of chromosome 12 show selective advantage during reprograming and increased tumorigenic potential [122-124]. Allelotyping has suggested increasing amounts of loss of heterozygosity in coexisting CIS, SE and EC. Concordance rates between the allelic patterns were higher for CIS vs SE and SE vs EC than for CIS vs EC. This suggests possible progression from CIS to SE to EC [125], a rare transition in mixed tumors (SE+NS) that has also been suggested by anecdotal evidence at the RNA and protein level [126].

Why does a germ cell turn to the dark side?

Genetic risk factors

A 76 fold risk increase in male twins [127] and a ≈25 fold risk increase in patients with a contralateral tumor [128] illustrates the genetic component in GCC pathogenesis (Table 1). Indeed, 25% of GCC susceptibility could be attributed to genetic factors according to an elegant epidemiological study by Czene and coworkers [129]. In targeted studies specific genomic risk loci

/ Single Nucleotide Polymorphisms (SNPs) were identified at the Y chromosome (gr/gr microdeletion) [130, 131] and in the FSH [132] / androgen receptors [133]. Risk loci were also found in PTEN [134] and loci related to genes involved in steroid hormone metabolism [135, 136], cAMP signaling [137], insulin like growth factor signaling [138], DNA damage response [139, 140], estrogen metabolism/signaling [141-143], gonadotropin regulation [144], TP53-KITL interaction [145], sex determination [146] and prostate/colorectal/breast cancer (8q24 locus) [147]. Larger scale GWAS studies consistently identified a number of (other) risk SNPs [148-155] which were validated in additional (targeted) studies: [156-158]. The results of genetic studies in subfertile patients [159] and familial TGCC [160, 161] populations did not significantly differ from the general population, implying a common genetic base in these risk groups and sporadic GCC cases. Recently all major GWAS studies were combined in a very diligent meta-analysis by Chung and coworkers, verifying 14 previously identified risk SNPs and identifying 5 additional ones [162] (Figure 2A). The genes associated with these SNPs are associated with a surprisingly condensed set of biological functions: KIT-KITL receptor tyrosine kinase signaling (reviewed below), telomere maintenance, cell cycle/division control (reviewed below) and establishment of the germ cell niche. More details are presented in the legend of Figure 2. In line with the proposed combination of genetic and environmental risk factors making up total TGCC risk, the per-allele odds-ratio for each SNP was relatively low, the KITL loci being the only ones with ORs > 2 or ORs < 0.5.

Table 1 Clinical risk factors for TGCC. The aim of this table is to give an overview of the risk factors and their impact as most frequently discussed in (recent) literature. Unless otherwise specified, studies were conducted with TGCC- & CIS-individuals ("healthy controls") as reference although in some studies individuals were under medical evaluation for other conditions. In some cases Dieckmann et al present min-max RR/ORs which should be interpreted carefully as no information about variation is presented. Abbreviations RR=relative risk, OR=odds ratio, SIR=standardized incidence ratio, CI=confidence interval, T=testosterone, ns=not significant, arrows=(significantly) higher/lower in TGCC patients, ?=not discussed in paper. CC=case control study; RC=retrospective cohort.

Factor	Risk Quantification	Magnitude	References	Remarks/study design
Familial GCC for review: [127]	2.15-12.3 Mono/dizygotic twin : 76.5 (11.2-518.0)/35.7 (5.2-244.7) Parent/sibling: 4.31 (2.05-7.95)/8.50 (6.01-11.68) Father/brother: 3.78 (2.16-6.16)/8.58 (6.41-11.26)	OR/RR (min-max) SIR (95%CI) SIR (95%CI) SIR (95%CI)	[128] [163] [164] [165]	Review (CC & cohort studies) Retrospective cohort of UK twins RC, Swedish Family Cancer Database RC, Swedish Family Cancer Database
Contralateral tumor	24.8-27.6	OR/RR (min-max)	[128]	Review (CC & cohort studies)
Urological/Andrological				
Cryptorchidism	3.5-17.1 4.1 (3.6-4.7) 3.71 (3.29-4.19) 3.18 (2.50-4.04)	OR/RR (min-max) RR (95%CI) RR (95%CI) OR (95%CI)	[128] [166] [167] [168]	Review (CC & cohort studies) Meta-analysis (cryptorchidism±TGCC) Retrospective cohort, Denmark Retrospective cohort, Sweden

Factor	Risk Quantification	Magnitude	References	Remarks/study design
FSH/LH/T/Inhibin-B	↑/↑/ns/-		[169]	Unilateral CIS vs bilateral CIS, testicular function after unilateral orchidectomy in Denmark.
	?/↑/↓/?		[170]	Orchidectomy for TGCC/benign lesions + healthy controls.
	ns/ns/ns/↓		[171]	RC in NL, microlithiasis±CIS.
Infertility	2.8 (1.16-6.72)	RR (95%CI)	[128]	New data in review (retrospective case-control, Germany)
	2.8 (1.5-4.8)	SIR (95%CI)	[172]	Retrospective (infertile patients vs general population, USA)
	1.6 (1.3-1.9)	SIR (95%CI)	[173]	Subfertile Danish RC; reference incidence in general population
	2.06 (0.64-7.18)	%	[174]	Incidence of CIS in biopsies from azoospermic men without spermatogenesis.
	2.4 (1.4-4.1)	%	[175]	Incidence of CIS in biopsies from infertile men.
Hypospadias	2.13 (1.26-3.61)	RR (95% CI)	[167]	RC, Denmark
	2.41 (1.27-4.57)	OR (95% CI)	[168]	RC, Denmark
Testicular aberrations				
DSD (gonadal dysgenesis)	25%	Life time risk	[128]	Review (CC & cohort studies)
Microlithiasis for review: [176, 177]	13.2 (8.3-21.5)	RR (95%CI)	[178]	Prospective cohort (inclusion: ultrasound for various reasons, UK)
	3.42 (1.82-6.01)	RR (95%CI)	[179]	Retrospective cohort (inclusion: ultrasound for various reasons, USA)
Atrophy	20.5 (2.7-153)	RR (95%CI)	[128]	Retrospective CC, Germany

Environmental risk factors

Apart from the limited magnitude of the observed genetic risk factors alone, the quickly increasing incidence in Western countries cannot be explained by genetic factors alone, drawing attention to the importance of factors related to the environment [2, 4]. This is strongly supported by first generation migrant studies showing a steep increase in incidence after migration to countries with a high TGCC incidence [180]. A comprehensive overview of all environmental and lifestyle factors that have been investigated with regard to TGCC risk lies outside of the scope of this review, has recently been presented by McGlynn and coworkers [181]. Examples include but are not limited to: birth weight, decreased gestational age, maternal smoking and parity and life style factors like dietary intake and physical exercise. In general, these associations with TGCC incidence are weak or inconsistent between studies, except for frequent marijuana use and NS [181-183]. The proposed common denominator in many risk factors is increased (fetal) exposure to estrogens and anti-androgens, supported by disorders observed in sons of women exposed to diethylstilbestrol (DES) during pregnancy [184, 185]. Endocrine disrupting chemicals as present in our daily life are known disrupt sperm function and fertility [186]. The evidence pointing to specific substances that

contribute to GCC risk is however weak [187]. Alkylphenols act like estrogens and stimulated proliferation in the SE cell line TCam-2, possibly via modification of the epigenetic machinery [188]. However, in vitro studies using anti-androgens present in everyday life showed little effect on androgen receptor function (reviewed in [189]). In studies investigating prenatal estrogen (progenitor) levels, the association between high prenatal estrogen exposure and GCC risk was inconsistent or non-significant [190, 191], although higher levels of androgens during adolescence showed a protective effect [192]. The relation of TGCC pathogenesis with (environmental exposure to) estrogens and anti-androgens remains a mechanistically tempting hypothesis, but additional large scale studies are needed to provide more consistent epidemiological evidence of this association and point to specific substances of interest which can be functionally tested.
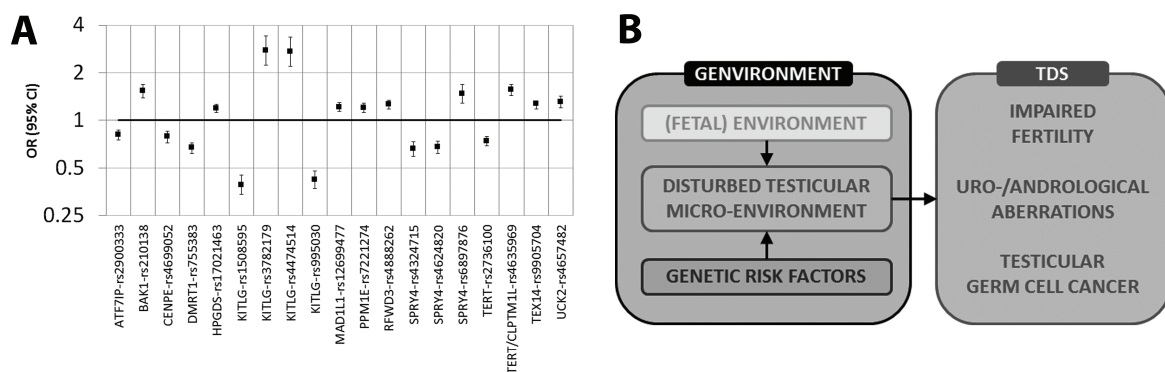
Figure 2. A genvironmental risk model for TGCC. (A) Genetic risk factors. Forest plot of GCC risk SNPs identified in [162]. Four SNPs were located in KITL (reviewed extensively below). Three SNPs were located in SPRY4 (MAP kinase pathway [193]) and one in BAK1 (apoptosis [194]), which are both downstream targets of KITLG/KIT signaling. SPRY4 responds to KIT inhibition by imatinib [195]. TERT and its transcription factor ATF7IP influence telomere maintenance (normally only active in germ cells) [196, 197]. The TERT-CLPTM1L locus is a cancer-hotspot ([149] for review) and CLPTM1L is involved in cisplatin sensitization [198]. DMRT1 influences sex determination and germ cell development [199, 200]. PPM1E is serine/threonine-protein phosphatase implicated as an anti-diabetic drug target [201]. CENPE [202], MAD1L1 [203] and TEX14 [204] are implicated in spindle assembly and progression of cell division as is RFWD3 (G1 checkpoint) [205]. UCK2 has been implicated in the efficiency of specific types of chemotherapeutic drugs [206]. HPGDS has been related to establishment of the fetal germ cell nice (e.g. disturbed nuclear SOX9 expression Sertoli cells in mice) [207]. A number of KITL and SPRY4 SNPs were in strong linkage disequilibrium (r^2>0.8; Broad Institute (SNP reference dataset: 1000 Genomes Pilot 1): KITL (rs4474514/rs3782179; rs4474514/rs1508595; rs3782179/rs1508595; rs4590952/rs1508595; rs4590952/rs3782179; rs4474514/rs4590952; rs995030/rs3782179; rs995030/rs4474514; rs995030/rs4590952) and SPRY4 (rs4624820/rs4324715; rs6897876/rs4624820). These SNPs should therefore not be interpreted as independent risk signals. (B) Global and schematic representation of the integrated role of genetic and environmental risk factors in the pathogenesis of TGCC inspired by the TDS model proposed by Sonne and colleagues [208].

The genvironmental hypothesis and its clinical manifestations

In the previous paragraphs numerous genetic and environmental (T)GCC risk factors have been described. Multiple studies have shown evidence of genetic and environmental factors to interact in the pathogenesis of (T)GCC, i.e. an combined genvironment risk model [94, 172, 209-212] (Figure 2B). In this perspective, epigenetic (de)regulation is a tempting bridge between genetics

and environment [94, 209, 213], even trans-generationally [212, 214]. Genvironmental disturbance results in a spectrum of clinical manifestations associated with increased risk of TGCC that has been baptized testicular dysgenesis syndrome (TDS) [208, 215] (Figure 2B). TDS is also associated with disorders of sex development (DSD). If the GBY region of the Y chromosome is present, DSD patients have 25% lifetime risk of TGCC [128, 216-218]. The spectrum of risk factors associated with GCC is reviewed in Table 1 and many of these are associated with decreased fertility. The (genetic) association between fertility and TGCC risk is emphasized further by lower fertility in TGCC patients [219] and their brothers [220]. Indeed, most studies report higher CIS prevalence in subfertile men as compared to the general population [174, 175, 221]. Taken together, TGCC risk is entwined with infertility and other urological/andrological aberrations and is determined by a complex set of interacting genvironmental modifiers of the germ cell micro-environment, each with individually moderate effect size, together making up the total TGCC risk (Figure 2, Table 1).

How to survive as a germ cell tumor: TGCC make the most of their onco-fetal phenotype

This section focusses on three genvironmentally modulated hallmarks of TGCC directly related to the oncofetal pathogenesis of TGCC: (1) maintenance of pluripotency, (2) cell cycle control/cisplatin sensitivity and (3) regulation of proliferation/migration/apoptosis by KIT-KITL mediated receptor tyrosine kinase signaling (Figure 3). These key factors show an inspiring overlap with the almost axiomatic hallmarks of cancer as defined by Hanahan and Weinberg [222, 223]. Other key associations with TGCC are sex determination and steroid/sex hormone regulation, in line with their fetal germ cell of origin. These and other TGCC related factors are excellently described in a number of recent reviews [2, 224, 225].

(Maintenance of) pluripotency

In the previous sections, the early embryonic origin of TGCC was reviewed. The fact that mainstream TGCC markers OCT3/4 and SOX2 represent two out of four Yamanaka factors required to create iPS cells emphasizes this [226, 227]. This is of interest, especially as specific iPS reprogramming protocols have been shown to induce GCC like tumors [228]. The parallel between GCCs (teratomas) and potential complications of stem cell therapy also illustrates this point [228-230]. OCT3/4, SOX2 and NANOG - together with a number of other factors - work in close harmony in regulating pluripotency. They bind to a large numbers of promoter regions, including their own [231, 232]. Apart from this complex crossover in auto-regulation of transcription they are all subjected to stringent post-translational control [230, 233, 234]. Key regulators of the pluripotency will be discussed in the context of TGCC in the following sections.
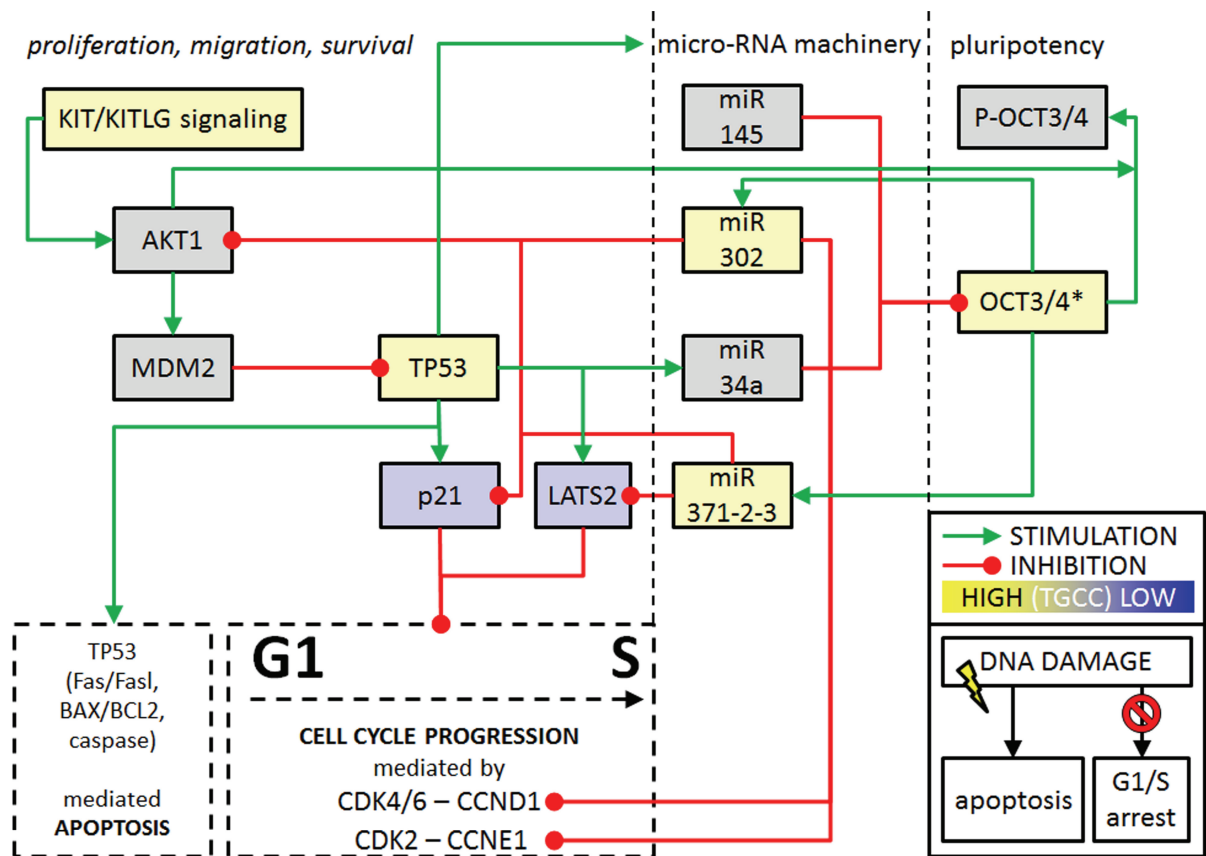
Figure 3. Crosstalk between pluripotency and cell cycle regulation is facilitated by (ES specific) microRNAs. This scheme fuses the evidence presented in the main text. Basically, two major onco-fetal mechanisms are at work in TGCC. The presence of pluripotency markers results in upregulation of ES specific miR families 371-2-3 and 302 which in turn (down) regulate G1/S checkpoint regulation to facilitate rapid transition through the cell cycle while at the same time preserving high sensitivity to DNA damaging agents (summarized schematically in the bottom right inlay). Along this integration of pluripotency and cell cycle control, KIT-KITL mediated receptor tyrosine kinase signaling also regulates proliferation, migration and survival during embryogenesis and germ cell development. * (miR) transcription is regulated by a complex of multiple transcription factors like OCT3/4, SOX2, NANOG, etc. For simplicity, only OCT3/4 is displayed. Colored boxes indicate known expression levels in TGCC (grey in no clear evidence).

OCT3/4 & NANOG

OCT3/4 (POU5F1) is detected in all stem cell components and CIS and is used as a mainstream diagnostic marker [2, 17, 235-238], but might also be present in maturation delayed (but non-malignant) germ cells [71, 72, 239-241] (Figure 1). OCT3/4 is a long standing master regulator of self-renewal, pluripotency and prevention of apoptosis in embryonic stem (ES) cells and in the early developing embryo, including PGCs [242-247]. Expression is lost during differentiation [248-252]. Reciprocally, OCT3/4 downregulation initiates differentiation [245] or, in PGCs, apoptosis [246, 253, 254].

The gene encoding the OCT3/4 transcription factor is located on 6p21 and consists of 5 exons [255-257]. Investigations of OCT3/4 expression and protein availability are hampered by the presence of similar transcripts with different or complimentary functions, e.g. non-protein coding

pseudogenes (n=5) [235, 243, 258, 259], anti-sense transcribed pseudogenes [260] and splice-variants (OCT4A: pluripotency, OCT4B: stress responseOCT4B1: possibly both) [237, 243, 261-266]. In addition, translation of OCT3/4 is mediated by micro-RNAs, specifically miR-145 [267, 268] and miR-34a [269]. Also, post-translational modification of OCT3/4, i.e. phosphorylation by AKT has been shown to change its affinity / specificity for DNA binding sites, potentially changing the functional impact of this transcription factor [270].

Like OCT3/4 and in line with their pluripotent phenotype, the homeodomain NANOG is only detectable in the TGCC stem cell components SE and EC and in CIS and not in differentiated NS derivatives [18, 271] (Figure 1). NANOG is physiologically responsible for maintaining pluripotency in the inner cell mass [272, 273] and is expressed in early germ cells [18]. Of interest, NANOG is located on chromosome 12 which frequently shows gain in TGCC [274]. Downregulation of NANOG causes differentiation into extra-embryonic lineages [275] or, in PGCs, apoptosis [276].

SOX2 & SOX17

Two members of the SRY-related HMG-box family are important TGCC markers associated with pluripotency: SOX17 in CIS / SE and SOX2 in EC [23, 24, 277] (Figure 1). SOX2 is expressed in pluripotent and multipotent embryonic and extra-embryonic cells [278]. However according to a number of studies SOX2 is not expressed in developing human germ cells as is the case in mice [277, 279, 280]. A single study showed SOX2 in normal germ cells/CIS/SE, but the authors stated that this is likely an artefact of a single antibody showing this result [281]. Instead of SOX2, human germ cell progenitors express SOX17 [224, 277, 279] which is reflected in SOX17 expression in CIS and SE [23, 24, 277]. SOX17 is involved in regulating differentiation into primitive endoderm [247, 282]. Interestingly, SOX2 and SOX17 both operate together with OCT3/4 in regulating transcription. However, the two protein pairs have different affinities for specific DNA binding motifs, presumably differentially regulating genes contributing to pluripotency or endodermal lineage commitment respectively [283].

microRNAs

Knockdown of the micro-RNA (miR) machinery in ES cells showed that miRs are crucial to ES cell cycle progression and differentiation [284-288]. This implicated these 22 nucleotide long (predominantly) inhibitors of translation as potential regulators of pluripotency. Indeed, induction of specific miRs have been shown to greatly enhance the efficiency of creating iPS [289, 290]. Reciprocally, master transcription factors regulating pluripotency (OCT3/4, SOX2, NANOG, TCF3) have been shown to regulate transcription of ES cell specific miRs [291]. Another example is the inhibition of the tumor suppressor miR let-7 by LIN28 [292, 293]. LIN28 is a WNT signaling associated maintainer of pluripotency which is normally highly expressed in developing germ cells and the stem cell components of TGCC [19]. An ES specific subset of these ≈22 long nucleotide

inhibitors of mRNA translation has been identified including the miR-302abdc/367 (302), miR-371-2-3, miR-290 and miR-17-19 clusters which show similarity in seed sequence [291, 294, 295]. Moreover, the miR-371-2-3 and 302 clusters are expressed highly in TGCC tumor tissues [13-15] and are detectable in serum of TGCC patients [14, 37-41]. These miRs will be discussed in more detail together with ES like cell cycle regulation and cisplatin sensitivity below.

Cell division, apoptosis and cisplatin chemosensitivity

ES cell cell cycle progression is mirrored in TGCC and is an important factor in its chemo sensitivity

In normal somatic tissue TP53 represents the proverbial break on unlimited proliferation / survival and is therefore a key target for oncogenic transformation in many cancers [296-298] as well as in iPS generation [299]. TGCC however only rarely show TP53 mutations but generally exhibit relatively high TP53 expression [300-303]. An explanation for the abundant TP53 in TGCC combined with their seemingly contradictory exquisite sensitivity to DNA damaging agents might lie in the onco-fetal origin of TGCC. The stem cell components and progenitors of TGCC exhibit pluripotent ES-like properties and are therefore equipped with high chemosensitivity by design [304, 305] (see below). Of note, this argues against therapeutic agents that induce differentiation as suggested before [306, 307]. Another illustration of the innate chemosensitivity of TGCC is the fact that mutations in cisplatin resistant tumors are much frequent (up to 25%) [308] than in TGCC in general (rare) [309, 310], illustrating that TGCCs need additional hits to escape their instrinsic, ES like apoptotic response to DNA damage.

ES cells exhibit very short gap phases in the cell cycle and rapidly progress through the G1/S checkpoint [311, 312]. This way they bypass TP53-p21 cell cycle mediated DNA damage response but instead choose a combination of rapid apoptosis as response to DNA damage and at the same time facilitate quick progression through the cell cycle. This allows rapid growth while protecting ES/germ line genomic integrity [224]. In TGCC this behavior is mirrored: high TP53 expression is associated with cisplatin sensitivity specifically in the presence of low p21 [302, 313-320]. A number of other downstream targets of G1/S cell cycle progression have been identified as mechanistic in TGCC cisplatin response. On the other hand, consistent defects in DNA damage response were only found in cisplatin resistant TGCC (reviewed in [224]) (Figure 3).

(Embryonic) miRs facilitate crossover between pluripotency and cell cycle progression

Two ES specific miR clusters both target cell cycle control at the G1/S checkpoint, but with different (quantitative) outcomes: miR-371-2-3 and miR-302abcd/367 (302). miR-371-2-3 has been shown to inhibit LATS2 and p21. These proteins normally facilitate TP53 mediated inhibition of cyclin dependent kinases involved in G1/S cell cycle progression and their inhibition allows rapid G1/S progression [13, 224, 321]. Interestingly, a positive feedback loop between TP53 and LATS2

prevents tetraploidization [322]. This is one of the earliest steps in TGCC pathogenesis, in this context possibly related to disruption of TP53 mediated LATS2 upregulation by miR-371-2-3.

The net influence of miR-302 on cell cycle progression is still under debate [323-326]. Validated miR-302 targets include many players in G1/S regulation. For example, CCND1 is a validated target and his closely related family member CCNE2 has also shown to be downregulated upon miR-302 expression [325]. Moreover, AKT - an upstream regulator of cell cycle progression - has been functionally validated as a miR-302 target leading to less MDM2 regulated inhibition of functionally active TP53 [326]. However, miR-302 has also been shown to interact directly with p21 mRNA, stimulating cell cycle progression [323, 324, 327]. Overall, cell cycle related ES specific miRs contribute to a complex, tightly controlled balance of rapid, but not uncontrolled proliferation.

miR-34a also influences the cross-over between pluripotency and cell cycle control, targeting OCT3/4 as discussed previously. miR 34-a is TP53 regulated and involved in cell cycle regulation [269]. This miR has been shown to increase iPS generation efficiency [328] and has been implicated in TP53 mediated inhibition of WNT signalling / epithelial to mesenchymal transition [329]. Its family member miR-34c, which has been shown to induce TP53 mediated apoptosis of germ cells in dairy goats [330] and is present in human germ cells [331]. Interestingly, TP53 itself has also been shown to interact with the miR production machinery, illustrating the complex interactions that guard the pluripotent homeostatic state [332, 333].

Because these miRs are at the TGCCs disposal [13-15], they can acquire rapid cell cycle progression without TP53 mutation, facilitating malignant transformation [334]. Indeed, mutations of the core regulator of the miR machinery (DICER1) are infrequent in sporadic and familial TGCC [335, 336]. The miR-371-2-3 family is highly expressed in TGCC as discussed above, but is not expressed in the only TP53 mutant GCC cell line NCCIT [337, 338] and only showed absence of expression in a fraction of the rare TGCC with a TP53 mutation [13]. Although expression levels are not necessarily associated with functional impact, the discrepancy between the levels of miR-371-2-3 and miR-302 in TGCCs (371-2-3>302x) might point to an evolutionary selection of miR-371-2-3, which most directly stimulates rapid proliferation and does not hypothetically stimulate apoptosis as well as is the case for miR-302 via AKT-MDM2-TP53 [224]. In summary, the ES like cell cycle progression facilitated by these fetal onco-miRs clearly illustrates the unique onco-fetal signaling that allows malignant transformation of totipotent embryonic germ cells into TGCC while retaining a chemo sensitive phenotype (Figure 3). Additional disturbances during progression might induce resistance to DNA damaging agents.

Receptor tyrosine kinase signaling: KIT-KITL

KIT/KITLG tyrosine kinase signaling has been consistently and crucially implicated in in embryogenesis, germ cell development and TGCC. The function of this pathway includes many of the prerequisites for a cancer cell: proliferation, migration and survival. KIT & KITLG establish their effects via a large number of interacting signaling routes amongst which are PI3-kinase-AKT, Src family kinases (SFK), mitogen-activated protein kinase (MAPK) pathways and phospholipases (reviewed in [339]) (Figure 3). These signaling pathways are crucial in physiological functioning of many tissue types / diseases including other types of cancer (reviewed in supplementary data). The human KIT gene (c-KIT/Piebald Trait Protein, PBT) encodes the cellular homolog of the viral oncogene v-kit or Hardy Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog. The genetic structure of KIT, its transcriptional regulation and degradation (turnover) has been excellently reviewed before [339]. Briefly, the KIT gene is located on 4q11-q12 and encodes a type III receptor tyrosine kinase with four isoforms. Another KIT variant is truncated, has no kinase activity, but was still proven to be functional in signaling (see below) [339]. The ligand of the KIT receptor is KITLG (stem cell factor, SCF) which causes dimerization of KIT monomers and subsequent activation of downstream signaling cascades [339, 340]. KITLG is located on chromosome 12q21.32 and is mainly expressed as two forms: membrane bound and soluble. The soluble form is initially bound to the cell membrane, but contains an additional 6th exon encoding an extracellular cleavage site. Both forms activate KIT, but with qualitative and quantitative differences [339].

KIT/KITLG in germ cell development

Fetal development

KIT and KITLG regulate PGC development and homing to the genital ridge [42, 339, 341-343]. In the embryonic phase the guidance of KIT positive PGCs from the hind gut epithelium to the gonads depends strongly on KITLG mediated chemo attraction [43, 342, 344, 345]. KIT or KITLG mutant mice show normal establishment of the germ cell lineage in het hind gut, but aberrant proliferation and unsuccessful migration during embryogenesis [42, 339]. Different mutations influence different phases in germ cell migration/proliferation [346]. KITLG or KIT mutant germ cells eventually go into apoptosis, effectively resulting in sterility (males worse than females) [42, 339]. This is in line with PI3K-AKT mediated inhibition of apoptosis being a major downstream effect of KIT signaling [347]. It also illustrates the characterization of KIT as a dependence receptor: without the ligand, the cells will go into apoptosis [339, 348, 349]. Upon arrival in the testis, estrogen signaling is an important regulator of KIT dependent embryonic gonadal development effecting both germ cell progenitors and supportive cells (Leydig cells) [350, 351]. More specifically, KIT phosphorylation is up regulated in PGCs by estrogen stimulation [350, 352]. Indeed, in the SE cell line TCam-2, representative of the malignant counterpart of committed germ cell progenitors [88], dependence on estrogen was also shown [353].

The postnatal testis

In the postnatal testis KIT-KITLG signaling takes place via paracrine signaling in the germline stem cell niche and is crucial for spermatogenesis from the pre-spermatogonial stage onwards [339, 342, 354-356]. Spermatogonia are positive for KIT, as are Leydig cells in which KIT signaling triggers testosterone production [342, 357]. This is differently regulated in mice. Sertoli cells on the other hand produce KITLG [339, 358, 359] under the influence of FSH [339, 360, 361]. This hormonal sensitivity is activated upon puberty (rat & primate models) [362, 363]. A feedback loop between KIT positive germ cells and KITLG production by Sertoli cells has been suggested because of increasing Sertoli cell KITLG-levels in men with decreasing fertility (highest: Sertoli cell only) [364-366]. This micro-environmental interplay is of crucial important for normal germ cell development, as was also shown in patients with non-obstructive azoospermia by (abnormal Sertoli cells) [367]. On a molecular level, KITLG contributes directly to PI3K mediated proliferation of germ cells as shown in cell cultures of spermatogonia [368]. Indeed, development of post-meiotic germ cells germ cells relies strongly on the presence of the PI3-kinase binding site in KIT, subsequently regulating AKT activation. If this p85 binding site is selectively mutated in mice, only the germ cell lineage suffers [369, 370]. Growth factors, hormones, cytokines and vitamins (e.g. *BMP4* & retinoic acid) regulating KIT expression in male germ cells as well as the molecular basis for regulation (e.g. testis specific transcription factors (*ZBTB16*, *AP-2γ*, *SOHLH1/2 [371]*, *SIX5*) and miR-221/222), has been reviewed before [342, 372, 373]. More mature mouse spermatids and spermatozoa express a C-terminal truncated form of KIT transcribed from an intronic promotor [358]. Transcription is presumably regulated by retinoic acid which is crucial to germ cell fate determination and timing of meiosis in mice [356, 366]. This truncated isoform lacks kinase activity via PI3K and does not have the extracellular/transmembrane components [374]. It is nevertheless able to activate egg activation of metaphase II activated mouse oocytes [358, 375-377]. It functions as a scaffold protein, interacting with Src-like tyrosine kinase Fyn [378]. In human testis, truncated KIT is expressed during spermatogenesis and maintained in ejaculated sperm. The protein is expressed in the equatorial region of human spermatozoa. This part enters the oocyte first, which is consistent with mouse oocyte activation by truncated KIT injection. Truncated KIT expression is correlated with sperm quality parameters [379].

KIT/KITLG in germ cell cancer

Analogous to its role in germ cell development and consistent with the importance of the germ cell micro-environment, disrupted KIT/KITLG signaling is crucial in the development of TGCC. Indeed both SE cells [380] (Sertoli cell feeder) and spermatogonia [368] depend on KITLG in culture. In mice, loss of functional transmembrane, but loss of soluble KITLG, leads to an increased incidence of TGCC like tumors (teratomas) [381]. Mechanistically, constitutive paracrine or autocrine activation of KIT/KITLG signaling is implicated to be a crucial initiating event for the malignant transformation of maturation arrested germ cell progenitors [2, 224]. In the early stages,

KITLG positivity is a hallmark of maturation arrested germ cells, CIS and intratubular SE [2, 69, 74, 382]. Progression into invasive SE is also strongly related to KIT/KITLG signaling while much less association with the NS phenotype has been shown [354, 383-386]. Indeed, KIT expression is significantly (19-fold) higher in SE as compared to NS (p<0.01, 2-sided t-test) while the inverse was true for KITLG (5-fold, p<0.01, 2-sided t-test, data from our group, published before in [72])). These results match with a previous study [387]. This suggests KIT dependence of SE and possibly a KITLG saturated micro-environment (threshold level for effect on viability already suggested for KIT in [383]). In the SE cell line TCam-2 [388-391] Goddard and coworkers identified a small but significant decrease in viability upon KIT knock down [383]. In a similar experiment from our group we could not replicate this result, nor did we detect a consistent effect of KITLG knock down or exogeneous KITL addition on TCam-2 viability (results in supplementary data). On the DNA level, KIT was the only consistently amplified gene in an amplicon present in 17-21% of all SE (rare in NS) [383, 392]. In a small series of adjacent precursor lesions, no amplification was identified [392] which, in contrast to KITLG positivity from the pre-malignant stage onwards, suggests KIT amplification or mutation to be a downstream event related to progression, coming into focus after the formation of the TGCC precursor lesion [383, 392].

Activating KIT mutations are identified in ca 13-60% of the SE (rare in NS) and result in constitutive kinase activity because of ligand independent dimerization and phosphorylation [392-397]. Of note, KIT mutations are much less frequent in high GCC-risk patients with Disorders of Sex Development, a major risk factor for GCC. In these cases an alternative, TSPY related pathogenesis has been suggested [398]. The effect of mutations depends both on the mutation and the cell type [339, 399]. In SE KIT mutations are predominantly found in exon 17 / D816V [395, 398], which might explain why many TGCC are innately resistant to the tyrosine kinase inhibitor imatinib [400, 401]. In bilateral germ cell tumors, somatic KIT mutations (codon D816X) were identified in 63-93% of the cases [402, 403], but this high frequency was not replicated in independent, smaller studies [383, 404, 405]. In addition, there is no indication of germ-line mutations of KIT in familiar GCC [406]. However, KITLG promoter methylation was significantly lower in blood of these patients [407] and SNPs in KITLG combined with aberrations in cAMP regulation were suggested to contribute to tumor risk in these patients [408]. In the end, no quantitative relation has been shown between overexpression and amplification / mutation status of KIT in SE [383, 392]. As reviewed under "genetic risk factors" and in Figure 2, many of the SNPs identified in GWAS studies investigating GCC risk loci were related to KIT-KITL signaling.

Concluding remarks

This review focused on the early embryonic origin of TGCC and the onco-fetal associations in their pathogenesis. Most importantly, TGCC exhibit identifiable stem cell components (SE and EC) and progenitors that show large and consistent similarities to PGCs, their presumed embryonic totipotent cells of origin. TGCC pathogenesis depends crucially on a complex interaction of genetic and (micro-)environmental, i.e. genvironmental risk factors that have only been partly elucidated despite significant effort. TGCC stem cell components show a high degree of similarity with ES cells and embryonic germ cells in the regulation of pluripotency and cell cycle control, directly related to their exquisite sensitivity to DNA damaging agents (cisplatin). Of note, (ES specific) micro-RNAs play a pivotal role in the crossover between cell cycle control, pluripotency and chemosensitivity. Moreover multiple consistent observations reported TGCC to be associated with KIT-KITL mediated receptor tyrosine kinase signaling, a pathway crucially implicated in proliferation, migration and survival during embryogenesis including germ cell development. In conclusion, TGCCs are a fascinating model for onco-fetal developmental processes especially with regard to studying cell cycle control, pluripotency maintenance and KIT-KITL signaling. The knowledge presented here contributes to better understanding of the molecular characteristics of TGCC pathogenesis, translating to identification of at risk individuals and enhanced quality of care for TGCC patients (diagnosis, treatment and follow-up).

Acknowledgements

Supporting information

Figure S1. No effect of KIT/KITL knock down and exogeneous KITL addition on the viability of SE cell line TCam-2. **(Abbreviations)** UNTR=untreated; TFR=transfection reagents only; siCtrl=control oligo; siKIT=siRNA directed against KIT; siKITLG=siRNA directed against KITLG. **(A)** y-axis depicts the expression level of KIT and both major isoforms of KITLG for all conditions (48hr knock down). dCt calculated as suggested by (reference-target) [409] **(B)** Western blot of KITLG and KIT for all conditions (72hr knock down). Loading control: beta-Actin. **(C)** Results of WST-1 proliferation assay (Roche, Woerden, The Netherlands) on TCam-2 cells cultured in 10% FCS supplemented medium. Y-axis: optical density, arbitrary measure indicating increasing vitality. Error bars depict the standard error of the mean of 8 independent quantifications (4-6 day knock down). **(D)** Idem C, but now for wild type TCam-2 cells cultured with varying amounts of FCS ± KITLG (50 ng/ml). Four replicate experiments are displayed. **(E)** Validation of the function of KITLG on MO7E cells which are KITLG (SCF) or GM-CSF (GM) dependent. Experiments were conducted in duplo. No GF=no growth factor added.

Additional review of related topics. (I) GCT animal models and cell lines. **(II)** KIT-KITL signaling in non-germ cell (pathy)physiology and development.

Supporting information is available at: http://dx.doi.org/10.1016/j.semcancer.2014.07.003

References

1.	Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5:210-222.
2.	Looijenga LH: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J Pathol* 2009, 218:146-162.
3.	Looijenga LH, Oosterhuis JW: **Pathobiology of germ cell tumors - applying the gossip test!** *The International journal of developmental biology* 2013, 57:289-298.
4.	Horwich A, Shipley J, Huddart R: **Testicular germ-cell cancer.** *Lancet* 2006, 367:754-765.
5.	Garner MJ, Turner MC, Ghadirian P, Krewski D: **Epidemiology of testicular cancer: an overview.** *Int J Cancer* 2005, 116:331-339.
6.	Alanee SR, Feldman DR, Russo P, Konety B: **Long-term mortality in patients with germ cell tumors: effect of primary cancer site on cause of death.** *Urol Oncol* 2014, 32:26 e29-15.
7.	Woodward PJ, Heidenreich A, Looijenga LHJ, al. e: **Testicular germ cell tumors.** In *World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs*. Edited by Eble JN, Sauter G, Epstein JI, Sesterhann IA, eds. Lyon: IARC Press; 2004: 17-278
8.	Oosterhuis JW, Stoop H, Honecker F, Looijenga LH: **Why human extragonadal germ cell tumours occur in the midline of the body: old concepts, new perspectives.** *International journal of andrology* 2007, 30:256-263; discussion 263-254.
9.	Fan R, Ulbright TM: **Does intratubular germ cell neoplasia, unclassified type exist in prepubertal, cryptorchid testes?** *Fetal and pediatric pathology* 2012, 31:21-24.
10.	Jorgensen N, Muller J, Giwercman A, Skakkebaek NE: **Clinical and biological significance of carcinoma in situ of the testis.** *Cancer surveys* 1990, 9:287-302.
11.	Giwercman A, Muller J, Skakkebaek NE: **Prevalence of carcinoma in situ and other histopathological abnormalities in testes from 399 men who died suddenly and unexpectedly.** *J Urol* 1991, 145:77-80.
12.	Honecker F, Stoop H, Mayer F, Bokemeyer C, Castrillon DH, Lau YF, Looijenga LH, Oosterhuis JW: **Germ cell lineage differentiation in non-seminomatous germ cell tumours.** *J Pathol* 2006, 208:395-400.
13.	Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, 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, Oosterhuis JW, Sun Y, Chen C, Guenther S, Sherlock J, Veltman I, Baeten J, et al: **High-throughput microRNAome analysis in human germ cell tumours.** *J Pathol* 2007, 213:319-328.
15.	Palmer RD, Murray MJ, Saini HK, van Dongen S, Abreu-Goodger C, Muralidhar B, Pett MR, Thornton CM, Nicholson JC, Enright AJ, Coleman N: **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, Nielsen JE, Bork-Jensen J, Almstrup K, Sonne SB, Kjems J, Rajpert-De Meyts E, Leffers H: **MicroRNA expression profiling of carcinoma in situ cells of the testis.** *Endocr Relat Cancer* 2012, 19:365-379.
17.	Looijenga LH, Stoop H, de Leeuw HP, de Gouveia Brazao CA, Gillis AJ, van Roozendaal KE, van Zoelen EJ, Weber RF, Wolffenbuttel KP, van Dekken H, et al: **POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.** *Cancer Res* 2003, 63:2244-2250.
18.	Hart AH, Hartley L, Parker K, Ibrahim M, Looijenga LH, Pauchnik M, Chow CW, Robb L: **The pluripotency homeobox gene NANOG is expressed in human germ cell tumors.** *Cancer* 2005, 104:2092-2098.
19.	Gillis AJ, Stoop H, Biermann K, van Gurp RJ, Swartzman E, Cribbes S, Ferlinz A, Shannon M, Oosterhuis JW, Looijenga LH: **Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis.** *International journal of andrology* 2011, 34:e160-174.
20.	Skotheim RI, Lind GE, Monni O, Nesland JM, Abeler VM, Fossa SD, Duale N, Brunborg G, Kallioniemi O, Andrews PW, Lothe RA: **Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development.** *Cancer Res* 2005, 65:5588-5598.
21.	Chieffi P, Chieffi S: **Molecular biomarkers as potential targets for therapeutic strategies in human testicular germ cell tumors: an overview.** *J Cell Physiol* 2013, 228:1641-1646.

22. Stoop H, Kirkels W, Dohle GR, Gillis AJ, den Bakker MA, Biermann K, Oosterhuis W, Looijenga LH: Diagnosis of testicular carcinoma in situ '(intratubular and microinvasive)' seminoma and embryonal carcinoma using direct enzymatic alkaline phosphatase reactivity on frozen histological sections. *Histopathology* 2011, 58:440-446.

23. Santagata S, Ligon KL, Hornick JL: Embryonic stem cell transcription factor signatures in the diagnosis of primary and metastatic germ cell tumors. *The American journal of surgical pathology* 2007, 31:836-845.

24. Emerson RE, Ulbright TM: Intratubular germ cell neoplasia of the testis and its associated cancers: the use of novel biomarkers. *Pathology* 2010, 42:344-355.

25. Pauls K, Jager R, Weber S, Wardelmann E, Koch A, Buttner R, Schorle H: Transcription factor AP-2gamma, a novel marker of gonocytes and seminomatous germ cell tumors. *Int J Cancer* 2005, 115:470-477.

26. Hong SM, Frierson HF, Jr., Moskaluk CA: AP-2gamma protein expression in intratubular germ cell neoplasia of testis. *Am J Clin Pathol* 2005, 124:873-877.

27. Weber S, Eckert D, Nettersheim D, Gillis AJ, Schafer S, Kuckenberg P, Ehlermann J, Werling U, Biermann K, Looijenga LH, Schorle H: Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biol Reprod* 2010, 82:214-223.

28. von Eyben FE, Liu FJ, Amato RJ, Fritsche HA: Lactate dehydrogenase isoenzyme 1 is the most important LD isoenzyme in patients with testicular germ cell tumor. *Acta oncologica (Stockholm, Sweden)* 2000, 39:509-517.

29. Sturgeon CM, Duffy MJ, Stenman UH, Lilja H, Brunner N, Chan DW, Babaian R, Bast RC, Jr., Dowell B, Esteva FJ, et al: National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. *Clin Chem* 2008, 54:e11-79.

30. Salem M, Gilligan T: Serum tumor markers and their utilization in the management of germ-cell tumors in adult males. *Expert review of anticancer therapy* 2011, 11:1-4.

31. Hessel M, Ramos L, Hulsbergen AF, D'Hauwers KW, Braat DD, Hulsbergen-van de Kaa CA: A novel cell-processing method 'AgarCytos' in conjunction with OCT3/4 and PLAP to detect intratubular germ cell neoplasia in non-obstructive azoospermia using remnants of testicular sperm extraction specimens. *Hum Reprod* 2013, 28:2608-2620.

32. Hoei-Hansen CE, Carlsen E, Jorgensen N, Leffers H, Skakkebaek NE, Rajpert-De Meyts E: Towards a non-invasive method for early detection of testicular neoplasia in semen samples by identification of fetal germ cell-specific markers. *Hum Reprod* 2007, 22:167-173.

33. Nielsen JE, Kristensen DM, Almstrup K, Jorgensen A, Olesen IA, Jacobsen GK, Horn T, Skakkebaek NE, Leffers H, Rajpert-De Meyts E: A novel double staining strategy for improved detection of testicular carcinoma in situ cells in human semen samples. *Andrologia* 2012, 44:78-85.

34. van Casteren NJ, Stoop H, Dohle GR, de Wit R, Oosterhuis JW, Looijenga LH: Noninvasive detection of testicular carcinoma in situ in semen using OCT3/4. *Eur Urol* 2008, 54:153-158.

35. Kawakami T, Okamoto K, Ogawa O, Okada Y: XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer. *Lancet* 2004, 363:40-42.

36. Ellinger J, Albers P, Perabo FG, Muller SC, von Ruecker A, Bastian PJ: CpG island hypermethylation of cell-free circulating serum DNA in patients with testicular cancer. *J Urol* 2009, 182:324-329.

37. Belge G, Dieckmann KP, Spiekermann M, Balks T, Bullerdiek J: Serum levels of microRNAs miR-371-3: a novel class of serum biomarkers for testicular germ cell tumors? *Eur Urol* 2012, 61:1068-1069.

38. Dieckmann KP, Spiekermann M, Balks T, Flor I, Loning T, Bullerdiek J, Belge G: MicroRNAs miR-371-3 in serum as diagnostic tools in the management of testicular germ cell tumours. *Br J Cancer* 2012, 107:1754-1760.

39. Gillis AJ, Rijlaarsdam MA, Eini R, Dorssers LC, Biermann K, Murray MJ, Nicholson JC, Coleman N, Dieckmann KP, Belge G, et al: Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle. *Molecular oncology* 2013, 7:1083-1092.

40. Murray MJ, Coleman N: Testicular cancer: a new generation of biomarkers for malignant germ cell tumours. *Nat Rev Urol* 2012, 9:298-300.

41. Murray MJ, Halsall DJ, Hook CE, Williams DM, Nicholson JC, Coleman N: 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.

42. McLaren A: Primordial germ cells in the mouse. *Dev Biol* 2003, 262:1-15.

43. Runyan C, Schaible K, Molyneaux K, Wang Z, Levin L, Wylie C: **Steel factor controls midline cell death of primordial germ cells and is essential for their normal proliferation and migration.** *Development* 2006, 133:4861-4869.

44. Donovan PJ: **Growth factor regulation of mouse primordial germ cell development.** *Current topics in developmental biology* 1994, 29:189-225.

45. Messerschmidt DM: **Should I stay or should I go: protection and maintenance of DNA methylation at imprinted genes.** *Epigenetics : official journal of the DNA Methylation Society* 2012, 7:969-975.

46. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W: **Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.** *Nature* 2010, 463:1101-U1126.

47. Reik W, Dean W, Walter J: **Epigenetic reprogramming in mammalian development.** *Science* 2001, 293:1089-1093.

48. Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W: **The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.** *Mol Cell* 2012, 48:849-862.

49. Seisenberger S, Peat JR, Reik W: **Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells.** *Curr Opin Cell Biol* 2013, 25:281-288.

50. Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W, et al: **Programming and inheritance of parental DNA methylomes in mammals.** *Cell* 2014, 157:979-991.

51. Matsui Y, Mochizuki K: **A current view of the epigenome in mouse primordial germ cells.** *Mol Reprod Dev* 2014, 81:160-170.

52. Saitou M, Kagiwada S, Kurimoto K: **Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells.** *Development* 2012, 139:15-31.

53. Payer B, Lee JT, Namekawa SH: **X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells.** *Hum Genet* 2011, 130:265-280.

54. Ohhata T, Wutz A: **Reactivation of the inactive X chromosome in development and reprogramming.** *Cellular and molecular life sciences : CMLS* 2013, 70:2443-2461.

55. Deuve JL, Avner P: **The coupling of X-chromosome inactivation to pluripotency.** *Annual review of cell and developmental biology* 2011, 27:611-629.

56. de Napoles M, Nesterova T, Brockdorff N: **Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells.** *PloS one* 2007, 2:e860.

57. Kristensen DG, Skakkebaek NE, Rajpert-De Meyts E, Almstrup K: **Epigenetic features of testicular germ cell tumours in relation to epigenetic characteristics of foetal germ cells.** *The International journal of developmental biology* 2013, 57:309-317.

58. Wermann H, Stoop H, Gillis AJ, Honecker F, van Gurp RJ, Ammerpohl O, Richter J, Oosterhuis JW, Bokemeyer C, Looijenga LH: **Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance.** *J Pathol* 2010, 221:433-442.

59. Almstrup K, Nielsen JE, Mlynarska O, Jansen MT, Jorgensen A, Skakkebaek NE, Rajpert-De Meyts E: **Carcinoma in situ testis displays permissive chromatin modifications similar to immature foetal germ cells.** *Br J Cancer* 2010, 103:1269-1276.

60. Wilhelm D, Koopman P: **The makings of maleness: towards an integrated view of male sexual development.** *Nature reviews* 2006, 7:620-631.

61. Polanco JC, Koopman P: **Sry and the hesitant beginnings of male development.** *Dev Biol* 2007, 302:13-24.

62. Payne CJ, Gallagher SJ, Foreman O, Dannenberg JH, Depinho RA, Braun RE: **Sin3a is required by sertoli cells to establish a niche for undifferentiated spermatogonia, germ cell tumors, and spermatid elongation.** *Stem Cells* 2010, 28:1424-1434.

63. Silvan U, Diez-Torre A, Moreno P, Arluzea J, Andrade R, Silio M, Arechaga J: **The spermatogonial stem cell niche in testicular germ cell tumors.** *The International journal of developmental biology* 2013, 57:185-195.

64. Baron D, Batista F, Chaffaux S, Cocquet J, Cotinot C, Cribiu E, De Baere E, Guiguen Y, Jaubert F, Pailhoux E, et al: **Foxl2 gene and the development of the ovary: a story about goat, mouse, fish and woman.** *Reproduction, nutrition, development* 2005, 45:377-382.

65. Ottolenghi C, Uda M, Crisponi L, Omari S, Cao A, Forabosco A, Schlessinger D: **Determination and stability of sex.** *BioEssays : news and reviews in molecular, cellular and developmental biology* 2007, 29:15-25.

66.   Cools M, van Aerde K, Kersemaekers AM, Boter M, Drop SL, Wolffenbuttel KP, Steyerberg EW, Oosterhuis JW, Looijenga LH: **Morphological and immunohistochemical differences between gonadal maturation delay and early germ cell neoplasia in patients with undervirilization syndromes.** *J Clin Endocrinol Metab* 2005, 90**:**5295-5303.

67.   Emerson RE, Cheng L: **Premalignancy of the testis and paratestis.** *Pathology* 2013, 45:264-272.

68.   Jorgensen N, Rajpert-De Meyts E, Graem N, Muller J, Giwercman A, Skakkebaek NE: **Expression of immunohistochemical markers for testicular carcinoma in situ by normal human fetal germ cells.** *Laboratory investigation; a journal of technical methods and pathology* 1995, 72**:**223-231.

69.   Honecker F, Stoop H, de Krijger RR, Chris Lau YF, Bokemeyer C, Looijenga LH: **Pathobiological implications of the expression of markers of testicular carcinoma in situ by fetal germ cells.** *J Pathol* 2004, 203**:**849-857.

70.   Rajpert-De Meyts E, Hanstein R, Jorgensen N, Graem N, Vogt PH, Skakkebaek NE: **Developmental expression of POU5F1 (OCT-3/4) in normal and dysgenetic human gonads.** *Hum Reprod* 2004, 19:1338-1344.

71.   Biermann K, Stoop H, Looijenga L: **c-KIT protein expression does not discriminate neoplastic from non-neoplastic intratubular germ cells.** *Histopathology* 2012, 60:1017-1019.

72.   Stoop H, Honecker F, van de Geijn GJ, Gillis AJ, Cools MC, de Boer M, Bokemeyer C, Wolffenbuttel KP, Drop SL, de Krijger RR, et al: **Stem cell factor as a novel diagnostic marker for early malignant germ cells.** *J Pathol* 2008, 216:43-54.

73.   Rajpert-De Meyts E: **Developmental model for the pathogenesis of testicular carcinoma in situ: genetic and environmental aspects.** *Human reproduction update* 2006, 12:303-323.

74.   Rajpert-de Meyts E, Hoei-Hansen CE: **From gonocytes to testicular cancer: the role of impaired gonadal development.** *Ann N Y Acad Sci* 2007, 1120**:**168-180.

75.   Hoei-Hansen CE, Rajpert-De Meyts E, Daugaard G, Skakkebaek NE: **Carcinoma in situ testis, the progenitor of testicular germ cell tumours: a clinical review.** *Ann Oncol* 2005, 16**:**863-868.

76.   Hoei-Hansen CE, Nielsen JE, Almstrup K, Sonne SB, Graem N, Skakkebaek NE, Leffers H, Rajpert-De Meyts E: **Transcription factor AP-2gamma is a developmentally regulated marker of testicular carcinoma in situ and germ cell tumors.** *Clin Cancer Res* 2004, 10:8521-8530.

77.   Schafer S, Anschlag J, Nettersheim D, Haas N, Pawig L, Schorle H: **The role of BLIMP1 and its putative downstream target TFAP2C in germ cell development and germ cell tumours.** *International journal of andrology* 2011, 34**:**e152-158; discussion e158-159.

78.   Mitchell RT, Camacho-Moll ME, Macdonald J, Anderson RA, Kelnar CJ, O'Donnell M, Sharpe RM, Smith LB, Grigor KM, Wallace WH, et al: **Intratubular germ cell neoplasia of the human testis: heterogeneous protein expression and relation to invasive potential.** *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2014.

79.   Almstrup K, Ottesen AM, Sonne SB, Hoei-Hansen CE, Leffers H, Rajpert-De Meyts E, Skakkebaek NE: **Genomic and gene expression signature of the pre-invasive testicular carcinoma in situ.** *Cell Tissue Res* 2005, 322**:**159-165.

80.   Yamada S, Kohu K, Ishii T, Ishidoya S, Hiramatsu M, Kanto S, Fukuzaki A, Adachi Y, Endoh M, Moriya T, et al: **Gene expression profiling identifies a set of transcripts that are up-regulated inhuman testicular seminoma.** *DNA research : an international journal for rapid publication of reports on genes and genomes* 2004, 11:335-344.

81.   Weikert S, Schrader M, Krause H, Schulze W, Muller M, Miller K: **The inhibitor of apoptosis (IAP) survivin is expressed in human testicular germ cell tumors and normal testes.** *Cancer Lett* 2005, 223:331-337.

82.   Juric D, Sale S, Hromas RA, Yu R, Wang Y, Duran GE, Tibshirani R, Einhorn LH, Sikic BI: **Gene expression profiling differentiates germ cell tumors from other cancers and defines subtype-specific signatures.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102:17763-17768.

83.   Chen YT, Chiu R, Lee P, Beneck D, Jin B, Old LJ: **Chromosome X-encoded cancer/testis antigens show distinctive expression patterns in developing gonads and in testicular seminoma.** *Hum Reprod* 2011, 26:3232-3243.

84.   Biermann K, Heukamp LC, Steger K, Zhou H, Franke FE, Guetgemann I, Sonnack V, Brehm R, Berg J, Bastian PJ, et al: **Gene expression profiling identifies new biological markers of neoplastic germ cells.** *Anticancer research* 2007, 27:3091-3100.

85.   Almstrup K, Hoei-Hansen CE, Nielsen JE, Wirkner U, Ansorge W, Skakkebaek NE, Rajpert-De Meyts E, Leffers H: **Genome-wide gene expression profiling of testicular carcinoma in situ progression into overt tumours.** *Br J Cancer* 2005, 92:1934-1941.

86.   Alagaratnam S, Lind GE, Kraggerud SM, Lothe RA, Skotheim RI: **The testicular germ cell tumour transcriptome.** *International journal of andrology* 2011, 34:e133-150; discussion e150-131.

87.   Alagaratnam S, Harrison N, Bakken AC, Hoff AM, Jones M, Sveen A, Moore HD, Andrews PW, Lothe RA, Skotheim RI: **Transforming pluripotency: an exon-level study of malignancy-specific transcripts in human embryonal carcinoma and embryonic stem cells.** *Stem cells and development* 2013, 22:1136-1146.

88.   Gashaw I, Dushaj O, Behr R, Biermann K, Brehm R, Rubben H, Grobholz R, Schmid KW, Bergmann M, Winterhager E: **Novel germ cell markers characterize testicular seminoma and fetal testis.** *Molecular human reproduction* 2007, 13:721-727.

89.   Port M, Schmelz HU, Stockinger M, Sparwasser C, Albers P, Pottek T, Abend M: **Gene expression profiling in seminoma and nonseminoma.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2005, 23:58-69.

90.   Bode PK, Barghorn A, Fritzsche FR, Riener MO, Kristiansen G, Knuth A, Moch H: **MAGEC2 is a sensitive and novel marker for seminoma: a tissue microarray analysis of 325 testicular germ cell tumors.** *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2011, 24:829-835.

91.   Biermann K, Heukamp LC, Steger K, Zhou H, Franke FE, Sonnack V, Brehm R, Berg J, Bastian PJ, Muller SC, et al: **Genome-wide expression profiling reveals new insights into pathogenesis and progression of testicular germ cell tumors.** *Cancer genomics & proteomics* 2007, 4:359-367.

92.   Okamoto K, Kawakami T: **Epigenetic profile of testicular germ cell tumours.** *International journal of andrology* 2007, 30:385-392; discussion 392.

93.   Okamoto K: **Epigenetics: a way to understand the origin and biology of testicular germ cell tumors.** *Int J Urol* 2012, 19:504-511.

94.   Van Der Zwan YG, Stoop H, Rossello F, White SJ, Looijenga LH: **Role of epigenetics in the etiology of germ cell cancer.** *The International journal of developmental biology* 2013, 57:299-308.

95.   van der Zwan YG, Rijlaarsdam MA, Rossello FJ, Notini AJ, de Boer S, Watkins DN, Gillis AJ, Dorssers LC, White SJ, Looijenga LH: **Seminoma and embryonal carcinoma footprints identified by analysis of integrated genome-wide epigenetic and expression profiles of germ cell cancer cell lines.** *PloS one* 2014, 9:e98330.

96.   Kristensen DG, Nielsen JE, Jorgensen A, Skakkebaek NE, Rajpert-De Meyts E, Almstrup K: **Evidence that active demethylation mechanisms maintain the genome of carcinoma in situ cells hypomethylated in the adult testis.** *Br J Cancer* 2014, 110:668-678.

97.   Smiraglia DJ, Szymanska J, Kraggerud SM, Lothe RA, Peltomaki P, Plass C: **Distinct epigenetic phenotypes in seminomatous and nonseminomatous testicular germ cell tumors.** *Oncogene* 2002, 21:3909-3916.

98.   Netto GJ, Nakai Y, Nakayama M, Jadallah S, Toubaji A, Nonomura N, Albadine R, Hicks JL, Epstein JI, Yegnasubramanian S, et al: **Global DNA hypomethylation in intratubular germ cell neoplasia and seminoma, but not in nonseminomatous male germ cell tumors.** *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2008, 21:1337-1344.

99.   Lind GE, Skotheim RI, Lothe RA: **The epigenome of testicular germ cell tumors.** *Apmis* 2007, 115:1147-1160.

100.   Di Vizio D, Cito L, Boccia A, Chieffi P, Insabato L, Pettinato G, Motti ML, Schepis F, D'Amico W, Fabiani F, et al: **Loss of the tumor suppressor gene PTEN marks the transition from intratubular germ cell neoplasias (ITGCN) to invasive germ cell tumors.** *Oncogene* 2005, 24:1882-1894.

101.   Gotovac K, Grubic Z, Kastelan Z, Stingl K, Kulis T, Krhen I, Hudolin T, Kastelan M, Zunec R: **Pilot study of the association between the HLA region and testicular carcinoma among Croatian patients.** *Urologia internationalis* 2011, 87:288-292.

102.   Goedert JJ, Purdue MP, McNeel TS, McGlynn KA, Engels EA: **Risk of germ cell tumors among men with HIV/acquired immunodeficiency syndrome.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2007, 16:1266-1269.

103.   Hvarness T, Nielsen JE, Almstrup K, Skakkebaek NE, Rajpert-De Meyts E, Claesson MH: **Phenotypic characterisation of immune cell infiltrates in testicular germ cell neoplasia.** *Journal of reproductive immunology* 2013, 100:135-145.

104. Hadrup SR, Braendstrup O, Jacobsen GK, Mortensen S, Pedersen LO, Seremet T, Andersen MH, Becker JC, Straten PT: **Tumor infiltrating lymphocytes in seminoma lesions comprise clonally expanded cytotoxic T cells.** *Int J Cancer* 2006, 119:831-838.

105. de Graaff WE, Oosterhuis JW, de Jong B, Dam A, van Putten WL, Castedo SM, Sleijfer DT, Schraffordt Koops H: **Ploidy of testicular carcinoma in situ.** *Laboratory investigation; a journal of technical methods and pathology* 1992, 66:166-168.

106. Ottesen AM, Larsen J, Gerdes T, Larsen JK, Lundsteen C, Skakkebaek NE, Rajpert-De Meyts E: **Cytogenetic investigation of testicular carcinoma in situ and early seminoma by high-resolution comparative genomic hybridization analysis of subpopulations flow sorted according to DNA content.** *Cancer genetics and cytogenetics* 2004, 149:89-97.

107. de Jong B, Oosterhuis JW, Castedo SM, Vos A, te Meerman GJ: **Pathogenesis of adult testicular germ cell tumors. A cytogenetic model.** *Cancer genetics and cytogenetics* 1990, 48:143-167.

108. Veltman I, Veltman J, Janssen I, Hulsbergen-van de Kaa C, Oosterhuis W, Schneider D, Stoop H, Gillis A, Zahn S, Looijenga L, et al: **Identification of recurrent chromosomal aberrations in germ cell tumors of neonates and infants using genomewide array-based comparative genomic hybridization.** *Genes Chromosomes Cancer* 2005, 43:367-376.

109. Rosenberg C, Schut TB, Mostert MC, Tanke HJ, Raap AK, Oosterhuis JW, Looijenga LH: **Comparative genomic hybridization in hypotriploid/hyperdiploid tumors.** *Cytometry* 1997, 29:113-121.

110. Korkola JE, Heck S, Olshen AB, Reuter VE, Bosl GJ, Houldsworth J, Chaganti RS: **In vivo differentiation and genomic evolution in adult male germ cell tumors.** *Genes Chromosomes Cancer* 2008, 47:43-55.

111. Frigyesi A, Gisselsson D, Hansen GB, Soller M, Mitelman F, Hoglund M: **A model for karyotypic evolution in testicular germ cell tumors.** *Genes Chromosomes Cancer* 2004, 40:172-178.

112. Skotheim RI, Autio R, Lind GE, Kraggerud SM, Andrews PW, Monni O, Kallioniemi O, Lothe RA: **Novel genomic aberrations in testicular germ cell tumors by array-CGH, and associated gene expression changes.** *Cellular oncology : the official journal of the International Society for Cellular Oncology* 2006, 28:315-326.

113. LeBron C, Pal P, Brait M, Dasgupta S, Guerrero-Preston R, Looijenga LH, Kowalski J, Netto G, Hoque MO: **Genome-wide analysis of genetic alterations in testicular primary seminoma using high resolution single nucleotide polymorphism arrays.** *Genomics* 2011, 97:341-349.

114. McIntyre A, Summersgill B, Lu YJ, Missiaglia E, Kitazawa S, Oosterhuis JW, Looijenga LH, Shipley J: **Genomic copy number and expression patterns in testicular germ cell tumours.** *Br J Cancer* 2007, 97:1707-1712.

115. Gilbert DC, McIntyre A, Summersgill B, Missiaglia E, Goddard NC, Chandler I, Huddart RA, Shipley J: **Minimum regions of genomic imbalance in stage I testicular embryonal carcinoma and association of 22q loss with relapse.** *Genes Chromosomes Cancer* 2011, 50:186-195.

116. Bergthorsson JT, Agnarsson BA, Gudbjartsson T, Magnusson K, Thoroddsen A, Palsson B, Bjornsson J, Stefansson K, Gulcher J, Einarsson GV, et al: **A genome-wide study of allelic imbalance in human testicular germ cell tumors using microsatellite markers.** *Cancer genetics and cytogenetics* 2006, 164:1-9.

117. van Echten J, Oosterhuis JW, Looijenga LH, van de Pol M, Wiersema J, te Meerman GJ, Schaffordt Koops H, Sleijfer DT, de Jong B: **No recurrent structural abnormalities apart from i(12p) in primary germ cell tumors of the adult testis.** *Genes Chromosomes Cancer* 1995, 14:133-144.

118. Mostert MC, Verkerk AJ, van de Pol M, Heighway J, Marynen P, Rosenberg C, van Kessel AG, van Echten J, de Jong B, Oosterhuis JW, Looijenga LH: **Identification of the critical region of 12p over-representation in testicular germ cell tumors of adolescents and adults.** *Oncogene* 1998, 16:2617-2627.

119. Looijenga LH, Zafarana G, Grygalewicz B, Summersgill B, Debiec-Rychter M, Veltman J, Schoenmakers EF, Rodriguez S, Jafer O, Clark J, et al: **Role of gain of 12p in germ cell tumour development.** *APMIS* 2003, 111:161-171; discussion 172-163.

120. Korkola JE, Houldsworth J, Chadalavada RS, Olshen AB, Dobrzynski D, Reuter VE, Bosl GJ, Chaganti RS: **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, 66:820-827.

121. de Jong B, van Echten J, Looijenga LH, Geurts van Kessel A, Oosterhuis JW: **Cytogenetics of the progression of adult testicular germ cell tumors.** *Cancer genetics and cytogenetics* 1997, 95:88-95.

122. Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, Plath K, Lowry WE, Benvenisty N: **Identification and classification of chromosomal aberrations in human induced pluripotent stem cells.** *Cell stem cell* 2010, 7:521-531.

123.   Werbowetski-Ogilvie TE, Bosse M, Stewart M, Schnerch A, Ramos-Mejia V, Rouleau A, Wynder T, Smith MJ, Dingwall S, Carter T, et al: **Characterization of human embryonic stem cells with features of neoplastic progression.** *Nat Biotechnol* 2009, 27:91-97.

124.   Draper JS, Smith K, Gokhale P, Moore HD, Maltby E, Johnson J, Meisner L, Zwaka TP, Thomson JA, Andrews PW: **Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells.** *Nat Biotechnol* 2004, 22:53-54.

125.   Miyai K, Yamamoto S, Iwaya K, Asano T, Tamai S, Tsuda H, Matsubara O: **Allelotyping analysis suggesting a consecutive progression from intratubular germ cell neoplasia to seminoma and then to embryonal carcinoma of the adult testis.** *Human pathology* 2013, 44:2312-2322.

126.   Hofer MD, Browne TJ, He L, Skotheim RI, Lothe RA, Rubin MA: **Identification of two molecular groups of seminomas by using expression and tissue microarrays.** *Clin Cancer Res* 2005, 11:5722-5729.

127.   Kratz CP, Mai PL, Greene MH: **Familial testicular germ cell tumours.** *Best Pract Res Clin Endocrinol Metab* 2010, 24:503-513.

128.   Dieckmann KP, Pichlmeier U: **Clinical epidemiology of testicular germ cell tumors.** *World journal of urology* 2004, 22:2-14.

129.   Czene K, Lichtenstein P, Hemminki K: **Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database.** *Int J Cancer* 2002, 99:260-266.

130.   Nathanson KL, Kanetsky PA, Hawes R, Vaughn DJ, Letrero R, Tucker K, Friedlander M, Phillips KA, Hogg D, Jewett MA, et al: **The Y deletion gr/gr and susceptibility to testicular germ cell tumor.** *American journal of human genetics* 2005, 77:1034-1043.

131.   Linger R, Dudakia D, Huddart R, Easton D, Bishop DT, Stratton MR, Rapley EA: **A physical analysis of the Y chromosome shows no additional deletions, other than Gr/Gr, associated with testicular germ cell tumour.** *Br J Cancer* 2007, 96:357-361.

132.   Ferlin A, Pengo M, Selice R, Salmaso L, Garolla A, Foresta C: **Analysis of single nucleotide polymorphisms of FSH receptor gene suggests association with testicular cancer susceptibility.** *Endocr Relat Cancer* 2008, 15:429-437.

133.   Vastermark A, Giwercman YL, Hagstromer O, Rajpert De-Meyts E, Eberhard J, Stahl O, Cedermark GC, Rastkhani H, Daugaard G, Arver S, Giwercman A: **Polymorphic variation in the androgen receptor gene: association with risk of testicular germ cell cancer and metastatic disease.** *Eur J Cancer* 2011, 47:413-419.

134.   Andreassen KE, Kristiansen W, Karlsson R, Aschim EL, Dahl O, Fossa SD, Adami HO, Wiklund F, Haugen TB, Grotmol T: **Genetic variation in AKT1, PTEN and the 8q24 locus, and the risk of testicular germ cell tumor.** *Hum Reprod* 2013, 28:1995-2002.

135.   Ferlin A, Ganz F, Pengo M, Selice R, Frigo AC, Foresta C: **Association of testicular germ cell tumor with polymorphisms in estrogen receptor and steroid metabolism genes.** *Endocr Relat Cancer* 2010, 17:17-25.

136.   Kristiansen W, Haugen TB, Witczak O, Andersen JM, Fossa SD, Aschim EL: **CYP1A1, CYP3A5 and CYP3A7 polymorphisms and testicular cancer susceptibility.** *International journal of andrology* 2011, 34:77-83.

137.   Horvath A, Korde L, Greene MH, Libe R, Osorio P, Faucz FR, Raffin-Sanson ML, Tsang KM, Drori-Herishanu L, Patronas Y, et al: **Functional phosphodiesterase 11A mutations may modify the risk of familial and bilateral testicular germ cell tumors.** *Cancer Res* 2009, 69:5301-5306.

138.   Chia VM, Sakoda LC, Graubard BI, Rubertone MV, Chanock SJ, Erickson RL, McGlynn KA: **Risk of testicular germ cell tumors and polymorphisms in the insulin-like growth factor genes.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2008, 17:721-726.

139.   Tsuchiya N, Mishina M, Narita S, Kumazawa T, Inoue T, Horikawa Y, Kakinuma H, Yuasa T, Matsuura S, Satoh S, et al: **Association of XRCC1 gene polymorphisms with the susceptibility and chromosomal aberration of testicular germ cell tumors.** *Int J Oncol* 2006, 28:1217-1223.

140.   Shiokawa M, Masutani M, Fujihara H, Ueki K, Nishikawa R, Sugimura T, Kubo H, Nakagama H: **Genetic alteration of poly(ADP-ribose) polymerase-1 in human germ cell tumors.** *Japanese journal of clinical oncology* 2005, 35:97-102.

141.   Starr JR, Chen C, Doody DR, Hsu L, Ricks S, Weiss NS, Schwartz SM: **Risk of testicular germ cell cancer in relation to variation in maternal and offspring cytochrome p450 genes involved in catechol estrogen metabolism.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2005, 14:2183-2190.

142.   Kristiansen W, Andreassen KE, Karlsson R, Aschim EL, Bremnes RM, Dahl O, Fossa SD, Klepp O, Langberg CW, Solberg A, et al: **Gene variations in sex hormone pathways and the risk of testicular germ cell tumour: a case-parent triad study in a Norwegian-Swedish population.** *Hum Reprod* 2012, 27:1525-1535.

143.   Brokken LJ, Lundberg-Giwercman Y, Rajpert De-Meyts E, Eberhard J, Stahl O, Cohn-Cedermark G, Daugaard G, Arver S, Giwercman A: **Association of polymorphisms in genes encoding hormone receptors ESR1, ESR2 and LHCGR with the risk and clinical features of testicular germ cell cancer.** *Mol Cell Endocrinol* 2012, 351:279-285.

144.   Purdue MP, Graubard BI, Chanock SJ, Rubertone MV, Erickson RL, McGlynn KA: **Genetic variation in the inhibin pathway and risk of testicular germ cell tumors.** *Cancer Res* 2008, 68:3043-3048.

145.   Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, Davies B, Peterse EF, Sacilotto N, Walker GJ, et al: **A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection.** *Cell* 2013, 155:410-422.

146.   Koster R, Mitra N, D'Andrea K, Vardhanabhuti S, Chung CC, Wang Z, Erickson RL, Vaughn DJ, Litchfield K, Rahman N, et al: **Pathway-based analysis of GWAs data identifies association of sex determination genes with susceptibility to testicular germ cell tumors.** *Hum Mol Genet* 2014.

147.   Cook MB, Graubard BI, Quraishi SM, Yeager M, Chanock SJ, Crenshaw A, Erickson RL, Rubertone MV, Thomas G, McGlynn KA: **Genetic variants in the 8q24 locus and risk of testicular germ cell tumors.** *Hum Genet* 2008, 123:409-418.

148.   Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, Ricketts M, Linger R, Nsengimana J, Deloukas P, et al: **Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer.** *Nat Genet* 2010, 42:604-607.

149.   Turnbull C, Rahman N: **Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour.** *International journal of andrology* 2011, 34:e86-96; discussion e96-87.

150.   Ruark E, Seal S, McDonald H, Zhang F, Elliot A, Lau K, Perdeaux E, Rapley E, Eeles R, Peto J, et al: **Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14.** *Nat Genet* 2013, 45:686-689.

151.   Rapley EA, Turnbull C, Al Olama AA, Dermitzakis ET, Linger R, Huddart RA, Renwick A, Hughes D, Hines S, Seal S, et al: **A genome-wide association study of testicular germ cell tumor.** *Nat Genet* 2009, 41:807-810.

152.   Rapley EA, Nathanson KL: **Predisposition alleles for Testicular Germ Cell Tumour.** *Current opinion in genetics & development* 2010, 20:225-230.

153.   Rapley E: **Susceptibility alleles for testicular germ cell tumour: a review.** *International journal of andrology* 2007, 30:242-250; discussion 250.

154.   Kanetsky PA, Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, Letrero R, Ciosek SL, Doody DR, Smith LM, Weaver J, et al: **Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer.** *Nat Genet* 2009, 41:811-815.

155.   Kanetsky PA, Mitra N, Vardhanabhuti S, Vaughn DJ, Li M, Ciosek SL, Letrero R, D'Andrea K, Vaddi M, Doody DR, et al: **A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility.** *Hum Mol Genet* 2011, 20:3109-3117.

156.   Schumacher FR, Wang Z, Skotheim RI, Koster R, Chung CC, Hildebrandt MA, Kratz CP, Bakken AC, Bishop DT, Cook MB, et al: **Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23.** *Hum Mol Genet* 2013, 22:2748-2753.

157.   Lessel D, Gamulin M, Kulis T, Toliat MR, Grgic M, Friedrich K, Zunec R, Balija M, Nurnberg P, Kastelan Z, et al: **Replication of genetic susceptibility loci for testicular germ cell cancer in the Croatian population.** *Carcinogenesis* 2012, 33:1548-1552.

158.   Karlsson R, Andreassen KE, Kristiansen W, Aschim EL, Bremnes RM, Dahl O, Fossa SD, Klepp O, Langberg CW, Solberg A, et al: **Investigation of six testicular germ cell tumor susceptibility genes suggests a parent-of-origin effect in SPRY4.** *Hum Mol Genet* 2013, 22:3373-3380.

159.   Ferlin A, Pengo M, Pizzol D, Carraro U, Frigo AC, Foresta C: **Variants in KITLG predispose to testicular germ cell cancer independently from spermatogenic function.** *Endocr Relat Cancer* 2012, 19:101-108.

160.   Kratz CP, Han SS, Rosenberg PS, Berndt SI, Burdett L, Yeager M, Korde LA, Mai PL, Pfeiffer R, Greene MH: **Variants in or near KITLG, BAK1, DMRT1, and TERT-CLPTM1L predispose to familial testicular germ cell tumour.** *Journal of medical genetics* 2011, 48:473-476.

161.   Kratz CP, Greene MH, Bratslavsky G, Shi J: **A stratified genetic risk assessment for testicular cancer.** *International journal of andrology* 2011, 34:e98-102.

162.   Chung CC, Kanetsky PA, Wang Z, Hildebrandt MA, Koster R, Skotheim RI, Kratz CP, Turnbull C, Cortessis VK, Bakken AC, et al: **Meta-analysis identifies four new loci associated with testicular germ cell tumor.** *Nat Genet* 2013.

163.   Swerdlow AJ, De Stavola BL, Swanwick MA, Maconochie NE: **Risks of breast and testicular cancers in young adult twins in England and Wales: evidence on prenatal and genetic aetiology.** *Lancet* 1997, 350:1723-1728.

164.   Dong C, Hemminki K: **Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families.** *Int J Cancer* 2001, 92:144-150.

165.   Hemminki K, Li X: **Familial risk in testicular cancer as a clue to a heritable and environmental aetiology.** *Br J Cancer* 2004, 90:1765-1770.

166.   Banks K, Tuazon E, Berhane K, Koh CJ, De Filippo RE, Chang A, Kim SS, Daneshmand S, Davis-Dao C, Lewinger JP, et al: **Cryptorchidism and testicular germ cell tumors: comprehensive meta-analysis reveals that association between these conditions diminished over time and is modified by clinical characteristics.** *Frontiers in endocrinology* 2012, 3:182.

167.   Schnack TH, Poulsen G, Myrup C, Wohlfahrt J, Melbye M: **Familial coaggregation of cryptorchidism, hypospadias, and testicular germ cell cancer: a nationwide cohort study.** *J Natl Cancer Inst* 2010, 102:187-192.

168.   Trabert B, Zugna D, Richiardi L, McGlynn KA, Akre O: **Congenital malformations and testicular germ cell tumors.** *Int J Cancer* 2013.

169.   Petersen PM, Giwercman A, Hansen SW, Berthelsen JG, Daugaard G, Rorth M, Skakkebaek NE: **Impaired testicular function in patients with carcinoma-in-situ of the testis.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1999, 17:173-179.

170.   Willemse PH, Sleijfer DT, Sluiter WJ, Schraffordt Koops H, Doorenbos H: **Altered Leydig cell function in patients with testicular cancer: evidence for bilateral testicular defect.** *Acta endocrinologica* 1983, 102:616-624.

171.   Elzinga-Tinke JE, Sirre ME, Looijenga LH, van Casteren N, Wildhagen MF, Dohle GR: **The predictive value of testicular ultrasound abnormalities for carcinoma in situ of the testis in men at risk for testicular cancer.** *International journal of andrology* 2010, 33:597-603.

172.   Walsh TJ, Croughan MS, Schembri M, Chan JM, Turek PJ: **Increased risk of testicular germ cell cancer among infertile men.** *Arch Intern Med* 2009, 169:351-356.

173.   Jacobsen R, Bostofte E, Engholm G, Hansen J, Olsen JH, Skakkebaek NE, Moller H: **Risk of testicular cancer in men with abnormal semen characteristics: cohort study.** *BMJ* 2000, 321:789-792.

174.   Mancini M, Carmignani L, Gazzano G, Sagone P, Gadda F, Bosari S, Rocco F, Colpi GM: **High prevalence of testicular cancer in azoospermic men without spermatogenesis.** *Hum Reprod* 2007, 22:1042-1046.

175.   McLachlan RI, Rajpert-De Meyts E, Hoei-Hansen CE, de Kretser DM, Skakkebaek NE: **Histological evaluation of the human testis--approaches to optimizing the clinical value of the assessment: mini review.** *Hum Reprod* 2007, 22:2-16.

176.   van Casteren NJ, Looijenga LH, Dohle GR: **Testicular microlithiasis and carcinoma in situ overview and proposed clinical guideline.** *International journal of andrology* 2009, 32:279-287.

177.   Tan IB, Ang KK, Ching BC, Mohan C, Toh CK, Tan MH: **Testicular microlithiasis predicts concurrent testicular germ cell tumors and intratubular germ cell neoplasia of unclassified type in adults: a meta-analysis and systematic review.** *Cancer* 2010, 116:4520-4532.

178.   Otite U, Webb JA, Oliver RT, Badenoch DF, Nargund VH: **Testicular microlithiasis: is it a benign condition with malignant potential?** *Eur Urol* 2001, 40:538-542.

179.   Bach AM, Hann LE, Hadar O, Shi W, Yoo HH, Giess CS, Sheinfeld J, Thaler H: **Testicular microlithiasis: what is its association with testicular cancer?** *Radiology* 2001, 220:70-75.

180.   Levine H, Afek A, Shamiss A, Derazne E, Tzur D, Zavdy O, Barchana M, Kark JD: **Risk of germ cell testicular cancer according to origin: a migrant cohort study in 1,100,000 Israeli men.** *Int J Cancer* 2013, 132:1878-1885.

181.   McGlynn KA, Cook MB: **Etiologic factors in testicular germ-cell tumors.** *Future Oncol* 2009, 5:1389-1402.

182.   Trabert B, Sigurdson AJ, Sweeney AM, Strom SS, McGlynn KA: **Marijuana use and testicular germ cell tumors.** *Cancer* 2011, 117:848-853.

183.   Daling JR, Doody DR, Sun X, Trabert BL, Weiss NS, Chen C, Biggs ML, Starr JR, Dey SK, Schwartz SM: **Association of marijuana use and the incidence of testicular germ cell tumors.** *Cancer* 2009, 115:1215-1223.

184. Sharpe RM, Skakkebaek NE: **Are oestrogens involved in falling sperm counts and disorders of the male reproductive tract?** *Lancet* 1993, 341:1392-1395.

185. Niculescu AM: **Effects of in utero exposure to DES on male progeny.** *Journal of obstetric, gynecologic, and neonatal nursing : JOGNN / NAACOG* 1985, 14:468-470.

186. Schiffer C, Muller A, Egeberg DL, Alvarez L, Brenker C, Rehfeld A, Frederiksen H, Waschle B, Kaupp UB, Balbach M, et al: **Direct action of endocrine disrupting chemicals on human sperm.** *EMBO Rep* 2014.

187. Richiardi L, Pettersson A, Akre O: **Genetic and environmental risk factors for testicular cancer.** *International journal of andrology* 2007, 30:230-240; discussion 240-231.

188. Ajj H, Chesnel A, Pinel S, Plenat F, Flament S, Dumond H: **An Alkylphenol Mix Promotes Seminoma Derived Cell Proliferation through an ERalpha36-Mediated Mechanism.** *PloS one* 2013, 8:e61758.

189. Kortenkamp A, Scholze M, Ermler S: **Mind the gap: can we explain declining male reproductive health with known antiandrogens?** *Reproduction* 2014, 147:515-527.

190. McGlynn KA, Graubard BI, Nam JM, Stanczyk FZ, Longnecker MP, Klebanoff MA: **Maternal hormone levels and risk of cryptorchism among populations at high and low risk of testicular germ cell tumors.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2005, 14:1732-1737.

191. Holl K, Lundin E, Surcel HM, Grankvist K, Koskela P, Dillner J, Hallmans G, Wadell G, Olafsdottir GH, Ogmundsdottir HM, et al: **Endogenous steroid hormone levels in early pregnancy and risk of testicular cancer in the offspring: a nested case-referent study.** *Int J Cancer* 2009, 124:2923-2928.

192. Trabert B, Sigurdson AJ, Sweeney AM, Amato RJ, Strom SS, McGlynn KA: **Baldness, acne and testicular germ cell tumours.** *International journal of andrology* 2011, 34:e59-67.

193. Sasaki A, Taketomi T, Kato R, Saeki K, Nonami A, Sasaki M, Kuriyama M, Saito N, Shibuya M, Yoshimura A: **Mammalian Sprouty4 suppresses Ras-independent ERK activation by binding to Raf1.** *Nat Cell Biol* 2003, 5:427-432.

194. Yan W, Samson M, Jegou B, Toppari J: **Bcl-w forms complexes with Bax and Bak, and elevated ratios of Bax/Bcl-w and Bak/Bcl-w correspond to spermatogonial and spermatocyte apoptosis in the testis.** *Mol Endocrinol* 2000, 14:682-699.

195. Frolov A, Chahwan S, Ochs M, Arnoletti JP, Pan ZZ, Favorova O, Fletcher J, von Mehren M, Eisenberg B, Godwin AK: **Response markers and the molecular mechanisms of action of Gleevec in gastrointestinal stromal tumors.** *Molecular cancer therapeutics* 2003, 2:699-709.

196. Schrader M, Burger AM, Muller M, Krause H, Straub B, Schostak M, Schulze W, Lauke H, Miller K: **The differentiation status of primary gonadal germ cell tumors correlates inversely with telomerase activity and the expression level of the gene encoding the catalytic subunit of telomerase.** *BMC cancer* 2002, 2:32.

197. Liu L, Ishihara K, Ichimura T, Fujita N, Hino S, Tomita S, Watanabe S, Saitoh N, Ito T, Nakao M: **MCAF1/AM is involved in Sp1-mediated maintenance of cancer-associated telomerase activity.** *J Biol Chem* 2009, 284:5165-5174.

198. Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y: **A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis.** *Biochemical and biophysical research communications* 2001, 280:1148-1154.

199. Barbaro M, Balsamo A, Anderlid BM, Myhre AG, Gennari M, Nicoletti A, Pittalis MC, Oscarson M, Wedell A: **Characterization of deletions at 9p affecting the candidate regions for sex reversal and deletion 9p syndrome by MLPA.** *European journal of human genetics : EJHG* 2009, 17:1439-1447.

200. Krentz AD, Murphy MW, Kim S, Cook MS, Capel B, Zhu R, Matin A, Sarver AL, Parker KL, Griswold MD, et al: **The DM domain protein DMRT1 is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106:22323-22328.

201. Voss M, Paterson J, Kelsall IR, Martin-Granados C, Hastie CJ, Peggie MW, Cohen PT: **Ppm1E is an in cellulo AMP-activated protein kinase phosphatase.** *Cellular signalling* 2011, 23:114-124.

202. Gui L, Homer H: **Spindle assembly checkpoint signalling is uncoupled from chromosomal position in mouse oocytes.** *Development* 2012, 139:1941-1946.

203. Ballister ER, Riegman M, Lampson MA: **Recruitment of Mad1 to metaphase kinetochores is sufficient to reactivate the mitotic checkpoint.** *J Cell Biol* 2014, 204:901-908.

204. Mondal G, Ohashi A, Yang L, Rowley M, Couch FJ: **Tex14, a Plk1-regulated protein, is required for kinetochore-microtubule attachment and regulation of the spindle assembly checkpoint.** *Mol Cell* 2012, 45:680-695.

205. Fu X, Yucer N, Liu S, Li M, Yi P, Mu JJ, Yang T, Chu J, Jung SY, O'Malley BW, et al: **RFWD3-Mdm2 ubiquitin ligase complex positively regulates p53 stability in response to DNA damage.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107:4579-4584.

206. Van Rompay AR, Norda A, Linden K, Johansson M, Karlsson A: **Phosphorylation of uridine and cytidine nucleoside analogs by two human uridine-cytidine kinases.** *Molecular pharmacology* 2001, 59:1181-1186.

207. Moniot B, Farhat A, Aritake K, Declosmenil F, Nef S, Eguchi N, Urade Y, Poulat F, Boizet-Bonhoure B: **Hematopoietic prostaglandin D synthase (H-Pgds) is expressed in the early embryonic gonad and participates to the initial nuclear translocation of the SOX9 protein.** *Developmental dynamics : an official publication of the American Association of Anatomists* 2011, 240:2335-2343.

208. Sonne SB, Kristensen DM, Novotny GW, Olesen IA, Nielsen JE, Skakkebaek NE, Rajpert-De Meyts E, Leffers H: **Testicular dysgenesis syndrome and the origin of carcinoma in situ testis.** *International journal of andrology* 2008, 31:275-287.

209. Mirbahai L, Chipman JK: **Epigenetic memory of environmental organisms: a reflection of lifetime stressor exposures.** *Mutation research Genetic toxicology and environmental mutagenesis* 2014, 764-765:10-17.

210. Looijenga LH, Van Agthoven T, Biermann K: **Development of malignant germ cells - the genvironmental hypothesis.** *The International journal of developmental biology* 2013, 57:241-253.

211. Hotaling JM, Walsh TJ: **Male infertility: a risk factor for testicular cancer.** *Nat Rev Urol* 2009, 6:550-556.

212. Skinner MK, Guerrero-Bosagna C, Haque M, Nilsson E, Bhandari R, McCarrey JR: **Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and the subsequent germ line.** *PloS one* 2013, 8:e66318.

213. Amatruda JF, Ross JA, Christensen B, Fustino NJ, Chen KS, Hooten AJ, Nelson H, Kuriger JK, Rakheja D, Frazier AL, Poynter JN: **DNA methylation analysis reveals distinct methylation signatures in pediatric germ cell tumors.** *BMC cancer* 2013, 13:313.

214. Skinner MK: **Epigenetic transgenerational toxicology and germ cell disease.** *International journal of andrology* 2007, 30:393-396; discussion 396-397.

215. Dalgaard MD, Weinhold N, Edsgard D, Silver JD, Pers TH, Nielsen JE, Jorgensen N, Juul A, Gerds TA, Giwercman A, et al: **A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation.** *Journal of medical genetics* 2012, 49:58-65.

216. Hersmus R, de Leeuw BH, Wolffenbuttel KP, Drop SL, Oosterhuis JW, Cools M, Looijenga LH: **New insights into type II germ cell tumor pathogenesis based on studies of patients with various forms of disorders of sex development (DSD).** *Mol Cell Endocrinol* 2008, 291:1-10.

217. Cools M, Looijenga LH: **Tumor risk and clinical follow-up in patients with disorders of sex development.** *Pediatric endocrinology reviews : PER* 2011, 9 Suppl 1:519-524.

218. Cools M, Drop SL, Wolffenbuttel KP, Oosterhuis JW, Looijenga LH: **Germ cell tumors in the intersex gonad: old paths, new directions, moving frontiers.** *Endocr Rev* 2006, 27:468-484.

219. Richiardi L, Akre O: **Fertility among brothers of patients with testicular cancer.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2005, 14:2557-2562.

220. Richiardi L, Akre O, Montgomery SM, Lambe M, Kvist U, Ekbom A: **Fecundity and twinning rates as measures of fertility before diagnosis of germ-cell testicular cancer.** *J Natl Cancer Inst* 2004, 96:145-147.

221. Olesen IA, Hoei-Hansen CE, Skakkebaek NE, Petersen JH, Rajpert-De Meyts E, Jorgensen N: **Testicular carcinoma in situ in subfertile Danish men.** *International journal of andrology* 2007, 30:406-411; discussion 412.

222. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, 144:646-674.

223. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, 100:57-70.

224. Boublikova L, Buchler T, Stary J, Abrahamova J, Trka J: **Molecular biology of testicular germ cell tumors: unique features awaiting clinical application.** *Critical reviews in oncology/hematology* 2014, 89:366-385.

225. Looijenga LH, Gillis AJ, Stoop H, Biermann K, Oosterhuis JW: **Dissecting the molecular pathways of (testicular) germ cell tumour pathogenesis; from initiation to treatment-resistance.** *International journal of andrology* 2011, 34:e234-251.

226. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.** *Cell* 2006, 126:663-676.

227.   Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA: **Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2.** *Nat Biotechnol* 2008, 26:1269-1275.

228.   Grad I, Hibaoui Y, Jaconi M, Chicha L, Bergstrom-Tengzelius R, Sailani MR, Pelte MF, Dahoun S, Mitsiadis TA, Tohonen V, et al: **NANOG priming before full reprogramming may generate germ cell tumours.** *European cells & materials* 2011, 22:258-274; discussio 274.

229.   Cunningham JJ, Ulbright TM, Pera MF, Looijenga LH: **Lessons from human teratomas to guide development of safe stem cell therapies.** *Nat Biotechnol* 2012, 30:849-857.

230.   Ramakrishna S, Kim KS, Baek KH: **Posttranslational modifications of defined embryonic reprogramming transcription factors.** *Cellular reprogramming* 2014, 16:108-120.

231.   Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, 122:947-956.

232.   Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, et al: **The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.** *Nat Genet* 2006, 38:431-440.

233.   Brumbaugh J, Hou Z, Russell JD, Howden SE, Yu P, Ledvina AR, Coon JJ, Thomson JA: **Phosphorylation regulates human OCT4.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109:7162-7168.

234.   Brumbaugh J, Russell JD, Yu P, Westphall MS, Coon JJ, Thomson JA: **NANOG Is Multiply Phosphorylated and Directly Modified by ERK2 and CDK1 In Vitro.** *Stem cell reports* 2014, 2:18-25.

235.   de Jong J, Looijenga LH: **Stem cell marker OCT3/4 in tumor biology and germ cell tumor diagnostics: history and future.** *Critical reviews in oncogenesis* 2006, 12:171-203.

236.   Cheng L, Sung MT, Cossu-Rocca P, Jones TD, MacLennan GT, De Jong J, Lopez-Beltran A, Montironi R, Looijenga LH: **OCT4: biological functions and clinical applications as a marker of germ cell neoplasia.** *J Pathol* 2007, 211:1-9.

237.   Rijlaarsdam MA, van Herk HA, Gillis AJ, Stoop H, Jenster G, Martens J, van Leenders GJ, Dinjens W, Hoogland AM, Timmermans M, Looijenga LH: **Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours.** *Br J Cancer* 2011, 105:854-863.

238.   Cheng CJ, Wu YC, Shu JA, Ling TY, Kuo HC, Wu JY, Chang EE, Chang SC, Huang YH: **Aberrant expression and distribution of the OCT-4 transcription factor in seminomas.** *Journal of biomedical science* 2007, 14:797-807.

239.   Jones TD, Ulbright TM, Eble JN, Cheng L: **OCT4: A sensitive and specific biomarker for intratubular germ cell neoplasia of the testis.** *Clin Cancer Res* 2004, 10:8544-8547.

240.   Jones TD, MacLennan GT, Bonnin JM, Varsegi MF, Blair JE, Cheng L: **Screening for intratubular germ cell neoplasia of the testis using OCT4 immunohistochemistry.** *The American journal of surgical pathology* 2006, 30:1427-1431.

241.   de Jong J, Stoop H, Dohle GR, Bangma CH, Kliffen M, van Esser JW, van den Bent M, Kros JM, Oosterhuis JW, Looijenga LH: **Diagnostic value of OCT3/4 for pre-invasive and invasive testicular germ cell tumours.** *J Pathol* 2005, 206:242-249.

242.   Pesce M, Scholer HR: **Oct-4: control of totipotency and germline determination.** *Mol Reprod Dev* 2000, 55:452-457.

243.   Wang X, Dai J: **Concise review: isoforms of OCT4 contribute to the confusing diversity in stem cell biology.** *Stem Cells* 2010, 28:885-893.

244.   Pesce M, Scholer HR: **Oct-4: gatekeeper in the beginnings of mammalian development.** *Stem Cells* 2001, 19:271-278.

245.   Niwa H, Miyazaki J, Smith AG: **Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.** *Nat Genet* 2000, 24:372-376.

246.   Kehler J, Tolkunova E, Koschorz B, Pesce M, Gentile L, Boiani M, Lomeli H, Nagy A, McLaughlin KJ, Scholer HR, Tomilin A: **Oct4 is required for primordial germ cell survival.** *EMBO Rep* 2004, 5:1078-1083.

247.   Le Bin GC, Munoz-Descalzo S, Kurowski A, Leitch H, Lou X, Mansfield W, Etienne-Dumeau C, Grabole N, Mulas C, Niwa H, et al: **Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst.** *Development* 2014, 141:1001-1010.

248. Botquin V, Hess H, Fuhrmann G, Anastassiadis C, Gross MK, Vriend G, Scholer HR: New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev* 1998, 12:2073-2090.

249. Hay DC, Sutherland L, Clark J, Burdon T: Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* 2004, 22:225-235.

250. Matin MM, Walsh JR, Gokhale PJ, Draper JS, Bahrami AR, Morton I, Moore HD, Andrews PW: Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. *Stem Cells* 2004, 22:659-668.

251. Velkey JM, O'Shea KS: Oct4 RNA interference induces trophectoderm differentiation in mouse embryonic stem cells. *Genesis* 2003, 37:18-24.

252. Zaehres H, Scholer HR: Induction of pluripotency: from mouse to human. *Cell* 2007, 131:834-835.

253. Boiani M, Kehler J, Scholer HR: Activity of the germline-specific Oct4-GFP transgene in normal and clone mouse embryos. *Methods Mol Biol* 2004, 254:1-34.

254. Sung B, Do HJ, Park SW, Huh SH, Oh JH, Chung HJ, Kang MJ, Kim JH, Kim NH, Kim JH: Regulation of OCT4 gene expression by liver receptor homolog-1 in human embryonic carcinoma cells. *Biochemical and biophysical research communications* 2012, 427:315-320.

255. Krishnan BR, Jamry I, Chaplin DD: Feature mapping of the HLA class I region: localization of the POU5F1 and TCF19 genes. *Genomics* 1995, 30:53-58.

256. Takeda J, Seino S, Bell GI: Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. *Nucleic Acids Res* 1992, 20:4613-4620.

257. Sturm RA, Herr W: The POU domain is a bipartite DNA-binding structure. *Nature* 1988, 336:601-604.

258. Liedtke S, Stephan M, Kogler G: Oct4 expression revisited: potential pitfalls for data misinterpretation in stem cell research. *Biological chemistry* 2008, 389:845-850.

259. Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu YM, Cao X, Asangani IA, Kothari V, Prensner JR, Lonigro RJ, et al: Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* 2012, 149:1622-1634.

260. Groen JN, Capraro D, Morris KV: The emerging role of pseudogene expressed non-coding RNAs in cellular functions. *Int J Biochem Cell Biol* 2014.

261. Jez M, Ambady S, Kashpur O, Grella A, Malcuit C, Vilner L, Rozman P, Dominko T: Expression and differentiation between OCT4A and its Pseudogenes in human ESCs and differentiated adult somatic cells. *PloS one* 2014, 9:e89546.

262. Asadi MH, Mowla SJ, Fathi F, Aleyasin A, Asadzadeh J, Atlasi Y: OCT4B1, a novel spliced variant of OCT4, is highly expressed in gastric cancer and acts as an antiapoptotic factor. *Int J Cancer* 2011.

263. Gao Y, Wang X, Han J, Xiao Z, Chen B, Su G, Dai J: The novel OCT4 spliced variant OCT4B1 can generate three protein isoforms by alternative splicing into OCT4B. *Journal of genetics and genomics = Yi chuan xue bao* 2010, 37:461-465.

264. Lee J, Kim HK, Rho JY, Han YM, Kim J: The human OCT-4 isoforms differ in their ability to confer self-renewal. *J Biol Chem* 2006, 281:33554-33565.

265. Atlasi Y, Mowla SJ, Ziaee SA, Gokhale PJ, Andrews PW: OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells* 2008, 26:3068-3074.

266. Papamichos SI, Kotoula V, Tarlatzis BC, Agorastos T, Papazisis K, Lambropoulos AF: OCT4B1 isoform: the novel OCT4 alternative spliced variant as a putative marker of stemness. *Molecular human reproduction* 2009, 15:269-270.

267. Wang L, Guo ZY, Zhang R, Xin B, Chen R, Zhao J, Wang T, Wen WH, Jia LT, Yao LB, Yang AG: Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma. *Carcinogenesis* 2013, 34:1773-1781.

268. Hu J, Guo H, Li H, Liu Y, Liu J, Chen L, Zhang J, Zhang N: MiR-145 regulates epithelial to mesenchymal transition of breast cancer cells by targeting Oct4. *PloS one* 2012, 7:e45965.

269. Ng WL, Chen G, Wang M, Wang H, Story M, Shay JW, Zhang X, Wang J, Amin AR, Hu B, et al: OCT4 as a target of miR-34a stimulates p63 but inhibits p53 to promote human cell transformation. *Cell Death Dis* 2014, 5:e1024.

270.   Lin Y, Yang Y, Li W, Chen Q, Li J, Pan X, Zhou L, Liu C, Chen C, He J, et al: **Reciprocal regulation of Akt and Oct4 promotes the self-renewal and survival of embryonal carcinoma cells.** *Mol Cell* 2012, 48:627-640.

271.   Hoei-Hansen CE, Almstrup K, Nielsen JE, Brask Sonne S, Graem N, Skakkebaek NE, Leffers H, Rajpert-De Meyts E: **Stem cell pluripotency factor NANOG is expressed in human fetal gonocytes, testicular carcinoma in situ and germ cell tumours.** *Histopathology* 2005, 47:48-56.

272.   Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S: **The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells.** *Cell* 2003, 113:631-642.

273.   Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A: **Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.** *Cell* 2003, 113:643-655.

274.   Korkola JE, Houldsworth J, Bosl GJ, Chaganti RS: **Molecular events in germ cell tumours: linking chromosome-12 gain, acquisition of pluripotency and response to cisplatin.** *BJU Int* 2009, 104:1334-1338.

275.   Pan G, Thomson JA: **Nanog and transcriptional networks in embryonic stem cell pluripotency.** *Cell Res* 2007, 17:42-49.

276.   Yamaguchi S, Kurimoto K, Yabuta Y, Sasaki H, Nakatsuji N, Saitou M, Tada T: **Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells.** *Development* 2009, 136:4011-4020.

277.   de Jong J, Stoop H, Gillis AJ, van Gurp RJ, van de Geijn GJ, Boer M, Hersmus R, Saunders PT, Anderson RA, Oosterhuis JW, Looijenga LH: **Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.** *J Pathol* 2008, 215:21-30.

278.   Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, et al: **Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.** *Nat Cell Biol* 2007, 9:625-635.

279.   Eini R, Dorssers LC, Looijenga LH: **Role of stem cell proteins and microRNAs in embryogenesis and germ cell cancer.** *The International journal of developmental biology* 2013, 57:319-332.

280.   Perrett RM, Turnpenny L, Eckert JJ, O'Shea M, Sonne SB, Cameron IT, Wilson DI, Rajpert-De Meyts E, Hanley NA: **The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture.** *Biol Reprod* 2008, 78:852-858.

281.   Sonne SB, Perrett RM, Nielsen JE, Baxter MA, Kristensen DM, Leffers H, Hanley NA, Rajpert-De-Meyts E: **Analysis of SOX2 expression in developing human testis and germ cell neoplasia.** *The International journal of developmental biology* 2010, 54:755-760.

282.   Kanai-Azuma M, Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PP, Hayashi Y: **Depletion of definitive gut endoderm in Sox17-null mutant mice.** *Development* 2002, 129:2367-2379.

283.   Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, Bogu GK, Teo R, Leng Ng CK, Herath W, Lili S, et al: **Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm.** *The EMBO journal* 2013, 32:938-953.

284.   Wang Y, Baskerville S, Shenoy A, Babiarz JE, Baehner L, Blelloch R: **Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation.** *Nat Genet* 2008, 40:1478-1483.

285.   Wang Y, Keys DN, Au-Young JK, Chen C: **MicroRNAs in embryonic stem cells.** *J Cell Physiol* 2009, 218:251-255.

286.   Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R: **DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal.** *Nat Genet* 2007, 39:380-385.

287.   Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K: **Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing.** *Genes Dev* 2005, 19:489-501.

288.   Hong Y, Stambrook PJ: **Restoration of an absent G1 arrest and protection from apoptosis in embryonic stem cells after ionizing radiation.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101:14443-14448.

289.   Lin SL, Chang DC, Chang-Lin S, Lin CH, Wu DT, Chen DT, Ying SY: **Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state.** *RNA* 2008, 14:2115-2124.

290.   Judson RL, Babiarz JE, Venere M, Blelloch R: **Embryonic stem cell-specific microRNAs promote induced pluripotency.** *Nat Biotechnol* 2009, 27:459-461.

291.   Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, et al: **Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.** *Cell* 2008, 134**:**521-533.

292.   Murray MJ, Saini HK, Siegler CA, Hanning JE, Barker EM, van Dongen S, Ward DM, Raby KL, Groves IJ, Scarpini CG, et al: **LIN28 Expression in malignant germ cell tumors downregulates let-7 and increases oncogene levels.** *Cancer Res* 2013, 73**:**4872-4884.

293.   Cai WY, Wei TZ, Luo QC, Wu QW, Liu QF, Yang M, Ye GD, Wu JF, Chen YY, Sun GB, et al: **The Wnt-beta-catenin pathway represses let-7 microRNA expression through transactivation of Lin28 to augment breast cancer stem cell expansion.** *Journal of cell science* 2013, 126**:**2877-2889.

294.   Stadler B, Ivanovska I, Mehta K, Song S, Nelson A, Tan Y, Mathieu J, Darby C, Blau CA, Ware C, et al: **Characterization of microRNAs involved in embryonic stem cell states.** *Stem cells and development* 2010, 19**:**935-950.

295.   Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, Cha KY, Chung HM, Yoon HS, Moon SY, et al: **Human embryonic stem cells express a unique set of microRNAs.** *Dev Biol* 2004, 270**:**488-498.

296.   Muller PA, Vousden KH: **Mutant p53 in cancer: new functions and therapeutic opportunities.** *Cancer Cell* 2014, 25**:**304-317.

297.   Muller PA, Vousden KH: **p53 mutations in cancer.** *Nat Cell Biol* 2013, 15**:**2-8.

298.   Solozobova V, Blattner C: **p53 in stem cells.** *World journal of biological chemistry* 2011, 2**:**202-214.

299.   Zhao Y, Yin X, Qin H, Zhu F, Liu H, Yang W, Zhang Q, Xiang C, Hou P, Song Z, et al: **Two supporting factors greatly improve the efficiency of human iPSC generation.** *Cell stem cell* 2008, 3**:**475-479.

300.   Kersemaekers AM, Mayer F, Molier M, van Weeren PC, Oosterhuis JW, Bokemeyer C, Looijenga LH: **Role of P53 and MDM2 in treatment response of human germ cell tumors.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2002, 20**:**1551-1561.

301.   Lutzker SG: **P53 tumour suppressor gene and germ cell neoplasia.** *APMIS* 1998, 106**:**85-89.

302.   Guillou L, Estreicher A, Chaubert P, Hurlimann J, Kurt AM, Metthez G, Iggo R, Gray AC, Jichlinski P, Leisinger HJ, Benhattar J: **Germ cell tumors of the testis overexpress wild-type p53.** *The American journal of pathology* 1996, 149**:**1221-1228.

303.   Schenkman NS, Sesterhenn IA, Washington L, Tong YA, Weghorst CM, Buzard GS, Srivastava S, Moul JW: **Increased p53 protein does not correlate to p53 gene mutations in microdissected human testicular germ cell tumors.** *J Urol* 1995, 154**:**617-621.

304.   Wu YC, Ling TY, Lu SH, Kuo HC, Ho HN, Yeh SD, Shen CN, Huang YH: **Chemotherapeutic sensitivity of testicular germ cell tumors under hypoxic conditions is negatively regulated by SENP1-controlled sumoylation of OCT4.** *Cancer Res* 2012, 72**:**4963-4973.

305.   Mueller T, Mueller LP, Luetzkendorf J, Voigt W, Simon H, Schmoll HJ: **Loss of Oct-3/4 expression in embryonal carcinoma cells is associated with induction of cisplatin resistance.** *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 2006, 27**:**71-83.

306.   Jorgensen A, Blomberg Jensen M, Nielsen JE, Juul A, Rajpert-De Meyts E: **Influence of vitamin D on cisplatin sensitivity in testicular germ cell cancer-derived cell lines and in a NTera2 xenograft model.** *The Journal of steroid biochemistry and molecular biology* 2013, 136**:**238-246.

307.   Bremmer F, Thelen P, Pottek T, Behnes CL, Radzun HJ, Schweyer S: **Expression and function of the vitamin D receptor in malignant germ cell tumour of the testis.** *Anticancer research* 2012, 32**:**341-349.

308.   Feldman DR, Iyer G, Van Alstine L, Patil S, Al-Ahmadie HA, Reuter VE, Bosl GJ, Chaganti RS, Solit DB: **Presence of Somatic Mutations within PIK3CA, AKT, RAS, and FGFR3 but not BRAF in Cisplatin-Resistant Germ Cell Tumors.** *Clin Cancer Res* 2014.

309.   Bignell G, Smith R, Hunter C, Stephens P, Davies H, Greenman C, Teague J, Butler A, Edkins S, Stevens C, 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**:**42-46.

310.   Chevalier N, Barlier A, Roche C, Mograbi B, Camparo P, Devouassoux-Shisheboran M, Michiels JF, Nebout M, Chevallier D, Benahmed M, et al: **RET gene mutations are not involved in the origin of human testicular seminoma.** *International journal of andrology* 2010, 33**:**848-852.

311.   Abdelalim EM: **Molecular mechanisms controlling the cell cycle in embryonic stem cells.** *Stem cell reviews* 2013, 9**:**764-773.

312.   Ballabeni A, Park IH, Zhao R, Wang W, Lerou PH, Daley GQ, Kirschner MW: **Cell cycle adaptations of embryonic stem cells.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108**:**19252-19257.

313. Koster R, di Pietro A, Timmer-Bosscha H, Gibcus JH, van den Berg A, Suurmeijer AJ, Bischoff R, Gietema JA, de Jong S: **Cytoplasmic p21 expression levels determine cisplatin resistance in human testicular cancer.** *The Journal of clinical investigation* 2010, 120:3594-3605.

314. Spierings DC, de Vries EG, Stel AJ, te Rietstap N, Vellenga E, de Jong S: **Low p21Waf1/Cip1 protein level sensitizes testicular germ cell tumor cells to Fas-mediated apoptosis.** *Oncogene* 2004, 23:4862-4872.

315. Gutekunst M, Oren M, Weilbacher A, Dengler MA, Markwardt C, Thomale J, Aulitzky WE, van der Kuip H: **p53 hypersensitivity is the predominant mechanism of the unique responsiveness of testicular germ cell tumor (TGCT) cells to cisplatin.** *PloS one* 2011, 6:e19198.

316. Gutekunst M, Mueller T, Weilbacher A, Dengler MA, Bedke J, Kruck S, Oren M, Aulitzky WE, van der Kuip H: **Cisplatin hypersensitivity of testicular germ cell tumors is determined by high constitutive Noxa levels mediated by Oct-4.** *Cancer Res* 2013, 73:1460-1469.

317. Kerley-Hamilton JS, Pike AM, Li N, DiRenzo J, Spinella MJ: **A p53-dominant transcriptional response to cisplatin in testicular germ cell tumor-derived human embryonal carcinoma.** *Oncogene* 2005, 24:6090-6100.

318. Mandoky L, Szende B, Geczi L, Bodrogi I, Kasler M, Bak M: **Apoptosis regulation and spontaneous apoptosis index of testicular germ cell tumors are associated with differentiation and resistance to systemic treatment.** *Anticancer research* 2008, 28:1641-1649.

319. Cavallo F, Feldman DR, Barchi M: **Revisiting DNA damage repair, p53-mediated apoptosis and cisplatin sensitivity in germ cell tumors.** *The International journal of developmental biology* 2013, 57:273-280.

320. Looijenga LH, Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW: **Chromosomes and expression in human testicular germ-cell tumors: insight into their cell of origin and pathogenesis.** *Ann N Y Acad Sci* 2007, 1120:187-214.

321. Qi J, Yu JY, Shcherbata HR, Mathieu J, Wang AJ, Seal S, Zhou W, Stadler BM, Bourgin D, Wang L, et al: **microRNAs regulate human embryonic stem cell division.** *Cell cycle (Georgetown, Tex* 2009, 8:3729-3741.

322. Aylon Y, Michael D, Shmueli A, Yabuta N, Nojima H, Oren M: **A positive feedback loop between the p53 and Lats2 tumor suppressors prevents tetraploidization.** *Genes Dev* 2006, 20:2687-2700.

323. Dolezalova D, Mraz M, Barta T, Plevova K, Vinarsky V, Holubcova Z, Jaros J, Dvorak P, Pospisilova S, Hampl A: **MicroRNAs regulate p21(Waf1/Cip1) protein expression and the DNA damage response in human embryonic stem cells.** *Stem Cells* 2012, 30:1362-1372.

324. Card DA, Hebbar PB, Li L, Trotter KW, Komatsu Y, Mishina Y, Archer TK: **Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells.** *Mol Cell Biol* 2008, 28:6426-6438.

325. Lin SL, Chang DC, Ying SY, Leu D, Wu DT: **MicroRNA miR-302 inhibits the tumorigenecity of human pluripotent stem cells by coordinate suppression of the CDK2 and CDK4/6 cell cycle pathways.** *Cancer Res* 2010, 70:9473-9482.

326. Cai N, Wang YD, Zheng PS: **The microRNA-302-367 cluster suppresses the proliferation of cervical carcinoma cells through the novel target AKT1.** *RNA* 2013, 19:85-95.

327. Liu L, Lian J, Zhang H, Tian H, Liang M, Yin M, Sun F: **MicroRNA-302a sensitizes testicular embryonal carcinoma cells to cisplatin-induced cell death.** *J Cell Physiol* 2013, 228:2294-2304.

328. Choi YJ, Lin CP, Ho JJ, He X, Okada N, Bu P, Zhong Y, Kim SY, Bennett MJ, Chen C, et al: **miR-34 miRNAs provide a barrier for somatic cell reprogramming.** *Nat Cell Biol* 2011, 13:1353-1360.

329. Cha YH, Kim NH, Park C, Lee I, Kim HS, Yook JI: **MiRNA-34 intrinsically links p53 tumor suppressor and Wnt signaling.** *Cell cycle (Georgetown, Tex* 2012, 11:1273-1281.

330. Li M, Yu M, Liu C, Zhu H, He X, Peng S, Hua J: **miR-34c works downstream of p53 leading to dairy goat male germline stem-cell (mGSCs) apoptosis.** *Cell proliferation* 2013, 46:223-231.

331. Bouhallier F, Allioli N, Lavial F, Chalmel F, Perrard MH, Durand P, Samarut J, Pain B, Rouault JP: **Role of miR-34c microRNA in the late steps of spermatogenesis.** *RNA* 2010, 16:720-731.

332. Suzuki HI, Miyazono K: **p53 actions on microRNA expression and maturation pathway.** *Methods Mol Biol* 2013, 962:165-181.

333. Hunten S, Siemens H, Kaller M, Hermeking H: **The p53/microRNA network in cancer: experimental and bioinformatics approaches.** *Advances in experimental medicine and biology* 2013, 774:77-101.

334. Looijenga LH, Gillis AJ, Stoop H, Hersmus R, Oosterhuis JW: **Relevance of microRNAs in normal and malignant development, including human testicular germ cell tumours.** *International journal of andrology* 2007, 30:304-314; discussion 314-305.

335.   Sabbaghian N, Bahubeshi A, Shuen AY, Kanetsky PA, Tischkowitz MD, Nathanson KL, Foulkes WD: **Germ-line DICER1 mutations do not make a major contribution to the etiology of familial testicular germ cell tumours.** *BMC Res Notes* 2013, 6:127.

336.   de Boer CM, Eini R, Gillis AM, Stoop H, Looijenga LH, White SJ: **DICER1 RNase IIIb domain mutations are infrequent in testicular germ cell tumours.** *BMC Res Notes* 2012, 5:569.

337.   Teshima S, Shimosato Y, Hirohashi S, Tome Y, Hayashi I, Kanazawa H, Kakizoe T: **Four new human germ cell tumor cell lines.** *Laboratory investigation; a journal of technical methods and pathology* 1988, 59:328-336.

338.   Andrews PW, Casper J, Damjanov I, Duggan-Keen M, Giwercman A, Hata J, von Keitz A, Looijenga LH, Millan JL, Oosterhuis JW, et al: **Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours.** *Int J Cancer* 1996, 66:806-816.

339.   Lennartsson J, Ronnstrand L: **Stem cell factor receptor/c-Kit: from basic science to clinical implications.** *Physiological reviews* 2012, 92:1619-1649.

340.   Yuzawa S, Opatowsky Y, Zhang Z, Mandiyan V, Lax I, Schlessinger J: **Structural basis for activation of the receptor tyrosine kinase KIT by stem cell factor.** *Cell* 2007, 130:323-334.

341.   Yabuta Y, Kurimoto K, Ohinata Y, Seki Y, Saitou M: **Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling.** *Biol Reprod* 2006, 75:705-716.

342.   Mithraprabhu S, Loveland KL: **Control of KIT signalling in male germ cells: what can we learn from other systems?** *Reproduction* 2009, 138:743-757.

343.   Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M, Clark AT: **The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation.** *Nat Cell Biol* 2013, 15:113-122.

344.   Mamsen LS, Brochner CB, Byskov AG, Mollgard K: **The migration and loss of human primordial germ stem cells from the hind gut epithelium towards the gonadal ridge.** *The International journal of developmental biology* 2012, 56:771-778.

345.   Farini D, La Sala G, Tedesco M, De Felici M: **Chemoattractant action and molecular signaling pathways of Kit ligand on mouse primordial germ cells.** *Dev Biol* 2007, 306:572-583.

346.   Mahakali Zama A, Hudson FP, 3rd, Bedell MA: **Analysis of hypomorphic KitlSl mutants suggests different requirements for KITL in proliferation and migration of mouse primordial germ cells.** *Biol Reprod* 2005, 73:639-647.

347.   Blume-Jensen P, Janknecht R, Hunter T: **The kit receptor promotes cell survival via activation of PI 3-kinase and subsequent Akt-mediated phosphorylation of Bad on Ser136.** *Curr Biol* 1998, 8:779-782.

348.   Bernet A, Mehlen P: **Dependence receptors: when apoptosis controls tumor progression.** *Bull Cancer* 2007, 94:E12-17.

349.   Goldschneider D, Mehlen P: **Dependence receptors: a new paradigm in cell signaling and cancer therapy.** *Oncogene* 2010, 29:1865-1882.

350.   La Sala G, Farini D, De Felici M: **Rapid estrogen signalling in mouse primordial germ cells.** *Experimental cell research* 2010, 316:1716-1727.

351.   La Sala G, Farini D, De Felici M: **Estrogenic in vitro assay on mouse embryonic Leydig cells.** *The International journal of developmental biology* 2010, 54:717-722.

352.   Correia S, Alves MR, Cavaco JE, Oliveira PF, Socorro S: **Estrogenic regulation of testicular expression of stem cell factor and c-kit: implications in germ cell survival and male fertility.** *Fertility and sterility* 2014.

353.   Guido C, Panza S, Santoro M, Avena P, Panno ML, Perrotta I, Giordano F, Casaburi I, Catalano S, De Amicis F, et al: **Estrogen receptor beta (ERbeta) produces autophagy and necroptosis in human seminoma cell line through the binding of the Sp1 on the phosphatase and tensin homolog deleted from chromosome 10 (PTEN) promoter gene.** *Cell cycle (Georgetown, Tex* 2012, 11:2911-2921.

354.   Devouassoux-Shisheboran M, Mauduit C, Tabone E, Droz JP, Benahmed M: **Growth regulatory factors and signalling proteins in testicular germ cell tumours.** *APMIS* 2003, 111:212-224; discussion 224.

355.   Oatley JM, Brinster RL: **The germline stem cell niche unit in mammalian testes.** *Physiological reviews* 2012, 92:577-595.

356.   Zhang L, Tang J, Haines CJ, Feng H, Lai L, Teng X, Han Y: **c-kit expression profile and regulatory factors during spermatogonial stem cell differentiation.** *BMC developmental biology* 2013, 13:38.

357.   Rothschild G, Sottas CM, Kissel H, Agosti V, Manova K, Hardy MP, Besmer P: **A role for kit receptor signaling in Leydig cell steroidogenesis.** *Biol Reprod* 2003, 69:925-932.

358.   Rossi P, Dolci S, Sette C, Geremia R: **Molecular mechanisms utilized by alternative c-kit gene products in the control of spermatogonial proliferation and sperm-mediated egg activation.** *Andrologia* 2003, 35:71-78.

359.   Rossi P, Albanesi C, Grimaldi P, Geremia R: **Expression of the mRNA for the ligand of c-kit in mouse Sertoli cells.** *Biochemical and biophysical research communications* 1991, 176:910-914.

360.   Taylor WE, Najmabadi H, Strathearn M, Jou NT, Liebling M, Rajavashisth T, Chanani N, Phung L, Bhasin S: **Human stem cell factor promoter deoxyribonucleic acid sequence and regulation by cyclic 3',5'-adenosine monophosphate in a Sertoli cell line.** *Endocrinology* 1996, 137:5407-5414.

361.   Rossi P, Dolci S, Albanesi C, Grimaldi P, Ricca R, Geremia R: **Follicle-stimulating hormone induction of steel factor (SLF) mRNA in mouse Sertoli cells and stimulation of DNA synthesis in spermatogonia by soluble SLF.** *Dev Biol* 1993, 155:68-74.

362.   Bhattacharya I, Pradhan BS, Sarda K, Gautam M, Basu S, Majumdar SS: **A switch in Sertoli cell responsiveness to FSH may be responsible for robust onset of germ cell differentiation during prepubartal testicular maturation in rats.** *American journal of physiology Endocrinology and metabolism* 2012, 303:E886-898.

363.   Majumdar SS, Sarda K, Bhattacharya I, Plant TM: **Insufficient androgen and FSH signaling may be responsible for the azoospermia of the infantile primate testes despite exposure to an adult-like hormonal milieu.** *Hum Reprod* 2012, 27:2515-2525.

364.   Sandlow JI, Feng HL, Cohen MB, Sandra A: **Expression of c-KIT and its ligand, stem cell factor, in normal and subfertile human testicular tissue.** *J Androl* 1996, 17:403-408.

365.   Malcher A, Rozwadowska N, Stokowy T, Jedrzejczak P, Zietkowiak W, Kurpisz M: **The gene expression analysis of paracrine/autocrine factors in patients with spermatogenetic failure compared with normal spermatogenesis.** *American journal of reproductive immunology* 2013, 70:522-528.

366.   Zhang L, Tang J, Haines CJ, Feng HL, Lai L, Teng X, Han Y: **c-kit and its related genes in spermatogonial differentiation.** *Spermatogenesis* 2011, 1:186-194.

367.   Ma M, Yang S, Zhang Z, Li P, Gong Y, Liu L, Zhu Y, Tian R, Liu Y, Wang X, et al: **Sertoli cells from non-obstructive azoospermia and obstructive azoospermia patients show distinct morphology, Raman spectrum and biochemical phenotype.** *Hum Reprod* 2013.

368.   Feng LX, Ravindranath N, Dym M: **Stem cell factor/c-kit up-regulates cyclin D3 and promotes cell cycle progression via the phosphoinositide 3-kinase/p70 S6 kinase pathway in spermatogonia.** *J Biol Chem* 2000, 275:25572-25576.

369.   Kissel H, Timokhina I, Hardy MP, Rothschild G, Tajima Y, Soares V, Angeles M, Whitlow SR, Manova K, Besmer P: **Point mutation in kit receptor tyrosine kinase reveals essential roles for kit signaling in spermatogenesis and oogenesis without affecting other kit responses.** *EMBO J* 2000, 19:1312-1326.

370.   Blume-Jensen P, Jiang G, Hyman R, Lee KF, O'Gorman S, Hunter T: **Kit/stem cell factor receptor-induced activation of phosphatidylinositol 3'-kinase is essential for male fertility.** *Nat Genet* 2000, 24:157-162.

371.   Barrios F, Filipponi D, Campolo F, Gori M, Bramucci F, Pellegrini M, Ottolenghi S, Rossi P, Jannini EA, Dolci S: **SOHLH1 and SOHLH2 control Kit expression during postnatal male germ cell development.** *Journal of cell science* 2012, 125:1455-1464.

372.   Rossi P: **Transcriptional control of KIT gene expression during germ cell development.** *The International journal of developmental biology* 2013, 57:179-184.

373.   Yang QE, Racicot KE, Kaucher AV, Oatley MJ, Oatley JM: **MicroRNAs 221 and 222 regulate the undifferentiated state in mammalian male germ cells.** *Development* 2012.

374.   Rossi P, Marziali G, Albanesi C, Charlesworth A, Geremia R, Sorrentino V: **A novel c-kit transcript, potentially encoding a truncated receptor, originates within a kit gene intron in mouse spermatids.** *Dev Biol* 1992, 152:203-207.

375.   Sette C, Dolci S, Geremia R, Rossi P: **The role of stem cell factor and of alternative c-kit gene products in the establishment, maintenance and function of germ cells.** *The International journal of developmental biology* 2000, 44:599-608.

376.   Sette C, Bevilacqua A, Geremia R, Rossi P: **Involvement of phospholipase Cgamma1 in mouse egg activation induced by a truncated form of the C-kit tyrosine kinase present in spermatozoa.** *J Cell Biol* 1998, 142:1063-1074.

377.   Sette C, Bevilacqua A, Bianchini A, Mangia F, Geremia R, Rossi P: **Parthenogenetic activation of mouse eggs by microinjection of a truncated c-kit tyrosine kinase present in spermatozoa.** *Development* 1997, 124:2267-2274.

378. Sette C, Paronetto MP, Barchi M, Bevilacqua A, Geremia R, Rossi P: **Tr-kit-induced resumption of the cell cycle in mouse eggs requires activation of a Src-like kinase.** *EMBO J* 2002, 21:5386-5395.

379. Muciaccia B, Sette C, Paronetto MP, Barchi M, Pensini S, D'Agostino A, Gandini L, Geremia R, Stefanini M, Rossi P: **Expression of a truncated form of KIT tyrosine kinase in human spermatozoa correlates with sperm DNA integrity.** *Hum Reprod* 2010, 25:2188-2202.

380. Berends JC, Schutte SE, van Dissel-Emiliani FM, de Rooij DG, Looijenga LH, Oosterhuis JW: **Significant improvement of the survival of seminoma cells in vitro by use of a rat Sertoli cell feeder layer and serum-free medium.** *J Natl Cancer Inst* 1991, 83:1400-1403.

381. Heaney JD, Lam MY, Michelson MV, Nadeau JH: **Loss of the transmembrane but not the soluble kit ligand isoform increases testicular germ cell tumor susceptibility in mice.** *Cancer Res* 2008, 68:5193-5197.

382. Rorth M, Rajpert-De Meyts E, Andersson L, Dieckmann KP, Fossa SD, Grigor KM, Hendry WF, Herr HW, Looijenga LH, Oosterhuis JW, Skakkebaek NE: **Carcinoma in situ in the testis.** *Scand J Urol Nephrol Suppl* 2000:166-186.

383. Goddard NC, McIntyre A, Summersgill B, Gilbert D, Kitazawa S, Shipley J: **KIT and RAS signalling pathways in testicular germ cell tumours: new data and a review of the literature.** *International journal of andrology* 2007, 30:337-348; discussion 349.

384. Izquierdo MA, Van der Valk P, Van Ark-Otte J, Rubio G, Germa-Lluch JR, Ueda R, Scheper RJ, Takahashi T, Giaccone G: **Differential expression of the c-kit proto-oncogene in germ cell tumours.** *J Pathol* 1995, 177:253-258.

385. Leroy X, Augusto D, Leteurtre E, Gosselin B: **CD30 and CD117 (c-kit) used in combination are useful for distinguishing embryonal carcinoma from seminoma.** *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2002, 50:283-285.

386. Strohmeyer T, Reese D, Press M, Ackermann R, Hartmann M, Slamon D: **Expression of the c-kit proto-oncogene and its ligand stem cell factor (SCF) in normal and malignant human testicular tissue.** *J Urol* 1995, 153:511-515.

387. Nikolaou M, Valavanis C, Aravantinos G, Fountzilas G, Tamvakis N, Lekka I, Arapantoni-Dadioti P, Zizi A, Ghiconti I, Economopoulos T, Pectasides D: **Kit expression in male germ cell tumors.** *Anticancer research* 2007, 27:1685-1688.

388. de Jong J, Stoop H, Gillis AJ, Hersmus R, van Gurp RJ, van de Geijn GJ, van Drunen E, Beverloo HB, Schneider DT, Sherlock JK, et al: **Further characterization of the first seminoma cell line TCam-2.** *Genes Chromosomes Cancer* 2008, 47:185-196.

389. Mizuno Y, Gotoh A, Kamidono S, Kitazawa S: **[Establishment and characterization of a new human testicular germ cell tumor cell line (TCam-2)].** *Nihon Hinyokika Gakkai zasshi The japanese journal of urology* 1993, 84:1211-1218.

390. Young JC, Jaiprakash A, Mithraprabhu S, Itman C, Kitazawa R, Looijenga LH, Loveland KL: **TCam-2 seminoma cell line exhibits characteristic foetal germ cell responses to TGF-beta ligands and retinoic acid.** *International journal of andrology* 2011, 34:e204-217.

391. Nettersheim D, Westernstroer B, Haas N, Leinhaas A, Brustle O, Schlatt S, Schorle H: **Establishment of a versatile seminoma model indicates cellular plasticity of germ cell tumor cells.** *Genes Chromosomes Cancer* 2012, 51:717-726.

392. McIntyre A, Summersgill B, Grygalewicz B, Gillis AJ, Stoop J, van Gurp RJ, Dennis N, Fisher C, Huddart R, Cooper C, et al: **Amplification and overexpression of the KIT gene is associated with progression in the seminoma subtype of testicular germ cell tumors of adolescents and adults.** *Cancer Res* 2005, 65:8085-8089.

393. Mol CD, Dougan DR, Schneider TR, Skene RJ, Kraus ML, Scheibe DN, Snell GP, Zou H, Sang BC, Wilson KP: **Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase.** *J Biol Chem* 2004, 279:31655-31663.

394. Agarwal S, Kazi JU, Ronnstrand L: **Phosphorylation of the activation loop tyrosine 823 in c-Kit is crucial for cell survival and proliferation.** *J Biol Chem* 2013, 288:22460-22468.

395. Fukushima S, Otsuka A, Suzuki T, Yanagisawa T, Mishima K, Mukasa A, Saito N, Kumabe T, Kanamori M, Tominaga T, et al: **Mutually exclusive mutations of KIT and RAS are associated with KIT mRNA expression and chromosomal instability in primary intracranial pure germinomas.** *Acta neuropathologica* 2014.

396. Willmore-Payne C, Holden JA, Chadwick BE, Layfield LJ: **Detection of c-kit exons 11- and 17-activating mutations in testicular seminomas by high-resolution melting amplicon analysis.** *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2006, 19:1164-1169.

397.   Kemmer K, Corless CL, Fletcher JA, McGreevey L, Haley A, Griffith D, Cummings OW, Wait C, Town A, Heinrich MC: **KIT mutations are common in testicular seminomas.** *The American journal of pathology* 2004, 164:305-313.

398.   Hersmus R, Stoop H, van de Geijn GJ, Eini R, Biermann K, Oosterhuis JW, Dhooge C, Schneider DT, Meijssen IC, Dinjens WN, et al: **Prevalence of c-KIT mutations in gonadoblastoma and dysgerminomas of patients with disorders of sex development (DSD) and ovarian dysgerminomas.** *PloS one* 2012, 7:e43952.

399.   Chen L, Faire M, Kissner MD, Laird DJ: **Primordial germ cells and gastrointestinal stromal tumors respond distinctly to a cKit overactivating allele.** *Hum Mol Genet* 2013, 22:313-327.

400.   Einhorn LH, Brames MJ, Heinrich MC, Corless CL, Madani A: **Phase II study of imatinib mesylate in chemotherapy refractory germ cell tumors expressing KIT.** *American journal of clinical oncology* 2006, 29:12-13.

401.   Frost MJ, Ferrao PT, Hughes TP, Ashman LK: **Juxtamembrane mutant V560GKit is more sensitive to Imatinib (STI571) compared with wild-type c-kit whereas the kinase domain mutant D816VKit is resistant.** *Molecular cancer therapeutics* 2002, 1:1115-1124.

402.   Looijenga LH, de Leeuw H, van Oorschot M, van Gurp RJ, Stoop H, Gillis AJ, de Gouveia Brazao CA, Weber RF, Kirkels WJ, van Dijk T, et al: **Stem cell factor receptor (c-KIT) codon 816 mutations predict development of bilateral testicular germ-cell tumors.** *Cancer Res* 2003, 63:7674-7678.

403.   Biermann K, Goke F, Nettersheim D, Eckert D, Zhou H, Kahl P, Gashaw I, Schorle H, Buttner R: **c-KIT is frequently mutated in bilateral germ cell tumours and down-regulated during progression from intratubular germ cell neoplasia to seminoma.** *J Pathol* 2007, 213:311-318.

404.   Sakuma Y, Matsukuma S, Yoshihara M, Sakurai S, Nishii M, Kishida T, Kubota Y, Nagashima Y, Inayama Y, Sasaki T, et al: **Mutations of c-kit gene in bilateral testicular germ cell tumours in Japan.** *Cancer Lett* 2008, 259:119-126.

405.   Coffey J, Linger R, Pugh J, Dudakia D, Sokal M, Easton DF, Timothy Bishop D, Stratton M, Huddart R, Rapley EA: **Somatic KIT mutations occur predominantly in seminoma germ cell tumors and are not predictive of bilateral disease: report of 220 tumors and review of literature.** *Genes Chromosomes Cancer* 2008, 47:34-42.

406.   Rapley EA, Hockley S, Warren W, Johnson L, Huddart R, Crockford G, Forman D, Leahy MG, Oliver DT, Tucker K, et al: **Somatic mutations of KIT in familial testicular germ cell tumours.** *Br J Cancer* 2004, 90:2397-2401.

407.   Mirabello L, Kratz CP, Savage SA, Greene MH: **Promoter methylation of candidate genes associated with familial testicular cancer.** *International journal of molecular epidemiology and genetics* 2012, 3:213-227.

408.   Azevedo MF, Horvath A, Bornstein ER, Almeida MQ, Xekouki P, Faucz FR, Gourgari E, Nadella K, Remmers EF, Quezado M, et al: **Cyclic AMP and c-KIT signaling in familial testicular germ cell tumor predisposition.** *J Clin Endocrinol Metab* 2013.

409.   Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods (San Diego, Calif* 2001, 25:402-408.

Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours

Martin A. Rijlaarsdam[1], Hilde A. van Herk[1], Ad J. Gillis[1], Hans Stoop[1],
Guido Jenster[2], John Martens[3], G.J. (Arno) van Leenders[1], Winand Dinjens[1],
A. Marije Hoogland[1], Mieke Timmermans[3], Leendert H.J. Looijenga[1]

Department of Pathology [1], Department of Urology [2], Department of Internal Medicine [3]
Erasmus MC - University Medical Center, Daniel den Hoed Cancer Center, Rotterdam, The Netherlands

Abstract

Background: OCT3/4 (POU5F1) is an established diagnostic immunohistochemical marker for specific histological variants of human malignant germ cell tumours (GCTs), including the seminomatous types and the stem cell component of nonseminomas, known as embryonal carcinoma. OCT3/4 is crucial for the regulation of pluripotency and the self-renewal of normal embryonic stem- and germ cells. Detection of expression of this transcription factor is complicated by the existence of multiple pseudogenes and isoforms. Various claims have been made about *OCT3/4* expression in non-germ cell tumours, possibly related to using non-specific detection methods. False positive findings undermine the applicability of OCT3/4 as a specific diagnostic tool in a clinical setting. In addition, false positive findings could result in misinterpretation of pluripotency regulation in solid somatic cancers and their stem cells. Of the three identified isoforms - *OCT4A*, *OCT4B* and *OCT4B1* - only OCT4A proved to regulate pluripotency. Until now, no convincing nuclear OCT4A protein expression has been shown in somatic cancers or tissues.

Methods: This study investigates expression of the various OCT3/4 isoforms in GCTs (both differentiated and undifferentiated) and somatic (non-germ cell) cancers, including representative cell lines and xenografts.

Results: Using specific methods, *OCT4A* and *OCT4B1* are shown to be preferentially expressed in undifferentiated GCTs. The *OCT4B* variant shows no difference in expression between GCTs (either differentiated or undifferentiated) and somatic cancers. In spite of the presence of *OCT4A* mRNA in somatic cancer-derived cell lines, no OCT3/4 protein is detected. Significant positive correlations between all isoforms of OCT3/4 were identified in both tumours with and without a known stem cell component, possibly indicating synergistic roles of these isoforms.

Conclusion: This study confirms that OCT4A GCTs, embryonal carcinoma and representative cell lines. Furthermore, it emphasizes that in order to correctly assess the presence of functional OCT3/4, both isoform specific mRNA and protein detection are required.

Introduction

OCT3/4 (also known as POU5F1) is a well-known marker for pluripotent stem cells, both physiologically and artificially induced (iPSC) [1, 2]. In addition, it is also expressed in primordial germ cells (PGCs) [1], the stem cell of gametogenesis later in life. OCT3/4 is a transcription factor involved in self-renewal and pluripotency [3, 4], and might counteract apoptosis in PGCs [5]. During further development (differentiation/maturation) of these types of embryonic cells, expression is down-regulated and finally lost in the differentiated derivatives. Owing to this specific pattern of expression during embryogenesis, which is retained during the process of malignant transformation, OCT3/4 is an established and highly informative diagnostic marker for defined types of malignant germ cell tumours (GCTs), especially those of the seminomatous cell type (seminoma, dysgerminoma and germinoma) and embryonal carcinoma [6-9].

The *OCT3/4* gene is located on human chromosome 6 band p21 and consists of 5 exons [10, 11]. It encodes a protein belonging to the family of octamer binding proteins that specifically binds to the conserved ATTTTGCAT motive in transcriptional control elements of genes. This sequence is recognized by the highly charged POU domain of the OCT3/4 protein, explaining its alternative name. The POU domain consists of two subdomains: a C-terminal homeodomain and an N-terminal POU-specific region separated by a short non-conserved linker [12].

Various investigations of OCT3/4 expression are reported, both on mRNA and protein level. Most are complicated by the existence of pseudogenes and splice variants (isoforms) [2, 7, 13], possibly leading to findings of false positive expression. Till date, five OCT3/4 pseudogenes have been identified. These will be amplified by most of the published primer sets, due to their high sequence similarity to OCT4A [14, 15]. However, proper DNase pre-treatment of the samples will exclude this technical artefact, which is a rather simple and straightforward step to include in the experimental setup.

In addition, three splice variants (isoforms) of *OCT3/4* have been identified [2]. The best known isoform is referred to as OCT4A. This variant is reported to be stem cell specific, while the function(s) of the other two variants, that is, OCT4B and OCT4B1, are still under investigation [2, 16, 17]. However, it has been demonstrated that OCT4B is unable to activate or repress transcription of known OCT4A responsive genes [18]. Therefore, OCT4B seems unlikely to be directly involved in transcriptional regulation of pluripotency and self renewal. OCT4B1 on the other hand, has recently been suggested to have a role in both regulation of pluripotency [16, 19, 20] and OCT4B mediated functions [17].

As indicated, OCT3/4 has mainly been linked to pluripotency, for which it is a well known and established marker. Pluripotency refers to the capacity of an (embryonic) stem cell to generate all different tissues (endo- ecto- and mesodermal differentiation). Indeed, differentiation induction is associated with down-regulation of OCT3/4 [21-25]. The other way around, OCT3/4 down-regulation results in loss of stem cells and induction of differentiation [4]. A high throughput immunohistochemical screen of many different types of human cancers demonstrated that OCT3/4 is a specific and highly informative diagnostic marker for seminomatous tumours, which are the malignant counterparts of PGCs/gonocytes, as well as embryonal carcinomas, the stem cell component of nonseminomas [6, 26]. This observation is confirmed by multiple independent studies, as reviewed before [9]. The overall findings resulted in the conclusion that OCT3/4 is an excellent, and currently successfully used, diagnostic marker for the detection of undifferentiated variants of so-called type II GCTs (UNDIF-GCTs = seminoma (SE) or dysgerminoma/germinoma, and the stem cell component of non-seminoma (NS, specifically embryonal carcinoma)), as well as their precursor stages (carcinoma *in situ* of the testis (CIS) and gonadoblastoma of dysgenetic gonads) [6-8, 26, 27]. Most recently, OCT3/4 protein detection has been used as a diagnostic tool for the non-invasive diagnosis of CIS [28].

In non-GCTs (N-GCTs), a highly heterogeneous expression pattern of OCT3/4, both mRNA and protein, is reported. This might be due to the use of non-specific primers detecting other isoforms, improper DNAse pre-treatment (resulting in amplification of pseudogenes) and incorrect interpretation of immunohistochemical stainings. The results of the various articles [29-35] to whether or not there is expression of (functional) OCT3/4 in solid cancers are therefore inconclusive. To further investigate this issue, the current study was undertaken. It investigates expression of the different isoforms of *OCT3/4* mRNA and protein in various types of solid cancers: undifferentiated GCTs (UNDIF-GCTs) and GCTs without an embryonic stem cell component (DIF-GCTs = yolk sac tumours and teratomas) as well as N-GCTs. Also, representative cell lines (-CL) of UNDIF-GCT and N-GCT are included.

Materials and Methods

Materials

The cell lines and tumour samples included in this study are indicated in Table 1. All samples/cell lines were obtained from different Departments in the Josephine Nefkens Institute (Erasmus MC - University Medical Center, the Netherlands). The prostate carcinoma cell lines and xenografts are extensively reviewed elsewhere [36, 37]. This also goes for the GCT-CLs [38-40].

Table 1. Samples and cell lines included in the study.

Group	Subgroup	Cell lines / Number of tumor samples
Undifferentiated GCT cell lines (UNDIF-GCT-CL)	Seminoma	TCam-2
	Embryonal carcinoma	TERA1, NCCIT, NT2
	Other	JKT-1
Non GCT cell lines (N-GCT-CL)	Oesophaguscarcinoma	ESO26, ESO51, I2425
	Cervixcarcinoma	HeLa
	Lungcarcinoma	H460
	Coloncarcinoma	H716, HCT116, SW620
	Prostatecarcinoma	Cell lines: LNCaP, 22Rv1, VCaP, LAPC-4, MDA PCa 2b Xenografts: PC324, PC329, PC339, PC346B, PC346C, PC374, PC133, PC82, PC135, PC295, PC310
	Breastcarcinoma	MDA175, ZR75
Undifferentiated GCTs (UNDIF-GCT)	Seminoma (SE)	5
	Embryonal Carcinoma (EC)	4
Differentiated GCTs (DIF-GCT)	Teratoma (TE)	5
	Yolk sac tumor	1
Non-GCT (N-GCT)	Bladdercarcinoma	5
	Lungcarcinoma	4
	Ovariuncarcinoma	5
	Prostatecarcinoma	5
	Rectalcarcinoma	4

RNA isolation

High-quality total RNA was extracted from the above mentioned cell lines and tumour samples using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. Samples were pre-treated with DNase-I, checked for residual DNA contamination by PCR, after which cDNA synthesis was performed as described before [38, 41]. For each sample, a no-reverse transcription (No-RT) control was used, and *HPRT* was used as reference level of expression. Quantitative PCR was performed using the Real-Time PCR HT7900® (Applied Biosystems, Foster City, CA, USA). Sequences for the OCT3/4 splice variant specific primers were as described before [19, 38]. These are highly specific for the different isoforms and even discriminate between OCT4A and its pseudogenes. The following forward (-F) and reverse (-R) primers used (annotation between brackets = annotation from [19]): HPRT: HPRT244-exon2-F, 5'-AATTATGGACAGGACTGAACGTC-3'; HPRT243-exon3-R, 5'-CGTGGGGTCCTTTTCACCAGCAAG-3'. OCT4A: OCT4A-F (OCT4-AF) 5'-CTTCTCGCCCCCTCCAGGT-3'; OCT4A-R (OCT4-RB1) 5'-AAATAGAACCCCCAGGGTGAGC-3'. OCT4B: OCT4B-F (OCT4-FB) 5'-AGACTATTCCTTGGGGCCACAC-3'; OCT4B-R (OCT4-RB5) 5'-GGCTGAATACCTTCCCAAATAGA-3. OCT4B1: OCT4B-F (OCT4-FB), 5'-AGACTATTCCTTGGGGCCACAC-3'; OCT4B1-R (OCT-RB4) 5'-CTTAGAGGGGAGATGCGGTCA-3'. The localization of the different primers is depicted in Figure 1. The efficiency and specificity of these primers was extensively tested before [19]. The specificity for human RNA is proven by the absence of any OCT4A/B/B1 expression in most of the xenografts, specifically in PC82 which has a large stromal component. Quantitative values were obtained from the Ct. *OCT3/4* mRNAs (A, B and B1) were quantified relative to *HPRT* (*OCT3/4* mRNA = $2^{(\text{mean Ct HPRT-mean Ct OCT3/4 (A,B or B1)})}$) as described before [42]. The *OCT4*B1 PCR-products were sequenced using OCT4B1-F and a primer in exon 5 (OCT4B1-R2: (OCT4-RB3) 5'-CCCCCTGTCCCCCATTCCTA -3') to verify the nature of this splice variant. MicroRNA expression was measured as described previously [43].

Immunohistochemistry

Immunohistochemistry was performed on paraffin-embedded tissue sections of 4 μm thickness. Endogenous peroxidase and biotin were blocked. A mouse monoclonal antibody directly against OCT3/4 was used to detect OCT3/4 protein (1:350; SC5279), Santa Cruz, USA), which recognizes amino acids 1-134 of the protein and therefore recognizes OCT4A more specifically than the polyclonal antibody. Expression of OCT3/4 protein was double checked for most samples using a polyclonal antibody (1:350; SC8629), Santa Cruz, USA). Previously [26], a similar specificity and sensitivity of these antibodies in GCT tumour diagnostics has been shown, but did not yet differentiate between the different OCT3/4 isoforms. The proteins of the different isoforms only differ at their N-terminus. Therefore, the monoclonal antibody is specific for OCT4A. However strong similarities still exist in this regions between OCT4A and OCT3/4 pseudogenes [2, 19]. Slides were incubated as described earlier [6]. For different tissues and cell lines known positive controls we used to verify tissue integrity. The following antibodies were used: E-cadherin (1:200; clone nch-38, DAKO, Denmark), Ki-67 (1:50; clone BIB-1, code M7240, DAKO, Denmark), AFP (1:100; code A008, DAKO, Denmark), Pankeratin (1:400; Cat #MS-743-P, Neomarkers, USA), ERG (1:100; clone EPR3864, Epitomics, USA), TTF1 (1:200; Cat #MS-699-P, Neomarkers, USA), SOX2 (1:250; AF2018, R&D systems, USA), NANOG (1:400; AF1979, R&D systems, USA), ER (1:50); clone 1D5, Neomarkers, USA).

Statistics

Differences in gene expression between the groups were evaluated using the Mann-Whitney U test, using VassarStats[a]. A p-value of less than 0.05 was considered statistically significant. Correlation analysis was performed by calculating the Pearson correlation coefficient using SPSS 15.0.1. SPSS was also used to design the logistic regression model predicting the presence of a malignant GCT stem cell component based on mRNA expression of the three *OCT3/4* variants.

Results

Isoform-specific expression of the various isoforms *OCT4A*, *OCT4B* and *OCT4B1* was analyzed in a series of UNDIF-GCTs and DIF-GCTs as well as N-GCTs. Moreover, expression of these isoforms was also investigated in a panel of cell lines (both GCT-CL [38, 39] and N-GCT-CL, the latter including the prostate xenografts). For this purpose, a highly specific set of verified primer pairs was used (see Materials & Methods section and Figure 1 for details). The primer pair used to identify *OCT4A* was specifically designed to avoid false positive results caused by sequence-based similarities between the *OCT4A* transcripts and *OCT3/4* pseudo-genes [19, 38]. The obtained results will be discussed in the following paragraphs for each *OCT4A*, *OCT4B* and *OCT4B1*

---

[a] http://vassarstats.net/

separately. Subsequently, correlations between the different variants and association with the presence of a malignant GCT stem cell component were investigated. Finally protein expression will be discussed and correlated to mRNA expression.
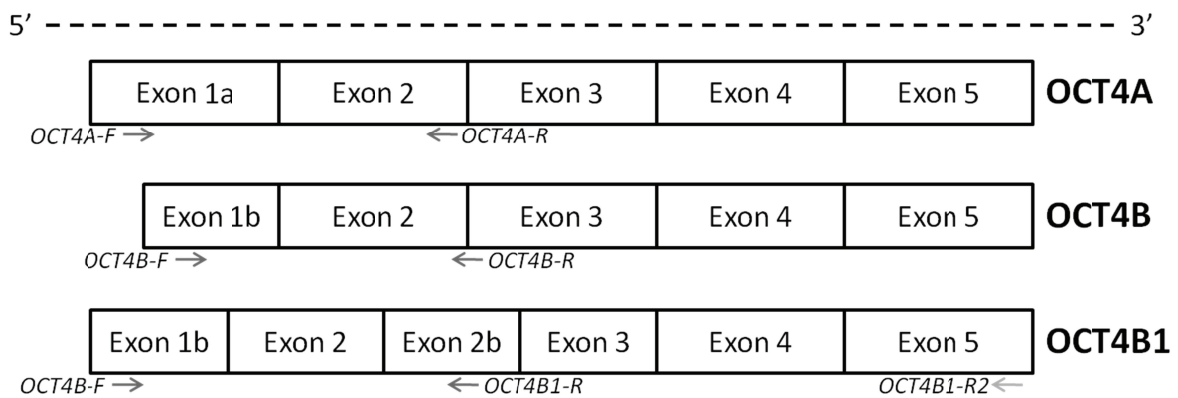

Figure 1. mRNA structure of the different *OCT3/4* splice variants. All *OCT3/4* splice variants have a similar 3' ends but differ in their 5' start (i.e. exon 1a and 1b). Moreover *OCT4B1* includes a previously identified intronic region now named exon 2b. For a detailed discussion see [19]. Arrows indicate forward (-F) and reverse (-R) primers from [19] used in the reverse transcription-polymerase chain reaction (RT-PCR) to specifically identify the different *OCT3/4* splice variants. OCT4B1-R2 denotes a specific reverse primer used to sequence *OCT4B1*.

OCT4A mRNA expression

OCT4A was significantly higher expressed in UNDIF-GCTs than in DIF-GCTs and N-GCTs (Figure 2A, Table 2). There was no significant difference between DIF-GCTs and N-GCTs. An overall higher level of expression was observed in the seminomas when compared to embryonal carcinomas (Figure 2B, Table 2). The DIF-GCTs, i.e., yolk sac tumours and teratomas, consistently showed virtually no expression of *OCT4A*. In the N-GCT group, tumours showed no or very low OCT4A expression (bladder-, prostate-, breast-, lung-, ovarian- and renal carcinomas, respectively) (Figure 2C, Table 2).

OCT4A was significantly higher expressed in GCT-CLs than in N-GCT-CLs (p=0.02) (Figure 3). All proven GCT-CLs consistently showed high expression levels of OCT4A. OCT4A expression is known to be absent in JKT-1, a not yet fully classified cell line suspected to be germ cell like, although not related to a seminoma [39]. Of note is that a few N-GCT-CLs showed a relatively high *OCT4A* expression level, defined as at least a ratio of 1.0 when compared to *HPRT*. However most of the GCT-CLs showed expression levels of >10. Cell lines with ratios of at least 1.0 included ESO51, HeLa, H460 and H716B. Ratios between 0.25 and 1.0 were found in one prostate and one mamma carcinoma cell line (22Rv1 and MDA175). All other N-GCT-CLs showed a low level or absence of OCT4A expression (Figure 3).
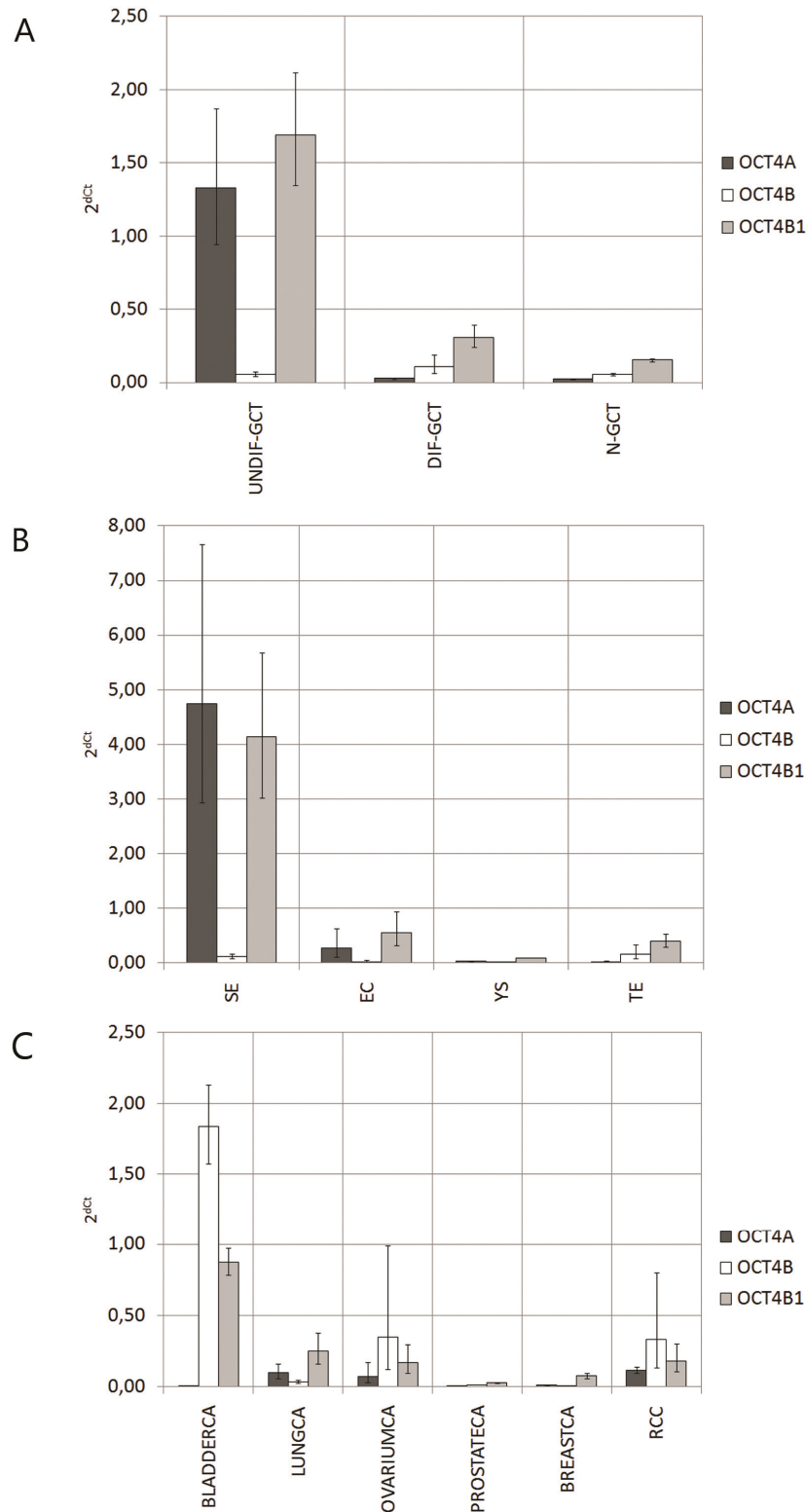
Figure 2. Expression patterns of OCT4A, OCT4B and OCT4B1 in UNDIF-GCT, DIF-GCT and N-GCT. X-axis: tumour samples/groups (see Materials & Methods section). Y-axis: 2^dCt (normalized to *HPRT*). Error bars depict SEM (Standard Error of Mean): variation within the groups (A-C). **(A)** Average expression for UNDIF-GCT, DIF-GCT and N-GCT groups. **(B)** Average expression for the UNDIF-GCT and DIF-GCT samples grouped per tumour type (SE = seminoma; EC = embryonal carcinoma; YS = yolk sac tumour; TE = teratoma). **(C)** Average expression for the N-GCT samples grouped per tumour type. CA = carcinoma.

OCT4B mRNA expression

OCT4B was expressed at equally low levels in UNDIF-GCTs and DIF-GCTs (Figure 2A, Table 2). No significant difference was detected between UNDIF-GCTs or DIF-GCTs and N-GCTs (Figure 2A, Table 2). Among the UNDIF-GCTs, seminomas expressed a low level of *OCT4B*, while *OCT4B* was virtually undetectable in embryonal carcinoma. The DIF-GCTs showed low expression in teratomas and practically absence of OCT4B in yolk sac tumours (Figure 2B, Table 2). Expression levels of *OCT4B* were highly variable within the N-GCT group. The bladder carcinoma samples showed the highest level of expression, which was rather similar between the different samples. The ovarian and renal carcinomas showed an intermediate level of expression, due to a number of high outliers within these groups. Almost no expression was found in lung, prostate and breast cancer samples (Figure 2C, Table 2).

The cell lines showed a highly variable expression of *OCT4B* (Figure 3). All UNDIF-GCT-CLs showed no or a very low level of *OCT4B* mRNA. No significant difference between N-GCT-CLs and UNDIF-GCT-CLs was detected (p=0.76). Most of the cell lines showed very low levels or absence of OCT4B expression. Relatively high levels of *OCT4B* were detected in HeLa, as well as in H716 and PC329. Moderate levels were detected in ESO51, VCaP and PC374 (Figure 3).

Table 2. Comparison of OCT4A/B/B1 expression in tumor samples. mRNA expression is scored as high (red, $2^\wedge$-dCt > 1), intermediate (black, $2^\wedge$-dCt 0.25-1) or low (green, $2^\wedge$-dCt < 0.25). The main tumor groups are tested for differential expression of the three *OCT3/4* variants: ns=not significant, arrow means relative overexpression of X in X vs. Y comparison. p-values: *OCT4A* (UNDIF-GCT vs. DIF-GCT p=0.02, UNIF-GCT vs. N-GCT p<0.01, DIF-GCT vs. N-GCT p=0.50); *OCT4B* (UNDIF-GCT vs. DIF-GCT p=0.95, UNDIF-GCT vs. N-GCT p=0.89, DIF-GCT vs. N-GCT p=0.89); *OCT4B1* (UNDIF-GCT vs. DIF-GCT p=0.04, UNDIF-GCT vs. N-GCT p<0.01, DIF-GCT vs. N-GCT p=0.39). Protein expression is scored as positive (high & nuclear or cytoplasmic) or negative (-,low) for absent or non-specific background staining. Staining with the monoclonal antibody was used for scoring.

	mRNA OCT4A	OCT4B	OCT4B1	Protein
UNDIF-GCT	high	low	high	nuclear
SE	high	low	high	nuclear
EC	intermediate	low	intermediate	nuclear
DIF-GCT	low	low	intermediate	-
YS	low	low	low	-
TE	low	low	intermediate	-
N-GCT	low	low	low	-
Bladdercarcinoma	low	high	intermediate	-
Lungcarcinoma	low	low	intermediate	-
Ovariumcarcinoma	low	intermediate	low	-
Prostatecarcinoma	low	low	low	-
Breastcarcinoma	low	low	low	-
Renal cell carcinoma	low	intermediate	low	-
Comparison between groups				
UNDIF-GCT vs. DIF-GCT	↑	ns	↑	
UNDIF-GCT vs. N-GCT	↑	ns	↑	
DIF-GCT vs. N-GCT	ns	ns	ns	

OCT4B1 mRNA expression

UNDIF-GCTs showed a significantly higher level of expression of *OCT4B1* than N-GCTs and DIF-GCTs (Figure 2A, Table 2). DIF-GCTs and N-GCTs exhibited no significant difference (Figure 2A, Table 2). Expression in the UNDIF-GCT group was high in seminoma and intermediate in embryonal carcinoma (Figure 2B, Table 2). Regarding DIF-GCTs, teratomas expressed intermediate levels of *OCT4B1* whereas expression in yolk sac tumours was low (see Figure 2B). Overall, expression of *OCT4B1* was low in N-GCTs. Bladder carcinomas showed, just as in the case of *OCT4B*, a relatively high expression level of *OCT4B1* when compared to other types of N-GCTs (Figure 2C).

Levels of *OCT4B1* expression varied between the cell lines. No significant difference between UNDIF-GCT-CLs and N-GCT-CLs was detected (p=0.92). High expression was observed in TCam-2, ESO51, HeLa, H716, 22Rv1, PC329, PC374. Moreover, intermediate expression levels were present in NCCIT, ESO26, SW620, VCaP, PC324, PC339, PC135, PC295. In many cases (ESO51, SW620, H716, VCaP, PC329, PC374) intermediate or high levels of OCT4B1 were combined with comparable levels of OCT4B (Figure 3).

During the sequencing process to confirm the PCR products for the different splice variants, a consistent TC insertion was found in exon 2B of *OCT4B1,* being a single nucleotide polymorphism (rs34631505). This SNP is located behind the stop codon and therefore has no consequence at the protein level.

Correlation between mRNA expression of different OCT3/4 isoforms and association of specific isoform expression and presence of a malignant germ cell component

When all samples were combined, OCT4A showed a strong positive correlation with OCT4B1 and a less strong correlation with OCT4B. OCT4B and OCT4B1 did not correlate significantly (Table 3). However, when the samples were split into undifferentiated (UNDIF-GCT) and differentiated tumours (DIF-GCT + N-GCT), strong, positive and highly significant correlations were found between all OCT3/4 variants. Overall, the strength of the correlation approached perfect positive correlation in the UNDIF-GCT. In general, the correlations were less strong in the differentiated tumours, but still highly significant and positive (Table 3). In a binary logistic regression model, OCT4A and OCT4B proved to be significant in predicting the presence of a malignant GCT stem cell component. OCT4A was strongly predictive for the presence a malignant GCT stem cell component (β=-4.92, p=0.045), while OCT4B proved to be associated with the absence of such a component, but this association was less strong (β=-1.28, p=0.048). *[correction: OCT4A was predictive for the presence a malignant GCT stem cell component (β=1.28, p=0.048), while OCT4B proved to be associated with the absence of such a component (β=-4.923, p=0.045).]*

Figure 3. Expression patterns of *OCT4A, OCT4B and OCT4B1* **in cell lines. (A)** mRNA levels in all investigated cell lines, both of GCT origin (GCT-CL) and non-GCT origin (N-GCT-CL). * = cervix carcinoma, ** = lung carcinoma. X-axis: cell lines (see Materials & Methods section) and corresponding tumour class. Y-axis: 2^dCt (normalized to *HPRT*). **(B)** Interpretation of the expression of *OCT3/4* isoforms relative to *HPRT*. mRNA expression is scored as high (red, 2^-dCt > 1), intermediate (black, 2^-dCt 0.25-1) or low (green, 2^-dCt < 0.25). CA = carcinoma. →
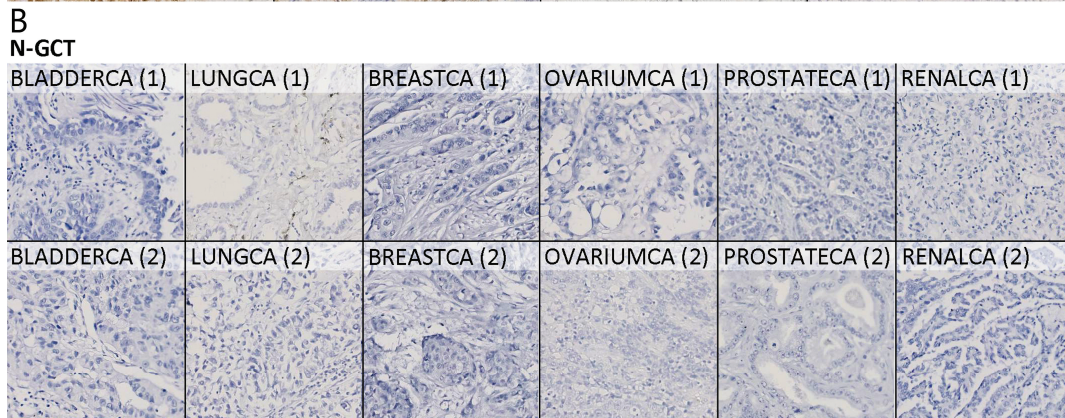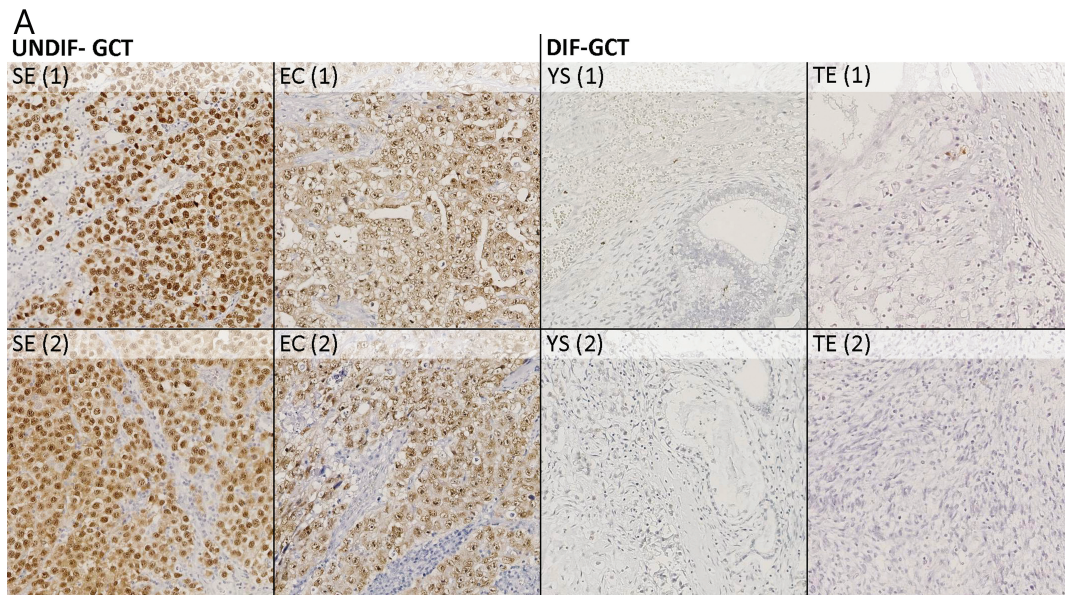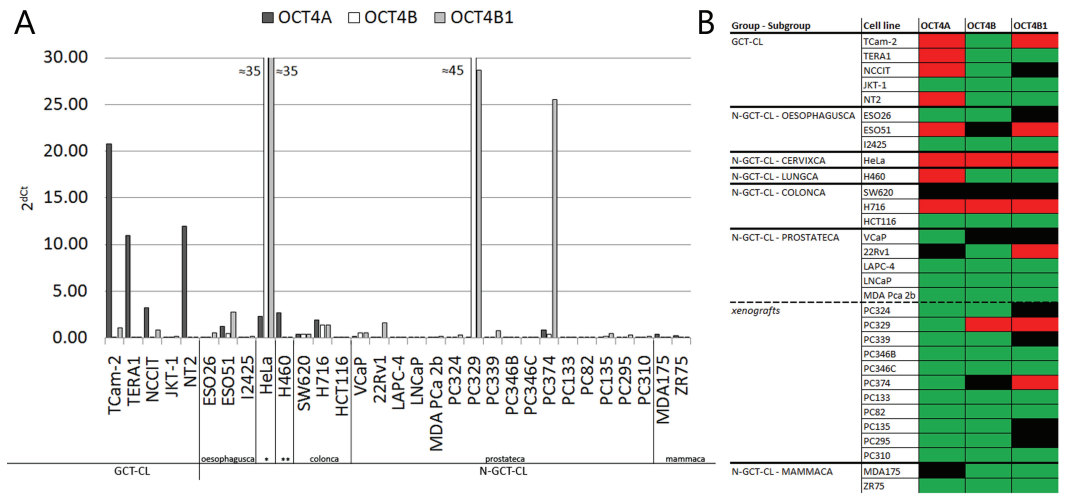
Figure 4. Immunohistochemical detection of OCT3/4 expression in tumour samples. For each tumour type, two different samples are shown. Magnification 100x. **(A)** Protein expression of OCT3/4 in UNDIF-GCTs and DIF-GCTs. Shown are two seminomas (SE), embryonal carcinomas (EC), yolk sac tumours (YS) and teratomas (TE), of which only the first two types are positive. **(B)** Protein expression of OCT3/4 in N-GCT tumour samples, including two carcinomas of the bladder, lung, breast, ovary, prostate and kidney, respectively, all are negative. All samples were stained using an antibody that is most specific for OCT4A (see Materials & Methods). →

Table 3. Correlation between OCT4A/B/B1 mRNA expression in tumor samples. Correlation was assessed using the Pearson correlation coefficient on the mRNA expression data of all tumor samples, the UNDIF-GCT group and the differentiated tumors (DIF-GCT and N-GCT). LEGEND: +++ = ρ > 0.75, ++ = ρ > 0.5, + = ρ < 0.5, ns = not significant. CORRELATION COEFICIENT/SIGNIFICANCE: All: $\rho_{OCT4A,OCT4B}$ = 0.37 (p=0.01), $\rho_{OCT4A,OCT4B1}$ = 0.89 (p<0.01), $\rho_{OCT4B,OCT4B1}$ = 0.12 (p=0.44). UNDIF-GCT: $\rho_{OCT4A,OCT4B}$ = 0.98 (p<0.01), $\rho_{OCT4A,OCT4B1}$ = 1.00 (p<0.01), $\rho_{OCT4B,OCT4B1}$ = 0.99 (p<0.01). DIF-GCT + N-GCT: $\rho_{OCT4A,OCT4B}$ = 0.63 (p<0.01), $\rho_{OCT4A,OCT4B1}$ = 0.97 (p<0.01), $\rho_{OCT4B,OCT4B1}$ = 0.68 (p<0.01).

	All		UNDIF-GCT		DIF-GCT + N-GCT	
	OCT4B	OCT4B1	OCT4B	OCT4B1	OCT4B	OCT4B1
OCT4A	+	+++	+++	+++	++	+++
OCT4B		ns		+++		++

Protein detection

Immunohistochemical staining of the various GCT samples and cell lines was performed to assess OCT3/4 protein expression. Clear nuclear staining of tumour cells was shown in UNDIF-GCTs. No expression was detected in the DIF-GCT components or the N-GCT samples (Figure 4). In accordance with the findings in the tumour samples, nuclear staining of tumour cells was shown in GCT-CLs, while being absent in all N-GCT-CLs and xenografts. Both the EC cell lines (NCCIT and NT2) and the SE cell line TCam-2 were OCT3/4 positive. The nonspecific staining in the xenografts was based on necrosis and again no cytoplasmic or nuclear staining was detected in these samples (Figure 5). Positive controls for all samples prove that all samples were suitable for immunohistochemistry (Figure S1AB, Figure S4A) and HE staining was used to assess tumour morphology (Figure S2AB). Finally, a double check for OCT3/4 expression was performed by staining the same tumour samples (Figure S3AB) and cell lines/xenografts (Figure S4B) with a second (polyclonal) antibody directed against OCT3/4. This confirmed our findings that OCT3/4 protein expression is specific to UNDIF-GCTs and the related cell line models (GCT-CLs). These data are completely in accordance with previously published findings [6, 26].

A

Legend: ■ OCT4A □ OCT4B ▨ OCT4B1

2^{dCt} (y-axis): 0.00 – 30.00

Annotations: ≈35, ≈35, ≈45

Cell lines (x-axis): TCam-2, TERA1, NCCIT, JKT-1, NT2, ESO26, ESO51, I2425, HeLa, H460, SW620, H716, HCT116, VCaP, 22Rv1, LAPC-4, LNCaP, MDA PCa 2b, PC324, PC329, PC339, PC346B, PC346C, PC374, PC133, PC82, PC135, PC295, PC310, MDA175, ZR75

Groups: GCT-CL | oesophagusca | * | ** | colonca | prostateca | mammaca | N-GCT-CL

B

Group - Subgroup	Cell line	OCT4A	OCT4B	OCT4B1
GCT-CL	TCam-2			
	TERA1			
	NCCIT			
	JKT-1			
	NT2			
N-GCT-CL - OESOPHAGUSCA	ESO26			
	ESO51			
	I2425			
N-GCT-CL - CERVIXCA	HeLa			
N-GCT-CL - LUNGCA	H460			
N-GCT-CL - COLONCA	SW620			
	H716			
	HCT116			
N-GCT-CL - PROSTATECA	VCaP			
	22Rv1			
	LAPC-4			
	LNCaP			
	MDA Pca 2b			
xenografts	PC324			
	PC329			
	PC339			
	PC346B			
	PC346C			
	PC374			
	PC133			
	PC82			
	PC135			
	PC295			
	PC310			
N-GCT-CL - MAMMACA	MDA175			
	ZR75			

A

UNDIF- GCT

SE (1)   EC (1)

DIF-GCT

YS (1)   TE (1)

SE (2)   EC (2)

YS (2)   TE (2)

B

N-GCT

BLADDERCA (1) | LUNGCA (1) | BREASTCA (1) | OVARIUMCA (1) | PROSTATECA (1) | RENALCA (1)

BLADDERCA (2) | LUNGCA (2) | BREASTCA (2) | OVARIUMCA (2) | PROSTATECA (2) | RENALCA (2)

GCT-CL

NT2	NCCIT	TCam-2

N-GCT-CL

ESO26 (OESOPHAGUSCA)	ESO51 (OESOPHAGUSCA)	I2425 (OESOPHAGUSCA)	ZR75 (BREASTCA)

HELA (CERVIXCA)	H460 (LUNGCA)	SW620 (COLONCA)	H716 (COLONCA)

HCT116 (COLONCA)	VCaP (PROSTATECA)	22Rv1 (PROSTATECA)	LAPC-4 (PROSTATECA)

LNCaP (PROSTATECA)	PC310 (PROSTATECA)	PC295 (PROSTATECA)	PC345C (PROSTATECA)

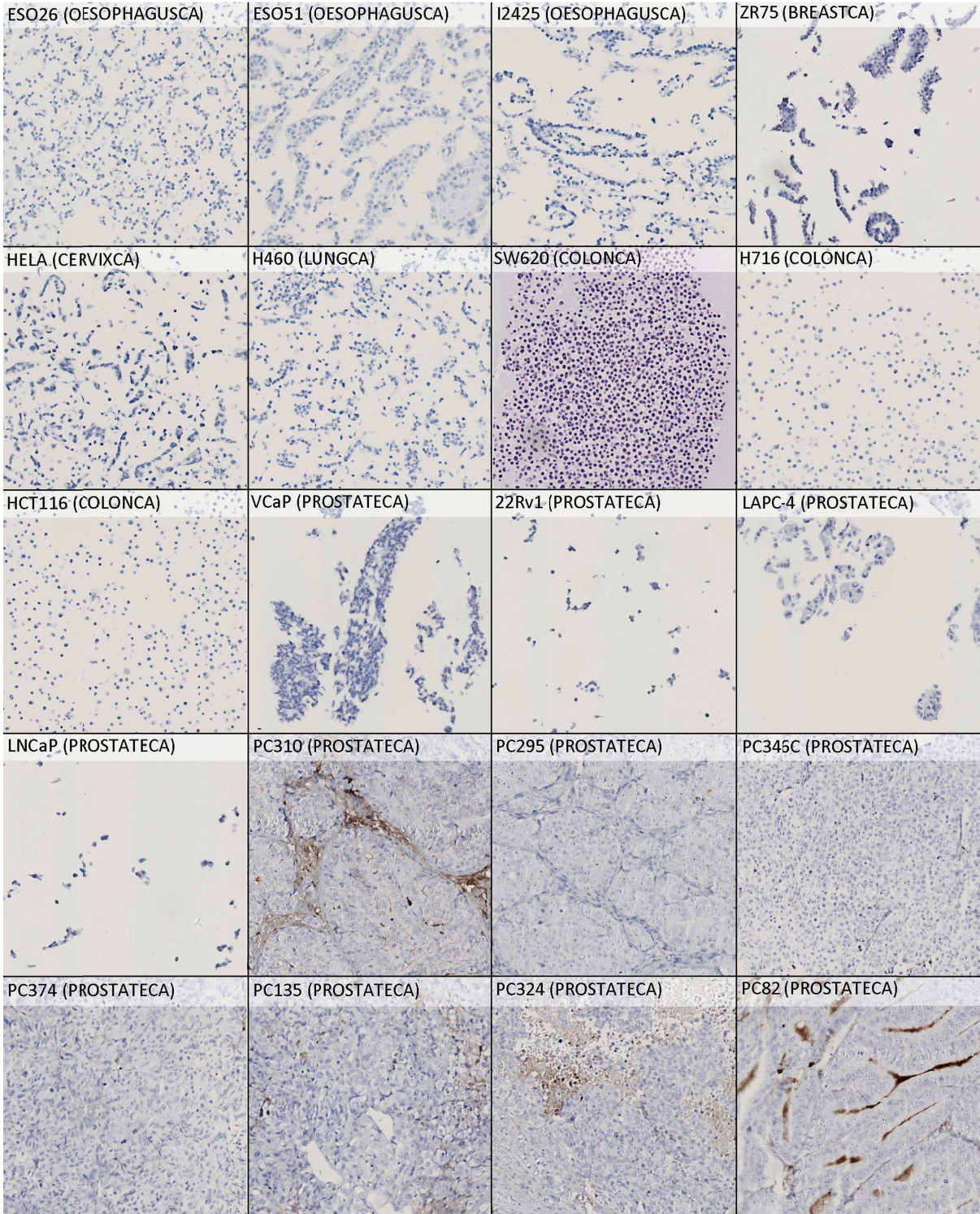| PC374 (PROSTATECA)	PC135 (PROSTATECA)	PC324 (PROSTATECA)	PC82 (PROSTATECA)

← Figure 5. **Immunohistochemical detection of OCT3/4 expression in cell lines and xenografts.** Stained are the undifferentiated GCT-CLs NT2, NCCIT and TCam-2 which are all positive. All somatic cancer cell lines (ESO26, ESO51, I2425, ZR75, HELA, H460, SW620, H716, HCT116, VCaP, 22Rv1, LAPC-4, LNCaP) and xenografts (PC310, PC295, PC346C, PC374, PC135, PC324, PC82) are negative (some nonspecific staining of necrosis). All samples were stained using an antibody that is most specific for OCT4A (see Materials & Methods). Magnification 200x GCT-CLs, 100xN-GCT-CLs. CA = carcinoma.

Discussion

Specificity: the pitfalls of pseudogenes and isoforms

Detection of OCT3/4 pseudogenes can and should be avoided by sufficient DNAse pre-treatment of the sample, because the respective sequences might be amplified based on their high level of similarity with the protein encoding variant (OCT4A) and the absence of introns. Therefore, their amplification in PCR might be falsely interpreted as actual OCT4A expression, suggesting possible translation into OCT4A protein. Moreover, specific PCR primer pairs followed by antibody-based analysis should be used to detect the different isoforms of OCT4 at the mRNA level and the presence of protein [19, 33, 38, 44].

By using such a validated, isoform-specific primer pair setup, this study shows that OCT4A is highly expressed in UNDIF-GCTs, which are known to have a pluripotent stem cell component, originating from PGCs/gonocytes [9, 45]. DIF-GCTs and N-GCT show virtually no OCT4A expression, which is in line with the notion that OCT4A is responsible for formation of the protein involved in regulation of pluripotency. In contrast, OCT4B is not differentially expressed between the three groups, while OCT4B1 is expressed significantly higher in the UNDIF-GCTs when compared to DIF-GCTs and N-GCTs, as found for OCT4A. These results indeed support the general consensus that OCT4A is the marker for stem cell populations in GCTs, and a similar specificity could be suggested for the OCT4B1 variant. However, OCT3/4 protein is only detected in GCTs and representative cell lines, also in this study. No specific signal could be detected in any of somatic cancers or cell lines investigated, irrespective of mRNA expression pattern. Because OCT4A protein is the only one of the three isoforms that directly regulates pluripotency, expression of this protein is a prerequisite for any cell that uses OCT3/4 as a regulator of pluripotency. Therefore, even specific mRNA detection of OCT3/4 isoforms does not yet conclusively prove the applicability of OCT4A, OCT4B or OCT4B1 in the detection of pluripotent cancer stem cells (see below) or somatic stem cells without protein confirmation.

Functions of OCT3/4 isoforms

OCT3/4 is a known marker for pluripotency and has been shown to have a role in regulation of pluripotency [1-3]. Moreover, it is an important diagnostic marker for specific types of GCTs [6, 7, 9, 26, 45]. Recently, investigations into the broader applicability of OCT4 as a marker in (cancer) stem cell biology showed that only OCT4A (and not OCT4B and OCT4B1) is specific to stem cell (like) populations [2, 19, 30, 32, 33, 46]. It has been shown that basic levels of *OCT3/4* mRNA (even OCT4A) and expression its pseudogenes, are detectable in somatic (tumour) cells [2, 47]. However, OCT4A protein expression has so far never been conclusively shown in non-pluripotent cells [2]. In addition, it has been described that OCT4A is primarily localised in the nucleus, while OCT4B(1) primarily resides in the cytoplasm.

OCT4B might have a role in stress response [2]. The role of OCT4B1 is more elusive. It has been associated with both pluripotency and tumorigenesis (via inhibition of apoptosis and cell cycle deregulation) [16]. Also, a recent report suggested OCT4B1 to be superior to OCT4A in the detection of stemness, at least in human embryonic stem cells [20]. These results are mainly based on statistical correlation and lack a biological explanation as to how OCT4B1 contributes to pluripotency, since this variant cannot be directly translated into a functional transcription factor [19]. Moreover, the presence of OCT4B1 expression in cancer tissues, which consists of mainly differentiated tissue, is not satisfactory linked to a specific hypothetical *in situ* population of cancer stem cells. In contrast, Gao and coworkers conclude that OCT4B1 can be alternatively spliced and subsequently be translated into all OCT4B protein forms [17], linking OCT4B1 to OCT4B mediated functions like stress respons.

The various OCT3/4 isoforms might also have an integrated function as interchangeable decoys in microRNA (miR) regulated OCT3/4 protein expression. miRs, specifically miR-145 which targets OCT3/4, have been suggested to play a role in the regulation of pluripotency in general and OCT3/4 translation specifically [48]. Competitive microRNA (miR) binding has been suggested as a biological function of pseudogenes [49]. This function might also apply to alternative splice variants that, such as the OCT3/4 variants share their 3'UTR. OCT4B and/or OCT4B1 might prevent translation inhibition of OCT4A mRNA in stem cell components of GCTs or the other way around in differentiated GCTs or somatic cancer cells. The latter could explain OCT4A mRNA expression without translation into detectable protein in somatic cancer cells. This hypothesis is supported by a correlation analysis on the OCT4A/B/B1 mRNA expression data (Table 3). We showed strong correlations between OCT4B and B1 (Table 3) (expected based on similar function [17]) and OCT4A and B/B1 (strongest in UNDIF-GCT where OCT4A is active at the protein level). However, functionally studies are required to support this hypothesis.

Finally, individuals homozygous for a polymorphism at the initiating codon of *OCT3/4* (rs3130932) are not able to transcribe OCT4B(1), and are therefore lacking the putative encoding proteins [11, 50]. Depending on race, minor allele frequencies of 23-33% are reported [50]. So far, no abnormalities have been found related to the absence of this protein in these individuals, but it would be interesting to investigate the relative frequency of this SNP in GCT patients, specifically with respect to OCT3/4 (protein) expression, tumour characteristics and clinical course.

OCT3/4 isoforms in (cancer) stem cells

Our results disprove the applicability of OCT3/4 mRNA [30] for the detection of pluripotent cells (possibly cancer stem cells [51, 52]) in solid cancers. Bladder carcinomas showed high mRNA expression of OCT4B and OCT4B1, but no OCT4A. This explains earlier reports of high non-specific OCT4 expression in this type of cancer [31], but does not indicate the presence of OCT4A positive cancer stem cells. Moreover, our analysis identified no OCT4A expression in prostate carcinoma, which has been reported before [33] using a specific primer set [44]. In contrast, low but detectable levels of OCT4A mRNA were found in lung-, ovary- and renal carcinoma samples. Also, some of the studied cell lines showed OCT4A mRNA expression, suggesting the presence of pluripotent cells in these cultures. However, no OCT3/4 protein expression could be identified in any of the N-GCT samples or N-GCT-CLs, using both monoclonal and polyclonal antibodies. Also the "stem cells" in cell lines do not necessarily represent *in vivo* cancer stem cells, illustrated by the absence of OCT4A mRNA in five lung carcinoma samples and the presence of OCT4A mRNA in lung carcinoma cell line H460.

OCT4A is therefore no marker of cancer stem-cellness in N-GCT, despite its undisputed crucial role in physiological (maintenance of) pluripotency (in germ cell precursors and their malignant counterparts) [1, 3]. OCT4B1 does have a significant tendency towards specificity for the pluripotent stem cell component of UNDIF-GCTs (Figure 2AB, Table 2, Table 3). It is however also (highly) expressed in differentiated tumours and cell lines (Figure 2C, Figure 3, Table 2, Table 3). OCT4B1 has been associated with detection of pluripotency before, but no functional relation has been proven and recent research has functionally linked OCT4B1 to OCT4B (stress response) rather than pluripotency [17].

Conclusion

This research confirms that different OCT4 isoforms (and pseudogenes) contribute to non-specific findings of OCT3/4 expression in various tissues and cell lines. This observation emphasizes the necessity of using highly specific primer sets and antibodies to investigate the presence of expression of functional (nuclear) OCT3/4 (protein). The presented data confirms the specificity of OCT4A as a marker for the seminomatous and the stem cell component of nonseminomatous GCTs and illustrates the varying mRNA expression levels of OCT3/4 isoforms in other types of solid

cancer and cell lines. OCT4A and OCT4B1 were both confirmed to have a significantly higher expression in tissues with a known stem cell component, but until now, only OCT4A can be directly functionally linked to pluripotency. Moreover, this study shows that OCT3/4 protein detection is of crucial importance, because of clear discrepancies between even isoform-specific mRNA expression and protein detection, possibly due to post-transcriptonal regulation. A synergistic role for the different OCT4 splice variants, possibly by competitive miR binding, might be an interesting model to investigate.

Acknowledgements

Supporting information

Figure S1. Positive controls for immunohistochemistry on GCT and N-GCT samples. To investigate the tissue integrity for immunohistochemistry, all samples were stained with a known positive control for each tumour type. For each tumour type, two different samples are shown. See legend of Figure 4 for explanation. The protein against which the staining was directed is displayed between brackets. Ki-67: proliferation-related Ki-67 antigen; TTF1: transcription termination factor; ER: estrogen receptor; pankeratin: staining of fibrous structural proteins (keratin family) present in ovarium carcinoma; ERG: a transcriptional regulator present in prostate carcinomas. Magnification 100x. (A) GCT samples. The SOX2 positive areas in the TE sections are EC components surrounded by mature teratoma. (B) N-GCT samples. CA = carcinoma.

Figure S2. Tumor morphology for GCT and N-GCT samples. Hematoxylin-eaosin staining. Magnification 100x. CA = carcinoma. (A) GCT samples. (B) N-GCT samples. See legend of Figure 4 for explanation.

Figure S3. Immunohistochemical detection of OCT3/4 expression in tumour samples: polyclonal antibody. A double check for OCT3/4 expression was performed by staining the same tumour samples with a second polyclonal (less specific) antibody directed against OCT3/4. This antibody is known to display more non-specific background staining. For each tumour type, two different samples are shown. See legend of Figure 4 for explanation. Magnification 100x. CA = carcinoma. (A) Protein expression of OCT3/4 in UNDIF-GCTs and DIF-GCTs. (B) Protein expression of OCT3/4 in N-GCT tumour samples. Prostate carcinoma samples were not included (shortage of material). The staining in breastcarcinoma-2 is nonspecific, non-nuclear and is absent in the monoclonal staining (Figure 4B).

Figure S4. Positive controls and confirmation of OCT3/4 expression in cell lines. (A) Positive controls for immunohistochemistry on GCT-CLs and N-GCT-CLs. To investigate the tissue integrity for immunohistochemistry, all cell lines were stained with a known positive control. See legend of Figure 5 for explanation. The protein against which the staining was directed is displayed between brackets. Magnification 200x GCT-CLs, 100xN-GCT-CLs. (B) Confirmation of OCT3/4 protein expression. A double check for OCT3/4 expression was performed by staining the N-GCT cell lines with a second polyclonal (less specific) antibody directed against OCT3/4. All were confirmed as negative. See legend of Figure 5 for explanation. Prostate carcinoma cell lines/xenografts and GCT-CLs were not included (shortage of material). CA = carcinoma.

Supporting information is available at: http://dx.doi.org/10.1038/bjc.2011.270

References

1. Pesce M, Scholer HR: **Oct-4: control of totipotency and germline determination.** *Mol Reprod Dev* 2000, 55:452-457.
2. Wang X, Dai J: **Concise review: isoforms of OCT4 contribute to the confusing diversity in stem cell biology.** *Stem Cells* 2010, 28:885-893.
3. Pesce M, Scholer HR: **Oct-4: gatekeeper in the beginnings of mammalian development.** *Stem Cells* 2001, 19:271-278.
4. Niwa H, Miyazaki J, Smith AG: **Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.** *Nat Genet* 2000, 24:372-376.
5. Kehler J, Tolkunova E, Koschorz B, Pesce M, Gentile L, Boiani M, Lomeli H, Nagy A, McLaughlin KJ, Scholer HR, Tomilin A: **Oct4 is required for primordial germ cell survival.** *EMBO reports* 2004, 5:1078-1083.
6. Looijenga LH, Stoop H, de Leeuw HP, de Gouveia Brazao CA, Gillis AJ, van Roozendaal KE, van Zoelen EJ, Weber RF, Wolffenbuttel KP, van Dekken H, et al: **POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.** *Cancer Res* 2003, 63:2244-2250.
7. de Jong J, Looijenga LH: **Stem cell marker OCT3/4 in tumor biology and germ cell tumor diagnostics: history and future.** *Critical reviews in oncogenesis* 2006, 12:171-203.
8. Cheng L, Sung MT, Cossu-Rocca P, Jones TD, MacLennan GT, De Jong J, Lopez-Beltran A, Montironi R, Looijenga LH: **OCT4: biological functions and clinical applications as a marker of germ cell neoplasia.** *J Pathol* 2007, 211:1-9.
9. Looijenga L: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J of Pathology* 2009, 218:146-162.
10. Krishnan BR, Jamry I, Chaplin DD: **Feature mapping of the HLA class I region: localization of the POU5F1 and TCF19 genes.** *Genomics* 1995, 30:53-58.
11. Takeda J, Seino S, Bell GI: **Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues.** *Nucleic Acids Res* 1992, 20:4613-4620.
12. Sturm RA, Herr W: **The POU domain is a bipartite DNA-binding structure.** *Nature* 1988, 336:601-604.
13. Liedtke S, Stephan M, Kogler G: **Oct4 expression revisited: potential pitfalls for data misinterpretation in stem cell research.** *Biological chemistry* 2008, 389:845-850.
14. Pain D, Chirn GW, Strassel C, Kemp DM: **Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification.** *J Biol Chem* 2005, 280:6265-6268.
15. Suo G, Han J, Wang X, Zhang J, Zhao Y, Zhao Y, Dai J: **Oct4 pseudogenes are transcribed in cancers.** *Biochemical and biophysical research communications* 2005, 337:1047-1051.
16. Asadi MH, Mowla SJ, Fathi F, Aleyasin A, Asadzadeh J, Atlasi Y: **OCT4B1, a novel spliced variant of OCT4, is highly expressed in gastric cancer and acts as an antiapoptotic factor.** *International journal of cancer* 2011, 128:2645-2652.
17. Gao Y, Wang X, Han J, Xiao Z, Chen B, Su G, Dai J: **The novel OCT4 spliced variant OCT4B1 can generate three protein isoforms by alternative splicing into OCT4B.** *Journal of genetics and genomics = Yi chuan xue bao* 2010, 37:461-465.
18. Lee J, Kim HK, Rho JY, Han YM, Kim J: **The human OCT-4 isoforms differ in their ability to confer self-renewal.** *J Biol Chem* 2006, 281:33554-33565.
19. Atlasi Y, Mowla SJ, Ziaee SA, Gokhale PJ, Andrews PW: **OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells.** *Stem Cells* 2008, 26:3068-3074.
20. Papamichos SI, Kotoula V, Tarlatzis BC, Agorastos T, Papazisis K, Lambropoulos AF: **OCT4B1 isoform: the novel OCT4 alternative spliced variant as a putative marker of stemness.** *Molecular human reproduction* 2009, 15:269-270.
21. Botquin V, Hess H, Fuhrmann G, Anastassiadis C, Gross MK, Vriend G, Scholer HR: **New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2.** *Genes Dev* 1998, 12:2073-2090.
22. Hay DC, Sutherland L, Clark J, Burdon T: **Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells.** *Stem Cells* 2004, 22:225-235.

23.   Matin MM, Walsh JR, Gokhale PJ, Draper JS, Bahrami AR, Morton I, Moore HD, Andrews PW: **Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells.** *Stem Cells* 2004, 22:659-668.

24.   Velkey JM, O'Shea KS: **Oct4 RNA interference induces trophectoderm differentiation in mouse embryonic stem cells.** *Genesis* 2003, 37:18-24.

25.   Zaehres H, Scholer HR: **Induction of pluripotency: from mouse to human.** *Cell* 2007, 131:834-835.

26.   de Jong J, Stoop H, Dohle GR, Bangma CH, Kliffen M, van Esser JW, van den Bent M, Kros JM, Oosterhuis JW, Looijenga LH: **Diagnostic value of OCT3/4 for pre-invasive and invasive testicular germ cell tumours.** *J Pathol* 2005, 206:242-249.

27.   de Jong J, Stoop H, Gillis AJ, van Gurp RJ, van de Geijn GJ, Boer M, Hersmus R, Saunders PT, Anderson RA, Oosterhuis JW, Looijenga LH: **Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.** *J Pathol* 2008, 215:21-30.

28.   van Casteren NJ, Stoop H, Dohle GR, de Wit R, Oosterhuis JW, Looijenga LH: **Noninvasive detection of testicular carcinoma in situ in semen using OCT3/4.** *European urology* 2008, 54:153-158.

29.   Ezeh UI, Turek PJ, Reijo RA, Clark AT: **Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma.** *Cancer* 2005, 104:2255-2265.

30.   Tai MH, Chang CC, Kiupel M, Webster JD, Olson LK, Trosko JE: **Oct4 expression in adult human stem cells: evidence in support of the stem cell theory of carcinogenesis.** *Carcinogenesis* 2005, 26:495-502.

31.   Atlasi Y, Mowla SJ, Ziaee SA, Bahrami AR: **OCT-4, an embryonic stem cell marker, is highly expressed in bladder cancer.** *Int J Cancer* 2007, 120:1598-1602.

32.   Lengner CJ, Camargo FD, Hochedlinger K, Welstead GG, Zaidi S, Gokhale S, Scholer HR, Tomilin A, Jaenisch R: **Oct4 expression is not required for mouse somatic stem cell self-renewal.** *Cell stem cell* 2007, 1:403-415.

33.   Sotomayor P, Godoy A, Smith GJ, Huss WJ: **Oct4A is expressed by a subpopulation of prostate neuroendocrine cells.** *The Prostate* 2009, 69:401-410.

34.   Chen YC, Hsu HS, Chen YW, Tsai TH, How CK, Wang CY, Hung SC, Chang YL, Tsai ML, Lee YY, et al: **Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells.** *PloS one* 2008, 3:e2637.

35.   Lai DM, Liu T, Huang Y, Wang LH, Zhang J, Cheng WW: **[Identification and characterization of ovarian cancer stem-like cells from primary tumor.].** *Zhonghua fu chan ke za zhi* 2009, 44:936-940.

36.   Marques RB, van Weerden WM, Erkens-Schulze S, de Ridder CM, Bangma CH, Trapman J, Jenster G: **The human PC346 xenograft and cell line panel: a model system for prostate cancer progression.** *Eur Urol* 2006, 49:245-257.

37.   van Weerden WM, de Ridder CM, Verdaasdonk CL, Romijn JC, van der Kwast TH, Schroder FH, van Steenbrugge GJ: **Development of seven new human prostate tumor xenograft models and their histopathological characterization.** *The American journal of pathology* 1996, 149:1055-1062.

38.   de Jong J, Stoop H, Gillis AJ, Hersmus R, van Gurp RJ, van de Geijn GJ, van Drunen E, Beverloo HB, Schneider DT, Sherlock JK, et al: **Further characterization of the first seminoma cell line TCam-2.** *Genes Chromosomes Cancer* 2008, 47:185-196.

39.   de Jong J, Stoop H, Gillis AJ, van Gurp RJ, van Drunen E, Beverloo HB, Lau YF, Schneider DT, Sherlock JK, Baeten J, et al: **JKT-1 is not a human seminoma cell line.** *International journal of andrology* 2007, 30:350-365.

40.   Andrews PW, Casper J, Damjanov I, Duggan-Keen M, Giwercman A, Hata J, von Keitz A, Looijenga LH, Millan JL, Oosterhuis JW, et al: **Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours.** *Int J Cancer* 1996, 66:806-816.

41.   Looijenga LH, Hersmus R, Gillis AJ, Pfundt R, Stoop HJ, van Gurp RJ, Veltman J, Beverloo HB, van Drunen E, van Kessel AG, 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, 66:290-302.

42.   Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods (San Diego, Calif* 2001, 25:402-408.

43.   Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW, Sun Y, Chen C, Guenther S, Sherlock J, Veltman I, Baeten J, et al: **High-throughput microRNAome analysis in human germ cell tumours.** *J Pathol* 2007, 213:319-328.

44.   Liedtke S, Enczmann J, Waclawczyk S, Wernet P, Kogler G: **Oct4 and its pseudogenes confuse stem cell research.** *Cell stem cell* 2007, 1:364-366.

45. Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5**:**210-222.

46. Cantz T, Key G, Bleidissel M, Gentile L, Han DW, Brenne A, Scholer HR: **Absence of OCT4 expression in somatic tumor cell lines.** *Stem Cells* 2008, 26**:**692-697.

47. Zhao S, Yuan Q, Hao H, Guo Y, Liu S, Zhang Y, Wang J, Liu H, Wang F, Liu K, et al: **Expression of OCT4 pseudogenes in human tumours: lessons from glioma and breast carcinoma.** *J Pathol* 2011, 223**:**672-682.

48. Xu N, Papagiannakopoulos T, Pan G, Thomson JA, Kosik KS: **MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.** *Cell* 2009, 137**:**647-658.

49. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP: **A coding-independent function of gene and pseudogene mRNAs regulates tumour biology.** *Nature* 2010, 465**:**1033-1038.

50. Hussain SK, Sequerra R, Bertucci C, Hastings NC, Rieder M, Schwartz SM: **Sequence variation in the human transcription factor gene POU5F1.** *BMC genetics* 2008, 9**:**15.

51. Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ: **Prospective identification of tumorigenic prostate cancer stem cells.** *Cancer Res* 2005, 65**:**10946-10951.

52. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, De Maria R: **Identification and expansion of human colon-cancer-initiating cells.** *Nature* 2007, 445**:**111-115.

CHAPTER **4**

miMsg: a target enrichment algorithm for predicted miR-mRNA interactions based on relative ranking of matched expression data

Martin A. Rijlaarsdam[1], David J. Rijlaarsdam[2,3], Ad J.M. Gillis[1],
Lambert C.J. Dorssers[1], Leendert H.J. Looijenga[1]

Department of Pathology, Erasmus MC - University Medical Center, Rotterdam, The Netherlands [1]
Department of Mechanical Engineering, Control Systems Technology,
Eindhoven University of Technology, Eindhoven, The Netherlands [2]
Department of Fundamental Electricity and Instrumentation,
Free University of Brussels (VUB), Brussels, Belgium [3]

Abstract

Motivation: Algorithms predicting miR-mRNA interactions generate high numbers of possible interactions, many of which might be non-existent or irrelevant in a certain biological context. It is desirable to develop a transparent, user-friendly, unbiased tool to enrich miR-mRNA predictions.

Results: The miMsg algorithm uses matched miR/mRNA expression data to enrich miR-mRNA predictions. It grades interactions by the number, magnitude and significance of misplacements in the combined ranking profiles of miR/mRNA expression assessed over multiple biological samples. miMsg requires minimal user input and makes no statistical assumptions. It identified 921 out of 56,262 interactions as top-scoring and significant in an actual germ cell cancer dataset. Twenty-eight miR-mRNA pairs were deemed of highest interest based on ranking by miMsg and supported by current knowledge about validated interactions and biological function. To conclude, miMsg is an effective algorithm to reduce a high number of predicted interactions to a small set of high confidence interactions for further study.

Availability and Implementation: Matlab source code and datasets available online[a]

---

[a] www.martinrijlaarsdam.nl/mimsg

Introduction

MicroRNAs (miRs) are a subclass of a large group of non-coding RNAs that play an important role in post transcriptional regulation of gene expression, crucially influencing development and physiology [1-5]. In addition, miRs also contribute to initiation and progression of diseases, including cancer [3, 6, 7]. Around 2000 miRs have been identified in mammals so far (miRBase[b], [8]), each possibly regulating (protein levels of) hundreds of different mRNAs according to available prediction algorithms [9]. Much research has been done to elucidate the regulation of mRNA levels by miRs. Initial results showed that in plants mRNA degradation occurs by fully complementary binding of miRs to open reading frames (ORFs) [10]. In animals, miRs were hypothesized to predominantly repress translation by binding of partially complementary sequences to their target's mRNAs 3'-untranslated region (UTR), hence not leading to mRNA degradation [3, 10-12]. However, mRNA degradation as a consequence of miR binding has also been demonstrated in animals [3, 10, 13-18]. In general, induction of miR expression results in reduced levels of their mRNA targets, while depletion of a specific miR will lead higher target levels [19-22]. Inactivation of the miR processing pathway also leads to higher overall levels of mRNA targets [10, 23]. Therefore, the current dogma acknowledges that both translational repression and mRNA degradation are present in plants and animals [3, 10, 13-18, 24]. These effects on mRNA levels are also detectable in gene expression data [10, 17].

Degradation of mRNAs by miRs can be detected in high-throughput expression data and can be successfully applied to enrich the many possible interactions suggested by sequence based prediction algorithms [25], e.g. TargetScan[c] [26], Miranda[d] [27] or PicTar[e] [28]. In this paper a novel method (miMsg) will be presented to enrich predicted miR-mRNA interactions within a specific biological context. The algorithm uses matched high throughput miR and mRNA expression data of biological samples. The miMsg algorithm is based on relative ranking profiles of the expression data and does not directly assess the absolute expression levels. Therefore, the algorithm can be applied to detect small effects of miRs on mRNA, inferring power by detecting patterns over multiple biological samples. In recent literature, various other computational methods have been applied to matched mRNA and miR expression data including regression, correlation analysis and Bayesian learning algorithms (detailed comparison and review of previously published related methods is provided in Supplementary data A and [25]). In comparison, miMsg is independent of the methods used to measure expression levels and the statistical properties of the data. It also only needs a minimal number of user-defined parameters and has built-in quality control. Finally, it yields a result which is directly related to the observed

---

[b] http://www.mirbase.org/
[c] http://www.targetscan.org/
[d] http://www.microrna.org/
[e] http://pictar.mdc-berlin.de/

expression patterns and also supplies a visualization tool for individual interactions. miMsg is extensively benchmarked against other methods using sixteen different (sub)sets of cancer related expression data. Benchmarking shows that non-parametric correlation methods are inferior to the other algorithms tested. For the remaining algorithms, results vary strongly depending on the dataset and overlap is minimal, identifying miMsg as a valuable alternative. As a proof of principle, miMsg is applied to germ cell tumor data, investigating mRNA degradation as most miRs have an inhibitory effect on mRNA/protein levels. In this deregulated (cancer) context, miMsg identified validated, relevant interactions and miRs/mRNAs.

System and methods

Biological samples, expression profiling, data normalization and statistics.

The following matched germ cell tumor (GCT) samples were included: five spermatocytic seminomas (SS, type III GCT), three seminomas (SE, type II GCT) and three embryonal carcinomas (EC, type II GCT) [29]. mRNA expression data (Human Genome U133 Plus 2.0 Array) was acquired, pre-processed and normalized as described in [30]. Matched miR expression data (multiplex qPCR, Applied Biosystems) was acquired as described in Supplementary Data M. Raw Ct values were normalized to RNU6b (results main text, Table S1) or their global average ([31], Table S2, significant overlap with RNU6b based normalization (30% overlap, $p<0.01$ based on simulation)). Unless specified otherwise, a t-test was used to compare groups (Excel 2010 / Matlab 2012a).

Annotation and target prediction matching

mRNA target predictions were downloaded from TargetScan's[f] website (conserved miR families). The most recent annotation files for the Affymetrix GeneChip HG-U133-Plus2 were downloaded from the manufacturer's website [g]. Data was carefully parsed to fit the miMsg format (Supplementary data, B). miR annotations were split and matched on miR family, class and member separately to prevent missed matches. Gene symbols were used to merge Affymetrix probe annotations with TargetScan's predictions. Only predictions for humans were included (speciesID=9606) and only probes marked as specific for one transcript were selected (suffix "_at"). In the end, 37,956 mRNA probes and 293 miRs were available for analysis. Of these 5,919 mRNA probes and 148 miRs were predicted to interact (inhibition, 56,262 interactions, Table 1).

---

[f] http://www.targetscan.org/
[g] http://www.affymetrix.com/

Algorithm

miMsg uses miRNA / mRNA expression and publicly available databases with predicted interactions to enrich miR-mRNA interactions within a specific biological context. The miMsg algorithm is based on the dogma that low mRNA target levels as a result of degradation are detectable in real targets of a specific miR and are absent in non-interacting predicted targets. miMsg is preeminently applicable to miR-mRNA interactions because the relevance of miR-mRNA interactions in a biologically coherent set of samples depends on how the samples interrelate (i.e. ranking) rather than individual expression levels. In an ideal interaction profile high expression of a miR in a tissue should be coupled with low mRNA expression. By observing this over multiple tissues, miMsg creates a combined ranking profile of miR/mRNA expression for each interaction (Relative Ranking, RR, Figure 1). The output of the miMsg algorithm is straightforward and focuses on biological interpretation. It detects interactions with similar or inverse RR depending on the predicted functional relation. The algorithm computes only three scores for each interaction. These are intuitively related to the observed RR and describe the quality and significance of the deviation from the ideal interaction profile. These scores are used to identify relevant and significant patterns, which can be further studied.
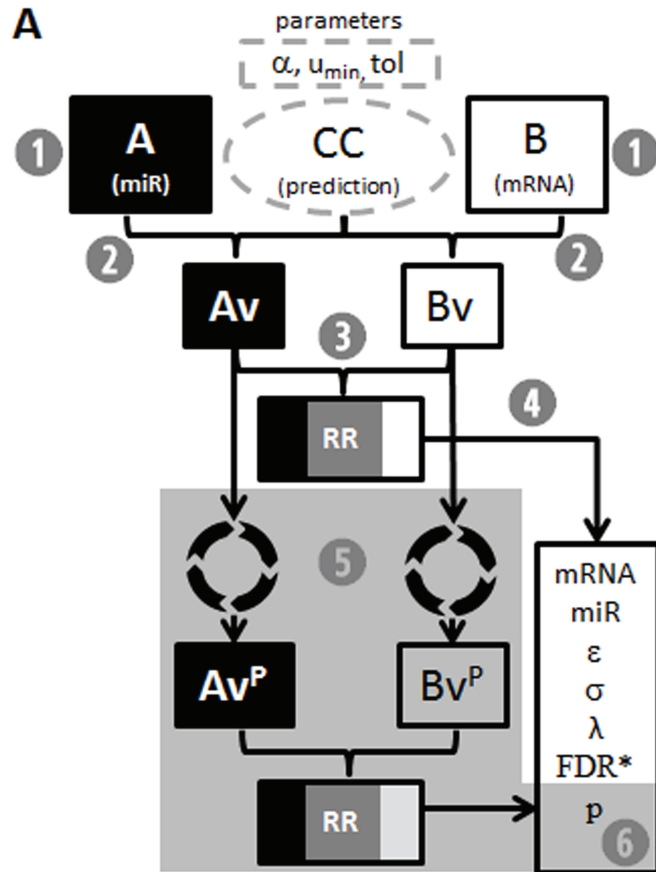
Definitions

As explained in Figure 1A, the miMsg algorithm uses two matched sets of expression data: A and B. CC contains the hypothesized interactions between A and B and their direction (inhibition for miR-mRNA interactions). The relative ranking of the expression levels is presented in RR_j for each interaction j and used to calculate the following quantities: (1) ε_j as an estimate for the number of misplacements in RR_j (Definition 1), (2) σ_j as a measure for the magnitude of these misplacements (Definition 2) and (3) λ_j (effect size) which combines the ε_j and σ_j (Definition 3). Also, associated levels of (empirically derived) significance and false discovery rate (FDR) are presented. Next, these quantities are formally defined (also see: Supplementary Data B-E). Let n denote the total number of interactions between the two datasets and m the total number of biological samples per interaction. Furthermore, let j=1,...,n denote the interaction and i=1,...,m the biological sample.

Definition 1: ε_j (relative number of misplacements)

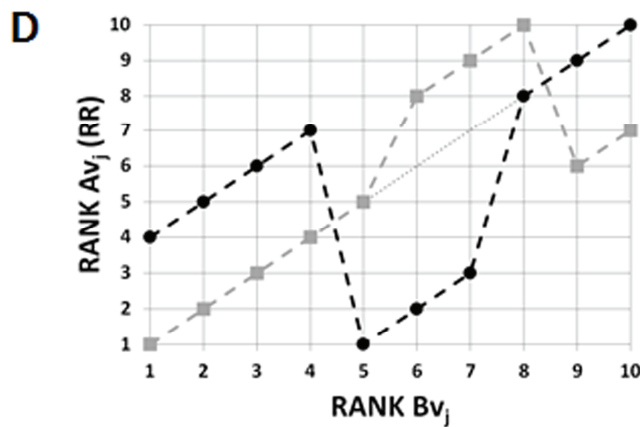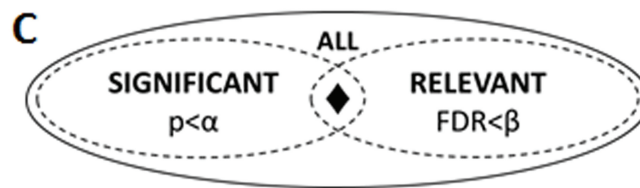*The relative number of misplacements εj is defined as follows:*

$$\varepsilon_j = \frac{\psi_j}{(m-1)} \in [0\ 1]$$

where ψ_j = 0,1,...,m-1 denotes the number of misplacements in the RR of interaction j involving m samples. A misplacement is defined as a difference between adjacent ranks being unequal to 1, i.e. ordering differs from ideal profile. (Supplementary Data, D)

← Figure 1. Flowchart of miMsg. **(A)** (User defined parameters) α, the desired level of significance (usually α ≤ 0.05); u_{min}, the minimal number of unique samples required in an interaction to qualify for analysis (u_{min} ≥ 3); tol, the accepted variation at convergence of the empirically derived cumulative distribution function (cdf) of (ε,σ) / λ (default: ≤ 10^{-4}). The following steps are executed: (1) Define two sets of measurements: A and B. Columns represent matched biological samples and rows contain measurements for miRs (A) or mRNAs (B). (2) A and B are linked using table CC containing predicted miR-mRNA interactions. This results in sets Av and Bv. (3) Using Av and Bv, the relative ranking (RR) for each interaction is calculated. This process is depicted graphically in Figure 1B. (4) Using RR, the measures ε, σ and λ (Definition 1-3) are calculated, describing the deviation of RR from the ideal interaction profile for each interaction. (5) An empirical 2-dimensional cdf for (ε,σ) is computed using iterative permuted versions of Av and Bv (Av^p and Bv^p, gray area). This cdf is then used to derive the significance of the scores for each interaction. (6) λ, the associated level of significance and the FDR (Definition 3-5) are used to identify biologically relevant and significant interactions. *An empirically derived cdf of λ is generated by permuting A and B. Using this cdf, the false discovery rate (FDR) of λ is calculated for all significant interactions to determine which values of λ are sufficiently low and thus represent relevant interaction profiles. **(B)** Example of the calculation of RR. 5 tissues (T1-T5) are analyzed for a single interaction j (hypothesis= inhibition). Values in A and B are equal to their rank x 100. Av_j/Bv_j are sorted according to Bv_j. The ranks of Avj in this sorted set constitute RR_j and are used to compute $ε_j$, $σ_j$, and $λ_j$: $Ψ_j$={3-2=1; 4-3=1; 1-4≠1; 5-1≠1}=2, $ε_j$=2/(5-1)=0.5; $σ_j$=(|3-2-1|+|4-3-1|+|1-4-1|+|5-1-1|)/ceil(5^2/2)=7/13, $λ_j$=0.5+7/13-(0.5*(7/13))=0.8 (Supplementary data B-G). **(C)** Graphical representation of the selection of interaction patterns that are both significant and relevant (♦). ALL = all interactions. Areas are not to scale. **(D)** Examples of two RRs (10 tissues) with larger (dashed black) or smaller (dashed, gray) displacements. Dotted gray line represents the ideal RR. $ε_{grey}$=$ε_{black}$=2/(10-1)=0.22; $σ_{grey}$ (|8-5-1|+|6-10-1|)/ceil(10^2/2)=7/50=0.14→ $λ_{grey}$=0.3229; $σ_{black}$=(|1-7-1|+|8-3-1|)/ceil(10^2/2)=11/50=0.22 →$λ_{black}$=0.3616. This illustrates definitions 2 and 3: larger displacements are penalized more severely. It also shows that miMsg is tailored to detect linear patterns with multiple "line segments", only penalizing discontinuities at their boundaries. Non-parametric correlation methods (Spearman & Kendall) identify the whole discontinuity as a disturbance of the overall pattern and repeatedly penalize all ranks involved (Supplementary data L). Because groups of biological samples can behave differently but still correlate well per group, this leads to underappreciated patterns.

The relative number of misplacements alone is an incomplete measure for how well the profile in RR_j correlates to the ideal profile. The magnitude of the difference between adjacent ranks is important as well. Small differences between adjacent ranks (small misplacements) are considered less serious than larger ones (Figure 1D). Therefore, consider the following measure for the magnitude of the misplacements in RR_j.

Definition 2: $σ_j$ (magnitude of misplacements)

The relative magnitude of the misplacements $σ_j$, measured by its relative variance, is defined as follows:

$$\sigma_j = \frac{\sum_{i=1}^{m-1}\left|RR_{j,i+1} - RR_{j,i} - 1\right|}{floor\left(\frac{m^2}{2}\right)} \in [0\ 1]$$

where $RR_{j,i}$ denotes the ith biological sample in the jth interaction and floor(m^2/2) is the maximum obtainable variance in a RR containing m biological samples.

To quantify the combined effect of the number and magnitude of the misplacements, a combined score (effect size) is defined.

Definition 3: λ_j (combined effect size)

The number and magnitude of misplacements (Definition 1,2) are combined in λ_j using:

$$\lambda_j = \varepsilon_j + \sigma_j - (\varepsilon_j \cdot \sigma_j) \in [0\ 1]$$

λ is a monotonically decreasing function of ε and σ which approaches zero if and only if both ε and σ approach zero (ε:[0 1] and σ:[0 1]). It is a measure for the combined effect size of ε and σ and is used to easily rank selected interactions. Low values of λ indicate a better match to the ideal RR (Figure 1D).

Finally, a statistical test is required to show relevance and statistical significance of the scores.

Significance & relevance

The final steps of the algorithm determine (1) which interaction patterns are statistically unlikely to occur by chance and (2) which values of λ should be considered relevant (=sufficiently low) (Figure 1C). These independent steps will be discussed below.

Significance

In the preceding computations, the absolute values of ε, σ and λ that result for each interaction may be coincidental. Hence, a measure of statistical significance is required. As no a priori knowledge about the statistical properties of the data is present, an empirically derived, 2-dimensional cumulative distribution function (cdf) is generated to assess the probability of coincidental (ε, σ) combinations. This cdf is iteratively approximated using permuted (shuffled) versions of the original expression data. After each iteration, the cdf is evaluated at all possible combinations (ε_k, σ_m) present in the real data, k=1,2...,K and m=1,2...,M. This yields a cdf $\rho(\varepsilon, \sigma)$ that can be compared between iterations at (ε, σ). The process of adding (ε, σ) combinations from permuted data is repeated until the difference between the current and previous iteration is neglectable (i.e. < tol). This difference is defined as the relative root mean square error (rrε):

$$rr\varepsilon_i = \sqrt{\frac{\sum_{k=1}^{K}\sum_{m=1}^{M}\left(\rho_i(\varepsilon_k, \sigma_m) - \rho_{i-1}(\varepsilon_k, \sigma_m)\right)^2}{\sum_{k=1}^{K}\sum_{m=1}^{M}\rho_i^2(\varepsilon_k, \sigma_m)}}$$

After convergence, the resulting $\rho_{def}(\varepsilon, \sigma)$ is evaluated at all combinations from the real data (ε_k, σ_m) to assess their respective p-values (Supplementary Data, E). $\rho_{def}(\varepsilon, \sigma)$ is uniquely generated for each group of interactions with the same number of unique biological samples as this number defines the theoretical variability of ε and σ (i.e. the shape of $\rho_{def}(\varepsilon, \sigma)$). (Supplementary Data, F for more information about repetitive values.) Formally, an interaction is deemed significant if the following holds.

Definition 4: statistical significance

Consider the jth interaction, the corresponding measures ε_j and σ_j (Definition 1, 2) and a user defined level of significance α. Moreover, consider the cdf $\rho_{def}(\varepsilon,\sigma)$ obtained using the procedure discussed in section Significance. Then, the jth interaction is defined to be significant if $\rho_{def}(\varepsilon,\sigma) \leq \alpha$.

Relevance

Apart from the significance of an interaction pattern, it is required to determine which values of λ should be considered low enough (i.e. relevant). High values of λ might still be statistically significant, but do not constitute relevant interaction patterns. By iteratively using permuted expression data as input for miMsg, a cdf of λ is generated, denoted by $\tau(\lambda)$ (analogous to the section Significance). Using τ, the FDR for all significant values of λ is determined. The researcher can determine a maximal acceptable FDR (β, usually $\leq$ 1% / 5%). See also Supplementary Data, E. Formally, an interaction is deemed relevant if the following holds.

Definition 5: relevance of the size of λ

Consider the jth interaction, the corresponding measure λ_j (Definition 3) and a user-defined maximal acceptable FDR β. Moreover, consider the cdf $\tau(\lambda)$ obtained using the procedure discussed in the section Relevance. Then, the jth interaction is defined to be relevant if $\tau_{def}(\lambda) \leq \beta$.

Output

The final output of miMsg is a table with all significant interactions and associated values of ε, σ and λ as well as the associated p-values (example: Table S1) and FDRs. Also, an overview of the raw expression levels is supplied. A detailed runfile with settings and properties of the input/output is generated as well as a visualization of the quality of the results. An additional tool is supplied to visualize individual significant interactions (example: Figure 2, see also Supplementary data G).

Table 1. Number of miRs/mRNAs in miMsg results (GCT[a]). [a] parameters: α=0.05, $u_{min} \geq 3$, tol=10^{-4} [b] miR-29a was measured twice using two independent primers on the qPCR. [c] validated = present in miRTarBase as validated interaction.

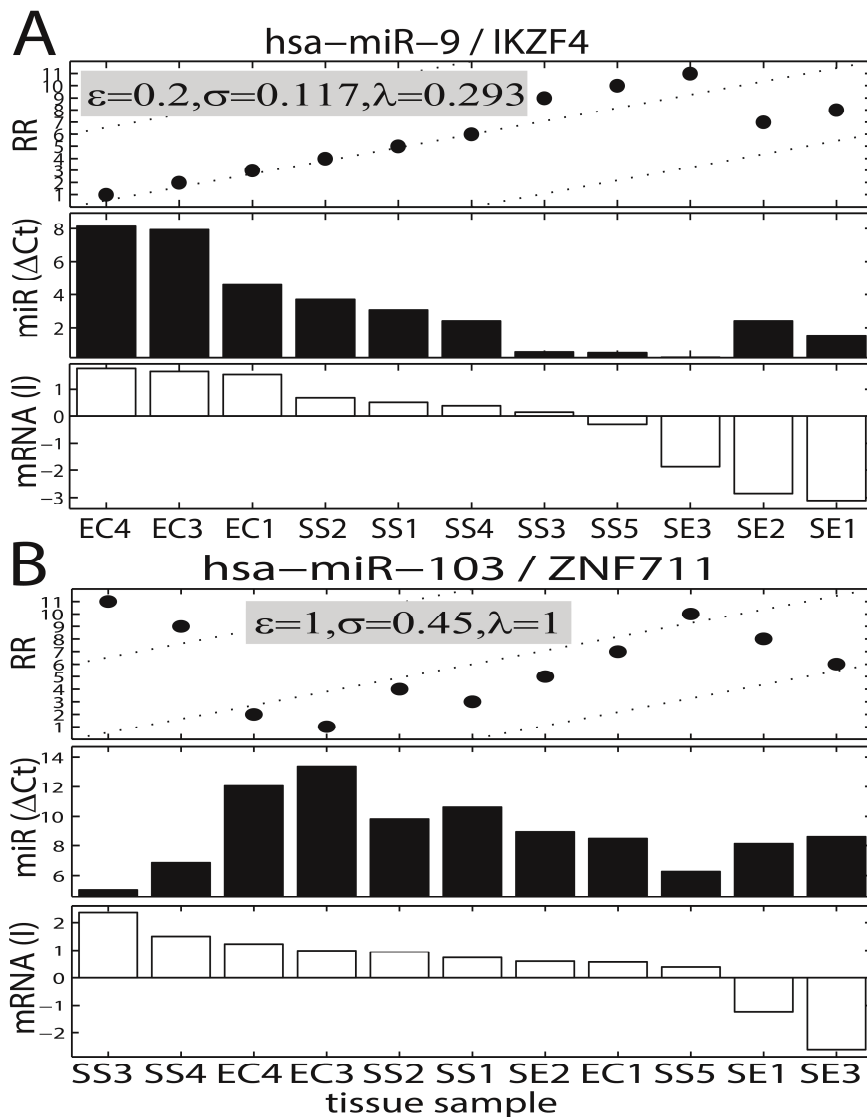| (Sub)set	miRs [b]	Probes (genes)	Unique interactions based on probes (Gene Symbol)	Validated [c]
All interactions	148	5,919 (2,762)	56,262 (23,429)	170
Significant interactions	146	2,447 (1,521)	6,093 (5,137)	51
Top-scoring (FDR 5%, $\lambda \leq 0.8750$)	145	1,891 (1,248)	3,473 (3,176)	39
Top scoring (FDR 1%, $\lambda \leq 0.8000$)	137	682 (560)	921 (896)	10
Top-100 (FDR 0.1%, $\lambda \leq 0.6867$)	56	93 (90)	100 (99)	2

Figure 2. Visual representation of the relative rankings (RR) and miR/mRNA expression levels. **(A)** Top-scoring significant interaction. **(B)** Worst scoring significant interaction. (RR-plot) The RR is plotted on the y-axis (see also: Figure 1B and D). A straight ascending line is expected when there is a perfect inverse relation between miR and mRNA expression. High values of λ (B) are associated with a noisier RR, not necessarily with inverse ranking patterns. **(miR/mRNA expression plots)** Bars represent relative mRNA levels (I, mean centered intensity) or miR expression (ΔCt values, high value=low expression). Ideally, identical descending patterns would be observed. X-axis: biological samples. Suffix of sample name is arbitrary.

Application & Benchmarking

A selection of histologically diverse GCTs was analyzed to illustrate the applicability and functionality of miMsg. Firstly, the performance of miMsg was studied by assessing the response to permuted or random data. Secondly, a selection of top-scoring and highly significant interactions was further investigated. Finally, extensive benchmarking was done including 16 datasets / subsets and 5 different algorithms for the selection of miR-mRNA interactions.

Germ cell tumor data

Response to permuted / random data and noise

To assess the robustness of the miMsg-algorithm and the relevance of the patterns detected, various tests with random, permuted and modified data were run. When compared to 800 runs with permuted or random expression data, miMsg identified significantly more and better scoring interactions in the real, biologically relevant data ($p<0.01$). This suggests detection of relevant patterns in biological data (Supplementary data, H). Replacing a fraction of the predicted interactions with random interactions did not prevent miMsg from identifying the originally top-scoring interactions. Moreover, random or targeted reduction of the dataset did not change miMsg's identification and grading of originally selected interactions. This shows that miMsg is relatively insensitive to (introduced) noise and the number of predicted targets (Supplementary data I).

Detection, selection & verification of interactions

High quality RRs were identified by both a low p-value (significant) and a low value of λ (relevant profile, Figure 2). A selection of 6,093 significant interactions was identified from the original pool of 56,262 predicted interactions (Table S1). Top-scoring sets with 1% / 5% FDR and a top-100 (FDR 0.1%) were defined based on the accumulated data of 400 runs with permuted expression data (Table 1). These top-scoring and highly significant subsets were matched with miRTarBase[h] to identify experimentally validated interactions. miRTarBase is a literature-based database of 3,969 experimentally validated interactions between 625 miRs and 2,433 genes in 14 species (version: April 15, 2011) [32]. Only human interactions were included (n=2,817). The fraction of validated interactions in all predicted interactions was relatively small as most of the predicted miR-mRNA interactions used as input are not yet experimentally validated. However, enrichment of validated interactions (%) in the selected subset was of interest as a quality measure (rather than the a priori % of validated interactions). Still, enrichment might not be present because interactions validated in non-GCT systems are not necessarily active in GCTs. However, from all algorithms tested in the benchmarking, miMsg showed the highest enrichment for validated interactions in the GCT dataset (37%, Supplementary data L: Figure 22).

---

[h] http://mirtarbase.mbc.nctu.edu.tw/

Benchmarking

miMsg was compared to existing algorithms for miR-mRNA interaction analysis using common non-parametric correlation methods (Spearman & Kendall), a method using Bayesian learning (GenmiR++ [33]) and a LASSO based regression algorithm [34] (Supplementary Data A). Sixteen dataset/subsets were investigated. Three were obtained from the original LASSO publication: Madison (nasopharyngeal cancer), Broad (various types and normal tissues, also used in the GenmiR publication) and MSKCC (prostate cancer). The other datasets included parallel miR-mRNA data from different genetic subtypes of multiple myeloma (MM, [35]) and our primary GCT dataset. Several subsets of the GCT and MM sets were analyzed. To start with, random subsets of interactions were constructed that were comparable to the (LASSO) datasets from [34] with regard to the number of (validated) interactions. Also, biologically coherent subsets were created including only validated interactions or only genes differentially expressed between tumor types (analogous to [34]).
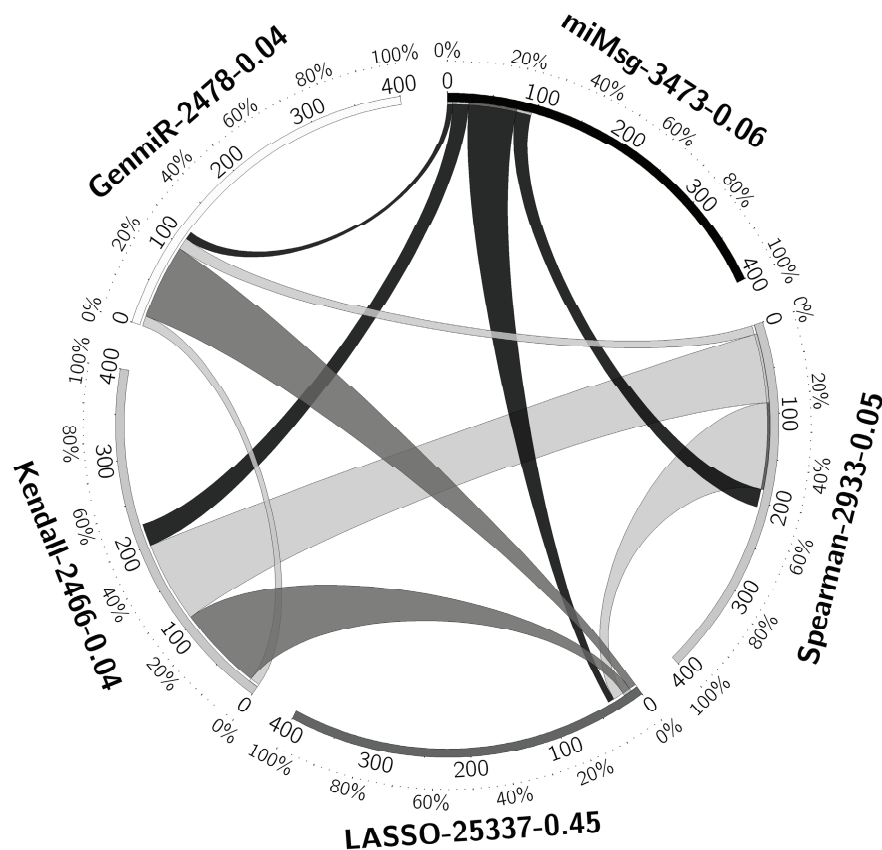

Figure 3. Overlap between interactions selected by the various algorithms (GCT: 56,262 interactions). Maximum total overlap between 5 algorithms is (5-1)x100%=400% (inner ticks). Outer ticks identify the % of this maximal overlap that is realized for a specific algorithm / dataset. The width of the ribbon identifies the overlap between two algorithms as a % of the number of selected interactions by each algorithm separately. Algorithm labels: "name – total nr. of selected interactions - % of all interactions selected." (Example of interpretation) Overall, miMsg shows ≈25% overlap with all other methods. Consider the ribbon from miMsg to LASSO: Of the 3,473 interactions selected by miMsg, 54% overlaps with the selection of the LASSO algorithm. On the hand, this overlap with miMsg constitutes only 7.5% of the 25,337 interactions selected by LASSO. Figure generated using Circos [36].

Algorithms were benchmarked based on several criteria (Table 2, Figure 3). For all algorithms, highly variable performance and very little overlap between results were observed between datasets. Overall, common non-parametric correlation methods (Spearman/Kendall) are clearly outperformed by miMsg, LASSO and GenmiR (also see Figure 1D). Between miMsg, LASSO and GenmiR there is no obvious best choice. Although we favor miMsg because of the limited number of interactions selected and the unbiased, distribution-free approach, application of different algorithms can aid the selection of relevant interactions. (Table 2, Figure 3, Benchmarking in detail: Supplementary data, L)

Table 2. Results of benchmarking. Supplementary data, L describes the benchmarking in detail.

Criterion	Motivation/Result
Fraction of interactions selected & enrichment for validated interactions	**Motivation:** Selection of a high fraction (=high number) of interactions is not feasible when applying the results to research. In addition, if validated interactions are active in the studied tissues, there should be a higher fraction of validated interactions in the selected interactions as compared to the complete dataset. **Result:** Different algorithms select quite constant fractions from almost all datasets. LASSO selects ca. 50% of all interactions; the other algorithms select ≈ 5%. Enrichment for validated interactions varies greatly between datasets with no clear trend toward one optimal algorithm (correlation = worst). In the very large datasets (GCT & MM), miMsg outperforms the other algorithms by achieving 37% and 33% enrichment for validated interactions ($\alpha \leq 0.05$, FDR≤ 0.05).
Performance • of all / selected interactions against permuted data • of (selected) validated interactions against permuted interactions	**Motivation:** Unpermuted data in general should perform better than random data. Also, validated interactions should outperform permuted data (ROC analysis). **Result:** The total set of (validated) interactions most likely contains a very high number of not interacting miR-mRNA pairs / validated interactions which are not active. These are not expected to perform better than random data. For most datasets, the algorithms therefore have difficulty outperforming permuted data when considering all (validated) interactions. GenmiR performs well on some datasets followed by miMsg; correlation analyses show the worst overall performance. When only the selected interactions are considered LASSO, miMsg and to a lesser extend GenmiR perform much better in assigning higher scores to unpermuted data and validated interactions.
Overlap between selected interactions	**Motivation:** Algorithms using different approaches, but identifying many of the same interactions might be more trustworthy **Results:** There is very limited overlap between the interactions selected by the various algorithms even when only very good scoring interactions (top-100 or less) are considered. Spearman & Kendall overlap ca. 100% because of high similarity in methodology.
Influence of dataset size and % of validated interactions	**Motivation:** The size of the dataset and the fraction of validated interactions might influence performance of the algorithms. **Results:** For all algorithms, performance does not consistently improve when using smaller, random subsets with a higher fraction of validated interactions.
Influence of pre-selecting interactions	**Motivation:** Pre-selection of relevant interactions based on (gene) expression / validated interactions might influence performance of the algorithms. **Results:** For all algorithms, performance does not consistently improve when using preselected / validated only sets of interactions.

Discussion

miMsg: considerations (input and application)

Predicted interactions. A sensible, trusted set of predicted interactions is a requirement to generate biological relevant output when using miMsg. Supplementary Data H illustrates the effect of random, non-sense interactions. Inverse expression patterns may occur by coincidence in expression data which is not permuted / randomized. miMsg might still identify top-scoring, significant interactions in these experiments. This effect disappears when expression data is permutated / randomized.

Biological samples. miMsg generates optimal results when using biological samples which are clearly related to the studied hypothesis and include multiple comparable as well as clinically/biologically diverse samples in roughly equally sized groups. More samples indicates a lower probability of coincidental RRs. Moreover, rankings become more sensible when the samples differ biologically/clinically (Figure 2A). However, clustering on biological/histological subtype may occur for many genes/miRs simultaneously, independent of miR/mRNA interaction. Multiple samples per cluster are required to reduce this non-specific effect.

Expression data. Adequate measurement precision of the data is of particular importance as miMsg relies heavily on ranking. Samples with identical expression levels for certain miRs/mRNAs due to rounding off will be excluded in the analysis of interactions involving these miRs/mRNAs (Supplementary Data, F). On the other hand, precision outside of the detection method's sensitivity / thresholds will not contribute to better results in miMsg.

Scope of the study, limitations & future plans. In this paper, miMsg is applied to mRNA inhibition by miRs. Although there is strong evidence that miRs influence mRNA levels, the combined effect of mRNA inhibition and translational repression without influencing mRNA levels [10] will not be detected by miMsg. Moreover, miMsg is applied to cancer related datasets. Cancer is a deregulated state in which interactions might be detected easier. Theoretically miMsg is also applicable to other datasets/directions (non-cancer, other than miR/mRNA interactions, stimulation). miMsg might also be used to create a reference database of miR-mRNA interaction signatures in different tissues / cells. These applications need to be investigated further.

Preliminary assessment of GCT results

miMsg was applied to expression data from a defined set of human GCTs. They represent the so called type II (seminoma (SE) and embryonal carcinoma (EC)) and type III (spermatocytic seminoma (SS)) subtypes. Despite a good response to systemic treatment, metastatic type II GCTs are a major cause of death in young adults. Both SE and EC originate from primordial germ cells (PGCs) / gonocytes. These cells have an intrinsic totipotent capacity and mimic totipotent ES cells. Type III GCTs originate from more differentiated germ cells and show no embryonic features [29, 37]. GCTs and their precursor lesions (either carcinoma in situ or gonadoblastoma) can be detected and classified by specific immunohistochemical markers and show distinct patterns / aberrations in DNA copy number and mRNA/miR expression [30, 38, 39].

Based on a systematic literature review top-scoring (top-100) and validated (FDR ≤ 5%) miRs/mRNAs associated with GCs/GCTs were shown to be involved in: (1) pluripotency / germ line development, (2) TP53 & TFGβ mediated apoptosis + therapy sensitivity, (3) fetal development (4) germ cell micro-environment, (4) proliferation and (5) carcinogenesis in general (extended review and references presented in Supplementary data K).

Moreover, miR-23ab and miR-27ab were highly enriched in the top-100 top-scoring subset (n=19/100, Figure 4). These miRs are part of the miR-23 paralog clusters: miR23a/27a/24-2 (cluster A, chromosome 19) and miR23b/27b/24-1 (cluster B, chromosome 9). The miRs in these clusters are not necessarily regulated identically, but are thought to be closely related based on their evolutionary conservation, sequence homology, genomic clustering and simultaneous expression level changes in human diseases including cancer [40, 41]. There was significant differential expression of these miR families, clusters and individual miRs between EC, SE and SS (p<0.05). The low expression in SE as compared to EC (both type II GCTs) is interesting because these tumor types differ in prognosis and biological behavior. Moreover, around half of the 19 top-scoring miR-23 paralog-targets were shown to be targeted by miRs in human ES cells [42]. (Also see Supplementary data K).

This overview does not imply proven functional relations, but shows that miMsg can be successfully applied to identify relevant targets for further study and validation in GCT-specific models.
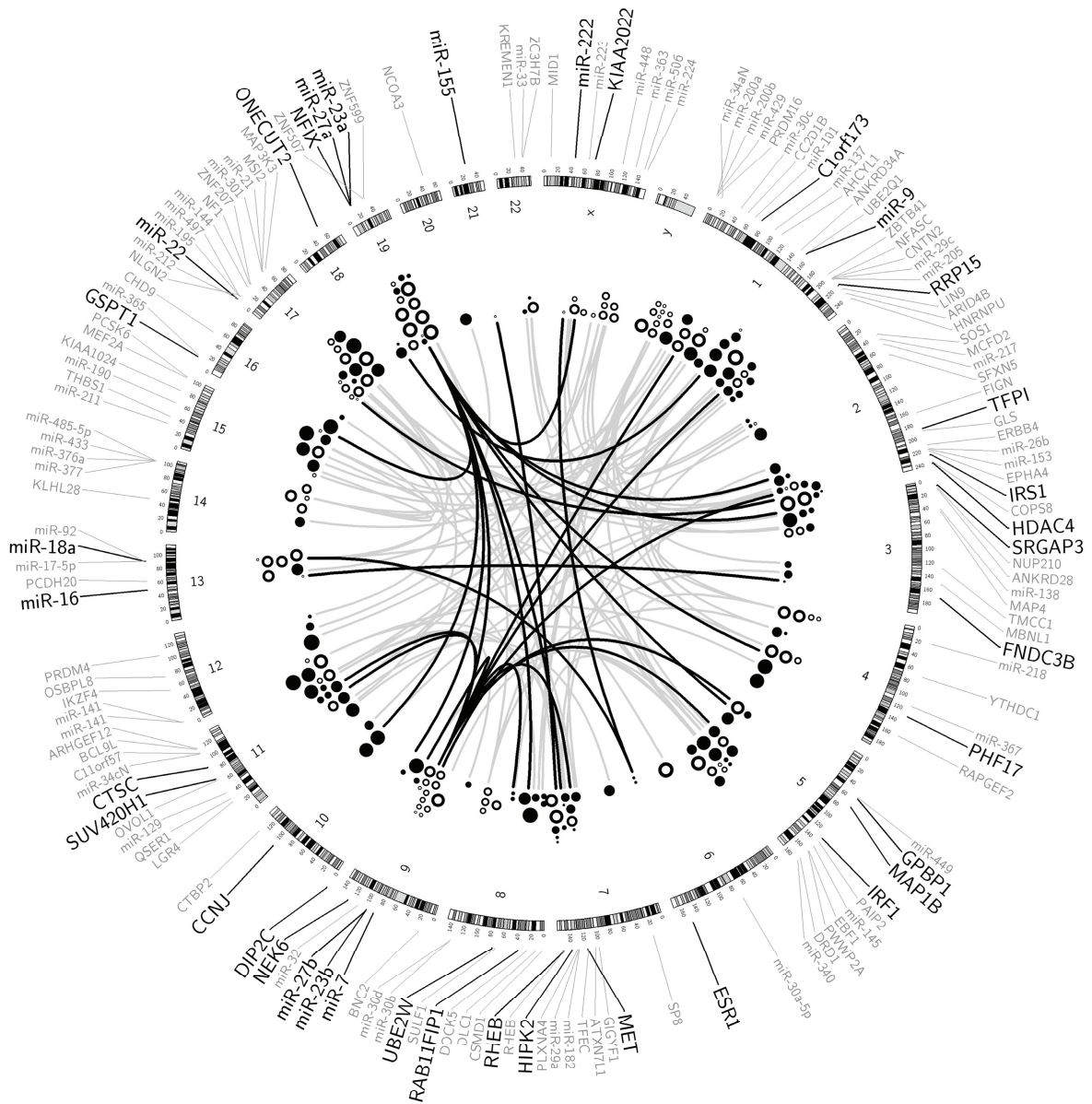
Figure 4. Visualization of top-scoring significant interactions in the GCT data. Circular representation of the human karyotype, generated using Circos [36]. Lines represent interactions between miRs and mRNAs. Black text/lines = miR-23ab/27ab cluster (top-100) or validated interactions in top-scoring significant interactions (FDR ≤1%). Gray text/lines= other miRs/mRNAs (top-100). Circles represent miRs (open circles) and targets (mRNAs, closed circles). Larger circles indicate better scoring interactions. Depending on the local density of the plot and the available space, the location of the circles cannot always exactly correspond to the genomic position of the miR/mRNA. Based on the known functions of the miRs/mRNAs, four top-scoring interactions were identified as most important. (*PHF17*/miR-23a) *PHF17* promotes apoptosis, is a known renal tumor suppressor and associates with OCT3/4 [43], a key protein in germ cell cancer [37]. Moreover, miR23a has been shown to function as a tumor suppressor-miR via repression by c-Myc. (*IRF1*/miR-23b) *IRF1* is involved in the pathogenesis of infertility in men with germ cell maturation arrest (validated in GCT cell line NT2) [44]. Moreover, miR-23b is associated with the hormone regulated process of spermiation in Sertoli cells [45]. (*IRS1*/miR-7) *IRS1* is present in peritubular myoid and interstitial cells [46], which migh suggest a role in the micro-environment of germ cells. miR-7 interferes with germ line differentiation in Drosophila [47]. (*HIPK2*/miR-27a) This interaction is validated in ovarian carcinoma cell lines when studying multidrug resistance [48]. miR-27a has also been shown to influence oncogenesis, proliferation and differentiation in various forms of cancer [40]. This interaction was present twice in the FDR ≤ 1% top set.

Conclusion

To conclude, the miMsg algorithm is an unbiased tool to enrich predicted miR-mRNA interactions for a specific biological context using matched high-throughput miR/mRNA expression data (Figure 1). As interactions are graded on scores directly related to combined ranking profiles of miR/mRNA expression, miMsg is very suitable to detect subtle miR-mRNA interactions (Figure 2). miMsg selected 0.18% / 1.64% / 6.17% (FDR 0.1% / 1% / 5%) highly relevant and significant interactions in the GCT data (Table 1, Figure 4). In the largest datasets (GCT/MM: 56,262/27,129 interactions) miMsg outperforms the other algorithms by achieving 37/33% enrichment for validated interactions. Overlap between interactions selected by different algorithms is very limited. Overall, the benchmarking results illustrate that there is no obvious best choice between miMsg, LASSO and GenmiR++. Spearman's and Kendall's correlation methods perform worst. Results depend strongly on the dataset used. Although we favor miMsg because of its limited fraction of selected interactions ($\approx$5% vs $\approx$50% (LASSO)) and unbiased, distribution free approach, application of different algorithms can aid selection of relevant interactions.

A systematic literature review of the GCT results showed association of top-scoring miRs/mRNAs with processes regulating germ cells (proliferation/differentiation) and GCT development. Moreover, interactions involving miR-23 paralogs were highly enriched in the top-100. 50% of the mRNAs involved in these interactions were shown to be miR-binding in ES cells. Based on: (1) top-100 ranking by miMsg / FDR $\leq$ 1%, (2) involvement of the enriched miR-23 paralog cluster, (3) known miR-mRNA interaction and/or (4) experimental validation, 28 high confidence interactions were selected. Literature review identified four of these as most relevant for further study (Figure 4). To conclude, miMsg was highly effective in reducing a high number of predicted miR-mRNA interactions in GCT data, generating a small high confidence set directly applicable to further research.

Acknowledgements

Supporting information

Additional review of literature, results an protocols:

A.   Review of bioinformatics methods to combine miR-mRNA data for target enrichment

B.   miMsg - Input & Parameters: meaning and format

C.   miMsg - Data transformation & calculation of RR

D.   miMsg - Computation and properties of ε, σ and λ

E.   miMsg – Significance & relevance

F.   miMsg - Handling of repetitive values

G.   miMsg - Output

H.   Permuted and random data / interactions (GCT data)

I.   Influence of dataset composition on selected top-scores (GCT data)

J.   Interactions in GCT datasets

K.   Biological interpretation (GCT data)

L.   Benchmarking

M.   48-plex TaqMan miRNA Assay Protocol v3

Tables S1/S2. GCT miR expression data (S1: RNU6b normalized, S2: global normalization)

Supporting information is available at: http://dx.doi.org/10.1093/bioinformatics/btt246

Source code and Matlab datasets (GCT and benchmarking data) are available at www.martinrijlaarsdam.nl/mimsg

References

1.   Pauli A, Rinn JL, Schier AF: **Non-coding RNAs as regulators of embryogenesis.** *Nature reviews* 2011, 12**:**136-149.
2.   Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS: **Non-coding RNAs: regulators of disease.** *J Pathol* 2010, 220**:**126-139.
3.   Carthew RW, Sontheimer EJ: **Origins and Mechanisms of miRNAs and siRNAs.** *Cell* 2009, 136**:**642-655.
4.   He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation.** *Nature reviews* 2004, 5**:**522-531.
5.   Mallory AC, Vaucheret H: **Functions of microRNAs and related small RNAs in plants.** *Nat Genet* 2006, 38 Suppl**:**S31-36.
6.   Krol J, Loedige I, Filipowicz W: **The widespread regulation of microRNA biogenesis, function and decay.** *Nature reviews* 2010, 11**:**597-610.
7.   Esquela-Kerscher A, Slack FJ: **Oncomirs - microRNAs with a role in cancer.** *Nat Rev Cancer* 2006, 6**:**259-269.
8.   Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Research* 2006, 34**:**D140-D144.
9.   Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, 136**:**215-233.
10.   Huntzinger E, Izaurralde E: **Gene silencing by microRNAs: contributions of translational repression and mRNA decay.** *Nature reviews* 2011, 12**:**99-110.
11.   Sobkowiak L, Szarzynska B, Szweykowska-Kulinska Z: **[Plant micro RNA biogenesis].** *Postepy biochemii* 2008, 54**:**308-316.
12.   Voinnet O: **Origin, biogenesis, and activity of plant microRNAs.** *Cell* 2009, 136**:**669-687.
13.   Eulalio A, Huntzinger E, Izaurralde E: **Getting to the root of miRNA-mediated gene silencing.** *Cell* 2008, 132**:**9-14.
14.   Filipowicz W, Bhattacharyya SN, Sonenberg N: **Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?** *Nature reviews* 2008, 9**:**102-114.
15.   Wu L, Belasco JG: **Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs.** *Mol Cell* 2008, 29**:**1-7.
16.   Huang JC, Morris QD, Frey BJ: **Bayesian inference of MicroRNA targets from sequence and expression data.** *J Comput Biol* 2007, 14**:**550-563.
17.   Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM: **Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.** *Nature* 2005, 433**:**769-773.
18.   Thomas M, Lieberman J, Lal A: **Desperately seeking microRNA targets.** *Nat Struct Mol Biol* 2010, 17**:**1169-1174.
19.   Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP: **The impact of microRNAs on protein output.** *Nature* 2008, 455**:**64-71.
20.   Guo H, Ingolia NT, Weissman JS, Bartel DP: **Mammalian microRNAs predominantly act to decrease target mRNA levels.** *Nature* 2010, 466**:**835-840.
21.   Hendrickson DG, Hogan DJ, McCullough HL, Myers JW, Herschlag D, Ferrell JE, Brown PO: **Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA.** *PLoS Biol* 2009, 7**:**e1000238.
22.   Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N: **Widespread changes in protein synthesis induced by microRNAs.** *Nature* 2008, 455**:**58-63.
23.   Schmitter D, Filkowski J, Sewer A, Pillai RS, Oakeley EJ, Zavolan M, Svoboda P, Filipowicz W: **Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells.** *Nucleic Acids Res* 2006, 34**:**4801-4815.
24.   Rajewsky N: **microRNA target predictions in animals.** *Nat Genet* 2006, 38 Suppl**:**S8-13.
25.   Barbato C, Arisi I, Frizzo ME, Brandi R, Da Sacco L, Masotti A: **Computational challenges in miRNA target predictions: to be or not to be a true target?** *J Biomed Biotechnol* 2009, 2009**:**803069.
26.   Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** *Cell* 2003, 115**:**787-798.

27. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS: **Human MicroRNA targets.** *PLoS biology* 2004, 2:e363.

28. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: **Combinatorial microRNA target predictions.** *Nature genetics* 2005, 37:495-500.

29. Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5:210-222.

30. Looijenga LH, Hersmus R, Gillis AJ, Pfundt R, Stoop HJ, van Gurp RJ, Veltman J, Beverloo HB, van Drunen E, van Kessel AG, 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, 66:290-302.

31. Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J: **A novel and universal method for microRNA RT-qPCR data normalization.** *Genome Biol* 2009, 10:R64.

32. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, et al: **miRTarBase: a database curates experimentally validated microRNA-target interactions.** *Nucleic Acids Res* 2011, 39:D163-169.

33. Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, Hughes TR, Blencowe BJ, Frey BJ, Morris QD: **Using expression profiling data to identify human microRNA targets.** *Nat Methods* 2007, 4:1045-1049.

34. Lu Y, Zhou Y, Qu W, Deng M, Zhang C: **A Lasso regression model for the construction of microRNA-target regulatory networks.** *Bioinformatics* 2011, 27:2406-2413.

35. Gutierrez NC, Sarasquete ME, Misiewicz-Krzeminska I, Delgado M, De Las Rivas J, Ticona FV, Ferminan E, Martin-Jimenez P, Chillon C, Risueno A, et al: **Deregulation of microRNA expression in the different genetic subtypes of multiple myeloma and correlation with gene expression profiling.** *Leukemia* 2010, 24:629-637.

36. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, 19:1639-1645.

37. Looijenga L: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J of Pathology* 2009, 217:146-162.

38. Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW, Sun Y, Chen C, Guenther S, Sherlock J, Veltman I, Baeten J, et al: **High-throughput microRNAome analysis in human germ cell tumours.** *J Pathol* 2007, 213:319-328.

39. Veltman I, Veltman J, Janssen I, Hulsbergen-van de Kaa C, Oosterhuis W, Schneider D, Stoop H, Gillis A, Zahn S, Looijenga L, et al: **Identification of recurrent chromosomal aberrations in germ cell tumors of neonates and infants using genomewide array-based comparative genomic hybridization.** *Genes Chromosomes Cancer* 2005, 43:367-376.

40. Chhabra R, Dubey R, Saini N: **Cooperative and individualistic functions of the microRNAs in the miR-23a~27a~24-2 cluster and its implication in human diseases.** *Mol Cancer* 2010, 9:232-248.

41. Tanzer A, Stadler PF: **Molecular evolution of a microRNA cluster.** *J Mol Biol* 2004, 339:327-335.

42. Lipchina I, Elkabetz Y, Hafner M, Sheridan R, Mihailovic A, Tuschl T, Sander C, Studer L, Betel D: **Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response.** *Genes Dev* 2011, 25:2173-2186.

43. Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J: **An expanded Oct4 interaction network: implications for stem cell biology, development, and disease.** *Cell stem cell* 2010, 6:382-395.

44. Lian J, Tian H, Liu L, Zhang XS, Li WQ, Deng YM, Yao GD, Yin MM, Sun F: **Downregulation of microRNA-383 is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation by targeting IRF1.** *Cell Death Dis* 2010, 1:e94.

45. Nicholls PK, Harrison CA, Walton KL, McLachlan RI, O'Donnell L, Stanton PG: **Hormonal regulation of sertoli cell micro-RNAs at spermiation.** *Endocrinology* 2011, 152:1670-1683.

46. Kokk K, Verajankorva E, Laato M, Wu XK, Tapfer H, Pollanen P: **Expression of insulin receptor substrates 1-3, glucose transporters GLUT-1-4, signal regulatory protein 1alpha, phosphatidylinositol 3-kinase and protein kinase B at the protein level in the human testis.** *Anat Sci Int* 2005, 80:91-96.

47. Pek JW, Lim AK, Kai T: **Drosophila maelstrom ensures proper germline stem cell lineage differentiation by repressing microRNA-7.** *Developmental cell* 2009, 17:417-424.

48. Li Z, Hu S, Wang J, Cai J, Xiao L, Yu L, Wang Z: **MiR-27a modulates MDR1/P-glycoprotein expression by targeting HIPK2 in human ovarian cancer cells.** *Gynecologic oncology* 2010, 119:125-130.

Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle

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

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

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 disease. Therefore, stable markers consistently identifying more GCC components, specifically the stem cell components SE and EC, are of interest. Expression of the embryonic stem cell miR-371-3 and miR-302/367 clusters in SE/EC/YS suggest possible application of these micro-RNAs as GCC tumor-markers. The TSmiR protocol constitutes a complete, quality-controlled pipeline for the detection of miRs in serum, based on magnetic bead-based purification and qPCR quantification. As a proof of principle, TSmiR was applied to five independent serum sample series including 80 GCCs, 47 controls, 11 matched pre/post orchidectomy samples and 12 no-GCC testicular masses. GCC serum samples showed a consistent, significant ($p \leq 0.00064$) increase of miR-371/372/373/367 levels. Analogous, serum levels returned to baseline after orchidectomy (stage-I disease). Moreover, there was a trend toward higher miR levels in patients with metastasis. These results imply suitability for diagnosis and follow-up. TSmiR showed an overall sensitivity of 98%, clearly outperforming the traditional serum markers AFP/hCG (36%/57%, sensitivity$_{AFP}$=3%/45%; sensitivity$_{hCG}$=62%/66%, SE/NS). TSmiR misclassified one tumor as a control. Serum AFP/hCG and TSmiR combined identified all T samples correctly. 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.

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, 9], NANOG [10], SOX2/SOX17 [11] and LIN28 [12, 13]). In addition, CIS/SE/EC/YS components, diagnosed at pediatric and adult age, express a specific set of embryonic stem cell related micro-RNAs (miRs), including the miR-371-372-373 (miR-371-3) and miR302a,b,c,d/367 (miR-302/367) clusters [14-17].

Depending on tumor stage and histology, pediatric and adult GCC (SE/EC/YS/CH) are effectively treatable by surgery, possibly followed by either irradiation and/or chemotherapy [18]. However, significant long term effects of the systemic treatment protocols have become evident [19, 20]. 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 for diagnosis, risk assessment and follow-up. Alpha Feto-Protein (AFP) is predominantly informative for YS and sporadically positive in EC and human Chorionic Gonadotrophine (hCG) is predominantly informative for CH with sporadic positivity in SE/EC [18]. LDH-1 has also been reported as a (less informative) serum marker [21]. So far, no consistent markers for the stem cell components SE and EC are available, which limits the use of serum markers 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 [22-24]. One study showed a decline to normal levels after

surgical removal of stage I GCC [23]. 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.

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. Using five independent sample series we show that TSmiR shows high sensitivity and specificity (GCC vs control). Also, the effect of metastasis / surgical removal of the tumor on miR levels, is investigated. This proof of concept indicates TSmiR as a potential additional tool for diagnosis & follow-up of GCC.
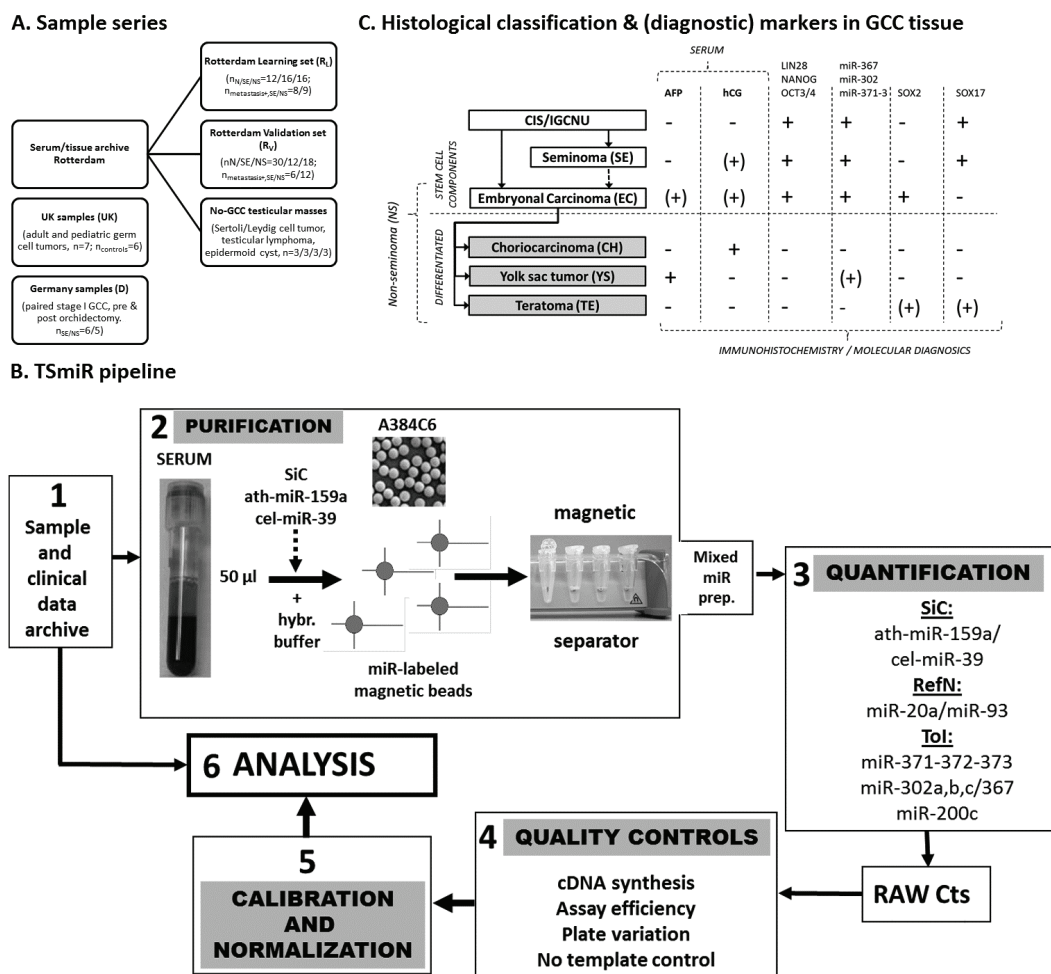

Figure 1. Sample series, histological classification / markers & TSmiR pipeline. (A) Sample series investigated. Five independent series were investigated: the Rotterdam-Learning (R_L) and Validation (R_V) sets including GCC cases and controls, a series of serum samples of patients with a testicular tumor other than GCC (no-GCC) and two previously published series: UK, including GCC samples (pediatric and adult) and controls [22, 24] and Germany (D), composed of pre- and post orchidectomy samples (local disease only, stage I) [23]. (B) Histological classification & overview of immunohistochemical, molecular and serum GCC markers [8, 10-17]. 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). (C) Complete pipeline of the TSmiR test. See Materials and Methods section for explanation. SiC=Spike in control; RefN=reference miRs/normalizers, ToI=Target of interest.

Materials and Methods

Patient and control serum samples

Detailed information on the composition of the various independent sample series (R_V/R_L/UK/D/no-GCC) is presented in Figure 1A. 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.

miR analysis of primary GCC and normal testis

Small RNA was prepared from 47 primary, independent GCC and three normal testis samples as described in [15]. Expression of 156 miRs was analyzed, quantified and normalized according to [25].

miRNA purification using anti-miRNA Magnetic Beads and quantification

For the purification, 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). This panel consists of superparamagnetic Dynabeads® covalently bound to a unique set of ~380 anti-miRNA oligonucleotides including endo-/exogenous controls, allowing for purification and quantification without pre-amplification step. 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 80×10^6 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 Micro-RNA 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 of all targets and samples 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 [28]) 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 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 [29]. Based on established algorithms for stability analysis (Normfinder [30] and geNorm [31]), miR-20a and miR-93 were the most stable combination for normalization in the R_L set after calibration, which was validated in the R_V set (Figure S3). Normalization was carried out using the mean of these normalizers according to the dCt method [32].

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 too stringent selection. 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 quality controlled pipeline for miR purification, recovery and quantification from serum (TSmiR) was developed as discussed extensively in the Materials and methods and Figure 1B. To ensure standardized results, the pipeline includes a number of quality control steps and spike in miRs to assess and compensate for differences in assay efficiency, inter-plate variability and cDNA/qPCR reaction efficiency. Regarding normalization, a panel of nine potential reference miRs was investigated, identifying miR-93 and miR-20a as the most stable combination for normalization in both a learning and a validation sample series (R_L, R_V, Figure 1A).

The miR-371-3 and miR-302/367 clusters are known to be expressed in SE, EC and YS (Figure 1C) [14-16]. In this study their serum levels were evaluated in five sample series (Figure 1A). First, the strength of the various miRs to distinguish between GCC samples (T) and controls (N) (or the presence/absence of metastasis) was investigated in the Rotterdam learning (R_L) series. These

results were verified in three independent series (Rotterdam validation (R_V, Erasmus MC Rotterdam Pathology), UK [24] and D [23]). Specificity of the miRs was verified in twelve serum samples from patients with testicular masses other than GCC (no-GCC). Next, The RL, Rv, UK and D series were combined to investigate the performance (sensitivity / specificity) of the most discriminative miRs. Finally, results from cases with and without metastasis and eleven matched stage I pre- and post orchidectomy serum samples (D series) were investigated to draw preliminary conclusions regarding the applicability of TSmiR to follow-up. Additional results, methods and raw data are presented in the supplementary data. By applying TSmiR to these sample series, the potential value in GCC diagnosis and follow-up of the serum levels of these GCC specific miRs was assessed in a controlled setup.

Serum miR-371-3/367 levels were significantly higher in tumor samples and were strongly correlated.

miR-371-3/367 showed significant high differences between tumor and control samples, also when stratified for histological subtype. These observations were consistent between the learning and validation sample series. miR-373 showed higher differences between N and T in the R_v as compared to the R_L series due to lower baseline levels in N samples and consistently higher levels in both the SE and the NS groups. Although there is no known difference between the samples in these series, this observation does not challenge the hypothesis that T samples exhibit higher miR levels than N samples. Even though miR-371 and miR-367 (uncorrected p-value) showed consistently higher levels in tumor versus normal, the results in the UK series were less consistent possibly due to the low number of samples and the inclusion of pediatric GCCs. miR-302a and to a lesser extend miR-302c showed significantly higher levels in in many comparisons of T vs N but the fold difference was lower than that of miR-371-3/367, limiting their usefulness (Figure 2AC). As expected miR-200c showed no increased levels in T vs N. There were no consistent significant differences between SE and NS. miR-371-3 and miR-302abc levels correlated most strongly within their own clusters but miR-367 correlated most strongly and consistently with the miR-371-3 cluster instead of its own miR-302abc cluster.

Figure 2. 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. ➔
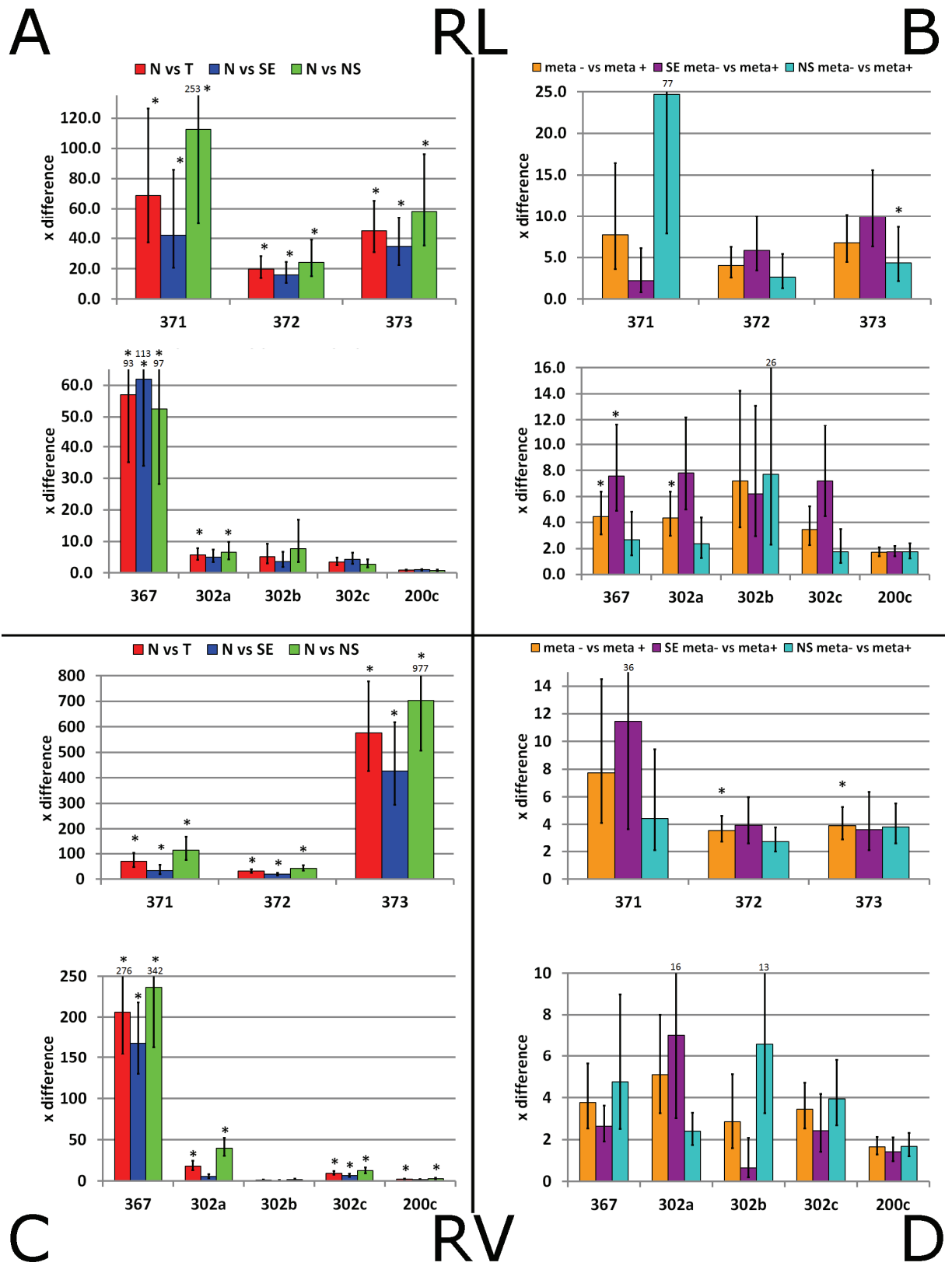
Combined serum miR-371-3/367 levels allowed for clear separation of tumor samples and controls.

Based on the consistent, significant and large fold differences in serum miR-371-3/367 levels between tumor and control samples, only these miRs were selected for further analysis. The power of these miRs to separate tumor from control samples was investigated using a Kohonen Self Organizing Map (SOM), principal component analysis (PCA) and hierarchical clustering / heat map visualization (Figures 3A, 3B and S4). The SOM was generated using the R_L data and then applied to the other datasets. The results of these different analyses were highly consistent. In the R_L series a tumor-only group and a second group containing all control and some tumor samples were distinguishable. Tumor and control samples were even more stringently separated in the R_V series. The UK series showed separation of most of the controls from the tumor samples and illustrated a trend in which the pediatric cases showed more similarity to the control samples than the adult GCC serum samples.

Serum levels of miR-371-3/302abc/367 were not increased in patients with testicular masses other than GCC.

A number of no-GCC testicular masses were investigated. Consistently, none of the miRs showed significant elevation in the serum compared to all pooled controls (N) samples. All no-GCC samples group together with the unaffected cases in PCA/SOM analysis. This confirmed the specificity of the investigated serum miR markers for GCC in the context of differentially diagnosing testicular masses (Figure S5).

Figure 3. (A) 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. R_L: In the R_L series, the right group only contained tumor samples, however the left group contained a mixed set of samples; R_V: 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; UK: 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. (B) Principal component analyses for all individual datasets. In all cases the first and second PC based on miR371-3 explained > 99% of the variance. 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; R_V: All T and N samples cluster separately. One N (testicular torsion) is situated on the edge of the T/N border; UK: all N samples and T samples cluster separately. The pediatric YS are situated between the N samples and the adult GCC clusters; D: all but two pre-orchidectomy samples cluster together. All post-orchidectomy samples cluster together. This cluster also included two pre-orchidectomy samples. →

A: Self organizing map
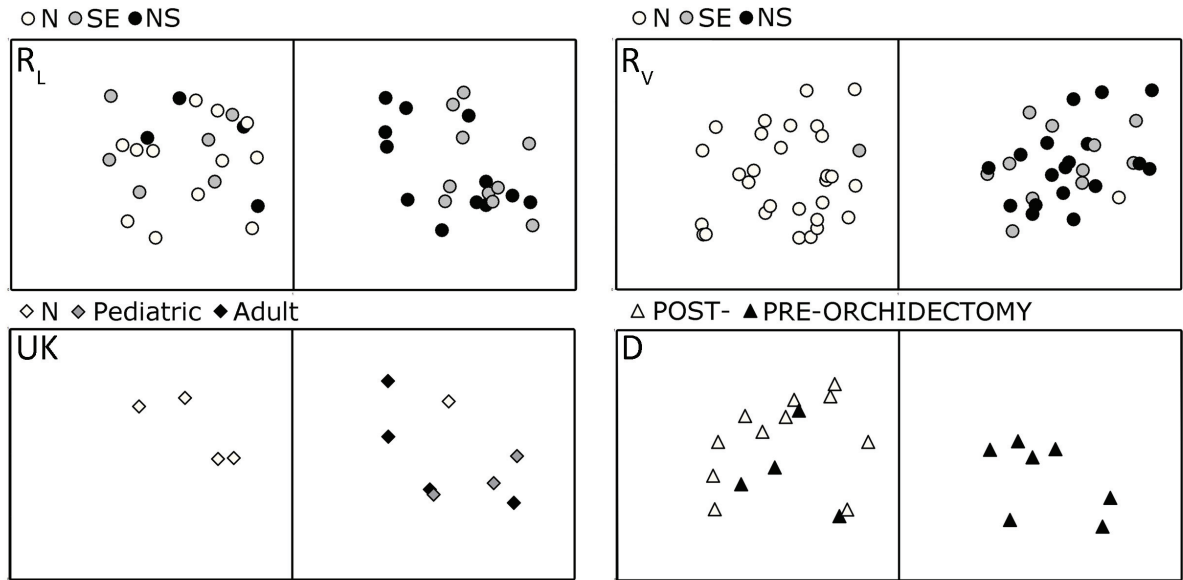
B: Principal Component Analysis
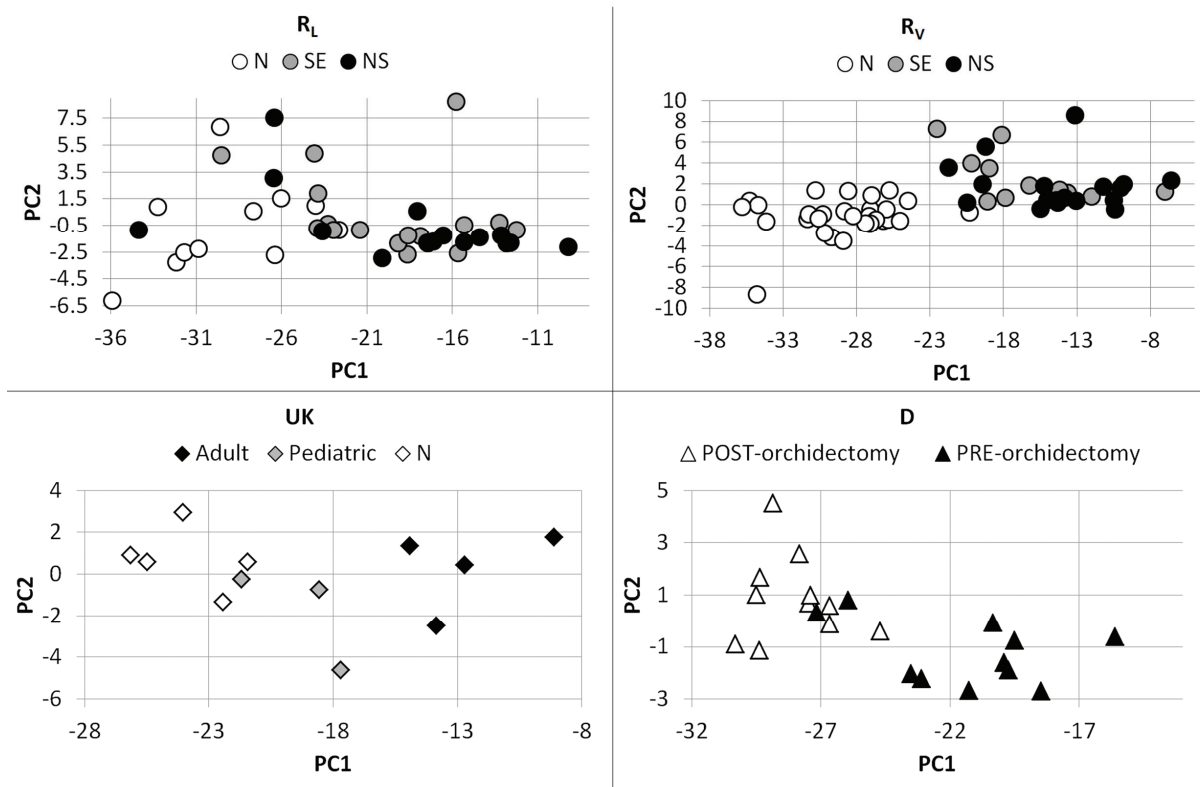
miR-371-3/367 levels in serum showed high sensitivity / specificity in the classification of tumor versus control samples and outperformed current markers (AFP/hCG).

The trade-off between specificity and sensitivity was assessed for all miRs by receiver operating curve (ROC) analysis using all samples (pooled). As expected, the discriminative power was very high for miR-371-3/367 (Figure 4A-E) and much lower for miR-302abc/200c. Classification/inclusion of the post-orchidectomy samples as part of the control group did not significantly influence the Area Under the Curve (AUC) and results only differed marginally when the SE or NS subgroups were analyzed separately. Cutoffs of miR levels were set for a sensitivity of 90%, preserving a specificity of 60-91% for individual miRs (Figure 4F). Combining the resulting classification (T/N) for miR-371-3/367 lead to a sensitivity of 98% (specificity$_{miR-371-3/367}$=48.3%). Addition of miR-302abc reduced the discriminative power and was therefore omitted (specificity$_{all\ miRs}$=8.6%).

Sensitivity was compared to markers currently used as golden standard (AFP and hCG, only available for tumor samples). miR-371-3/367 levels were more informative than AFP and hCG, particularly for SE (Figure 4C-E). Even when the YS subgroup was investigated specifically, investigation of the miR levels showed higher sensitivity than AFP. Overall, application of the serum markers AFP and hCG alone resulted in ≈50% false-negative cases. Investigation of only the miR-levels resulted in one false negative NS serum sample (sensitivity=98%). Combined interpretation of the established serum markers AFP/hCG and miR-371-3/367 levels led to the correct classification of all cases with a malignancy.
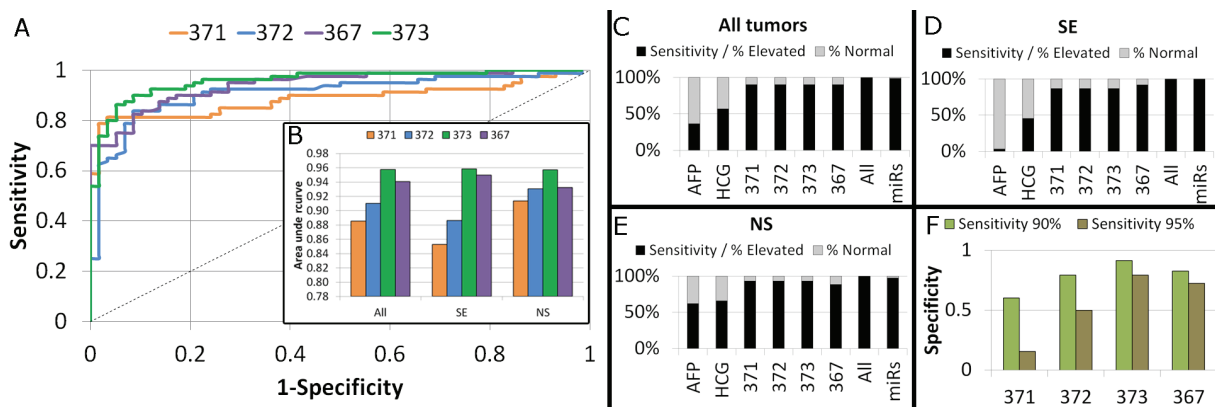

Figure 4. (A) ROC curves were generated for the four most differentiating miRs (371-3/367). Samples from the R$_L$/R$_V$/UK/D series were pooled for analysis.; (B) The AUC varied from 0.89 to 0.96 for all ROCs. Separate ROCs generated for only the SE or the NS showed minor differences between the quality 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% (green) in the ROC analysis lead to an acceptable remaining specificity (y-axis). A sensitivity of 95% (brown) 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.

miR levels returned to baseline after surgical removal of a local tumor & there was a trend towards higher GCC-specific miR levels in metastasized disease.

A series of matched sera from patients with local GCC was analyzed before and after orchidectomy to assess the reduction in serum miR levels after surgery. As expected, miR-371-3/367 showed significantly higher serum levels in the pre-orchidectomy samples. miR-302ac showed a similar trend without correction for multiple testing. SOM, PCA and clustering analysis illustrated that the majority of the pre- and post-orchidectomy cases could be clearly separated based on serum miR-371-3/367 levels (Figure 3), analogous to the results of the T vs N comparisons discussed above. After surgery serum miR-371-3/367 levels returned to baseline, i.e. they were not significantly different from the levels in all pooled controls (Figure 5A). Consistent with the observations described above, miR-371-3/367 showed an overall steeper reduction in serum levels than miR 302abc (Figure 5B).

When metastasized disease was compared to local GCC, miR-371-3/367/302ac showed overall higher levels in metastasized cases. However, the fold difference in serum miR levels was much lower than in the tumor vs control comparisons, results were not always consistent between the different sample series and the statistical strength was limited. miR-372-3/367/302a remained significant after correction for multiple testing (Figure 2BD).

These results indicate possible applicability of serum miR-371-3/367 levels not only for diagnosis as discussed in the previous sections, but also for follow-up. However, these observations need to be validated and extended in a larger, prospective setup with longer follow-up period.
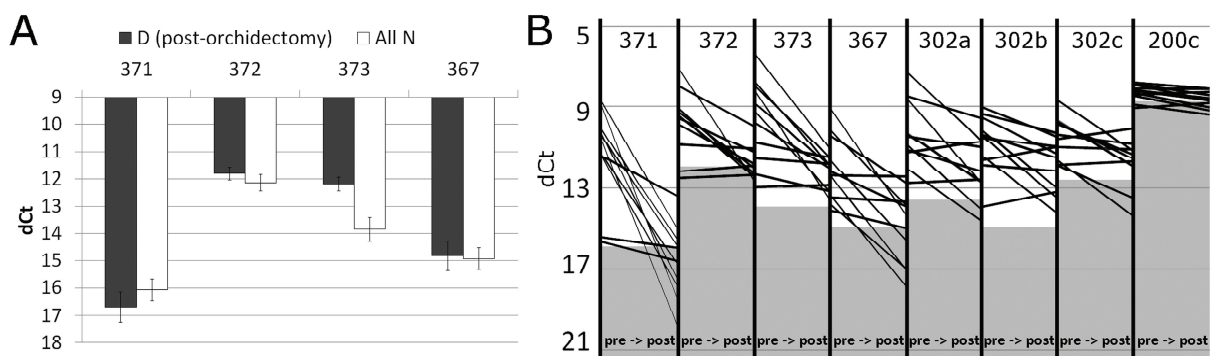
Figure 5. (A) When all N from the R_L, R_V and UK series are pooled and compared to the post-orchidectomy D samples, there is no significant difference between the serum-miR levels in both groups for the most differentiating miRs (371-3/367). X-axis: most consistently differentially present miR (between 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 4).

Discussion

GCC patients require intensive follow-up after primary diagnosis and treatment [18]. For diagnosis and follow-up, various approaches are currently used, including the established diagnostic/follow-up serum-markers AFP and hCG (Figure 1C). 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 GCC patients 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 [33] (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). They belong to the embryonic stem cell miR-371-3 and miR-302/367 clusters and were validated in multiple independent GCC series [14-16].

Levels of these miRs have proven to be detectable in the serum in the serum of GCC patients [22-24]. Although exosomes have been suggested as a miR-transporters / -repository, the mechanism behind miR release into serum remains to be elucidated [34, 35]. The protocols applied in earlier 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. The TSmiR protocol addresses these quality control issues, resulting in a robust and informative pipeline to detect the levels of these miRs in serum of GCC patients and controls (Figure 1B). Specific quality control steps for miR purification and recovery artifacts were implemented and suitable targets for normalization of Ct values in serum identified. 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 sets of independent serum samples from controls and GCC patients. miR-371-3 and miR-367 were most informative (Figure 2). At time of diagnosis, these specific serum miR levels allowed almost complete separation of control- and patient samples (Figure 3). 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 worse (max ≈60%), especially in the SE (max ≈40%) samples (Figure 4). 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 2B and D) and complete normalization of levels after complete resection of a local tumor (Figure 5) suggest an important role for TSmiR in follow-up as well as diagnosis (prospective setup currently initiated).

Conclusions

TSmiR constitutes a pipeline with stringent QC for the detection of miRs in serum, based on magnetic bead-based purification and qPCR quantification. In five independent GCC-sample series, TSmiR shows high sensitivity and specificity (GCC/control), especially for GCC stem cell components not consistently identified by current markers. Moreover, higher miR levels in patients with metastasis were identified as well as normalization of miR levels after surgical removal of local disease. 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

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. 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.

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_L 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.

Table S1. (A) Constitution of sample series & fold differences between groups. Numbers indicate times (x) difference between two group identified as "A vs B" in column headers, for example N vs T). Numbers >1 indicate relative higher levels in group B. Numbers between brackets indicate SEM (Standard Error of Mean), asymetric because of exponential transformation of ddCT 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). (B) P-values of pairewise comparisons.

Table S2. dCt values

Table S3. Correlation between miRs (R_L & R_V)

Supporting information is avaliable at: http://dx.doi.org/10.1016/j.molonc.2013.08.002

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, Oosterhuis JW, McLeod DG, Moller H: **Testicular germ cell tumors.** In *World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs.* Edited by Eble JN, Sauter G, Epstein JI, Sesterhann IA. Lyon: IARC Press; 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, Pukkala E, Cuninkova M, Moller H: **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, Bokemeyer C, Castrillon DH, Chris Lau YF, Looijenga LHJ, Oosterhuis JW: **Germ cell lineage differentiation in nonseminomatous germ cell tumors.** *J Pathol* 2006, 208:395-400.
8.   Looijenga LHJ, Stoop H, De Leeuw PJC, De Gouveia Brazao CA, Gillis AJM, Van Roozendaal KEP, Van Zoelen EJJ, Weber RFA, Wolffenbuttel KP, Van Dekken H, et al: **POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.** *Cancer Res* 2003, 63:2244-2250.
9.   Rijlaarsdam MA, van Herk HA, Gillis AJ, Stoop H, Jenster G, Martens J, van Leenders GJ, Dinjens W, Hoogland AM, Timmermans M, Looijenga LH: **Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours.** *Br J Cancer* 2011, 105:854-863.
10.   Hart AH, Hartley L, Parker K, Ibrahim M, Looijenga L, .H.J., Pauchnick M, C.W. C, Robb L: **The pluripotency homeobox gene NANOG is expressed in human germ cell tumors.** *Cancer* 2005, 104:2092-2098.
11.   Looijenga LH: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J Pathol* 2009, 218:146-162.
12.   Gillis AJ, Stoop H, Biermann K, van Gurp RJ, Swartzman E, Cribbes S, Ferlinz A, Shannon M, Oosterhuis JW, Looijenga LH: **Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis.** *International journal of andrology* 2011, 34:e160-174.
13.   Looijenga LHJ: **Human testicular (non)seminomatous germ cell tumors:Patho-biology and optimal diagnostics.** *J Pathol* 2008, 218:146-162.
14.   Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, et al: **A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.** *Cell* 2006, 124:1169-1181.
15.   Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW, Sun Y, Chen C, Guenther S, Sherlock J, Veltman I, Baeten J, et al: **High-throughput microRNAome analysis in human germ cell tumours.** *J Pathol* 2007, 213:319-328.
16.   Palmer RD, Murray MJ, Saini HK, van Dongen S, Abreu-Goodger C, Muralidhar B, Pett MR, Thornton CM, Nicholson JC, Enright AJ, 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.
17.   Novotny GW, Belling KC, Bramsen JB, Nielsen JE, Bork-Jensen J, Almstrup K, Sonne SB, Kjems J, Rajpert-De Meyts E, Leffers H: **MicroRNA expression profiling of carcinoma in situ cells of the testis.** *Endocr Relat Cancer* 2012, 19:365-379.
18.   Horwich A, Shipley J, Huddart R: **Testicular germ-cell cancer.** *Lancet* 2006, 367:754-765.
19.   Van Leeuwen FE, Stiggelbout AM, Van den Belt-Dusebout AW, Noyon R, Eliel MR, Van Kerkhoff EHM, Delemarre JFM, Somers R: **Second cancer risk following testicular cancer: A follow-up study of 1,909 patients.** *JournalofClinicalOncology* 1993, 11:415-424.
20.   Haugnes HS, Wethal T, Aass N, Dahl O, Klepp O, Langberg CW, Wilsgaard T, Bremnes RM, Fossa SD: **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.

21.   von Eyben FE, Liu FJ, Amato RJ, Fritsche HA: **Lactate dehydrogenase isoenzyme 1 is the most important LD isoenzyme in patients with testicular germ cell tumor.** *Acta oncologica (Stockholm, Sweden)* 2000, 39:509-517.

22.   Murray MJ, Halsall DJ, Hook CE, Williams DM, Nicholson JC, Coleman N: **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.

23.   Belge G, Dieckmann KP, Spiekermann M, Balks T, Bullerdiek J: **Serum Levels of MicroRNAs miR-371-3: A Novel Class of Serum Biomarkers for Testicular Germ Cell Tumors?** *Eur Urol* 2012.

24.   Murray MJ, Coleman N: **Testicular cancer: A new generation of biomarkers for malignant germ cell tumours.** *Nat Rev Urol* 2012.

25.   Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J: **A novel and universal method for microRNA RT-qPCR data normalization.** *Genome Biol* 2009, 10:R64.

26.   Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D, Diehn M, Liu H, Panula SP, Chiao E, 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, Wiersma J, Te Meerman G, Schraffordt Koops H, Sleijfer DT, De Jong B: **No recurrent structural abnormalities in germ cell tumors of the adult testis apart from i(12p).** *Genes Chromosom & Cancer* 1995, 14:133-144.

28.   de Jong J, Stoop H, Gillis AJ, Hersmus R, van Gurp RJ, van de Geijn GJ, van Drunen E, Beverloo HB, Schneider DT, Sherlock JK, et al: **Further characterization of the first seminoma cell line TCam-2.** *Genes Chromosomes Cancer* 2008, 47:185-196.

29.   Zubakov D, Boersma AW, Choi Y, van Kuijk PF, Wiemer EA, Kayser M: **MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation.** *Int J Legal Med* 2010, 124:217-226.

30.   Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.** *Genome Biol* 2002, 3:RESEARCH0034.

31.   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.

32.   Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods (San Diego, Calif* 2001, 25:402-408.

33.   Kawakami T, Okamoto K, Ogawa O, Okada Y: **XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer.** *Lancet* 2004, 363:40-42.

34.   Chen X, Liang H, Zhang J, Zen K, Zhang CY: **Horizontal transfer of microRNAs: molecular mechanisms and clinical applications.** *Protein Cell* 2012, 3:28-37.

35.   Somasundaram R, Herlyn M: **Melanoma exosomes: messengers of metastasis.** *Nature medicine* 2012, 18:853-854.

CHAPTER **6**

Identification of known and novel germ cell cancer specific (embryonic) miRs in serum by high throughput profiling

Martin A. Rijlaarsdam[1,*], Ton van Agthoven[1,*], Ad J.M. Gillis[1], Sunali Patel[2], Kathleen Hayashibara[2], Kathy Y. Lee[2], Leendert H.J.Looijenga[1]

* both authors contributed equally to the work

Department of Pathology, Erasmus MC Cancer Institute – University Medical Center, Rotterdam, The Netherlands [1]
Thermo Fisher Scientific, South San Francisco, California, USA [2]

Abstract

MicroRNAs (miRs) are short non-coding RNA molecules (≈21 nucleotides) involved in regulation of translation. As such they are crucial for normal cell development and differentiation as well as cellular maintenance. Dysregulation of miRs has been reported in various diseases, including cancer. Interestingly, miRs can be informative as tumor classifiers and disease biomarkers. Recent studies demonstrated the presence of miRs in body fluids like serum, thus providing a putative non-invasive tool to study and monitor disease state. Earlier targeted studies by several independent groups identified specific embryonic miRs as characteristic for germ cell tumors (miR-371-2-3 & miR-302/367 clusters). This study reports a high throughput miR profiling (≈750 miRs) approach on serum from testicular germ cell tumor patients (14 seminoma and 10 non-seminoma) and controls (n=11), aiming at independent identification of miRs as candidate biomarkers for testicular germ cell tumors. A magnetic bead capture system was used to isolate miRs from serum. Subsequently, the TaqMan®Array Card 3.0 platform was used for profiling. The previously identified miRs 371 and 372 were confirmed to be specifically elevated in serum from germ cell tumor patients. In addition, several novels miRs were identified that were discriminative between GCC and controls: miR-511, -26b, -769, -23a, -106b, -365, -598, -340 and let-7a. In conclusion, this study validates the power of the embryonic miRs 371 and 372 in detecting malignant germ cell tumors (SE and NS) based on serum miR levels and identifies several potential novel miR targets.

Introduction

Germ cell tumors (GCT) are a heterogeneous group of malignancies originating from developing germ cells. Five subtypes have been identified (I-V). The type II GCTs are also referred to as germ cell cancer (GCC) and occur most frequently. GCC comprises 60% of all cancers in Caucasian males between 20 and 40 years of age and 1% of all male cancers in general. Testicular GCC arise from carcinoma *in situ* cells (CIS) and closely resemble embryonic germ cells, either primordial germ cells or gonocytes. CIS can progress to either seminoma (SE) or embryonal carcinoma (EC). EC is the stem cell component of all non-seminomas (NS), i.e. teratoma (embryonic lineages, TE) and yolk sac tumor (YS) / choriocarcinoma (CH) (extra-embryonic lineages) [1-6]. GCC show an exquisite sensitivity to current treatment protocols. Over 90% of patients with SE are cured independent of stage (100% in case of local disease). For NS patients the cure rate is more variable (95% for stage I and II disease, about 70% for stage III and IV disease) [7, 8]. Intensive follow-up of these young patients is required, e.g. surveillance of stage I disease after orchidectomy alone and follow-up for (late) recurrences.

A number of serum markers are available for GCC follow-up, including α-fetoprotein (AFP), human chorionic gonadotropin (hCG), and to a lesser extend lactate dehydrogenase (LDH) [7, 9-11]. Approximately 80% of the NS, and 20% of the SE show increased levels of these markers. AFP and hCG are primarily related to the presence of respectively a YS or CH component while they show little sensitivity for the stem cell components SE and EC. AFP levels might be elevated due to liver damage/disease, while hCG can be induced by chemotherapy [10, 11]. In addition, serum levels of these markers can change during disease progression, e.g. because of the development of different histological components. Therefore, better biomarkers are welcome, particularly markers with a high sensitivity for SE and EC.

microRNAs (miRs) have been shown to be an alternative to conventional serum markers. miRs are short non-coding RNA molecules regulating translation. miRs are highly tissue specific and play a key role in cellular differentiation and maintenance of tissue identity. Dysregulation of miRs has been linked to cancer development and tumor progression [12]. A number of independent recent studies showed that serum levels of the miR-371-2-3 and miR-302/367 clusters are predictive for the presence of malignant germ cell tumors, including (testicular) GCC. These miR clusters are normally only expressed during embryogenesis (embryonic germ/stem cells) [13-15]. These observations enhance the close relation of GCC to early/fetal germ cell development. Indeed, these embryonic miRs are expressed by SE and EC [16, 17]. Palmer and colleagues profiled miRs in tissues from malignant pediatric GCT patients and controls, and also showed that the miR-371-2-3 and miR-302 clusters were some of the few very highly overexpressed miRs in GCT [18]. In a case study, Murray and colleagues showed elevation of the levels of these miRs in serum of a 4-year-old boy presenting with a YS [19]. Levels dropped during chemotherapy and returned to normal

during an uneventful clinical follow-up [19], which was confirmed in a larger case series/cohort (eight patients versus six controls) [20]. This rapid drop in serum levels was confirmed by Belge and colleagues, who examined sera of six SE and five NS patients with stage I disease before and after orchidectomy. Only two of these patients initially presented with elevated levels of the classical serum tumor markers [21]. Dieckmann and colleagues verified these findings in an additional 24 pre- and post orchidectomy GCT patients and additionally showed that miR levels in the tumor do not necessarily correlate with serum levels [22].

In a larger study we presented the Targeted Serum miR (TSmiR) test [23]. A para-magnetic bead capture system was used to isolate and purify miRs from serum of 80 GCC patients, 47 controls, 11 matched pre- and post-orchidectomy samples and 12 non-GCC testicular masses. miR levels were quantified using targeted RT-PCR. Pre- and post-orchidectomy samples concerned samples also described in [22]. The high specificity of the miR-371-2-3 cluster and miR-367 was confirmed, as was the drop in serum levels after orchidectomy. A trend towards higher miR levels in patients with disseminated disease was observed. While application of the markers AFP and hCG resulted in about 50% false-negative cases, the TSmiR test showed an overall sensitivity of 98%, and a specificity of 48% (control versus tumor) [23]. These findings were validated in a recent independent study [24].

The studies reviewed above imply that specific miRs are potentially suitable as markers for the diagnosis and follow-up of GCC patients. So far however, studies have predominantly been limited to selected miR targets. However, other miRs might also be of interest for this application, which is the subject of this study. Here, we investigate serum levels of ≈750 miRs in 24 GCC samples and 11 controls using a high throughput approach to validate the previously identified markers and identify novel potential miRs as serum biomarkers for testicular GCC.

Materials and methods

Patient samples

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[a]" developed by the Dutch Federation of Medical Scientific Societies (FMWV[b] (Version 2002, update 2011). 35 serum samples were profiled for ≈750 miRs. These consisted of serum from 10 patients with NS (2 YS: NS-1/2, 2 TE: NS-3/4, 2

---

[a] http://www.federa.org/codes-conduct
[b] http://www.federa.org/

EC: NS-5/6, 2 EC+YS: NS-7/8 and 2 EC+TE: NS-9/10) and 14 pure SE cases. Controls (C) consisted of serum from 3 individuals with normal testis tissue showing adequate spermatogenesis (Johnson Score 8-10) (CONTROL-9/10/11) and serum from 8 patients with non-germ cell tumor related testicular masses. The latter group contained one epidermoid cyst (CONTROL-1), three epidermis cysts (CONTROL-2/3/4), one low grade liposarcoma (CONTROL-5), two neuro-endocrine tumors (CONTROL-7) and a case of paratesticular hemorrhage (CONTROL-8).

miR isolation from serum

miRs were purified from serum sample using TaqMan® ABC Purification Kit - Human Panel A and B (Life Technologies, PN 4473087 and 4473088). All reagents are provided in the kit. These panels consist of superparmagnetic Dynabeads® covalently bound to a unique set of ~380 anti-miR oligonucleotides for each panel. The miRs in each panel match the miRs in Megaplex™ Pools A and B described below. Each panel includes exogenous and endogenous controls. Briefly, for each panel 100ul of lysis buffer was added to 50 µl of serum. 2 µl of 1nM of an external control (ath-miR-159a) was added to monitor the extraction process, followed by the addition of 80 µl of beads (80 X 10^6 beads). The tubes were shaken at 1200 rpm at 30°C for 40 minutes then washed three times with wash buffer using a magnetic bead separator. The bound miRs were eluted from the beads with 100 µl elution buffer and incubation at 70°C for 3 minutes. The eluted miR pool was stored at -80°C until ready to use.

For miR profiling, Megaplex™ Primer Pools A and B v3.0 (PN 4444750) were used in conjunction with their matching TaqMan® MicroRNA Array Cards A and B (PN 4444913). Each Pool contains either reverse transcription primers or pre-amplification primers that target 377 unique miRs and additional control targets and allows for profiling up to of 754 unique miRs across the two pools. Each sample was run separately with Pool A and B according to an optimized protocol for blood plasma outlined in the Life Technologies Application Note: "Optimized blood plasma protocol for profiling human miRs using the OpenArray® Real-Time PCR System", 2011, PN 4399721 rev C. Reaction volumes, dilutions and the number of pre-amplification cycles were optimized to improve sensitivity for miRNA detection in blood serum samples. Briefly, 3 µl of the miR sample that was isolated with the ABC kit was reverse transcribed with the Megaplex™ RT Primer Pool A or B in a 10 µl final volume. The RT reaction was thermal cycled (2 min at 16°C, 1 min at 42°C, 1 sec at 50°C, for 40 cycles) and the enzyme was inactivated at 85°C for 5 min. The entire RT reaction was combined with its matching Megaplex™ PreAmp Primer Pool and TaqMan® PreAmp Master Mix (PN 4391128) in a final volume of 50 µl. Pre-amplification was run using the following cycling conditions: 10 min at 95°C; 2 min at 55°C; 2 min at 72°C; 15s at 95°C, 4 min at 60°C for 14 cycles; 99°C for 10 min. The final pre-amplification product was diluted 1:100 in 1X TaqMan® Universal Master Mix II; loaded onto the matching TaqMan® MicroRNA Array Card and run on a QuantStudio™ 12K Flex System (using a TaqMan® Array Card Block, universal cycling conditions).

miR quantification

Raw data files (.eds) were imported and analyzed in using ExpressionSuite v1.0.4 (Life Technologies), a software data analysis tool that can easily import and analyze large raw data files. In these experiments the quantification cycle (Cq) is defined as the fractional cycle at which the amplification plot crosses the fluorescence threshold (Ct). The baseline was set automatically and the threshold was manually set at 0.2. To capture as many differentially expressed miRs as possible the threshold was set at 40 instead of 30-32 which is generally recommended by Life Technologies for relatively high miR levels. A larger number of replicates or low standard deviation of Ct increases the confidence and power of results in this range. Statistical analysis was applied to identify the most relevant and consistent (i.e. significant) results.

Undetermined values were replaced with the maximal number of cycles (40). Non-human RNAs were removed from the dataset before normalization. Each assay was treated as an independent target. Ct values were normalized using global normalization according to [25]. RNU44, RNU48 and U6 snRNA were removed from the dataset before further analysis because these have been shown to be absent in serum [23]. Batch effects were corrected using Combat[c] as implemented in the sva[d] Bioconductor package (v3.10.0). All raw and processed data is available at GEO (accession ID: GSE59520[e])

Software and statistics

Data was extracted using ExpressionSuite v1.0.4 (Life Technologies/Applied Biosystems). Data pre-processing was performed in Microsoft Excel 2010 & GenEx 5.3.7.332. Further processing, was carried out in R 3.1.0 alpha, 64 bits. The ggplot2[f] 2.1.0.0 and gplots[g] 2.13.0 packages were used for visualization. The gplots:heatmap.2 function was used to generate the heatmaps and dendrograms using the default settings. To build the randomForest[h] classifier in R, the implementation according to the method of Breiman and Cutler was used (v4.6) because this algorithm also supplies information about the importance of the miRs in the classifier besides determining the classification strength, does not overfit and is relatively insensitive to the choice of its parameters [26, 27]. Comparison of miR levels (dCt) between the tumor and control groups was performed using a Mann-Whitney U test. If not otherwise specified, significant differences are indicated by a Benjamini-Hochberg adjusted p value of less than or equal to 5% [28]. A median |dCt| of 2 between the tumor (T) and control (C) groups was considered relevant.

---

[c] http://www.bu.edu/jlab/wp-assets/ComBat/Abstract.html
[d] http://www.bioconductor.org/packages/release/bioc/html/sva.html
[e] http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59520
[f] http://cran.r-project.org/web/packages/ggplot2/index.html
[g] http://cran.r-project.org/web/packages/gplots/index.html
[h] http://cran.r-project.org/web/packages/randomForest/index.html

Results

This study aims to verify known miR markers for testicular GCC and identify new targets by comparing the profile of ≈750 miRs in serum of 14 NS and 10 NS patients with the serum-miR profile of healthy individuals and patients with non-GCC testicular lesions ($n_{controls}$=11).

Discriminative power of the microRNA profiles of GCC

The discriminative power of the miR profile between GCC (tumor=T=SE+NS) and controls (C) was evaluated using principal component analysis (Figure 1). No distinction between SE and NS was observed and some of the controls were indistinguishable from the tumors. However, many controls were positioned at distance of the tumor cluster. This indicates a difference between T and C. A random forest approach was applied to investigate the value of the miR profiles in classifying samples as T or C (Table 1). The random forests™ method[i] constructs many decision trees using bootstrap samples from the original data. For each bootstrap sample a decision tree is trained and used to obtain class predictions (T or C) for the unseen samples by taking the majority vote from all individual decision trees [26]. In the miR classifier, the estimated misclassification error was 20% (out of bag error rate). In the tumor group 8% (n=2) of the samples were misclassified as controls. These concerned NS1 (yolk sac tumor, 31 year old male) and SE6 (seminoma). In line with our previous findings the sensitivity of the classifier was much higher than the specificity: 91.7% [80.5%-98.4%] versus 45.5% [21.1-60.2%] (95% confidence intervals between square brackets) [23].

Table 1 Random forest classifier. The number of random variables per tree was set to the square root of the total number of variables (miRs) in the dataset. The number of trees was set to 20 times the ratio of [the total number of variables / the **number** of variables per tree] to ensure that each variable is included in several trees. Classification was stable when the number of variables per tree was multiplied by 2 or ½ (out of bag error: 20% and 23% respectively). C=control, T=tumor (SE or NS). Of the tumors only NS1 (yolk sac) and SE6 (seminoma) were misclassified.

	Classified as C	Classified as T	Classification error
True C	6	5	45%
True T	2	22	8%

---

[i] http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#features
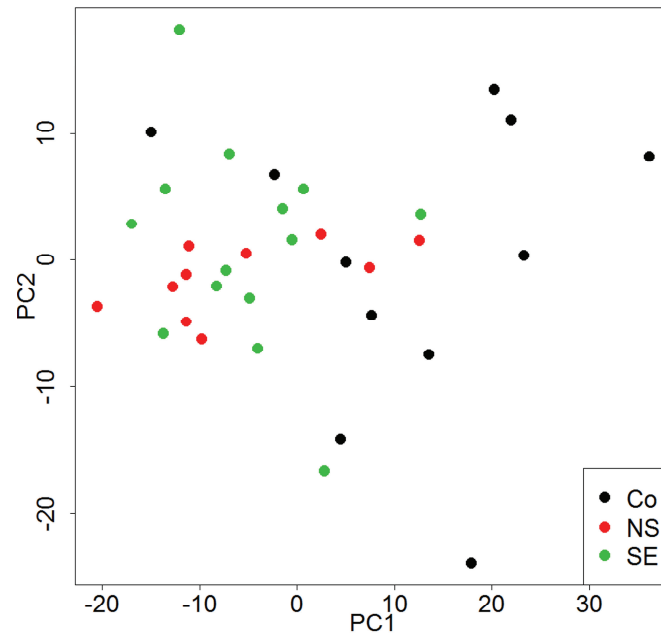
Figure 1. Principal Component Analysis on serum levels of all miRs. Co=controls, SE=seminoma, NS=non-seminoma. The first 2 principal components (PC) explained 24.38% + 9.38% = 33.76 % of the total variance. The third principal component explained 6.59% of the variance and did not show any more separation of the sample classes than PC1 and PC2 (data not shown).
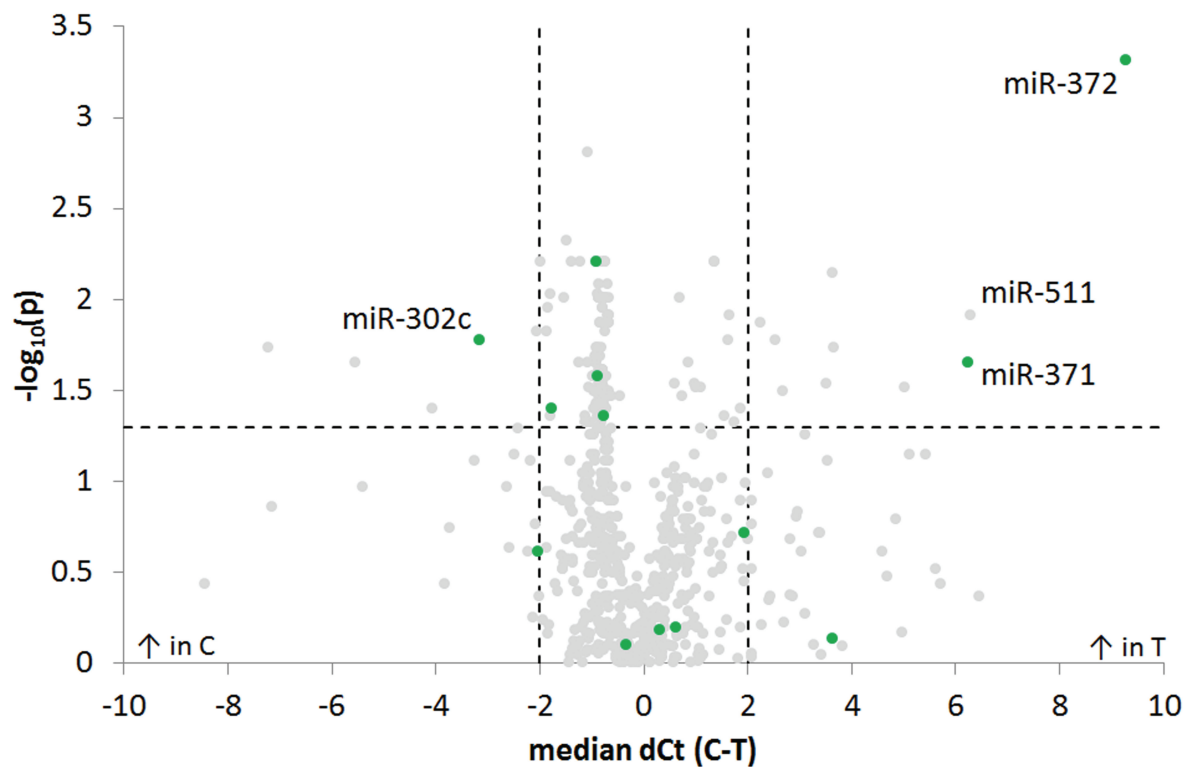
Figure 2. Differentiating miRs between tumor and control groups. Difference in median dCt between tumor (T: SE or NS) and control (C) is plotted against the -log$_{10}$ of the Benjamini-Hochberg adjusted p-value resulting from a Mann-Whitney U test comparing these two groups. An adjusted p<5% was considered significant and a difference in median |dCt| > 2 was considered relevant. Grey dots, all miRs; green dots: known GCC associated miRs (miR-371-2-3 and miR-302d/367 clusters).

GCC specific miRs in serum

Following up on the observed differences in global miR-profile between T and C, specific differentiating miRs were identified. When comparing the distribution of the serum levels between T and C, fifteen miRs showed significant and relevant differential serum levels (Figure 2, left and right upper areas). Ten miRs showed higher serum levels in T (4.77-621.29 fold difference). miR-372 showed the strongest difference (621.29 fold), followed by miR-511 (79.00 fold) and miR-371 (76.32 fold). miR-372 and -371 are known informative GCC miRs. Five miRs showed higher serum levels in C (4.14-147.57 fold difference) (Figure 3A, S1A). Of interest these included the GCC specific miR-302c with a moderate 8.83 fold difference. This is in line with the miR-302 family showing less consistent results and small differences between patients and controls than the miR-371-2-3 cluster in previous studies [23]. Even though it is a member of the miR-302 cluster, miR-367 showed high levels in serum from GCC patients in previous studies [23, 24]. In line with miR-302c however, miR-367 was expressed moderately lower in tumor than in control cases in the current study (Δmedian_dCT$_{N-T}$=-0.88764, p<0.01).

miR relevance can also be evaluated based on its importance in the decision trees that make up the random forest classifier. Between each downstream node (parent-child) in each tree the reduction in incorrect classification (impurity) can be estimated. By averaging this value for all nodes containing a certain miR, the relative importance of this miR in the total classifier can be estimated [26]. Analogous to the fifteen differentially expressed miRs discussed above, the top-fifteen miRs that contributed strongest to the classifier were selected (Figure 3B, S1B). Two miRs overlapped between both approaches: miR-372 and miR-511 (Figure 3, 4). Expression levels of miRs previously associated with GCC are shown in Figure 3C and S1C. Most convincing is miR-372 with very high levels in almost all T samples and low levels in all controls. In the two pure TEs (NS-3/4) and one EC+TE (NS-9) sample which consisted mainly of TE based on the pathology report, miR-372 showed only moderate levels. These samples were therefore clustered with the controls. miR-511 was moderately expressed in these samples, in line with the tumor samples. A single seminoma sample (SE-12) also clustered with the controls based on low miR-511 expression (Figure 4). miR-371 was less consistent and miR-302c showed the aberrant pattern described above (Figure S1). Additional differentiating miRs are presented in Figure 3A, 3B, 4A and 4B and more details are presented in Table S1 and Figure S1.
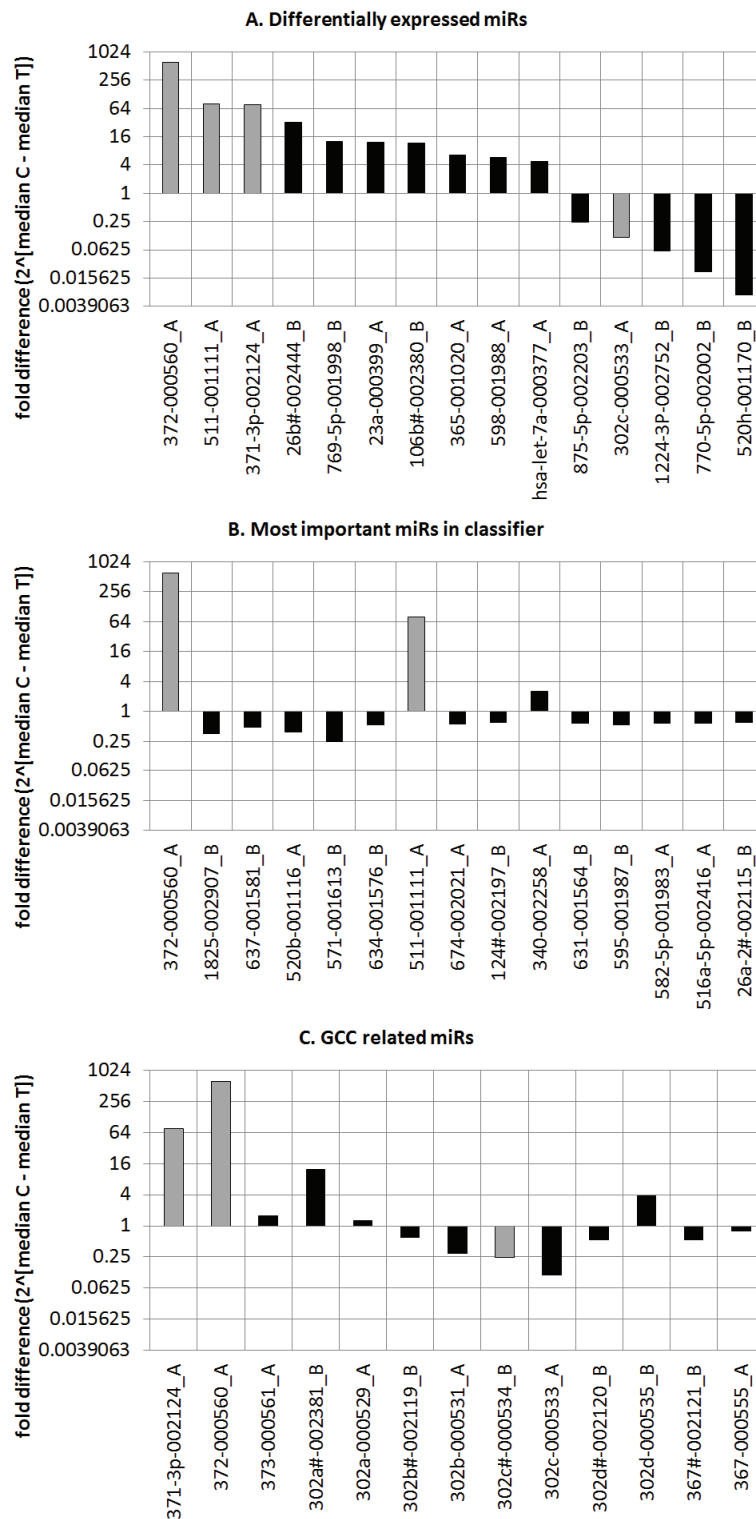
Figure 3. Fold difference between tumors and controls. (A) Differentially expressed miRs. These miRs are depicted in the upper-right and upper-left areas in the volcano plot (Figure 2). Order: fold difference between T and C (decreasing) (B) Most important miRs in the random forest classifier. Please see main text. Order: variable importance (decreasing). (C) GCC related miRs from previous studies. miR-371-2-3 and miR-302/367 clusters. Order: Genetic cluster. (all) X-axis: 2^(median C – median T). Positive values indicate higher serum levels in the T group. Y-axis: miRs, prefix "hsa-miR-" is omitted. ID after miR name indicates the assay. A/B indicates one of the TaqMan® MicroRNA Array Cards. C: controls. T: SE or NS. Overlapping miRs between panels A, B and/or C are indicated in grey.

Figure 4. Heat map of the most discriminative miRs. Red=high levels. Green=low levels. CONTROL=sera from controls, SE=seminoma, NS=non-seminoma. Details of the histological composition of the NS and CONTROL samples is presented in the Materials and Methods section. Detailed heatmaps of all miRs mentioned in Figure 3 are presented in Figure S1.

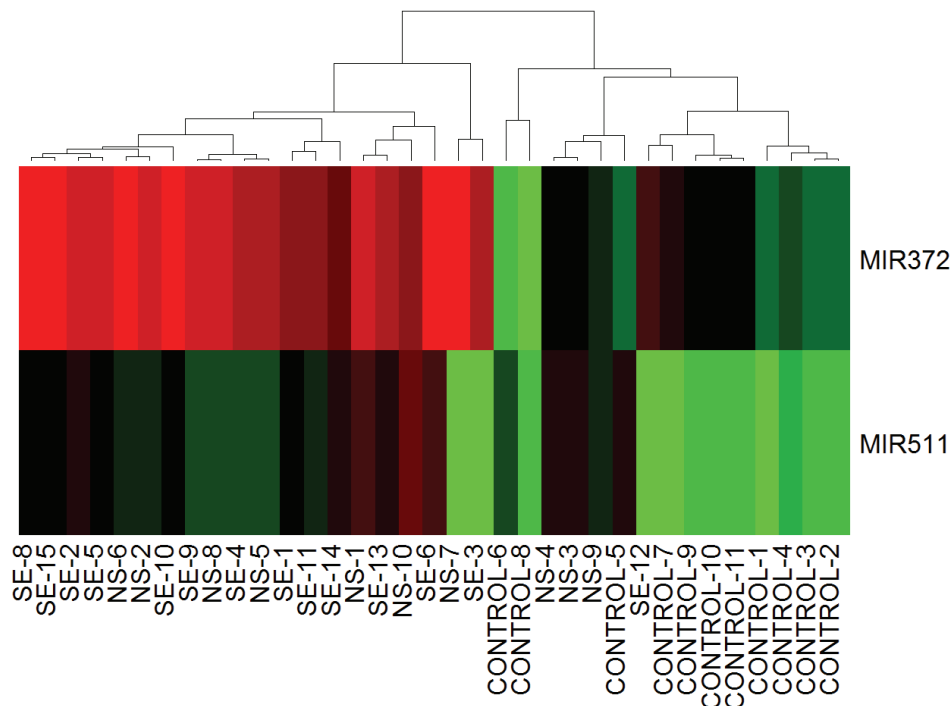# Discussion & Conclusion

Serum of 24 GCC cases and 11 controls were profiled with regard to the levels of ≈750 miRs using a miR specific magnetic bead purification system followed by profiling with Megaplex Pools and TaqMan Array cards. miR profiles did show moderate global differences between tumors (T) and control (C) samples (Figure 1). A random forest classifier was able to identify T samples with high sensitivity and moderate specificity (Table 1), indicating that the miR profiles include information to distinguish between T and C.

The germ cell cancer specific miRs identified in this study contained a number of earlier described miRs. The top differentiating miR between controls and tumors was the embryonic miR-372, in line with previous studies, closely followed by its cluster member miR-371 (Figure 2, 3). The absence of other previously reported GCC related miRs as well as the aberrant observation that a single GCC related miR (miR-302c) was moderately, but relevantly higher in C compared to T is potentially be related to very low levels of these miRs in the serum samples. This could be associated with selective release of specific miRs into the blood (e.g. 371 > 302c). Also, less sensitive detection in a high throughput setting as compared to earlier targeted studies might contribute to these findings [19-24].

A number of new differentiating targets was also identified which might be of future interest. Most markedly miR-511 was consistently higher in serum samples from GCC patients (Figures 3A, 3B and 4). This miR has been reported to have possible tumor suppressing activity in adenocarcinoma and modulates tumor associated immune response [29-31]. Other potential targets that showed significantly higher levels in T or were important in the classifier included miR-26b, -769, -23a, -106b, -365, -598, -340 and let-7a (Figure 2, 3A, 3B, S1A, S1B). Enrichment of miR-23a and its targets has previously been shown in an integrated genome wide study into GCC gene and miR expression, possibly indicating a function of this miR in GCC pathogenesis [32]. miR-23a and miR-106b have been reported to be related to spermatogenesis in dogs [33] and mice [34].

Pure TE have previously been shown not to express the embryonal set of germ cell cancer specific miRs in tumor tissue or serum from patients (miR-371-2-3 and miR-302 clusters) [18, 22]. It has to be kept in mind however that the testis in which the TE is diagnosed might contain CIS and EC components which do express these miRs [1-6, 35]. Clustering based on the most differentiating miRs (372 and 511) showed consistent clustering of TE samples (NS-3 and NS-4) with the controls, although miR-511 showed levels comparable to the tumors. A sample consisting of TE and a small EC component also clustered with the controls based on low miR-372 expression (NS-9, Figure 4). In contrast however, the two tumor samples that were misclassified as controls in the random forest classifier did not concern the TE samples but NS-1 (yolk sac tumor) and SE-6 (seminoma) (Table 1). In contrast to tissue findings, our results suggest that TE samples show an overall serum expression pattern of all miRs that makes them more similar to GCC than to controls. This is further supported by the finding of CIS in at least NS-3 and NS-9 by the pathologist. However, this does not contradict the fact that the targeted studies mentioned above as well as this study were unable to shown high levels of specific embryonal miRs (miR-371-2-3, miR-302/367) in tumor tissue or serum of TE patients. This potentially allows distinction of pure TE cases from mixed tumors including a teratoma component based on individual miRs.

In conclusion, this study shows the discriminative power of miR profiles in separating serum from GCC patients and controls. It also validated the strength of members of the embryonic miR-371-2-3 cluster in detecting GCC, both of the SE and NS subtype. As shown previously, high serum levels of miR-371 and miR-372 are predominantly related to the presence of SE and EC. Several other discriminating miRs with high serum levels in GCC cases were identified. Targeted validation of these novel targets and optimization of the sensitivity of the detection system will be crucial next steps. Investigation of differences in miR profile between the various histological components would also be of interest, but requires more power. Currently we are also investigating the correlation between miR profiles in serum and matched frozen & FFPE tumor samples.

Acknowledgements

Supplementary Information

Figure S1. Heatmaps of all miRs of interest. Analogous to Figure 4 depicting all miRs mentioned in Figure 3.

Table S1. Results for all miRs. For each miR the following results (columns) are presented: (i) the detected serum level level in C and T, (ii) the fold difference between C and T, (iii) the (adjusted) p-values from the group comparison and (iv) the importance in the random forest classifier. miR identifier is constructed by combining the miR name, the assay ID and the TaqMan® MicroRNA Array Card (A or B) .

Supplementary information will be made available upon publication.

References

1. Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5:210-222.
2. Looijenga LH: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J Pathol* 2009, 218:146-162.
3. Boublikova L, Buchler T, Stary J, Abrahamova J, Trka J: **Molecular biology of testicular germ cell tumors: unique features awaiting clinical application.** *Critical reviews in oncology/hematology* 2014, 89:366-385.
4. Rijlaarsdam MA, Looijenga LH: **An oncofetal and developmental perspective on testicular germ cell cancer.** *Seminars in cancer biology* 2014.
5. Vasdev N, Moon A, Thorpe AC: **Classification, epidemiology and therapies for testicular germ cell tumours.** *The International journal of developmental biology* 2013, 57:133-139.
6. Woodward PJ, Heidenreich A, Looijenga LHJ, al. e: **Testicular germ cell tumors.** In *World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs.* Edited by Eble JN, Sauter G, Epstein JI, Sesterhann IA, eds. Lyon: IARC Press; 2004: 17-278
7. Horwich A, Shipley J, Huddart R: **Testicular germ-cell cancer.** *Lancet* 2006, 367:754-765.
8. Beyer J, Albers P, Altena R, Aparicio J, Bokemeyer C, Busch J, Cathomas R, Cavallin-Stahl E, Clarke NW, Classen J, et al: **Maintaining success, reducing treatment burden, focusing on survivorship: highlights from the third European consensus conference on diagnosis and treatment of germ-cell cancer.** *Ann Oncol* 2013, 24:878-888.
9. von Eyben FE, Liu FJ, Amato RJ, Fritsche HA: **Lactate dehydrogenase isoenzyme 1 is the most important LD isoenzyme in patients with testicular germ cell tumor.** *Acta oncologica (Stockholm, Sweden)* 2000, 39:509-517.
10. Sturgeon CM, Duffy MJ, Stenman UH, Lilja H, Brunner N, Chan DW, Babaian R, Bast RC, Jr., Dowell B, Esteva FJ, et al: **National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers.** *Clin Chem* 2008, 54:e11-79.
11. Salem M, Gilligan T: **Serum tumor markers and their utilization in the management of germ-cell tumors in adult males.** *Expert review of anticancer therapy* 2011, 11:1-4.
12. Esquela-Kerscher A, Slack FJ: **Oncomirs - microRNAs with a role in cancer.** *Nat Rev Cancer* 2006, 6:259-269.
13. Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, Cha KY, Chung HM, Yoon HS, Moon SY, et al: **Human embryonic stem cells express a unique set of microRNAs.** *Dev Biol* 2004, 270:488-498.
14. Barroso-del Jesus A, Lucena-Aguilar G, Menendez P: **The miR-302-367 cluster as a potential stemness regulator in ESCs.** *Cell cycle (Georgetown, Tex* 2009, 8:394-398.
15. Kuo CH, Deng JH, Deng Q, Ying SY: **A novel role of miR-302/367 in reprogramming.** *Biochemical and biophysical research communications* 2012, 417:11-16.
16. Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW, Sun Y, Chen C, Guenther S, Sherlock J, Veltman I, Baeten J, et al: **High-throughput microRNAome analysis in human germ cell tumours.** *J Pathol* 2007, 213:319-328.
17. Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, et al: **A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.** *Cell* 2006, 124:1169-1181.
18. Palmer RD, Murray MJ, Saini HK, van Dongen S, Abreu-Goodger C, Muralidhar B, Pett MR, Thornton CM, Nicholson JC, Enright AJ, Coleman N: **Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets.** *Cancer Res* 2010, 70:2911-2923.
19. Murray MJ, Halsall DJ, Hook CE, Williams DM, Nicholson JC, Coleman N: **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.
20. Murray MJ, Coleman N: **Testicular cancer: a new generation of biomarkers for malignant germ cell tumours.** *Nat Rev Urol* 2012, 9:298-300.
21. Belge G, Dieckmann KP, Spiekermann M, Balks T, Bullerdiek J: **Serum levels of microRNAs miR-371-3: a novel class of serum biomarkers for testicular germ cell tumors?** *Eur Urol* 2012, 61:1068-1069.

22. Dieckmann KP, Spiekermann M, Balks T, Flor I, Loning T, Bullerdiek J, Belge G: **MicroRNAs miR-371-3 in serum as diagnostic tools in the management of testicular germ cell tumours.** *Br J Cancer* 2012, 107:1754-1760.

23. Gillis AJ, Rijlaarsdam MA, Eini R, Dorssers LC, Biermann K, Murray MJ, Nicholson JC, Coleman N, Dieckmann KP, Belge G, et al: **Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle.** *Molecular oncology* 2013, 7:1083-1092.

24. Syring I, Bartels J, Holdenrieder S, Kristiansen G, Muller SC, Ellinger J: **Circulating serum microRNA (miR-367-3p, miR-371a-3p, miR-372-3p, miR-373-3p) as biomarkers for patients with testicular germ cell cancers.** *J Urol* 2014.

25. Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J: **A novel and universal method for microRNA RT-qPCR data normalization.** *Genome Biol* 2009, 10:R64.

26. Breiman L: **Random Forests.** *Machine Learning* 2001, 45:5-32.

27. Liaw A, Wiener M: **Classification and Regression by randomForest.** *R News* 2002, 2:18-22.

28. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B (Methodological)* 1995, 57:289-300.

29. Zhang HH, Pang M, Dong W, Xin JX, Li YJ, Zhang ZC, Yu L, Wang PY, Li BS, Xie SY: **miR-511 induces the apoptosis of radioresistant lung adenocarcinoma cells by triggering BAX.** *Oncology reports* 2014, 31:1473-1479.

30. Zhang C, Chi YL, Wang PY, Wang YQ, Zhang YX, Deng J, Lv CJ, Xie SY: **miR-511 and miR-1297 inhibit human lung adenocarcinoma cell proliferation by targeting oncogene TRIB2.** *PloS one* 2012, 7:e46090.

31. Squadrito ML, Pucci F, Magri L, Moi D, Gilfillan GD, Ranghetti A, Casazza A, Mazzone M, Lyle R, Naldini L, De Palma M: **miR-511-3p modulates genetic programs of tumor-associated macrophages.** *Cell reports* 2012, 1:141-154.

32. Rijlaarsdam MA, Rijlaarsdam DJ, Gillis AJ, Dorssers LC, Looijenga LH: **miMsg: a target enrichment algorithm for predicted miR-mRNA interactions based on relative ranking of matched expression data.** *Bioinformatics* 2013, 29:1638-1646.

33. Kasimanickam VR, Kasimanickam RK, Dernell WS: **Dysregulated microRNA Clusters in Response to Retinoic Acid and CYP26B1 Inhibitor Induced Testicular Function in Dogs.** *PloS one* 2014, 9:e99433.

34. Tong MH, Mitchell DA, McGowan SD, Evanoff R, Griswold MD: **Two miRNA clusters, Mir-17-92 (Mirc1) and Mir-106b-25 (Mirc3), are involved in the regulation of spermatogonial differentiation in mice.** *Biol Reprod* 2012, 86:72.

35. Novotny GW, Belling KC, Bramsen JB, Nielsen JE, Bork-Jensen J, Almstrup K, Sonne SB, Kjems J, Rajpert-De Meyts E, Leffers H: **MicroRNA expression profiling of carcinoma in situ cells of the testis.** *Endocr Relat Cancer* 2012, 19:365-379.

DMRforPairs: identifying Differentially Methylated Regions between unique samples using array based methylation profiles

Martin A. Rijlaarsdam[1,*], Yvonne G. van der Zwan[1,*],
Lambert C.J. Dorssers[1], Leendert H.J. Looijenga[1]

* both authors contributed equally to the work

Department of Pathology, Laboratory for Experimental Patho-Oncology,
Erasmus MC - University Medical Center, Rotterdam, The Netherlands [1]

Abstract

Background: Array based methylation profiling is a cost-effective solution to study the association between genome methylation and human disease & development. Available tools to analyze the Illumina Infinium HumanMethylation450 BeadChip focus on comparing methylation levels per locus. Other tools combine multiple probes into a range, identifying differential methylated regions (DMRs). These tools require groups of samples to compare. However, comparison of unique, individual samples is essential in situations where larger sample sizes are not possible.

Results: DMRforPairs was designed to compare regional methylation status between unique samples. It identifies probe dense genomic regions and quantifies / tests their (difference in) methylation level between the samples. As a proof of concept, DMRforPairs is applied to public data from four human cell lines: two lymphoblastoid cell lines from healthy individuals and the cancer cell lines A431 and MCF7 (including 2 technical replicates each). DMRforPairs identified an increasing number of DMRs related to the sample phenotype when biological similarity of the samples decreased. DMRs identified by DMRforPairs were related to the biological origin of the cell lines.

Conclusion: To our knowledge, DMRforPairs is the first tool to identify and visualize relevant and significant differentially methylated regions between unique samples.

Background

Epigenetic (de)regulation, including DNA (CpG) methylation, is associated with development, differentiation and many human diseases [1-3] including the initiation and progression of various cancers [3-8]. While the primary DNA sequence is mostly stable during the lifetime of an individual, the epigenome is highly dynamic and responsive. Because of this, it provides valuable information about (past) (micro-)environmental conditions in the context of human disease and development [9,10].

DNA CpG methylation is routinely investigated on a genome wide scale [2,3]. The methylation profile can be assessed using micro-arrays or sequencing by applying (1) methylation-sensitive restriction enzymes or immunoprecipitation (anti-5mC) or (2) bisulfite-based treatment, which converts unmethylated cytosines into uracils [11]. The Illumina Infinium HumanMethylation450 BeadChip (450 K) is a bisulfite-based, cost-effective, two-color array querying over 480,000 independent genomic positions (99% Refseq genes, 96% CpG islands) [12-14]. Various tools are available to pre-process and analyze the 450 K data, but differential methylation is primarily detected per locus or by comparing differential patterns across regions using groups of samples [15]. The latter is implemented in IMA[a] and bumphunter[b]. Indeed, IMA offers region based analysis [16], but it does not work when using unique samples. Bumphunter identifies regional changes in the regression coefficient between methylation status and phenotype. Therefore, bumphunter (like IMA) requires groups of samples of sufficient size to estimate this coefficient for each probe [17]. However, when analyzing small numbers of samples with unique characteristics (e.g. normal and affected tissue of a clinically unique patient or a manipulated cell model), large series of samples are not available and current methods cannot be applied. Although larger series of samples are preferred (biological replicates or more patients), comparison of unique samples is desired in such a situation. DMRforPairs was designed to address this problem by comparing regional methylation status between unique samples.

---

[a] http://ima.r-forge.r-project.org/
[b] http://www.bioconductor.org/packages/release/bioc/html/bumphunter.html

Implementation

The algorithm consists of a number of phases (Figure 1A) with fully customizable parameters which will be discussed below:

1. Recoding of the probe classes
2. Identification of regions with sufficient probe density
3. Quantification and testing of (difference in) methylation status.

Data import and pre-processing

As input DMRforPairs requires the methylation percentage of each CpG site in each sample. It was originally designed for the 450 K array, but is applicable to any platform that generates a methylation percentage per CpG site and has sufficient coverage. For example, Additional file 1 illustrates the algorithm's applicability to data generated using Nimblegen microarrays and the McrBC protocol (CHARM). DMRforPairs does not provide functions to import, filter (cross-hybrization, SNPs in probe sequence) or pre-process 450 K data because of the existence of a number of excellent, well maintained pre-processing R-pipelines [11,15,16,18-22]. In the package documentation examples are provided on how to extract 450 K data for DMRforPairs using the lumi[c], IMA[d] and minfi[e] pipelines. The output of these pipelines serves as input for DMRforPairs.

Recoding of the probe classes

Illumina assigns the majority of probes to eleven specific classes according to their association to one or more functional regions (relation to gene: Body, 5'UTR, 3'UTR, 1st exon, TSS1500, TSS200; relation to CpG island: Island, Northern/Southern Shelf & Shore [12]). Highly detailed classification may result in too low probe density per class as DMRforPairs investigates probes in close proximity to each other within each class individually. DMRforPairs therefore allows custom grouping and / or selection of classes. Three commonly used schemes are hard-coded in the software: (0) retain all eleven classes, (1) group on relation to gene/transcription start site/CpG island or (2) put all probes in one class. The last option ignores the assigned classes as it might be desirable to just let DMRforPairs identify DMRs without providing information about probes that belong to the same functional class. This option can also be used in case this functional classification is unknown.

---

[c] http://www.bioconductor.org/packages/release/bioc/html/lumi.html
[d] http://ima.r-forge.r-project.org/
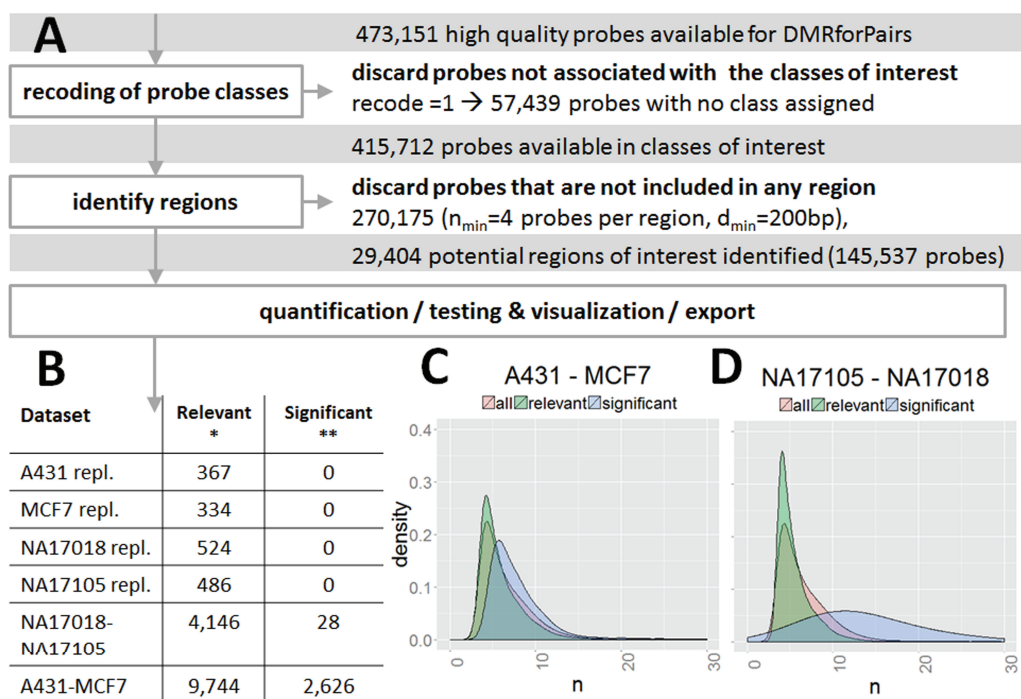[e] http://www.bioconductor.org/packages/2.12/bioc/html/minfi.html

Figure 1. Flowchart and overview of DMRforPairs results of the Illumina data. (A) The subsequent steps of recoding probe classes, identifying regions, quantifying and testing methylation differences and exporting the results are described in detail in the main text. Briefly, 473,151 probes remained after quality control. Subsequently, probes not associated to any of the 11 classes or not included in any of the regions are discarded. 145,537 probes (35%) were included in 29,404 potential regions of interest. Finally, these were assessed for methylation differences. (B) Number of regions identified in the various pairwise analyses. * = relevant indicates regions with $|\Delta M| > 1.4$, ** = significant indicates relevant + $p_{adjusted} \leq$ 0.05. "repl." indicates technical replicates. (C, D) The density plots illustrate the distribution of all/the relevant/the significant regions with regard to the number of probes in each region. Only the comparisons of the two cancer cell lines (C) and the pair of lymphoblastoid cell lines (D) are depicted as the technical replicates yielded no significant DMRs.

Identification of regions with sufficient probe density

A region of interest meets the following criteria:

1. Neighboring probes lay within d_{min} bp of each other (default = 200),
2. The number of probes per region $\geq n_{min}$ (default = 4), and
3. All probes are annotated to the same functional class (please see above).

Default settings of d_{min} are based on decreasing correlation between methylation status of adjacent loci when evaluated at inter-locus distances between 0 and 1 kb (200 bp is reported to correlate well) [11,23]. The default value for n_{min} is based on the theoretical minimal number of 2x4 observations required for statistical testing using Mann–Whitney U test. Probes annotated to more than one class are included in multiple regions and fully identical regions from different classes are merged into one region with a combined class. Figure 2 illustrates the number of regions identified for various settings of d_{min} and n_{min} and the fraction of all probes included in the regions. A function is available in DMRforPairs to generate these benchmarking results for specific data sets and tune the settings of n_{min} and d_{min}.

143

Quantification and testing of methylation status

As recommended, the methylation percentage β and the M-values ($\text{logit}_2(\beta)$) were used for visualization and statistical computations respectively [24]. Descriptive statistics are computed by DMRforPairs for all regions and samples (optional parallelization). These consist of median methylation levels (M and β values) and pairwise differences in median methylation level between all samples. If the median difference in M value between any pair of samples is sufficiently large in a specific region (> |ΔM|), the difference is formally tested using the Mann Whitney U or Kruskal-Wallis test. Pairwise testing is performed for more than two samples if indicated ($p_{Kruskal-Wallis} \leq 0.05$). An α of 0.05 after adjustment for multiple testing (Bejamini & Hochberg (FDR) [25]) is used to select significant regions (default settings). α and the method to correct for multiple testing can be specified by the user.

Several issues need to be kept in mind when choosing the algorithm's parameters and (test) results. In general, setting the algorithm's parameter more stringently (|ΔM|↑,n_{min}↑,d_{min}↓) reduces the amount of regions to be tested, but also discards potential DMRs that are less optimally covered by the probes on the array. Concerning the |ΔM| threshold it is important to be aware that the default setting (1.4) lies at the upper bound of the range (0.4-1.4) recommended by Du *et al.* A less stringent setting might result in a higher detection rate but reduces the true positive rate and increases the amount of multiple testing performed by DMRforPairs [24]. Also, correlation of methylation levels of CpG sites located closely together on the genome should be kept in mind. The potential presence of correlation warrants careful evaluation of statistical test results related to the independency assumption even though methylation levels at specific sites are technically (different probes) and biologically (different genomic positions) independent. Finally, comparisons with a higher number of probes per region have a higher power and are more likely to survive multiple testing. Therefore, the list of significant DMRs is theoretically biased towards regions with more probes (i.e. larger sample size). This bias was limited in a comparison of samples which are derived from a strongly biologically different origin (Figure 1B, C). When comparing the more similar samples (28 DMRs) there was some overrepresentation of regions with a high number of probes (Figure 1B, D).

Visualization, export and exploration

HTML tables listing all, only relevant (median difference $\geq \Delta$M) and only significant regions are generated with links to genome browsers (Figure 3A, application of the R2HTML[f] package [26]). Links are also provided to images depicting observed methylation profiles and a text file with additional descriptive statistics (Figure 3). Pairwise plots are generated in case of more than two biological samples. For relevant and significant DMRs an extended output can be generated including thumbnails in the HTML tables and visualizations that also depict transcripts annotated

---

[f] http://cran.r-project.org/web/packages/R2HTML/index.html

(close) to the region (Figure 3C, application of the Gviz[g] & GenomicRanges[h] packages [27,28]). In addition, DMRforPairs includes a number of functions to further inquire the data. Methylation status of genes of interest, regions identified by DMRforPairs and custom genomic intervals can be visualized, annotated and quantified/tested.

A. Number of regions identified

$\downarrow d_{min} / n_{min} \rightarrow$	4	5	6	7	8	9	10
100	15,699	10,572	7,327	4,644	2,825	1,713	1,041
200	29,404	20,640	15,289	11,329	8,383	5,950	4,102
300	38,524	28,608	21,733	16,737	12,971	9,732	7,163
400	43,684	33,803	26,304	20,704	16,511	12,907	9,900
500	46,599	37,116	29,425	23,545	19,076	15,265	11,970
600	48,340	39,227	31,547	25,484	20,801	16,831	13,404
700	49,442	40,610	32,987	26,840	22,071	18,011	14,460
800	50,239	41,568	33,992	27,795	22,960	18,832	15,247
900	50,709	42,239	34,688	28,449	23,604	19,427	15,827
1,000	51,016	42,706	35,253	29,020	24,142	19,969	16,278

B. Fraction of probes included in regions

$\downarrow d_{min} / n_{min} \rightarrow$	4	5	6	7	8	9	10
100	19%	14%	11%	8%	6%	4%	3%
200	35%	28%	24%	20%	17%	14%	11%
300	47%	40%	34%	30%	26%	22%	19%
400	54%	48%	42%	37%	34%	30%	26%
500	59%	54%	48%	43%	39%	36%	32%
600	63%	58%	52%	48%	44%	40%	36%
700	65%	60%	55%	51%	47%	43%	39%
800	67%	63%	58%	53%	49%	46%	42%
900	69%	64%	59%	55%	51%	47%	44%
1,000	70%	66%	61%	56%	53%	49%	45%

Figure 2. Tuning of the of d_{min} and n_{min} parameters. (A) Number of regions identified and (B) fraction of all probes included in these regions using different settings of d_{min} and n_{min}. d_{min} denotes the maximal distance in bp allowed between two adjacent probes to be accepted in the same region. n_{min} denotes the minimal number of probes in a region (per sample). All runs of the algorithm were done using the 415,712 probes annotated to at least one Illumina class grouped according to gene/transcription start site/CpG island (recode parameter = 1). These benchmark statistics can be generated using the *tune_parameters* function in the algorithm (optional parallelization).
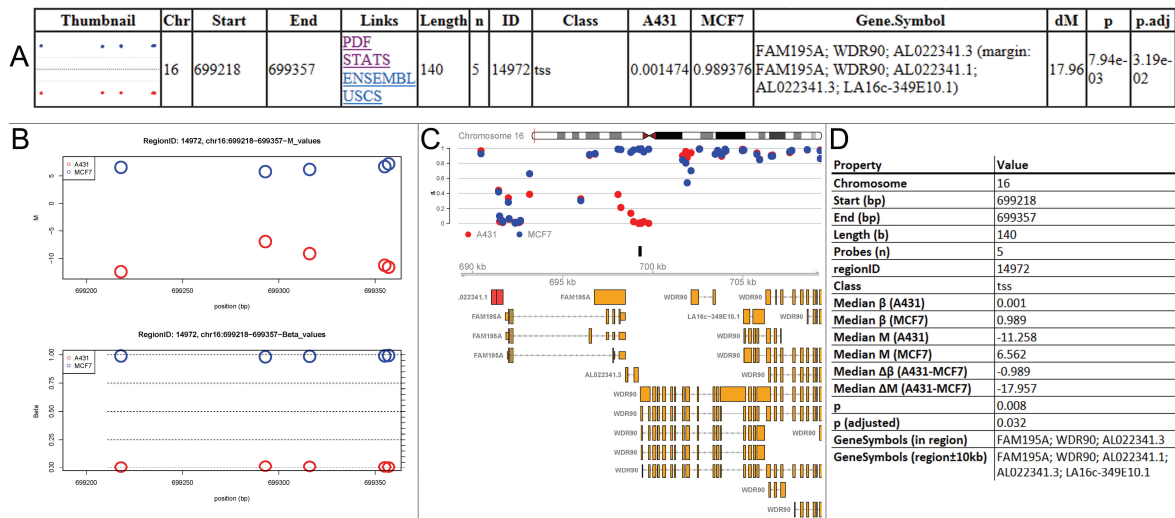

Figure 3. DMRforPairs output. (A) One row of the HTML table describing one DMR. Thumbnail, genomic annotation and descriptive statistics regarding (the difference between) the samples are presented as well as links to figures/tables illustrating the methylation patterns in the samples in detail. Direct links to the genomic region in two genome browsers are also provided (Ensembl & UCSC). Region IDs are generated on the fly by the *regionfinder* function and are specific to a dataset and to a set of DMRforPairs parameters. They are therefore not interchangeable between datasets/experiments and serve mainly as identifiers during exploration of the dataset. (B) Methylation level per probe (M and β values) plotted against its genomic position. These plots are generated for all relevant and significant regions. (C) Annotated visualization of DMRs (β values) ±10 kb. Black box indicates the DMR. Transcripts overlapping/near the region are retrieved from Ensembl. These plots are optionally generated for all relevant and significant regions. (D) Additional statistics (STATS link in table) as provided for each region.

---

[g] http://www.bioconductor.org/packages/release/bioc/html/Gviz.html
[h] http://www.bioconductor.org/packages/release/bioc/html/GenomicRanges.html

Results and discussion

Dataset

As a proof of concept, DMRforPairs is applied to a public dataset including two commercially available EBV transfected lymphoblastoid cell lines from healthy individuals (NA17105 (African American male) and NA17018 (Chinese female), Coriell Institute for Medical Research[i] (NJ, USA)). The dataset also includes the breast cancer cell line MCF7 [29] and HPV negative squamous-cell vulva carcinoma cell line A431 [30,31]. Data is available at Illumina's website[j] [12,13] and was processed in GenomeStudio V2011.1 and R 3.0.1 (Windows 7 x64) and 2.15.2 (Redhat Linux x64) using Illumina's annotation manifest (v. 1.1[k]) . Import and pre-processing was carried out using the LUMI[l] package [19] following the optimized "lumi: QN + BMIQ" pipeline [11]. This includes exclusion of poorly performing probes (p < 0.01, n = 713), color adjustment, quantile normalization and correction for probe type bias (Infinium I vs II) using the BMIQ algorithm [20]. Differentially methylated regions were identified by applying the DMRforPairs algorithm using the default settings (Figure 1, d_{min} = 200, n_{min} = 4, ΔM = 1.4, recode = 1, α = 0.05, correction for multiple testing = Benjamini Hochberg (FDR)). The networks/enrichment analyses were performed in IPA (Ingenuity® Systems, www.ingenuity.com, Core analysis; default settings).

Results

In the Illumina manifest, 12% of the probes were not assigned to any of the 11 categories (discarded in this analysis with recode parameter set to 1). 35% of the remaining probes was included in one or more regions, leading to 29,404 potential regions of interest. Samples were compared pairwise in descending order of biological similarity: technical duplicates, lymphoblastoid cell lines and cancer cell lines (average of duplicates) (Figure 1, Additional file 2).

As expected, no DMRs were identified when comparing the pairs of technical replicates (Figure 1B). In the two lymphoblastoid cell lines, 28 DMRs were identified (Figure 1B,D). Fitting with the Chinese and African American origin of the cell lines, top DMRs were associated with regions encoding human leucocyte antigens involved in immune response and known to be differently methylated between populations [32] (e.g. HLA-DRB1 (rank 2), HCG27 (rank 4), HLA-K / HCG4B (rank 7)). Enrichment / network analysis in IPA showed significant overrepresentation of genes associated with immunological diseases. This concerned various auto-immune diseases and lymphoma (9 genes, p = 0.000271-0.0293 depending on the subcategory; ACTA1, CHST8, GABR1, HCG27, HLA-DRB1, IGF2-AS, POU5F1, ZNF165, VTRNA2-1).

---

[i] http://ccr.coriell.org/
[j] http://support.illumina.com/downloads/genomestudio_software_20111.ilmn
[k] http://support.illumina.com/downloads/humanmethylation450_15017482_v12.ilmn
[l] http://www.bioconductor.org/packages/release/bioc/html/lumi.html

Between A431 and MCF7 2,626 DMRs were identified (Figure 1B,C). On top of the list was FAM195A a gene with known low expression [33] and complete methylation in MCF7. In A431, the region showed complete demethylation, but no public expression data was available for this cell line. The rest of the top-5 consisted of homeobox genes which are frequently methylated in breast cancer and active in squamous cell carcinoma [34,35]. Cancer was by far the strongest overrepresented disease category in the enrichment / network analysis (989 genes, p = 1.31E-19 - 2.71E-4). Enriched subcategories included breast cancer (n = 234, p = 2.06E-10), head and neck (squamous cell) carcinoma (n = 131, p = 1.30E-7) and genital tumor (n = 192, p = 1.94E-7).

Conclusions

DMRforPairs defines genomic regions using local probe density and optionally functional homogeneity. It quantifies, tests, annotates and visualizes (differential) methylation patterns between unique samples including pairwise comparison of samples if n > 2. Here, it is shown that in two lymphoblastoid cell lines from healthy individuals and cancer cell lines A431 and MCF7 (including 2 technical replicates each), DMRforPairs was able to identify an increasing number of DMRs related to the sample phenotype when biological similarity of the samples decreased. DMRs identified by DMRforPairs were related to the biological origin of the cell lines. In addition, DMRforPairs has been applied successfully in the analysis of integrated genome-wide epigenetic and expression profiles of germ cell cancer cell lines [36].

Availability & Requirements

Project home page: bioconductor.org[m] or martinrijlaarsdam.nl[n]

Operating system(s) & programming language: R, platform independent

Other requirements: R 2.15.2 or higher. Bioconductor packages: Gviz (> = 1.2.1) [27], R2HTML (> = 2.2.1) [26], GenomicRanges (> = 1.10.7) [28] and parallel. The lumi [19] package is suitable to import and pre-process 450 K data for use with DMRforPairs.

License: GPLv3, no restrictions to use by non-academics

---

[m] http://bioconductor.org/packages/release/bioc/html/DMRforPairs.html
[n] http://www.martinrijlaarsdam.nl/DMRforPairs/

Acknowledgements

Supporting Information

Additional file 1. R script illustrating the use of DMRforPairs with CHARM instead of 450 K data.

Additional file 2. DMRforPairs output for the comparison of A431-MCF7 and NA17018-NA17105. Please start from the HTML file in each folder.

Supporting information is available via http://dx.doi.org/10.1186/1471-2105-15-141

° http://www.eurospe.org/

References

1. Portela A, Esteller M: **Epigenetic modifications and human disease.** *Nat Biotechnol* 2010, **28**(10):1057–1068.
2. Rakyan VK, Down TA, Balding DJ, Beck S: **Epigenome-wide association studies for common human diseases.** *Nat Rev Genet* 2011, 12(8):529–541.
3. Suva ML, Riggi N, Bernstein BE: **Epigenetic reprogramming in cancer.** *Science* 2013, 339(6127):1567–1570.
4. De Carvalho DD, Sharma S, You JS, Su SF, Taberlay PC, Kelly TK, Yang X, Liang G, Jones PA: **DNA methylation screening identifies driver epigenetic events of cancer cell survival.** *Cancer Cell* 2012, 21(5):655–667.
5. Esteller M: **Epigenetics in cancer.** *N Engl J Med* 2008, 358(11):1148–1159.
6. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR, Ren B: **Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.** *Cell Stem Cell* 2010, 6(5):479–491.
7. Jones PA, Baylin SB: **The epigenomics of cancer.** *Cell* 2007, 128(4):683–692.
8. Van Der Zwan YG, Stoop H, Rossello F, White SJ, Looijenga LH: **Role of epigenetics in the etiology of germ cell cancer.** *Int J Dev Biol* 2013, 57(2-3-4):299–308.
9. Mirbahai L, Chipman JK: **Epigenetic memory of environmental organisms: A reflection of lifetime stressor exposures.** *Mutat Res* 2014, 764-765:10-7.
10. Skinner MK, Haque CG-BM, Nilsson E, Bhandari R, McCarrey JR: **Environmentally Induced Transgenerational Epigenetic Reprogramming of Primordial Germ Cells and the Subsequent Germ Line.** *PloS one* 2013, 8(7): e66318.
11. Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerstrom-Billai F, Jagodic M, Sundberg CJ, Ekstrom TJ, Teschendorff AE, Tegner J, *et al*: **An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform.** *Epigenetics* 2013, 8(3):333–346.
12. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Scroth GP, Gunderson KL, Fan JB, Shen R: **High density DNA methylation array with single CpG site resolution.** *Genomics* 2011, 98(4):288–295.
13. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL: **Genome-wide DNA methylation profiling using Infinium(R) assay.** *Epigenomics* 2009, 1(1):177–200.
14. Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, Lehmann U: **Quantitative cross-validation and content analysis of the 450 k DNA methylation array from Illumina, Inc.** *BMC Res Notes* 2012, 5:210.
15. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, Marsit CJ, Houseman EA, Brown R: **Review of processing and analysis methods for DNA methylation array data.** *Br J Cancer* 2013, 109(6):1394–1402.
16. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, Johnson CS, Smiraglia DJ, Liu S: **IMA: an R package for high-throughput analysis of Illumina's 450 K Infinium methylation data.** *Bioinformatics* 2012, 28(5):729–730.
17. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA: **Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies.** *Int J Epidemiol* 2012, 41(1):200–209.
18. Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempi G, Fuks F: **A comprehensive overview of Infinium HumanMethylation450 data processing.** *Briefings in bioinformatics* 2013.
19. Du P, Kibbe WA, Lin SM: **lumi: a pipeline for processing Illumina microarray.** *Bioinformatics* 2008, 24(13):1547–1548.
20. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S: **A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.** *Bioinformatics* 2013, 29(2):189–196.
21. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, Robinson WP, Kobor MS: **Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array.** *Epigenetics Chromatin* 2013, 6(1):4.

22. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R: **Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.** *Epigenetics* 2013, 8(2)**:**203–209.

23. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S,: **DNA methylation profiling of human chromosomes 6, 20 and 22.** *Nat Genet* 2006, 38(12)**:**1378–1385.

24. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM: **Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.** *BMC Bioinformatics* 2010, **11:**587.

25. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Roy Statist Soc Ser B (Methodological)* 1995, **57**(1)**:**289–300.

26. Lecoutre E: **The R2HTML Package.** *R News* 2003, 3(3), http://cran.r-project.org/web/packages/R2HTML/index.html.

27. Hahne F, Durinck S, Ivanek R, Mueller A, Lianoglou S: **Gviz: Plotting data and annotation information along genomic coordinates.** *R package version 145,* http://www.bioconductor.org/packages/devel/bioc/html/Gviz.html.

28. Aboyoun P, Pages H, Lawrence M: **GenomicRanges: Representation and manipulation of genomic intervals.** *R package version 1125,* http://www.bioconductor.org/packages/release/bioc/html/GenomicRanges.html.

29. Soule HD, Vazguez J, Long A, Albert S, Brennan M: **A human cell line from a pleural effusion derived from a breast carcinoma.** *J Natl Cancer Inst* 1973, 51(5)**:**1409–1416.

30. Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, Parks WP: **In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors.** *J Natl Cancer Inst* 1973, 51(5)**:**1417–1423.

31. Hietanen S, Grenman S, Syrjanen K, Lappalainen K, Kauppinen J, Carey T, Syrjanen S: **Human papillomavirus in vulvar and vaginal carcinoma cell lines.** *Br J Cancer* 1995, 72(1)**:**134–139.

32. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, Monk D, Hata K, Marques-Bonet T, Wang L, Esteller M: **DNA methylation contributes to natural human variation.** *Genome Res* 2013, 23(9)**:**1363–1372.

33. Rae JM, Johnson MD, Scheys JO, Cordero KE, Larios JM, Lippman ME: **GREB 1 is a critical regulator of hormone dependent breast cancer growth.** *Breast Cancer Res Treat* 2005, 92(2)**:**141–149.

34. Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP: **Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer.** *Breast Cancer Res* 2009, 11(1)**:**R14.

35. Rodini CO, Xavier FC, Paiva KB, De Souza Setubal Destro MF, Moyses RA, Michaluarte P, Carvalho MB, Fukuyama EE, Tajara EH, Okamoto OK, Nunes FD: **Homeobox gene expression profile indicates HOXA5 as a candidate prognostic marker in oral squamous cell carcinoma.** *Int J Oncol* 2012, 40(4)**:**1180–1188.

36. van der Zwan YG, Rijlaarsdam MA, Rossello FJ, Notini AJ, de Boer S, Watkins DN, Gillis AJM, Dorssers LCJ, White SJ, Looijenga LHJ: **Seminoma and embryonal carcinoma footprints identified by analysis of integrated genome-wide epigenetic and expression profiles of germ cell cancer cell lines.** *PLoS One* 2014, 9(6)e98330.

CHAPTER **8**

Seminoma and embryonal carcinoma footprints identified by analysis of integrated genome-wide epigenetic and expression profiles of germ cell cancer cell lines

Yvonne G. van der Zwan[1,*], Martin A. Rijlaarsdam[1,*], Fernando J. Rossello[2], Amanda J. Notini[3], Suzan de Boer[3], D. Neil Watkins[2], Ad J.M. Gillis[1], Lambert C.J. Dorssers[1], Stefan J. White[3,^], Leendert H.J. Looijenga[1,^]

joint first (*) or last (^) authors

Department of Pathology, Erasmus MC - University Medical Center, Rotterdam, The Netherlands [1]
Centre for Cancer Research [2] and Centre for Genetic Diseases [3]
MIMR-PHI Institute of Medical Research, Monash University, Clayton, Victoria, Australia

Abstract

Background: Originating from Primordial Germ Cells/gonocytes and developing via a precursor lesion called Carcinoma *In Situ* (CIS), Germ Cell Cancers (GCC) are the most common cancer in young men, subdivided in seminoma (SE) and non-seminoma (NS). During physiological germ cell formation/maturation, epigenetic processes guard homeostasis by regulating the accessibility of the DNA to facilitate transcription. Epigenetic deregulation through genetic and environmental parameters (i.e. genvironment) could disrupt embryonic germ cell development, resulting in delayed or blocked maturation. This potentially facilitates the formation of CIS and progression to invasive GCC. Therefore, determining the epigenetic and functional genomic landscape in GCC cell lines could provide insight into the pathophysiology and etiology of GCC and provide guidance for targeted functional experiments.

Results: This study aims at identifying epigenetic footprints in SE and EC cell lines in genome-wide profiles by studying the interaction between gene expression, DNA CpG methylation and histone modifications, and their function in the pathophysiology and etiology of GCC. Two well characterized GCC-derived cell lines were compared, one representative for SE (TCam-2) and the other for EC (NCCIT). Data were acquired using the Illumina HumanHT-12-v4 (gene expression) and HumanMethylation450 BeadChip (methylation) microarrays as well as ChIP-sequencing (activating histone modifications (H3K4me3, H3K27ac)). Results indicate known germ cell markers not only to be differentiating between SE and NS at the expression level, but also in the epigenetic landscape.

Conclusion: The overall similarity between TCam-2 / NCCIT support an erased embryonic germ cell arrested in early gonadal development as common cell of origin although the exact developmental stage from which the tumor cells are derived might differ. Indeed, subtle difference in the (integrated) epigenetic and expression profiles indicate TCam-2 to exhibit a more germ cell-like profile, whereas NCCIT shows a more pluripotent phenotype. The results provide insight into the functional genome in GCC cell lines.

Introduction

Type II (testicular) germ cell tumors, here referred to as Germ Cell Cancers (GCC), are the most common malignancy in Caucasian adolescents and young adults, and their incidence is still rising [1-3]. GCC originate from primordial germ cells or gonocytes, and are subdivided into seminomas (SE) and non-seminomas (NS), with carcinoma *in situ* (CIS) of the testis as their common precursor lesion[1], also known as Intratubular Germ Cell Neoplasia Unclassified (IGCNU) [3]. In contrast to CIS and SE, the stem cell component of NS (i.e., embryonal carcinoma, EC) is characterized by pluripotent potential [4]. EC can differentiate into somatic lineages and extra–embryonic tissues (teratoma vs yolk sac tumor and choriocarcinoma, respectively), including the germ cell lineage [4]. Various clinical, environmental and genetic risk factors for GCC have been identified, although the exact role of these factors is not completely clear. Clinical risk factors constitute urological/andrological/gonadal aberrations [5-8], while environmental factors focus on endocrine disruptors and androgen - estrogen balance [9-11]. Genetic risk factors include a number of susceptibility Single Nucleotide Polymorphisms, likely related to early gonadal development [12-15] and an association with familial predisposition [16]. Somatic mutations are rarely found in GCC [17]. There are strong indications that the micro-environment of the developing testis is of significant importance in the pathogenesis of GCC. Patients with Testicular Dysgenesis Syndrome (TDS) and specific forms of Disorders of Sex Development (DSD) are known to have an increased risk of developing GCC due to abnormal gonadal development, i.e. hypovirilization [18].

Epigenetic processes have a clear role in both the initiation and protection of pluripotency [19]. Deregulation of these tightly controlled processes is known to be involved in the formation and progression of various cancer types [20-24], including GCC [25]. During physiological germ cell formation and maturation, epigenetic processes (e.g. DNA methylation, histone modifications) guard homeostasis by regulating the accessibility of the DNA to facilitate transcription [25, 26]. The epigenome is highly dynamic, and changes occur depending on cell type and developmental stage, influenced by / reflecting the (micro-) environment. In spite of this knowledge, little is known about the role of histone modifications and DNA methylation regarding gene expression in GCC in general, and the possible similarities and differences between SE and EC [25, 27]. Epigenetic deregulation through genetic and environmental parameters (referred to as genvironment) could disrupt physiological embryonic germ cell development, resulting in delayed or blocked maturation, thereby facilitating the formation of CIS, and potentially progression to an invasive GCC [25, 28-30]. Therefore, determining the epigenetic and functional genomic landscape in GCC cell lines could provide insight into the pathophysiology and etiology of GCC. The results could provide guidance for targeted functional experiments.

In this study, epigenetic footprints of SE and EC cell lines were identified by studying the interaction between gene expression, DNA methylation and histone modifications. Two well characterized GCC-derived cell lines were used, one representative for SE (TCam-2) [31, 32] and the other for EC (NCCIT) [33]. Two types of epigenetic modifications were investigated and related to genome wide expression analysis: CpG DNA methylation status, and enrichment of activating histone marks (H3K4me3, H3K27ac).

Methods

Cell culture

TCam-2 [31, 32, 34] and NCCIT [33] cells were cultured in DMEM medium (#31966-021, Thermo Fisher Scientific / Life Technologies, Carlsbad, CA, USA) containing 10% fetal calf serum (FCS, GE Healthcare Life Sciences, HyClone Laboratories, Utah, USA) in T75 cm^2 flasks to 75-90% confluence. For RNA preparation, fresh medium was added 24 hours before harvest. Cells were washed once with Hanks balanced Salt Solution (HBSS, #14175-053, Thermo Fisher Scientific / Life Technologies, Carlsbad, CA, USA), and lysed with 7ml of ice-cold RNA-Bee (#Cs-105B, TEL-TEST Inc, Friendswood, Texas, USA). For methylation, gene expression (biological duplicates) and ChIP-seq analyses, different cultures of cells from a single source were used (LEPO lab, Department of Pathology, Erasmus MC Rotterdam). Biological replicates were started as independent cultures at different days and processed similarly.

Methylation profiling

DNA was isolated using the DNeasy kit according to manufacturer's instructions (#69504, QIAGEN, Hilden, Germany). Bisulfite conversion (EZ DNA Methylation Gold Kit, Zymo Research, Irvine, CA, USA) and methylation detection was performed at ServiceXS (ServiceXS B.V., Leiden, The Netherlands). Illumina's HumanMethylation450 BeadChip was used (Illumina, Inc., San Diego, CA, U.S.A, processing and hybridization according to the manufacturer's instructions). Image processing took place on the iScan system and the data was extracted using GenomeStudio, using default analysis settings (including background correction and normalization based on internal controls) and v1.2[a] of the annotation manifest. Further processing was carried out in R using the LUMI[b] package [35] following the optimized "lumi: QN+BMIQ" pipeline [36] This includes exclusion of poorly performing probes (p<0.01), color adjustment, quantile normalization and correction for probe type bias (Infinium I vs II) using the BMIQ algorithm [37]. All raw and processed data files are

---

[a] http://support.illumina.com/downloads/humanmethylation450_15017482_v12.ilmn
[b] http://www.bioconductor.org/packages/release/bioc/html/lumi.html

submitted as a GEO SuperSeries and accessible via GSE56454[c]. Differentially methylated regions were identified using the DMRforPairs[d] algorithm using the default settings [38]. DMRforPairs is available via Bioconductor. Briefly, DMRforPairs defines genomic regions using local probe density and optionally functional homogeneity (e.g. all probes in a region should be gene associated). It quantifies, tests and visualizes (differential) methylation patterns between unique samples. Differences were calculated as NCCIT versus TCam-2.

Gene expression profiling

Approximately 20 µg of RNA was treated with RNase-free DNaseI (#2238, Ambion, Ambion Inc., Austin, TX, U.S.A.) for 30 minutes at 37°C and subsequently purified using the RNeasy mini kit (#74104, Qiagen, Hilden, Germany) according to the manufacturer's instructions. Pure RNA was eluted in 50µl of water, and quantified using a Nanodrop (Thermo Scientific). Quality control, RNA labeling, hybridization and data extraction were performed at ServiceXS B.V. (Leiden, The Netherlands) according to their in-house protocol. Biotinylated cRNA was prepared using the Illumina TotalPrep RNA Amplification Kit (#AMIL1791, Ambion Inc., Austin, TX, USA) according to the manufacturer's specifications with an input of 200ng total RNA. Per sample, 750 ng of the biotinylated cRNA was hybridized onto the Illumina HumanHT-12 v4 (Illumina, Inc., San Diego, CA, U.S.A.) according to the Illumina Manual "Direct Hybridization Assay Guide". Image processing took place on the iScan system and the data was extracted using GenomeStudio (default settings). Further processing was carried out in R using the LUMI[e] package [35]. Following the guidelines presented in [39], robust spline normalization was applied to the log2 transformed intensity values. Probes with $p_{detection}>0.05$ in >50% of the samples were excluded from the analysis (n=27,964 out of 47,323). Average log2 intensities of biological replicates and per gene were used to assess expression levels (GEO accession number GSE56454[f]). Log2 ratios (R) of the average intensities in the two cell lines (NCCIT/TCam-2) were used to identify significantly differentially expressed genes. Genes with expression levels outside the 99% confidence interval (CI) of this log ratio were identified as differentially expressed between NCCIT and TCam-2.

Histone modification profiling (H3K27ac, H3K4me3)

The ChIP assay was performed according to the low cell number ChIP protocol from Diagenode (Liege, Belgium), with minor modifications. In brief, 1×10^6 cells were cross-linked for eight minutes by addition of formaldehyde to a final concentration of 1%, followed by neutralization with 1.25M glycine. The cells were then lysed, and chromatin was sheared to ~500 bp fragments using the Covaris sonicator under the following conditions; duty cycle 20%, peak incident power 200 watts,

---

[c] http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56454
[d] http://www.bioconductor.org/packages/release/bioc/html/DMRforPairs.html
[e] http://www.bioconductor.org/packages/release/bioc/html/lumi.html
[f] http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56454

cycles/burst 200, time 5min, temperature 4°C. Protein A-coated Dynabeads (#10002D, Invitrogen, Thermo Fisher Scientific / Life Technologies, Carlsbad, CA, USA) were incubated with 7 µg of the following antibodies: H3K4me3 (Diagenode pAb-003-050) or H3K27ac (Ab4729, Abcam, Cambridge, UK). The beads were combined with chromatin from 1×10^6 cells overnight on a rotating wheel. The immunobeads were washed, and DNA was purified using the iPure DNA purification kit (AL-100-0100, Diagenode, Liège, Belgium) according to manufacturer's instructions. DNA fragments were sheared a second time using a Covaris sonicator (duty cycle 10%, peak incidence power 175 watts, cycles/burst 200, time 5 minutes, temperature 4°C). Massively parallel sequencing of ChIP DNA (ChIP-Seq) was performed using the 5500xl SOLiD™ sequencing platform (Applied Biosystems, Foster City, CA, USA) at the Monash Health Translation Precinct Medical Genomics Facility. The sequencing experiments were single-end with 50nt read length (300nt average fragment size). Sequencing reads were aligned to the complete hg19 human genome (UCSC version, February 2009) using LifeScope™ Genomic Analysis Software v2.5[g]. ChIP-Seq experimental samples were normalized to a total of 10^7 uniquely mapped sequencing tags. Data was processed in HOMER[h] ([40]) to detect peaks and motif enrichments using the default settings (except "fold enrichment over input"; used 2; threshold for p-value 0.01) (GEO: GSE56454). Peak heights from HOMER were corrected for background (lowest peak height detectable). Heights were then summed per gene (ΣP) as annotated by HOMER and genes without any detectable peak were set to 0. The difference in summed peak heights ($\Delta\Sigma P = \Sigma P_{TCam-2} - \Sigma P_{NCCIT}$) was used to quantify differences between the cell lines. Genes with significantly differential histone modification patterns were identified for both marks separately (outside 99% CI of $\Delta\Sigma P$). Association of a peak with TSS was used as annotated by HOMER (1kb upstream of the TSS - 100bp downstream).

MLPA-DNaseI analysis

MLPA probes were designed following previously described criteria [41]. Based on differential modification patterns in NCCIT and TCam-2 probes were designed for the following loci: NCCIT: chr3:181425532-181425720, chr5:146699813-146699953, chr3:181577755-181577830, chr6:15240050-15240160, chr15:93191596-93191696, chr3:178908801-178908966, chr5:101550991-101551125, TCam-2: chr11:10613134-10613220, chr1:201278163-201278251, chr12:3091402-3091483, chr19:13985162-13985322, chr2:38323813-38323931, chr9:843018-843110, chr5:140762378-140762475. MLPA-DNaseI was performed as previously described [42], with minor modifications. In brief, nuclei from 1×10^6 cells were isolated, and treated with a range of DNaseI concentrations (0; 2 and 5 Units). Digested genomic DNA was purified, and 50-100 ng was used in an MLPA reaction. Following PCR amplification of ligated probes, products were separated on an ABI3700 DNA sequencer. Data was analyzed as previously described [43], with a reduction to 75% of peak height in undigested DNA used as a threshold for defining DNaseI-hypersensitivity.

---

[g] http://www.lifetechnologies.com/lifescope
[h] http://homer.salk.edu/homer/chipseq/

Software

Analyses were performed in R 3.0.1 (Windows 7 x64) and 2.15.2 (Redhat Linux x64). The networks/enrichment analyses were performed in IPA (Ingenuity® Systems, www.ingenuity.com). Genomic positions reported in this manuscript are based on the GRch37/hg19 assembly.

Results

To investigate epigenetic characteristics of SE and EC and their relationship to gene expression, genome-wide histone modification and DNA methylation patterns were investigated in the cell lines TCam-2 (SE) and NCCIT (EC), and matched to gene expression profiles. The differences between the two cell lines with regard to histone modification and DNA methylation status were first investigated separately. Subsequently, the resulting datasets were integrated to identify (target) genes with a strong relationship between primed DNA configuration and higher expression levels.

Histone modification

Histone modification patterns were assessed using chromatin immunoprecipitation combined with high throughput sequencing (ChIP-seq). Data analysis was performed as described in the materials and methods section. Alterations in H3K4me3 and H3K27ac were investigated, which are markers associated with promoter activation (transcription start site (TSS), H3K4me3 and H3K27ac) and enhancer activation (primarily H3K27ac) [44, 45]. In addition to the analysis of (differential) modification patterns, motif enrichment of the modified regions was investigated and compared between the cell lines.

H3K4me3 and H3K27ac do not show differential enrichment near transcription start sites and their peak heights correlated within genes.

Depending on the cell line, 10.2/11.3% of the H3K27ac enriched loci were located within 1 kb of a TSS against a comparable 14.7/16.8% of the H3K4me3 loci (TCam-2/NCCIT). This is in line with observations in other cell types showing that, even though H3K4me3 is directly related to promoter activation, a large majority of the H3K4me3 loci are located distally of the TSS [46]. H3K27ac has no reported preferential localization to TSS [47]. The level of summed peaks per gene (ΣP, see Methods) was used to compare histone modification patterns between the two histone marks. There was significant correlation in both cell lines between the peak levels at genes where both were present ($\rho_{TCam-2}=0.62$, $p_{TCam-2}<0.001$, $n_{TCam-2}=1837$; $\rho_{NCCIT}=0.37$, $p_{NCCIT}<0.001$, $n_{NCCIT}=746$; Spearman's ρ). This is in line with their overlapping function: open chromatin configuration is associated with both marks [19, 22] and allowed us to combine the histone modification results in the subsequent analysis.

CHAPTER 8

*H3K4me3 and H3K27ac enrichment patterns in TCam-2 and NCCIT are in accordance with known
SE / EC markers specificity.*

We previously showed that active chromatin modification patterns for *SOX17* and *SOX2* in the cell
lines TCam-2 and NCCIT match the expected pattern, based on gene and protein expression and
histological constitution (*SOX17* active in TCam-2, *SOX2* active in NCCIT) [25]. In line with this,
SOX17 and *SOX2* were differentially enriched for both H3K4me3 and H3K27ac in TCam-2 and
NCCIT respectively (Figure 1). *OCT3/4* showed no enrichment within the coding sequence,
however there was enrichment of both markers close to the TSS in NCCIT and TCam-2. This is
consistent with known OCT3/4 mRNA and protein expression in both cell lines [31, 48]. NANOG
was more enriched for both markers in TCam-2, in line with differences in expression level (see
below).
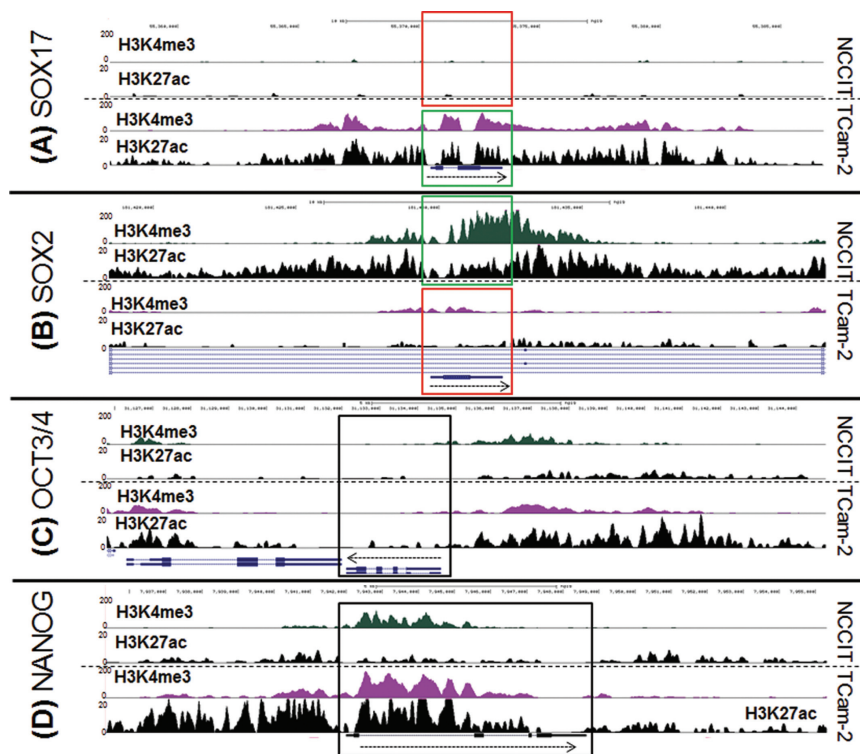
Figure 1. Display of H3K4me3 and H3K27ac tracks for both NCCIT and TCam-2. **(A)** *SOX17,* **(B)** *SOX2,* **(C)** *OCT3/4*
(*POU5F1*), **(D)** *NANOG*. Arrows indicate direction of transcription. Green boxes indicate markers specific for the
histological subtype represented by the cell line. Black boxes = no difference between the cell lines; red boxes = not a
marker for that cell type. Note the different ranges on the y-axis for H3K4me3 and H3K27ac.

On a genome-wide scale, 29,428 H3K4me3 enriched loci were identified in TCam-2 and 19,015 in
NCCIT. 25-41% more enriched loci were identified for H3K27ac than for H3K4me3 (n_{TCam-2}=41,569,
n_{NCCIT}=23,763). Genes with significant differences in summed peak height per gene (ΔΣP, see
Methods) were selected for further analysis. For TCam-2, genes showing differential histone
modifications were higher in number for H3K27ac (n_{TCam-2}=433, N_{NCCIT}=28). For NCCIT, there were
substantially more genes selected for H3K4me3 compared to H3K27ac (n_{TCam-2}=215, N_{NCCIT}=325).

For both marks, 86 / 11 genes overlapped between top differentiating lists in TCam-2 and NCCIT respectively. These included the SE marker *SOX17* in TCam-2 and the EC marker *SOX2* in NCCIT (Figure S1, Table S1). Functionally, the gene lists of both cell lines showed significant enrichment for (embryonic) stem cell maintenance/pluripotency (Table S2). Enrichment of biological functions in TCam-2 indicated similarity to more mature germ cells, which was lacking in the list of NCCIT (GO categories TCam-2 included development of normal testis morphology and germ cell maintenance). Moreover, two germ cell-specific canonical pathways IGF1 signaling (log_p=3.42) and germ cell-Sertoli cell Junction Signaling (log_p=2.11)) showed enrichment. In TCam-2, two functional networks were identified incorporating the AR pathway and lipid metabolism (Table S2).

Germ cell markers AP-2α and AP-2γ are top enriched motifs in TCam-2, while embryonic stem cell specific motifs SOX2/OCT4 /TCF/NANOG are enriched in both cell lines.

Significantly enriched motifs were identified for each cell line and histone mark (HOMER tool, see Methods). There was strong overlap between the top-ranked enriched motifs in NCCIT and TCam-2. This was true for both activating markers. For example, for H3K4me3 the top enriched motif was MAZ for both TCam-2 and NCCIT, a transcription factor associated with MYC (binds to two sites in its promoter) and known to be involved in transcription initiation as well as termination [49] (Figure 2A,B; Table S3).

A limited number of markers showed differences in enrichment between the cell lines (Figure 2C, 2D, Table S3). For H3K4me3, five motifs were identified which showed sufficient difference. Four were higher ranked in TCam-2 and one higher in NCCIT. The top ranked TCam-2 motifs presented in this differentiating list were EBF1 (role in developmental processes), AP-2α and AP-2γ (known germ cell markers) and OCT4/SOX2/TCF/NANOG (pluripotency motif). For NCCIT, E2F1 (cell cycle control, action of tumor suppressor proteins, cell proliferation) was ranked higher. For H3K27ac, there were four differentiating motifs, of which three were ranked higher in TCam-2 and one higher in NCCIT. The OCT4/SOX2/TCF/NANOG motif was the most significantly enriched motif for H3K27ac in NCCIT (ranked 20 in TCam-2). This motif is known to be predominantly enriched in embryonic stem (ES) cells [50] as well as embryonic germ cells, which is in line with the stem cell-like origin (EC) of NCCIT and the germ cell-like origin of TCam-2 respectively. Moreover, for H3K27ac, AP-2α and AP-2γ were ranked as 3rd and 4th most enriched motif in TCam-2 (compared to 29th and 31st in NCCIT) reflecting their (embryonic) germ cell origin (Figure 2C,D). For ES cells the enrichment rankings for these two motifs were 87th and 105th [50]. These observations fit with the proposed more differentiated (germ cell lineage) cell of origin of SE as compared to EC [28], and are in line with the findings of related histone peaks (see Discussion section).

(A) H3K4me3 top enriched motifs

Transcription factor	Family	Consensus	Rank TCam-2	Rank NCCIT
Maz	Zf	GGGGGGGG	1	1
Elk4	ETS	NRYTTCCGGY	2	4
E2F4	E2F	GGCGGGAAAH	8	2
GFY-Staf	na	RACTACAATTCCCAGAAKGC	3	3
Fli1	ETS	NRYTTCCGGH	4	7
Elk1	ETS	HACTTCCGGY	5	5
EBF1	EBF	GTCCCCWGGGGA	6	30
E2F6	E2F	GGCGGGAARN	16	6
BMYB	HTH	NHAACBGYYV	7	10
ELF1	ETS	AVCCGGAAGT	14	8
YY1	Zf	CAAGATGGCGGC	10	9
FOXP1	Forkhead	NYYTGTTTACHN	9	15
YY1	Zf	CAAGATGGCGGC	10	9

(B) H3K27ac top enriched motifs

Transcription factor	Family	Consensus	Rank TCam-2	Rank NCCIT
Maz	Zf	GGGGGGGG	1	9
OCT4-SOX2-TCF-NANOG	POU/Homeobox/HMG	ATTTGCATAACAATG	20	1
EBF1	EBF	GTCCCCWGGGGA	2	19
Elk4	ETS	NRYTTCCGGY	8	2
TFAP2G (AP2γ)	AP2	HHTGSCCTSAGGSCA	3	31
Elk1	ETS	HACTTCCGGY	12	3
TFAP2A (AP2α)	AP2	ATGCCCTGAGGC	4	29
Fli1	ETS	NRYTTCCGGH	5	4
GABPA	ETS	RACCGGAAGT	10	5
Sox2	HMG	BCCATTGTTC	6	8
Sox3	HMG	CCWTTGTY	7	6
ERG	ETS	ACAGGAAGTG	9	7
ETV1	ETS	AACCGGAAGT	11	10

(C) H3K4me3 top differentiating motifs

Transcription factor	Family	Motif	Score	Rank TCam-2	Rank NCCIT
EBF1	EBF		0.24	6	30
TFAP2A (AP2α)	AP2		0.85	12	98
OCT4-SOX2-TCF-NANOG	POU/Homeobox/HMG		0.42	18	58
E2F1	E2F		0.33	72	19
TFAP2G (AP2γ)	AP2		-	20	Undetermined

(D) H3K27ac top differentiating motifs

Transcription factor	Family	Motif	Score	Rank TCam-2	Rank NCCIT
OCT4-SOX2-TCF-NANOG	POU/Homeobox/HMG		0.12	20	1
TFAP2G (AP2γ)	AP2		0.18	3	31
TFAP2A (AP2α)	AP2		0.16	4	29
FOXP1	Forkhead		0.31	19	67

Figure 2. Motif enrichment in histone modification data. All motifs were significantly enriched in target over background sequences (p<0.01). Fold enrichment is indicated relative to background. **(A,B)** Top ranking motifs in both cell lines showed strong overlap (top 10). **(C,D)** Motifs that differed strongly between the cell lines with regard to their enrichments were selected. A motif was assessed favorably if its ranking was high (≤20) for one cell line and low for the other cell line (or was absent in the other list of enriched motifs). Score: The difference in ranking was assessed based on the difference in relative position in the list ($|1-(r_{TCam-2}/n_{TCam-2}) - 1-(r_{NCCIT}/n_{NCCIT})| \geq 15\%$, n=nr of enriched motifs, r is the rank of a specific motif in the list of enriched motifs for either cell line).

Verification of the H3K4me3 & H3K27ac enrichment-based open chromatin configuration was independently confirmed using DNaseI-hypersensitivity.

As the investigated chromatin marks (H3K4me3 & H3K27ac) are considered to be associated with active chromatin, we explored whether their presence was associated with another characteristic of active chromatin: DNaseI-hypersensitivity [51]. Using a DNaseI-MLPA approach [43] we targeted 14 regions which showed the greatest differences in either H3K4me3 and/or H3K27ac enrichment between the same two cell lines. Six of seven enriched regions in NCCIT (Figure S2A: N1,N2,N3,N4,N6,N7), and five of seven enriched regions in TCam-2 (Figure S2B: S1,S2,S5,S6,S7), showed significant DNaseI-hypersensitivity in the respective cell lines. In contrast, only one H3K27ac-negative locus in NCCIT (Figure S2A, S1), and no H3K27ac-negative loci in TCam-2 (Figure S2B), showed DNaseI-hypersensitivity.

CpG Methylation

The DMRforPairs algorithm was used to identify differentially methylated regions (DMR) [38]. The algorithm was set to detect strong differences, i.e. using stringent settings. Regions containing a minimum of four probes within 200 bp distance of each other were considered for further analysis (n=30,306). Regions in which median methylation levels (M-values) between the samples differed at least |1.4| (n=5,139) were tested for statistical significance (significant: $p<0.05$; Bonferroni adjusted, Wilcoxon-rank-sum test, n=143) (Output DMRforPairs: File S1).

Methylation patterns at DMRs are in line with marker positivity in SE and EC.

Because of the activating histone modifications investigated, we focused on hypo- or absence of DNA methylation. Global methylation levels were in line with the hyper- and hypomethylated global status of NS and SE respectively, and the previously shown intermediate status of TCam-2 (Figure S3) [52, 53]. After DMR identification using DMRforPairs, a total of 99 DMRs (annotating to 170 unique gene symbols) were hypomethylated in TCam-2, compared to 44 in NCCIT (annotating to 64 unique gene symbols) (Table S1). In line with the histone modifications (see above), the *SOX2* promoter region was found to be strongly hypomethylated in NCCIT (Figure 3A). *SOX2* was partly methylated in TCam-2, in line with findings illustrating that TCam-2 can differentiate and become *SOX2* positive after extra-gonadal injection in mice [54]. A 220bp region directly upstream of the TSS of *SOX2* (chr3:181429712) has previously been shown to be completely hypomethylated in TCam-2 [55]. This is in line with our findings as a consistently hypomethylated region (chr3:181429233-181430485) is shown directly upstream of the TSS while a 652bp long DMR (chr3:181428046-181428697) between NCCIT and TCam-2 is detected by DMRforPairs in a region ca. 800bp upstream of the region sequenced by Nettersheim et al. *SOX17* did not show a significant differential methylation pattern, indicating that it is, in principle, accessible for transcription in both cell types (Figure 3B). Indeed, *SOX17* expression can be induced in NCCIT

(unpublished observation). In line with known gene expression in both cell lines, *OCT3/4* showed an inconsistent, but non-differential methylation pattern (Figure 3C). The TSS of *NANOG* was hypomethylated in both cell lines (Figure 3D), in line with the expression data and previous reports [56]. In the list of top DMRs, the miR-371/2/3 cluster stood out by significant differential hypermethylation in NCCIT (Figure 3E). The promoter region of *GATA4* was significantly hypermethylated in TCam-2 (Figure 3F). In general, 33% (47/143) DMRs were annotated to TSSs (File S1, according to Illumina's manifest) which is similar to the fraction of TSS associated regions identified by DMRforPairs (12,652/30,306). Functionally, the DMR list of both cell lines showed enrichment for (embryonic) stem cell maintenance/pluripotency. Biological functions indicating similarity to more mature germ cells were enriched in TCam-2 (Table S2).
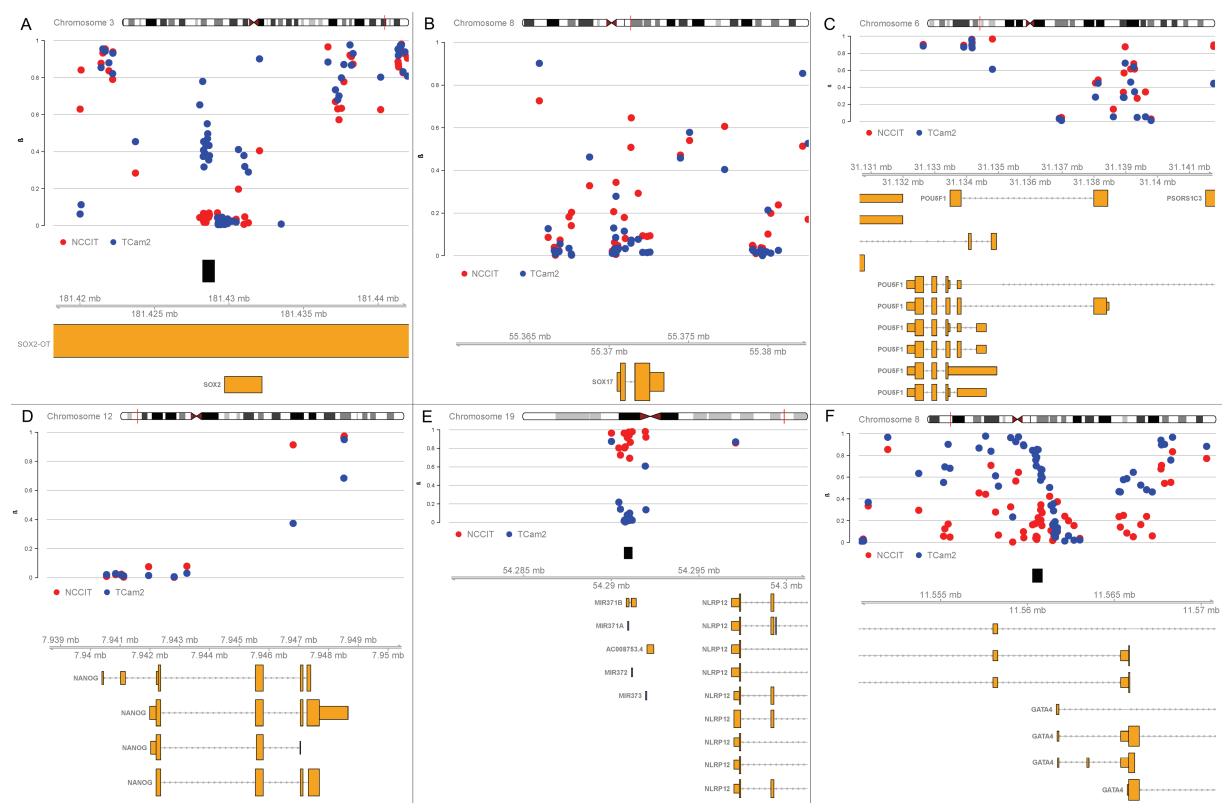
Figure 3. **Methylation patterns of known germ cell markers (A-D)** and significant DMRs for both cell lines **(E, F)**. Dots depict individual CpGs and black boxes denote DMRs identified by DMRforPairs. Percentages below indicate average CG density in the plotted regions (calculated using the Repitools[i] R package, gcContentCalc function). (A) *SOX2* [44%], (B) *SOX17* [46%], (C) *OCT3/4* (*POU5F1*), [52%] (D) *NANOG*, [44%] (E) *miR-371/2/3* cluster, [49%] (F) *GATA4* [53%].

---

[i] http://www.bioconductor.org/packages/devel/bioc/html/Repitools.html

DMRs were significantly enriched for imprinted genes, and 59% (51/86) of all imprinted genes showed loss of methylation around their TSS in one or both cell lines.

From the list of verified imprinted human genes (n=88 retrieved from geneimprint.com), 82 were also annotated in Illumina's manifest (total of 21,243 unique gene symbols annotated) and 10 were present in the top DMRs between TCam-2 and NCCIT. This overrepresentation of imprinted genes in the list of DMRs was significant (p<0.0001, χ^2 test). When investigating the region surrounding the TSS of the 86 imprinted genes with known genomic localization, 14 showed a differential status between the cell lines (8 / 6 hypomethylated in NCCIT and TCam-2 respectively). In total, 37 imprinted genes displayed a hypomethylated status in both cell lines, compared with 12 hypermethylated genes (Figure 4). In summary, these results indicate an overall erased status of the imprinted regions in both cell lines. Regarding genes with a differential methylation status, there is no clear difference in number of hypermethylated genes that would indicate a difference in maturation status or environmental disruption.

Expression

Expression levels of markers matched histological origin of both cell lines.

In total, 257 genes were expressed higher in TCam-2, compared to 149 in NCCIT (Table S1). Greater than 3.65 fold difference in expression level (99% confidence interval (CI) of log2-ratio of intensities) was considered significant. The expression levels were in agreement with the classification of the cell lines: *SOX17* was higher in TCam-2 compared with NCCIT, with the opposite observed for *SOX2*[57] (Figure 5). *OCT3/4*, a general marker for the stem cell components (SE/EC) of GCC, was expressed at equal levels in both cell lines. *NANOG* was expressed higher in TCam-2, which is in line with the open histone configuration (Figure 1D, transcription possible in both cell lines). Functionally, the gene lists of both cell lines showed enrichment for (embryonic) stem cell maintenance/pluripotency, and Wnt/β-catenin signaling. Enrichment of biological functions consistent with more mature germ cells was present in TCam-2, and absent in NCCIT. In addition, network analysis revealed the androgen pathway in TCam-2, represented by both the AR and testosterone, thus showing major overlap between the networks found by genes that had differential histone modification patterns (Table S2).
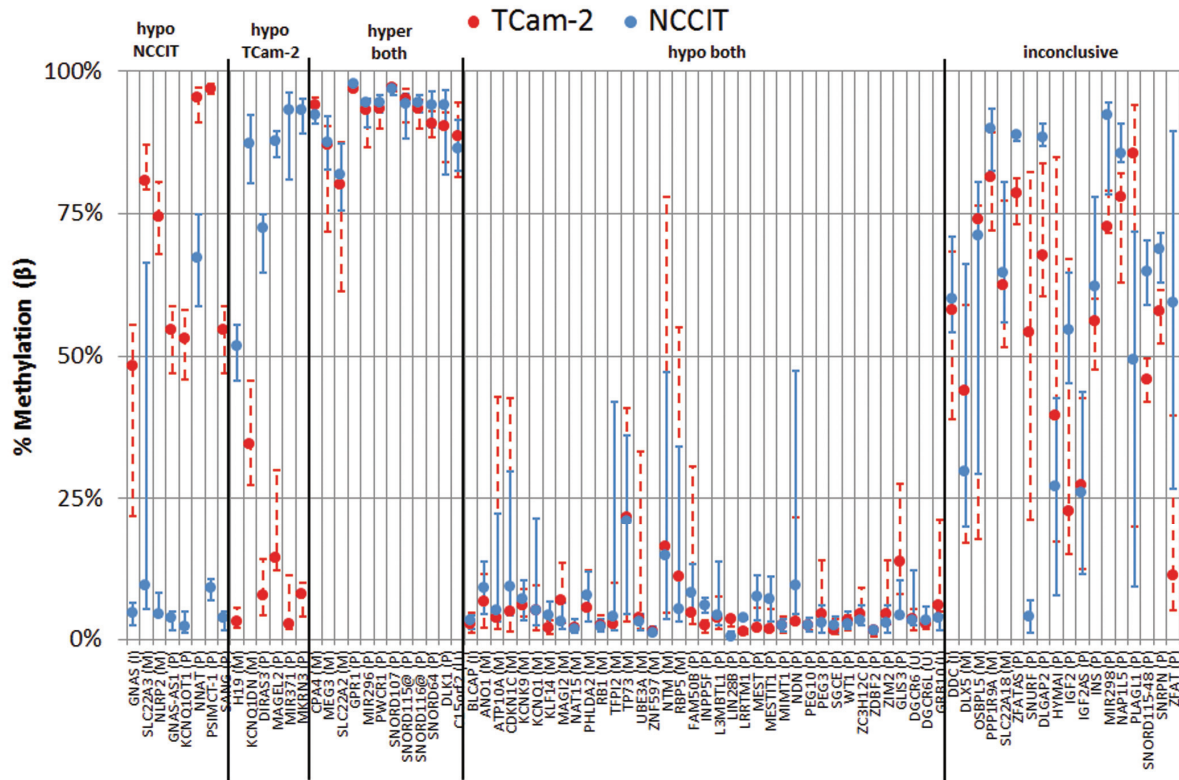
Figure 4. Methylation patterns in promoter regions of imprinted genes. Localization of TSS was retrieved from ensembl.org and manually corrected for genes with multiple transcripts to select a region with representative coverage on the Illumina BeadChip (n_{probes} varied between the genes: median=10, inter-quartile range 6-21). Promoter region was defined as TSS-1000 – TSS+100 (or opposite on reverse strand). Genes with > 0.25 difference in median β and a consistent (stable) methylation pattern were identified as differentially methylated at the TSS between the two cell lines. Median methylation <25% for both cell lines was interpreted as a hypomethylated state in both cell line. Median methylation >75% was interpreted as hypermethylation. No probes were annotated around the TSS of TCEB3C (M), RNU5D (P), SNORD108 (P), SNORD109A (P) and SNORD109B (P). SANG and GNAS-AS1 are known aliases but separate entities on geneimprint.com and are depicted separately to preserve consistency. P=paternally imprinted/expressed, M=maternally imprinted/expressed, I=isoform-dependent imprinting.

Figure 5. Expression levels of known GCC markers match the histological subtype from which the cell lines originate.

Integration of epigenetic and expression data

Differential gene lists from histone modification, methylation and expression data showed limited overlap.

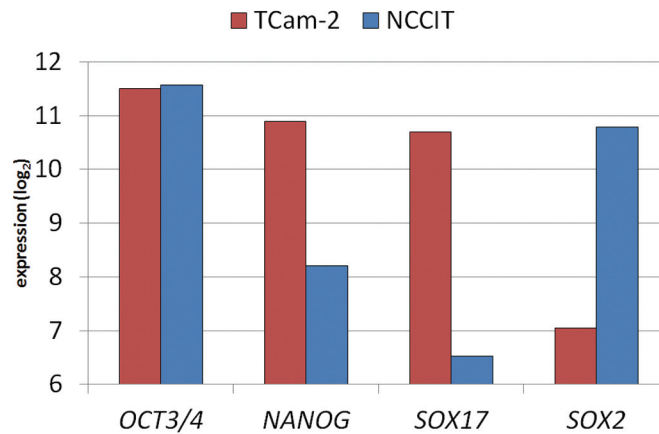Differentiating gene lists from the separate analyses discussed above were matched based on Gene Symbol to assess the relationship between active histone modifications, the absence of CpG methylation and gene expression. Overlap between gene expression and one of the epigenetic regulatory mechanisms is of interest as expression of a specific gene does not need to be regulated by both mechanisms. Figure 6 shows the overlap of the different variables for the differentiating gene lists between TCam-2 and NCCIT. In general, little overlap between relative hypomethylation / histone marker enrichment and relatively high expression is observed, but this overlap was significant (Figure S4). In TCam-2, one gene, *PRAME*, was present in all three differential lists. *H19* and *CHCHD5* were differentially hypomethylated in TCam-2 and showed high expression compared to NCCIT, but no differential enrichment for H3K27ac or H3K4me3. There were 62 genes with overlapping active chromatin marks and expression, including *SOX17* and *NANOG*. In NCCIT, three genes showed overlap between the histone marks, hypomethylation and expression. Significantly, one of these genes was *SOX2*. An additional 18 genes showed overlap between active chromatin marks and expression.
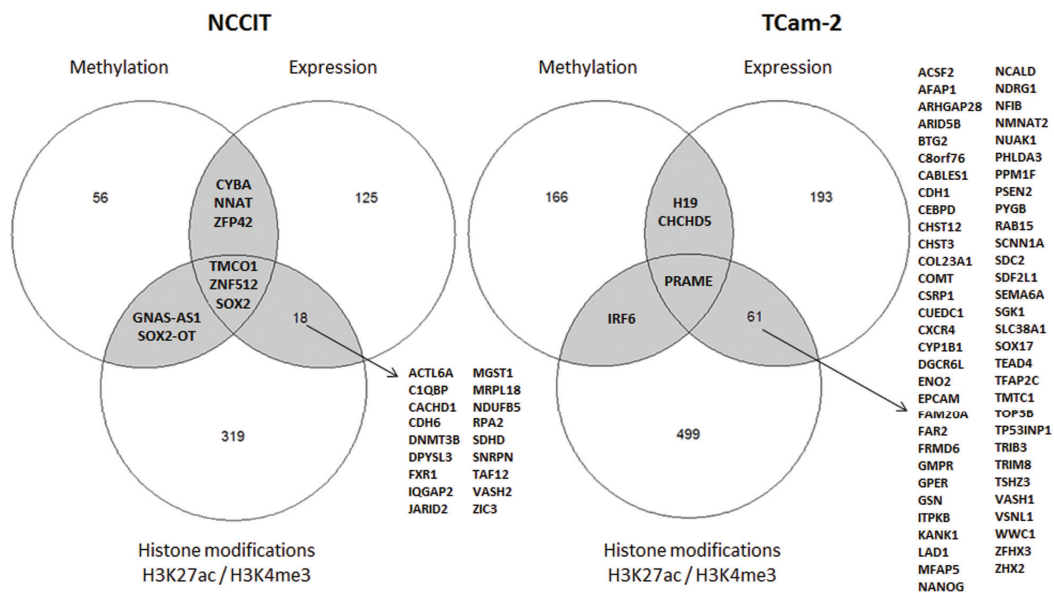
Figure 6. **Overlap between top differentiating genes (methylation/histone modification/gene expression).** (Hypo) methylation, (high) gene expression and histone marker (enrichment) should be interpreted relative to the other cell line. Criteria for selection are described in the main text. Briefly, significant differential methylation of regions with sufficient probe density was identified by DMRforPairs (frequently, but not necessarily close to, the TSS). The difference in histone modification enrichment was assessed by significant differences in summed peak heights between the cell lines. Finally, a fold difference of 3.65 (boundary of 99% CI) was used as cutoff for differential gene expression. Gene lists are presented in Table S1, and overlap was determined based on matching gene symbol.

Enrichment of both histone marks in general and absence of DNA methylation around the TSS is related to higher expression levels.

Higher expression levels are present when genes are more enriched for either histone mark. Quantification of the fraction of genes with higher than median expression at various intervals of summed peak heights confirmed the general trend towards higher expression at higher levels of enrichment (Figure 7A). DNA methylation levels were only correlated to expression around the TSS: a fully methylated TSS is predominantly associated with low gene expression levels while low methylation status is not predictive of expression level (Figure 7B).
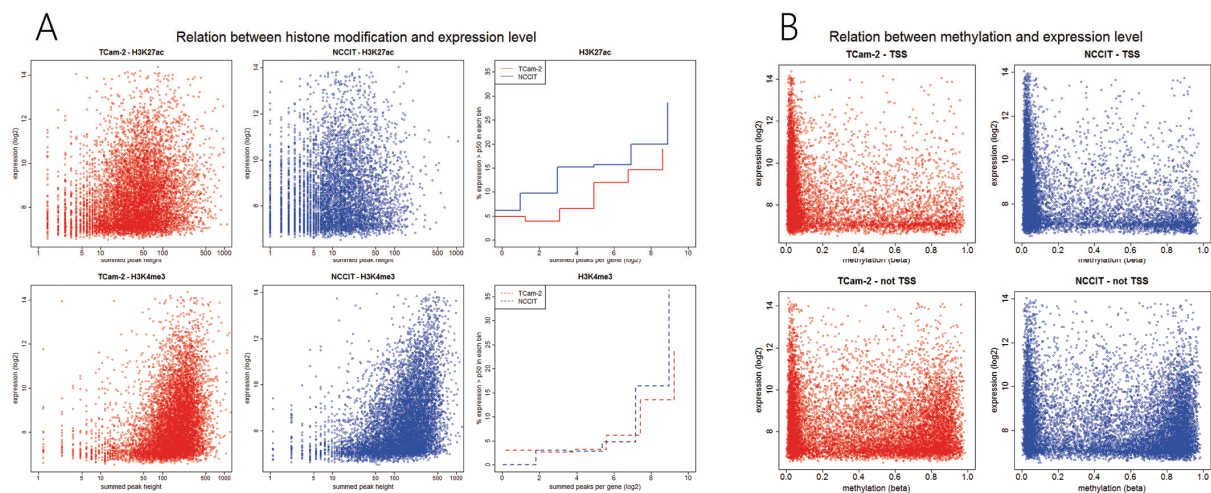
Figure 7. (A) Relation between histone modification level (summed peak heights per gene) and expression level. Top and bottom right images depict the percentage of highly (>p50) expressed genes calculated for an interval of summed peaks. For example, 5% of genes with a log2(summed peak height) of ≈5.5-7.5 were highly expressed. **(B)** Relation between CpG methylation (TSS/no TSS) and gene expression.

NCCIT and TCam-2 show a largely overlapping epigenetically open network with specific elements that are differentially regulated based on cell type.

Histone modification, methylation and gene expression data were analyzed together using hierarchical clustering. This unsupervised clustering procedure revealed specific groups of genes with a profile poised for transcription. These gene clusters showed active histone marks, combined with an activating methylation landscape and are hypothesized to contain genes accessible for transcription (e.g. epigenetically ON=transcription possible, Figure 8). Functionally, the ON network for both cell lines showed a large degree of overlap (related to the androgen pathway, lipid metabolism and pluripotency) (Table S2). There was considerable overlap between the genes found in this separate analysis and the differential gene lists (Figure 6, overlap indicated by gene symbols in Figure 8).
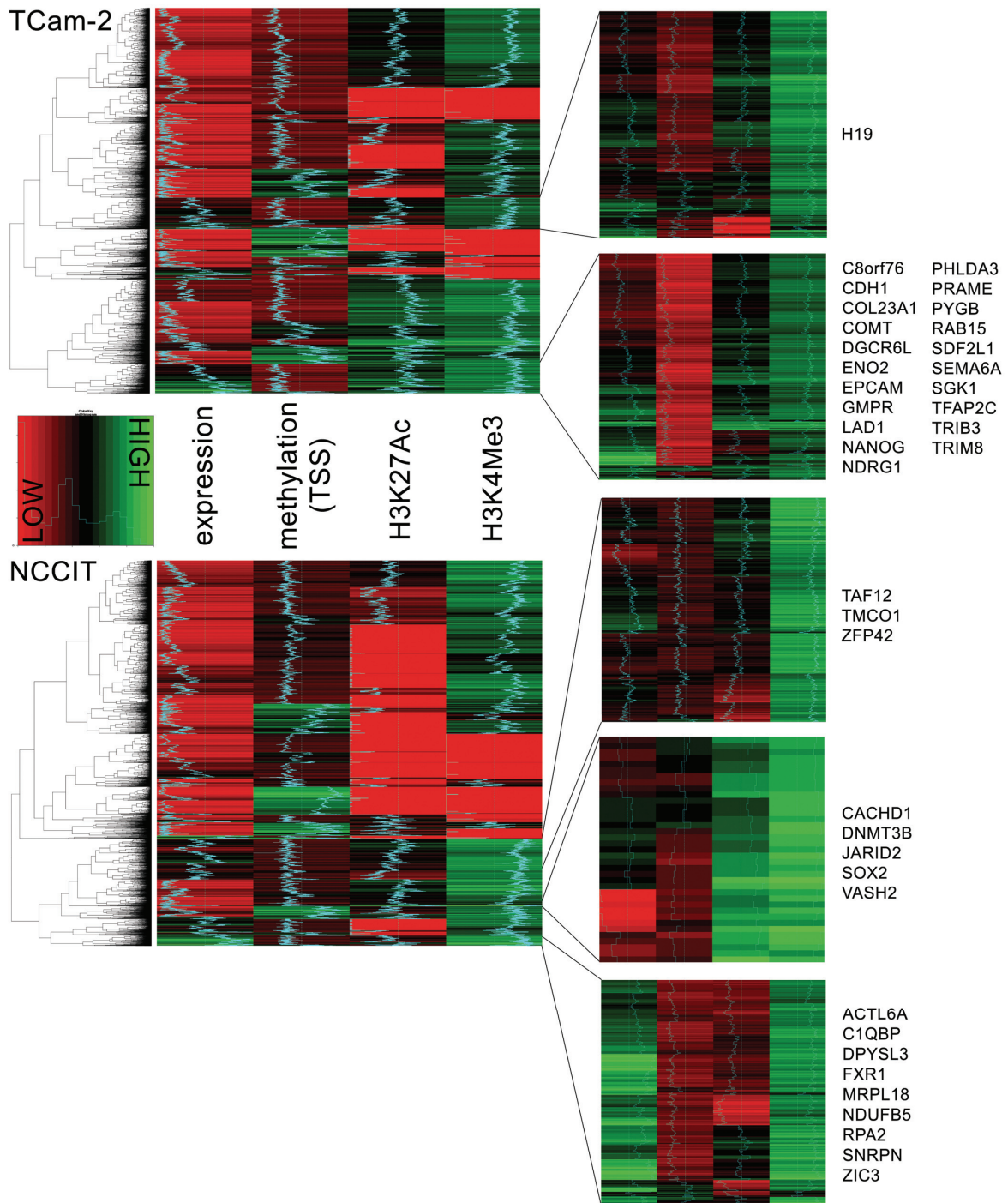
Figure 8. Heat map of epigenetic markers and gene expression profiles. Genes with quantified methylation status around their TSS (based on Illumina annotation) and valid (see Materials & Methods) measurement of their expression level were included (n=11,620). Log-2 summed peak heights per gene were used as an estimate of histone marker enrichment. Variables scaled between 0 and 1. Hierarchical clustering was performed using complete linkage. Clusters of interest were identified based on a consistent enriched state for one or more of the active histone markers and a hypomethylated state around the TSS. Number of genes in the displayed right panel (zoomed in heat maps, top→bottom): 899/892 (TCam-2) and 1,224/37/308 (NCCIT). Gene expression was allowed to vary within clusters, but clusters with almost absent expression levels (completely red) were not selected. Gene symbols indicate genes that overlap with the analysis of top differentiating genes between the cell lines (Figure 6). Gene symbols are listed alphabetically. An indication of the level for each gene in each column is presented by the color/shade and a blue line (for each column: left=0, right=1).

Discussion

Histone and DNA methylation signatures were studied to explore the epigenetic differences between representative cell lines for the GCC histological subtypes and their relation to expression. TCam-2 and NCCIT cells were used as representatives of SE and EC respectively. Our study includes the generation, integration and interpretation of genome-wide profiles for histone marks (H3K4me3 and H3K27ac), DNA methylation and gene expression. H3K4me3 and H3K27ac are well-characterized markers for active promoter and enhancer sites, respectively. Analysis of the histone marks matched the classification of the cell-lines: *SOX17* was strongly enriched for H3K4me3 and H3K27ac in TCam-2 compared to NCCIT cells, whereas the opposite pattern was observed for *SOX2*. Motifs for germ cell-specific transcription factors AP-2α/γ were enriched in TCam-2, but not in NCCIT. Methylation profiling showed *SOX2* to be in the top DMRs, being more methylated in TCam-2 as expected. In addition, *SOX2* and *SOX17* expression levels confirmed and matched the previously described patterns in SE and EC [2]. General SE/EC markers (*NANOG* and *OCT3/4*) showed expression patterns compatible with their epigenetic configurations [58]. Moreover, imprinting patterns confirm the suggested overall erased status of genomic imprinting in both cell lines, in line with the situation in early germ cells [59].

The cell lines were studied for their differences with regard to epigenetic marks and expression. In line with the early germ cell origin of both cell lines (NCCIT more stem cell-like than TCam-2) canonical pathways related to stem cell maintenance and regulation were significantly overrepresented for all three variables studied (NCCIT more pronounced than TCam-2). In TCam-2 (differential histone modification) there was strong overrepresentation of genes involved in IGF1 signaling, a pathway that is implicated in maintenance of spermatogonia [60]. Moreover, genes involved in germ cell – Sertoli cell junction signaling were significantly overrepresented in this cell line, fitting with a more mature type of germ cell depending on the Sertoli cell niche [2]. In the list of top DMRs, the miR-371/2/3 cluster stood out by significant differential hypermethylation in NCCIT. These miRs have shown to be specific biomarkers for GCC in serum and tumor tissue [61-65]. Even though GCCs are reported to express these embryonic miRs [63, 66, 67], NCCIT has been shown to exhibit low expression levels due to the absence of a functional TP53 pathway (i.e. lacking the need to inactivate this pathway by miR-372/3 expression via LATS2 inhibition) [67].

Pathway analysis using IPA revealed a network including the androgen receptor (AR) and testosterone targets enriched for open chromatin configuration marks in TCam-2. Such enrichment was also identified based on the expression data of TCam-2, and to a lesser extent in NCCIT. Despite of this observation, no differential AR expression was present between the cell lines, and no enrichment of DMRs in AR targets was identified. Future experiments are needed to validate the differential role of the AR pathway between NCCIT and TCam-2 and their *in vivo* counterparts. A second network was closely related, focused on lipid metabolism (LEP central, late germ cell

differentiation and survival [68]). Androgens are well-known environmental and physiological factors that influence epigenetic marks and phenotypes. They are related to GCC risk and are considered to be crucial for the progression of germ cell development [28] Ammerpohl *et al.* found significant enrichment of hypermethylated AR target genes in androgen insensitivity syndrome (AIS) patients versus controls [69]. This is consistent with earlier reports that diminished gene activation (due to an *AR* mutation) results in subsequent increased DNA methylation of target genes [70, 71], linking DSD and GCC at the epigenetic level.

On the other hand, functional analysis for NCCIT predominantly revealed genes involved in embryonic stem cell maintenance (less pronounced in TCam-2). More specifically, an interaction between *SOX2* and *DMRT1* was the most important network identified. Both are involved in stem cell maintenance in embryonic (mouse) germ cells [28]. It has been reported that Dmrt1 can bind the mouse Sox2 promoter, and Dmrt1 controls expression of Sox2 and other pluripotency genes (such as Nanog and Oct3/4) in the embryonic testis, in part via transcriptional repression [72]. In addition, *DMRT1* has a role in sex determination, as it prevents female reprogramming in the postnatal mammalian testis [73]. A study by Murphy et al confirmed the influence of *DMRT1* on *SOX2* expression while no change in expression of *SOX17* was observed [74]. However, these observations in mice are not necessarily representative for the human situation. No consistent networks related to progression of germ cell differentiation (as with AR in TCam-2) were identified in NCCIT.

As stated above, motif enrichment based on the histone modification data showed that the AP-2α and AP-2γ motifs were enriched in TCam-2 only. The ETS family was enriched in both TCam-2 and NCCIT. AP-2γ is a known germ cell marker, abundantly expressed in CIS and SE, and heterogeneously expressed in NS and somatic tumors [75, 76]. *AP-2γ* and *KIT* are co-expressed in gonocytes [76], which could point to a direct regulation loop which supports proliferation, and agrees with the observation that TCam-2 is more germ cell-like than NCCIT. AP-2γ would then serve as a molecule that keeps fetal germ cells in a pluripotent state by suppressing differentiation and supporting proliferation [76, 77]. *AP-2γ* expression is induced by estrogens [78] and *AP-2α* and *AP-2γ* are able to induce changes in the chromatin structure known to be associated with *ERα* (*ESR1*) transcription [79]. The importance of the androgen-estrogen balance is also indicated by the strongly androgen/estrogen-centered gene networks identified in this study. Additionally, the ETS family was present in the top motif enrichments for both TCam-2 and NCCIT. Recently it was shown that overexpression of ETS, combined with loss of *PTEN*, increases AR binding and restores AR transcriptional activity in prostate [80]. Indeed, disruption of the PTEN pathway has been suggested to be part of the pathogenesis of GCC [81].

Conclusions

In conclusion, this study provides an integrated analysis of the functional genome in GCC cell lines. Our data show that known germ cell markers are not only present and differentiating between SE and NS at the expression level, but also in the epigenetic landscape. The overall similarity between TCam-2 / NCCIT support an erased embryonic germ cell arrested in early gonadal development as common cell of origin although the exact developmental stage from which the tumor cells are derived might differ. Indeed, subtle difference in the (integrated) epigenetic and expression profiles indicate TCam-2 to exhibit a more germ cell-like profile, whereas NCCIT shows a more pluripotent phenotype. Future research has already been initiated to investigate primary cancer samples from patients to confirm and further expand the integrated epigenetic EC and SE footprints identified in this study.

Acknowledgements

Supplementary information

Figure S1. Number of top differentially modified regions between TCam-2 and NCCIT, and their overlap between H3K27ac and H3K4me3 based on associated genes. Genes with significant differences (outside 99% confidence interval) in summed peak height per gene (ΔΣP) were identified as top-differentially modified.

Figure S2. Normalized ratios for each of the 14 loci tested in the MLPA-DNaseI assay. A threshold of < 0.75 was defined for DNaseI-hypersensitivity. N=enriched in ChIP-seq analysis in NCCIT (non-seminoma cell model), S=enriched in ChIP-seq analysis in TCam-2 (seminoma model). (A) Analysis of NCCIT cells. (B) Analysis of TCam-2 cells. (C) Overview of loci and interpretation of results. DNaseI hypersensitivity is indicated if present in the cell line in which marker enrichment was also found in the ChIP-seq analysis.

Figure S3. Visualization of global methylation patterns in both cell lines. Depicted is a violin plot of the distribution of methylation values (β) for both cell lines. In general, NS are considered globally hypermethylated in comparison to SE but TCam-2 is known to show an intermediate phenotype with regard to global methylation status (see Wermann et al 2010 and Netto et al 2008 in reference list). Indeed, significantly lower methylation levels were detected in TCam-2 but the quantitative difference in methylation distribution was very moderate ($p<0.01$, Mann Whitney U test, median$_\beta$ (1st-3rd quantile$_\beta$): 51%$_{TCam-2}$ (46%-84%) versus 63%$_{NCCIT}$ (58%-87%)).

Figure S4. Venn diagrams analogous to Figure 6 corrected for gene symbols that are not represented by valid measurements in the expression or methylation data (histone modification = genome wide assessment). NCCIT: 33 genes differentially methylated were not annotated in the expression data. 17 overexpressed genes were not annotated in the methylation data. 97 genes showing differential histone modifications were not present in the expression or methylation data. For TCam-2 these numbers were 101/28/198. Based on an empirical probability distribution we assessed random overlap using 10,000 draws from simulated genelists with $n_{expression\ (EXPR)}$=14,525, $n_{histone-modification\ (HM)}$=22,000 and $n_{methylation\ (MEHTY)}$=21,243 genes. These numbers correspond with the number of genes with valid measurements on the arrays (histone modification: genome wide proxy). Significant overlap indicates more overlapping genes identified in these venn diagrams than would we expected based on random subsets of genes. Significant overlap is indicated with a * ($p<0.05$). p-values TCam-2: p_{EXPR_HM} = <0.0001, p_{EXPR_METHY} = 0.0370-0.1604, p_{HM_METHY} = 0.3229-0.6860, p_{all3} = 0.0003-0.0167. p-values NCCIT: p_{EXPR_HM} = 0-0, p_{EXPR_METHY} = 0.0001-0.0007, p_{HM_METHY} = 0.3110-1.0000, p_{all3} = <0.0001. (P-values are ranges if in the repeated random draws used to construct the empirical cumulative distribution function a specific count of overlapping genes occurred more than once.)

Table S1. Top differentiating genes in histone modification, CpG methylation and gene expression analyses.

Table S2. Results of IPA / functional analysis of (differentiating) gene lists. IPA was performed using the default settings including "testis" as specific tissue of interest, and only incorporating experimentally observed evidence. Green fill indicates overlap between TCam-2 and NCCIT. Reported log(p) values are the result of IPA's internal enrichment tests.

Table S3. Detailed results of motif enrichment analysis (HOMER) per cell line and per histone mark.

File S1. ZIP file containing DMRforPairs output for significant regions. Please start from the html files.

References

1.   Looijenga LH: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J Pathol* 2009, 218:146-162.
2.   Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5:210-222.
3.   Woodward PJ, Heidenreich A, Looijenga LHJ, Oosterhuis JW, McLeod DG, Moller H: **Testicular germ cell tumors.** In *World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs.* Edited by Eble JN, Sauter G, Epstein JI, Sesterhann IA. Lyon: IARC Press; 2004: 217-278
4.   Looijenga LH, Gillis AJ, Stoop H, Biermann K, Oosterhuis JW: **Dissecting the molecular pathways of (testicular) germ cell tumour pathogenesis; from initiation to treatment-resistance.** *International journal of andrology* 2011, 34:e234-251.
5.   Banks K, Tuazon E, Berhane K, Koh CJ, De Filippo RE, Chang A, Kim SS, Daneshmand S, Davis-Dao C, Lewinger JP, et al: **Cryptorchidism and testicular germ cell tumors: comprehensive meta-analysis reveals that association between these conditions diminished over time and is modified by clinical characteristics.** *Frontiers in endocrinology* 2012, 3:182.
6.   Dieckmann KP, Pichlmeier U: **Clinical epidemiology of testicular germ cell tumors.** *World journal of urology* 2004, 22:2-14.
7.   Elzinga-Tinke JE, Sirre ME, Looijenga LH, van Casteren N, Wildhagen MF, Dohle GR: **The predictive value of testicular ultrasound abnormalities for carcinoma in situ of the testis in men at risk for testicular cancer.** *International journal of andrology* 2010, 33:597-603.
8.   Horwich A, Shipley J, Huddart R: **Testicular germ-cell cancer.** *Lancet* 2006, 367:754-765.
9.   Czene K, Lichtenstein P, Hemminki K: **Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database.** *Int J Cancer* 2002, 99:260-266.
10.   Del-Mazo J, Brieno-Enriquez MA, Garcia-Lopez J, Lopez-Fernandez LA, De-Felici M: **Endocrine disruptors, gene deregulation and male germ cell tumors.** *The International journal of developmental biology* 2013, 57:225-239.
11.   Godmann M, Lambrot R, Kimmins S: **The dynamic epigenetic program in male germ cells: Its role in spermatogenesis, testis cancer, and its response to the environment.** *Microsc Res Tech* 2009, 72:603-619.
12.   Kanetsky PA, Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, Letrero R, Ciosek SL, Doody DR, Smith LM, Weaver J, et al: **Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer.** *Nat Genet* 2009, 41:811-815.
13.   Kanetsky PA, Mitra N, Vardhanabhuti S, Vaughn DJ, Li M, Ciosek SL, Letrero R, D'Andrea K, Vaddi M, Doody DR, et al: **A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility.** *Hum Mol Genet* 2011, 20:3109-3117.
14.   Rapley EA, Turnbull C, Al Olama AA, Dermitzakis ET, Linger R, Huddart RA, Renwick A, Hughes D, Hines S, Seal S, et al: **A genome-wide association study of testicular germ cell tumor.** *Nat Genet* 2009, 41:807-810.
15.   Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, Ricketts M, Linger R, Nsengimana J, Deloukas P, et al: **Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer.** *Nat Genet* 2010, 42:604-607.
16.   Kratz CP, Mai PL, Greene MH: **Familial testicular germ cell tumours.** *Best Pract Res Clin Endocrinol Metab* 2010, 24:503-513.
17.   Bignell G, Smith R, Hunter C, Stephens P, Davies H, Greenman C, Teague J, Butler A, Edkins S, Stevens C, et al: **Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults.** *Genes Chromosomes Cancer* 2006, 45:42-46.
18.   Cools M, Drop SL, Wolffenbuttel KP, Oosterhuis JW, Looijenga LH: **Germ cell tumors in the intersex gonad: old paths, new directions, moving frontiers.** *Endocr Rev* 2006, 27:468-484.
19.   Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, Lee LK, Ngo QM, Klugman S, Ching KA, Edsall LE, Ye Z, et al: **Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency.** *Cell Res* 2011, 21:1393-1409.
20.   De Carvalho DD, Sharma S, You JS, Su SF, Taberlay PC, Kelly TK, Yang X, Liang G, Jones PA: **DNA methylation screening identifies driver epigenetic events of cancer cell survival.** *Cancer Cell* 2012, 21:655-667.
21.   Esteller M: **Epigenetics in cancer.** *The New England journal of medicine* 2008, 358:1148-1159.
22.   Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, et al: **Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.** *Cell stem cell* 2010, 6:479-491.
23.   Jones PA, Baylin SB: **The epigenomics of cancer.** *Cell* 2007, 128:683-692.
24.   Suva ML, Riggi N, Bernstein BE: **Epigenetic reprogramming in cancer.** *Science* 2013, 339:1567-1570.

25.   Van Der Zwan YG, Stoop H, Rossello F, White SJ, Looijenga LH: **Role of epigenetics in the etiology of germ cell cancer.** *The International journal of developmental biology* 2013, 57**:**299-308.

26.   Seisenberger S, Peat JR, Reik W: **Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells.** *Curr Opin Cell Biol* 2013, 25**:**281-288.

27.   Amatruda JF, Ross JA, Christensen B, Fustino NJ, Chen KS, Hooten AJ, Nelson H, Kuriger JK, Rakheja D, Frazier AL, Poynter JN: **DNA methylation analysis reveals distinct methylation signatures in pediatric germ cell tumors.** *BMC cancer* 2013, 13**:**313.

28.   Looijenga LH, Van Agthoven T, Biermann K: **Development of malignant germ cells - the genvironmental hypothesis.** *The International journal of developmental biology* 2013, 57**:**241-253.

29.   Mirbahai L, Chipman JK: **Epigenetic memory of environmental organisms: a reflection of lifetime stressor exposures.** *Mutation research Genetic toxicology and environmental mutagenesis* 2014, 764-765**:**10-17.

30.   Skinner MK, Guerrero-Bosagna C, Haque M, Nilsson E, Bhandari R, McCarrey JR: **Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and the subsequent germ line.** *PloS one* 2013, 8**:**e66318.

31.   de Jong J, Stoop H, Gillis AJ, Hersmus R, van Gurp RJ, van de Geijn GJ, van Drunen E, Beverloo HB, Schneider DT, Sherlock JK, et al: **Further characterization of the first seminoma cell line TCam-2.** *Genes Chromosomes Cancer* 2008, 47**:**185-196.

32.   Mizuno Y, Gotoh A, Kamidono S, Kitazawa S: **[Establishment and characterization of a new human testicular germ cell tumor cell line (TCam-2)].** *Nihon Hinyokika Gakkai zasshi The japanese journal of urology* 1993, 84**:**1211-1218.

33.   Teshima S, Shimosato Y, Hirohashi S, Tome Y, Hayashi I, Kanazawa H, Kakizoe T: **Four new human germ cell tumor cell lines.** *Laboratory investigation; a journal of technical methods and pathology* 1988, 59**:**328-336.

34.   Eckert D, Nettersheim D, Heukamp LC, Kitazawa S, Biermann K, Schorle H: **TCam-2 but not JKT-1 cells resemble seminoma in cell culture.** *Cell Tissue Res* 2008, 331**:**529-538.

35.   Du P, Kibbe WA, Lin SM: **lumi: a pipeline for processing Illumina microarray.** *Bioinformatics* 2008, 24**:**1547-1548.

36.   Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerstrom-Billai F, Jagodic M, Sundberg CJ, Ekstrom TJ, Teschendorff AE, Tegner J, Gomez-Cabrero D: **An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform.** *Epigenetics : official journal of the DNA Methylation Society* 2013, 8**:**333-346.

37.   Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S: **A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.** *Bioinformatics* 2013, 29**:**189-196.

38.   Rijlaarsdam MA, van der Zwan YG, Dorssers LC, Looijenga LH: **DMRforPairs: identifying differentially methylated regions between unique samples using array based methylation profiles.** *BMC Bioinformatics* 2014, 15**:**141.

39.   Schmid R, Baum P, Ittrich C, Fundel-Clemens K, Huber W, Brors B, Eils R, Weith A, Mennerich D, Quast K: **Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3.** *BMC Genomics* 2010, 11**:**349.

40.   Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.** *Mol Cell* 2010, 38**:**576-589.

41.   White SJ, Breuning MH, den Dunnen JT: **Detecting copy number changes in genomic DNA: MAPH and MLPA.** *Methods in cell biology* 2004, 75**:**751-768.

42.   Ohnesorg T, Eggers S, Leonhard WN, Sinclair AH, White SJ: **Rapid high-throughput analysis of DNaseI hypersensitive sites using a modified Multiplex Ligation-dependent Probe Amplification approach.** *Bmc Genomics* 2009, 10**:**412.

43.   Ohnesorg T, Eggers S, White SJ: **Detecting DNaseI-hypersensitivity sites with MLPA.** *Methods Mol Biol* 2012, 786**:**201-210.

44.   Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, 470**:**279-283.

45.   Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, Lee JH, Skalnik D, Bird A: **Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells.** *Genes Dev* 2012, 26**:**1714-1728.

46.   Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA: **A chromatin landmark and transcription initiation at most promoters in human cells.** *Cell* 2007, 130**:**77-88.

47.   Rada-Iglesias A, Wysocka J: **Epigenomics of human embryonic stem cells and induced pluripotent stem cells: insights into pluripotency and implications for disease.** *Genome Med* 2011, 3**:**36.

48.   de Jong J, Stoop H, Dohle GR, Bangma CH, Kliffen M, van Esser JW, van den Bent M, Kros JM, Oosterhuis JW, Looijenga LH: **Diagnostic value of OCT3/4 for pre-invasive and invasive testicular germ cell tumours.** *J Pathol* 2005, 206**:**242-249.

49.   Bossone SA, Asselin C, Patel AJ, Marcu KB: **MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination.** *Proceedings of the National Academy of Sciences of the United States of America* 1992, 89:7452-7456.

50.   Consortium EP: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, 489:57-74.

51.   Shu W, Chen H, Bo X, Wang S: **Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains.** *Nucleic Acids Res* 2011, 39:7428-7443.

52.   Wermann H, Stoop H, Gillis AJ, Honecker F, van Gurp RJ, Ammerpohl O, Richter J, Oosterhuis JW, Bokemeyer C, Looijenga LH: **Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance.** *J Pathol* 2010, 221:433-442.

53.   Netto GJ, Nakai Y, Nakayama M, Jadallah S, Toubaji A, Nonomura N, Albadine R, Hicks JL, Epstein JI, Yegnasubramanian S, et al: **Global DNA hypomethylation in intratubular germ cell neoplasia and seminoma, but not in nonseminomatous male germ cell tumors.** *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2008, 21:1337-1344.

54.   Nettersheim D, Westernstroer B, Haas N, Leinhaas A, Brustle O, Schlatt S, Schorle H: **Establishment of a versatile seminoma model indicates cellular plasticity of germ cell tumor cells.** *Genes Chromosomes Cancer* 2012, 51:717-726.

55.   Nettersheim D, Gillis A, Biermann K, Looijenga LH, Schorle H: **The seminoma cell line TCam-2 is sensitive to HDAC inhibitor depsipeptide but tolerates various other chemotherapeutic drugs and loss of NANOG expression.** *Genes Chromosomes Cancer* 2011, 50:1033-1042.

56.   Nettersheim D, Biermann K, Gillis AJ, Steger K, Looijenga LH, Schorle H: **NANOG promoter methylation and expression correlation during normal and malignant human germ cell development.** *Epigenetics : official journal of the DNA Methylation Society* 2011, 6:114-122.

57.   de Jong J, Stoop H, Gillis AJ, van Gurp RJ, van de Geijn GJ, Boer M, Hersmus R, Saunders PT, Anderson RA, Oosterhuis JW, Looijenga LH: **Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.** *J Pathol* 2008, 215:21-30.

58.   Rijlaarsdam MA, van Herk HA, Gillis AJ, Stoop H, Jenster G, Martens J, van Leenders GJ, Dinjens W, Hoogland AM, Timmermans M, Looijenga LH: **Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours.** *Br J Cancer* 2011, 105:854-863.

59.   MacDonald WA, Mann MR: **Epigenetic regulation of genomic imprinting from germ line to preimplantation.** *Mol Reprod Dev* 2014, 81:126-140.

60.   Huang YH, Chin CC, Ho HN, Chou CK, Shen CN, Kuo HC, Wu TJ, Wu YC, Hung YC, Chang CC, Ling TY: **Pluripotency of mouse spermatogonial stem cells maintained by IGF-1- dependent pathway.** *FASEB J* 2009, 23:2076-2087.

61.   Belge G, Dieckmann KP, Spiekermann M, Balks T, Bullerdiek J: **Serum levels of microRNAs miR-371-3: a novel class of serum biomarkers for testicular germ cell tumors?** *Eur Urol* 2012, 61:1068-1069.

62.   Dieckmann KP, Spiekermann M, Balks T, Flor I, Loning T, Bullerdiek J, Belge G: **MicroRNAs miR-371-3 in serum as diagnostic tools in the management of testicular germ cell tumours.** *Br J Cancer* 2012, 107:1754-1760.

63.   Gillis AJ, Rijlaarsdam MA, Eini R, Dorssers LC, Biermann K, Murray MJ, Nicholson JC, Coleman N, Dieckmann KP, Belge G, et al: **Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle.** *Molecular oncology* 2013, 7:1083-1092.

64.   Murray MJ, Coleman N: **Testicular cancer: a new generation of biomarkers for malignant germ cell tumours.** *Nat Rev Urol* 2012, 9:298-300.

65.   Murray MJ, Halsall DJ, Hook CE, Williams DM, Nicholson JC, Coleman N: **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.

66.   Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW, Sun Y, Chen C, Guenther S, Sherlock J, Veltman I, Baeten J, et al: **High-throughput microRNAome analysis in human germ cell tumours.** *J Pathol* 2007, 213:319-328.

67.   Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, et al: **A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.** *Cell* 2006, 124:1169-1181.

68.   Bhat GK, Sea TL, Olatinwo MO, Simorangkir D, Ford GD, Ford BD, Mann DR: **Influence of a leptin deficiency on testicular morphology, germ cell apoptosis, and expression levels of apoptosis-related genes in the mouse.** *J Androl* 2006, 27:302-310.

69.   Ammerpohl O, Bens S, Appari M, Werner R, Korn B, Drop SL, Verheijen F, van der Zwan Y, Bunch T, Hughes I, et al: **Androgen receptor function links human sexual dimorphism to DNA methylation.** *PloS one* 2013, 8:e73288.

70.   Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al: **DNA-binding factors shape the mouse methylome at distal regulatory regions.** *Nature* 2011, 480**:**490-495.

71.   Clark SJ, Melki J: **DNA methylation and gene silencing in cancer: which is the guilty party?** *Oncogene* 2002, 21**:**5380-5387.

72.   Krentz AD, Murphy MW, Kim S, Cook MS, Capel B, Zhu R, Matin A, Sarver AL, Parker KL, Griswold MD, et al: **The DM domain protein DMRT1 is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106**:**22323-22328.

73.   Matson CK, Murphy MW, Sarver AL, Griswold MD, Bardwell VJ, Zarkower D: **DMRT1 prevents female reprogramming in the postnatal mammalian testis.** *Nature* 2011, 476**:**101-104.

74.   Murphy MW, Sarver AL, Rice D, Hatzi K, Ye K, Melnick A, Heckert LL, Zarkower D, Bardwell VJ: **Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog DMRT1 in the juvenile testis.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107**:**13360-13365.

75.   Hoei-Hansen CE, Nielsen JE, Almstrup K, Sonne SB, Graem N, Skakkebaek NE, Leffers H, Rajpert-De Meyts E: **Transcription factor AP-2gamma is a developmentally regulated marker of testicular carcinoma in situ and germ cell tumors.** *Clin Cancer Res* 2004, 10**:**8521-8530.

76.   Weber S, Eckert D, Nettersheim D, Gillis AJ, Schafer S, Kuckenberg P, Ehlermann J, Werling U, Biermann K, Looijenga LH, Schorle H: **Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance.** *Biol Reprod* 2010, 82**:**214-223.

77.   Schemmer J, Arauzo-Bravo MJ, Haas N, Schafer S, Weber SN, Becker A, Eckert D, Zimmer A, Nettersheim D, Schorle H: **Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice.** *PloS one* 2013, 8**:**e71113.

78.   Orso F, Cottone E, Hasleton MD, Ibbitt JC, Sismondi P, Hurst HC, De Bortoli M: **Activator protein-2gamma (AP-2gamma) expression is specifically induced by oestrogens through binding of the oestrogen receptor to a canonical element within the 5'-untranslated region.** *The Biochemical journal* 2004, 377**:**429-438.

79.   Schuur ER, McPherson LA, Yang GP, Weigel RJ: **Genomic structure of the promoters of the human estrogen receptor-alpha gene demonstrate changes in chromatin structure induced by AP2gamma.** *J Biol Chem* 2001, 276**:**15519-15526.

80.   Demichelis F, Attard G: **A step toward functionally characterized prostate cancer molecular subtypes.** *Nature medicine* 2013, 19**:**966-967.

81.   Andreassen KE, Kristiansen W, Karlsson R, Aschim EL, Dahl O, Fossa SD, Adami HO, Wiklund F, Haugen TB, Grotmol T: **Genetic variation in AKT1, PTEN and the 8q24 locus, and the risk of testicular germ cell tumor.** *Hum Reprod* 2013, 28**:**1995-2002.

Genome wide DNA methylation profiles provide clues to the origin and pathogenesis of germ cell tumors

Martin A. Rijlaarsdam[1], David M.J. Tax[2], Ad J.M. Gillis[1], Lambert C.J. Dorssers[1], Devin C. Koestler[3], Jeroen de Ridder[2], Leendert H.J. Looijenga[1]

Department of Pathology, Erasmus MC Cancer Institute - University Medical Center, Rotterdam, The Netherlands [1]
Faculty of Electrical Engineering, Mathematics and Computer Science Intelligent Systems,
Delft Bioinformatics Lab, Technical University of Delft, Delft, The Netherlands [2]
Department of Biostatistics, University of Kansas Medical Center, Kansas City, United States of America [3]
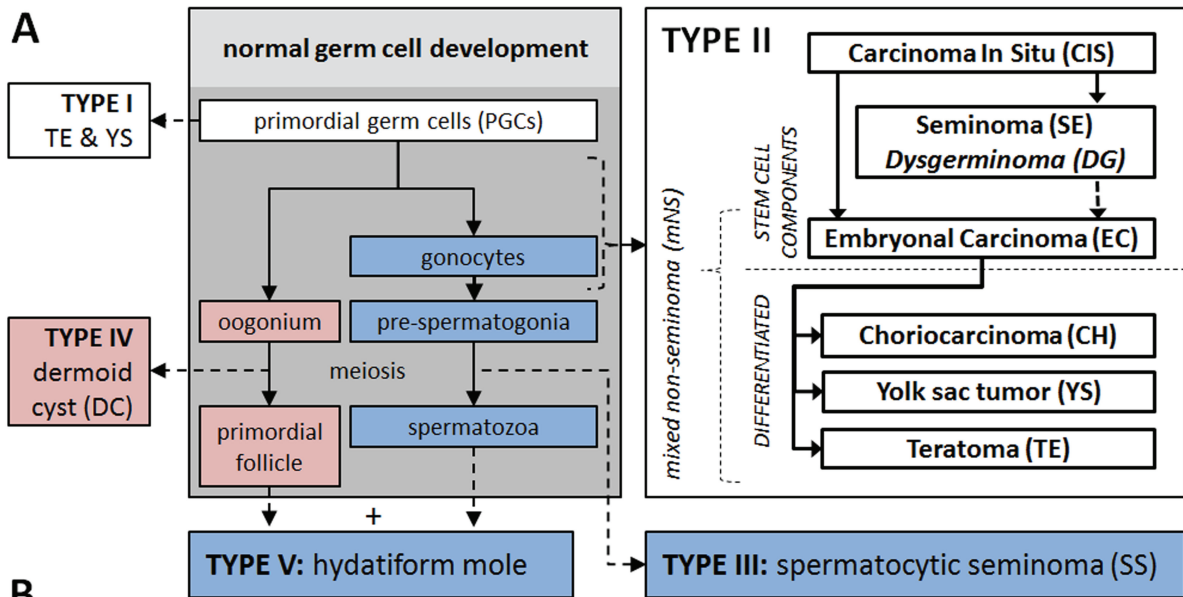
Abstract

The cell of origin (CoO) of the five subtypes of germ cell tumors (GCT,I-V) are assumed to be germ cells from different maturation stages. This is (potentially) reflected in their methylation status as fetal maturing primordial germ cells are globally demethylated during migration from the yolk sac to the gonad. Imprinted regions are erased in the gonad and later become uniparentally imprinted according to fetal sex. Here, 91 GCTs (I-IV) and four cell lines were profiled (Illumina's HumanMethylation450 BeadChip). Data was pre-processed controlling for cross hybridization, SNPs, poor detection rate, probe-type bias and batch effects. The annotation was extended, covering snRNAs/microRNAs, repeat elements and imprinted regions. A Hidden Markov Model-based genome segmentation was devised to identify differentially methylated genomic regions. Methylation profiles allowed for separation of clusters of non-seminomas (type II, NS), seminomas/dysgerminomas (II, SE/DG), spermatocytic seminomas (III, SS) and teratomas/dermoid cysts (I, TE/IV, DC). SE/DG/SS were globally hypomethylated, in line with previous reports and the demethylated state of their precursor. Differential methylation between subtypes reflected the presumed CoO as did imprinting status. Ovarian type I TE and DC: (partial) sex specific uniparental maternal imprinting. SS: uniparental paternal imprinting. Testicular TE: partial imprinting erasure. Somatic imprinting in type II GCT may indicate an earlier CoO than previously described: after global demethylation but before imprinting erasure. This is in line with the totipotent/embryonic stem cell like potential of type II GCTs and rare extra-gonadal localization. The results support the common origin of the type I TEs and show strong similarity between ovarian type I TE and DC. In conclusion, we identified specific and global methylation differences between GCT subtypes, providing insight into the developmental timing and underlying developmental biology of GCTs and their CoO, relating them closely to (early) developing embryonic germ cells. Data and extended annotation manifest deposited at GEO:GSE58538,GPL18809.

Introduction

During fetal development primordial germ cells (PGC) migrate from the yolk sac, via the hindgut to the genital ridge and enter the gonad where they undergo further maturation into the sex specific lineage, i.e. oogonia for females and spermatogonia for males. During migration and maturation an epigenetic "reset" takes place. This includes global DNA CpG demethylation during the early phases of migration. Specific areas like imprinted regions remain methylated until the PGCs arrive in the developing gonads where imprinting is subsequently gradually erased. After these maturing gonadal germ cells reach mitotic (male) or meiotic (female) arrest, *de novo* methylation is initiated and uniparental sex specific imprinting is acquired [1-8]. Another informative marker of developmental stage is X chromosome reactivation which occurs in female germ cells before the initiation of oogenesis. Studies report varying results regarding the exact timing of the various steps of the epigenetic reset, i.e. during migration or after arrival in the gonads. However, PGCs with an XX chromosomal constitution have been shown to lack X chromosome reactivation if they never reach the gonad [9-12]. For ethical reasons, most of these data have been experimentally investigated and validated in mice. Even though germ cell development differs between mice and men [13], methylation patterns during germ cell development are reported to be highly similar [14, 15].

Germ cell tumors (GCT) originate from germ cells at different developmental stages and are thought to inherit their methylation profile from their ancestors. The WHO classification supports five GCT subtypes. Each subtype has specific molecular, clinical and histopathological properties [16-19]. GCT subtypes have been put in context of normal germ cell development (Figure 1A) based on gene/microRNA expression, (targeted) epigenetic analysis and genomic constitution as described below and reviewed extensively elsewhere [13, 16, 17, 20-22]. Most of these studies were targeted at specific genes/genomic regions or concerned a subset of the GCT subtypes only, most prominently type I or II.

Type I ("infantile") GCTs manifest clinically as teratoma (TE) and/or yolk sac tumor (YS) along the migration route of developing PGCs, i.e. the midline of the body. Extra-gonadal, sacral TEs occur most frequently and are mostly benign. Typically these rare tumors (incidence 0.12/100 000) arise before the age of 6 and no *Carcinoma In Situ* (CIS, see below) is found. They show global methylation patterns that are reminiscent of their embryonic stem cell progenitor (i.e. bimodal with modes at ≈0 and ≈100% methylation). These tumors showed somatic/biparental (≈50%) imprinting status in earlier studies. Therefore, type I GCTs have been suggested to originate from PGCs at an early stage, prior to global demethylation and imprinting erasure [16-18, 23-25].

A

| normal germ cell development | TYPE II |

TYPE I
TE & YS

TYPE IV
dermoid cyst (DC)

normal germ cell development:
- primordial germ cells (PGCs)
- gonocytes
- oogonium
- pre-spermatogonia
- meiosis
- primordial follicle
- spermatozoa
- +

TYPE V: hydatiform mole

TYPE III: spermatocytic seminoma (SS)

TYPE II:
- Carcinoma In Situ (CIS)
- Seminoma (SE) / *Dysgerminoma (DG)*
- Embryonal Carcinoma (EC)
- Choriocarcinoma (CH)
- Yolk sac tumor (YS)
- Teratoma (TE)

mixed non-seminoma (mNS)

STEM CELL COMPONENTS

DIFFERENTIATED

B

GCT subtype	n	
I: TE	14	6 sacrum (male:3 [I.TE.s.m], female: 3 [I.TE.s.f]), 4 testis [I.TE.t], 4 ovary [I.TE.o]
II: EC	30	
II: mNS	24 (3 pure type II TE)	
II: SE	12	
II: DG	4	
III: SS	4	
IV: DC	3	
Cell lines (type)	TCam-2 (SE), NT2 (EC), 2102EP (EC), NCCIT (EC)	

C

Gene

IMPR_P/M1500 *
IMPR_P/M200 *
TSS200 (56,807)
5'UTR *
EXON1
TSSAssociated (127,037)
(35,836)
GENE.BODY (158,372)
3'UTR (17,851)

-1500 | -200 | 0 | 5'UTR | Exon 1 | Gene Body | 3'UTR
TSS
transcription regulating (tr) | *non-tr*

LINE (17,984)
SINE (21,243)
ICR_P (253)
ICR_M (18)

MicroRNA
MIR200 (1,343)
MIR1500 (2,946)

-1500 | -200 | 0 | Micro-RNA
TSS

CpG Islands
ISLAND *
IslandAssociated (280,656)

Shelves/Shores | Island | Shelves/Shores

← Figure 1. Tumor types/samples and cell lines analyzed and schematic visualization of genomic functional categories of interest. This figure presents an overview of (A) GCT subtypes in the hypothesized context of normal germ cell development as proposed in earlier studies (grey box) and (B) the samples included in this study. Panel C presents a reference to (abbreviations of) the functional genomic regions as mentioned in the rest of the manuscript. (A) Developmental schemes are indicated in blue (male), red (female) or when possible in both sexes (white). DG does not originate from CIS but is indicated together with SE for reasons of consistency. (B) Compilation of the dataset. Abbreviations match Figure 1A and roman numbers indicate the GCT type to which the histological subtypes belongs. n indicates the number of tumor samples per group. All samples are from male patients except the DGs, DCs and a subset of the type I TEs. Please note that when only TE is denoted, this indicates the group of all type I TEs together. Otherwise II.TE (type II pure TE) or the abbreviations for specific localizations are used as indicated in this figure. Four GCT cell lines were included; tumor of origin between brackets. (C) Functional genomic categories. Probes were classified according to their relation to gene coding regions, micro-RNA (MIR) coding regions, CpG islands and/or transposon elements (LINE/SINE). The distance to the transcription start site (TSS) was used in accordance with the Illumina manifest: 200 or 1500 bp. Of note, the TSSAssociated category contains all probes with a distance < 1500 bp to the TSS in contrast to the TSS1500 category from Illumina which is only contains probes 200-1500bp from the TSS. Probes within imprinting associated regions were classified as (1) mapped inside a known imprinting control region (ICR) or (2) either mapped inside an ICR or mapped close to the TSS of a transcript of an imprinted gene (200/1500bp upstream, not mutually exclusive). P/M indicates the expressed allele, i.e. paternal/maternal respectively. Numbers between brackets indicate the number of valid probes within each specific category (total number of valid probes: 437,881). *The visualization did not permit including the probe count for all categories. The counts for the empty categories are: 5'UTR=59,338; ISLAND=136,339; IMPR_P200=638; IMPR_P1500=1,659; IMPR_M200=610; IMPR_M1500=2,265.

Type II GCTs present most frequently in the gonads and are also called germ cell cancer (GCC). The incidence of these tumors peaks between 25-35 years of age depending on the subtype [16, 17, 19]}. They comprise ≈1% of all solid cancers in Caucasian males and are responsible for 60% of all malignancies diagnosed in men between 20 and 40 years with increasing incidence in the last decades [26] (Dutch Caner Registration (IKNL)[a]). Risk factors have been thoroughly investigated and are integrated in a genvironmental risk model, in which risk is determined by a combination of micro/macro-environmental and (epi)genetic factors [19, 26-32]. A common precursor lesion called CIS or intratubular germ cell neoplasia unclassified (IGCNU, WHO definition [18]) is identified for type II GCT [16, 17, 33, 34]. Because of the non-epithelial origin these tumors, CIS is technically not a proper term but will be used throughout this article in the interest of consistency with existing literature. Type II GCT consist of non-seminomatous (NS) and seminomatous (SE) tumors (Figure 1A), which differ in clinical behavior and molecular profile. SE and embryonal carcinoma (EC) are the stem cell components of type II GCT and EC can further differentiate in the other NS subtypes: TE, YS and choriocarcinoma (CH) [16, 17]. Type II GCT originate from maturation arrested, germ line committed PGCs or gonocytes and historically have been suggested to exhibit erasure of genomic imprinting [13, 16-19, 22, 35]}

Type III, IV and V GCTs originate from more differentiated germ cell progenitor cells. Type III GCTs are also known as spermatocytic seminoma (SS) and occur solely in the testis. They arise after the age of 50 and are generally benign and rare (incidence: 0.2/100000). Their presentation in elderly males, morphology and immunohistochemical profile separates SS from SE. They originate from germ cells around the spermatogonium stage and are paternally imprinted [16, 36-40]. Type IV

---

[a] www.cijfersoverkanker.nl

tumors are historically hypothesized to originate from a maternally imprinted, committed female germ cell. Type V GCT were excluded from this study because they show an independent pathogenesis. They originate from the fertilization of an empty ovum by two sperm cells, resulting in a completely paternally imprinted genomic constitution. This explains their mono-directional lineage of differentiation, unrelated to the germ cell origin [16-18].

This study aims to identify specific and global differences between the genome-wide methylation profiles of GCT subtypes. Type I, II, III and IV GCTs and four cell lines representative of type II GCTs are investigated (Figure 1A,B). Differences in methylation profile provides insight into the developmental timing and underlying biology of GCTs. The findings ultimately relate GCT subtypes to specific stages of (early) developing (embryonic) germ cells. Emphasis was placed on combining the results with the available literature and on providing extensive accompanying data to supply an integrated, hypothesis generating data source for future research.

Results

Methylation differences are investigated, starting from global methylation profiles, followed by functional enrichment analysis (functional genomic regions: Figure 1C). Next, specific differentially methylated probes are identified (DMPs). Probes represent individual CpG sites. Also, differentially methylated regions (DMRs) containing multiple adjacent probes are identified. Finally, imprinting status is evaluated. Please note that differential methylation indicates a statistically significant difference after correction for multiple testing, unless specifically stated otherwise. For details about the statistical procedures, please see the materials and methods section (analysis protocol).

### SS and SE/DG show global hypomethylation when compared to EC/mNS and TE

Figure 2A shows the methylation distributions for all probes, probes associated with the TSS, 3' UTR, LINES, microRNAs and CpG Islands, respectively. The distributions of the remaining functional categories are presented in Figure S2A. SS showed global hypomethylation (Figure 2A), i.e. a large

Figure 2. Methylation patterns in GCT subtypes and cell lines. To illustrate differences in methylation status between histological GCT subtypes two (visualization) methods were applied. Firstly, the methylation pattern over the whole genome and specific functional categories (Figure 1C) is visualized using the distribution of the methylation percentage β in all samples of a certain GCT subtype. Next, the discriminatory power of the methylation pattern for each individual sample is shown using principal component analysis. **(A)** Distribution of methylation percentage. Violin plots: grey areas indicate a kernel density plot of the methylation percentage (β) of all probes in all samples in a certain category. The boxplot indicates the interquartile range (black bars) and median (white squares). X-axis labels indicate histological subgroup according to Figure 1A,B. TE indicates type I TE only. **(B)** Principal Component Analysis. The first two principal components (PC) are plotted to evaluate the discriminative power of the methylation pattern between the subtypes. Abbreviations of histological subtypes are explained in Figure 1A. CL indicates cell lines. Please note that in legend of the PCA the TE group is subdivided based on gender and localization: I=type I; II=type II/formally part of the mNS group, s=sacrum, t=testis, o=ovary, m=male, f=female. A more detailed visualization of the TE classes is provided in Figure S2, which also includes the full series of 18 functional categories, bootstrap validation of the PCA and an estimation of the variance explained by the first two principal components. →

A. Distribution of methylation percentage per GCT subtype
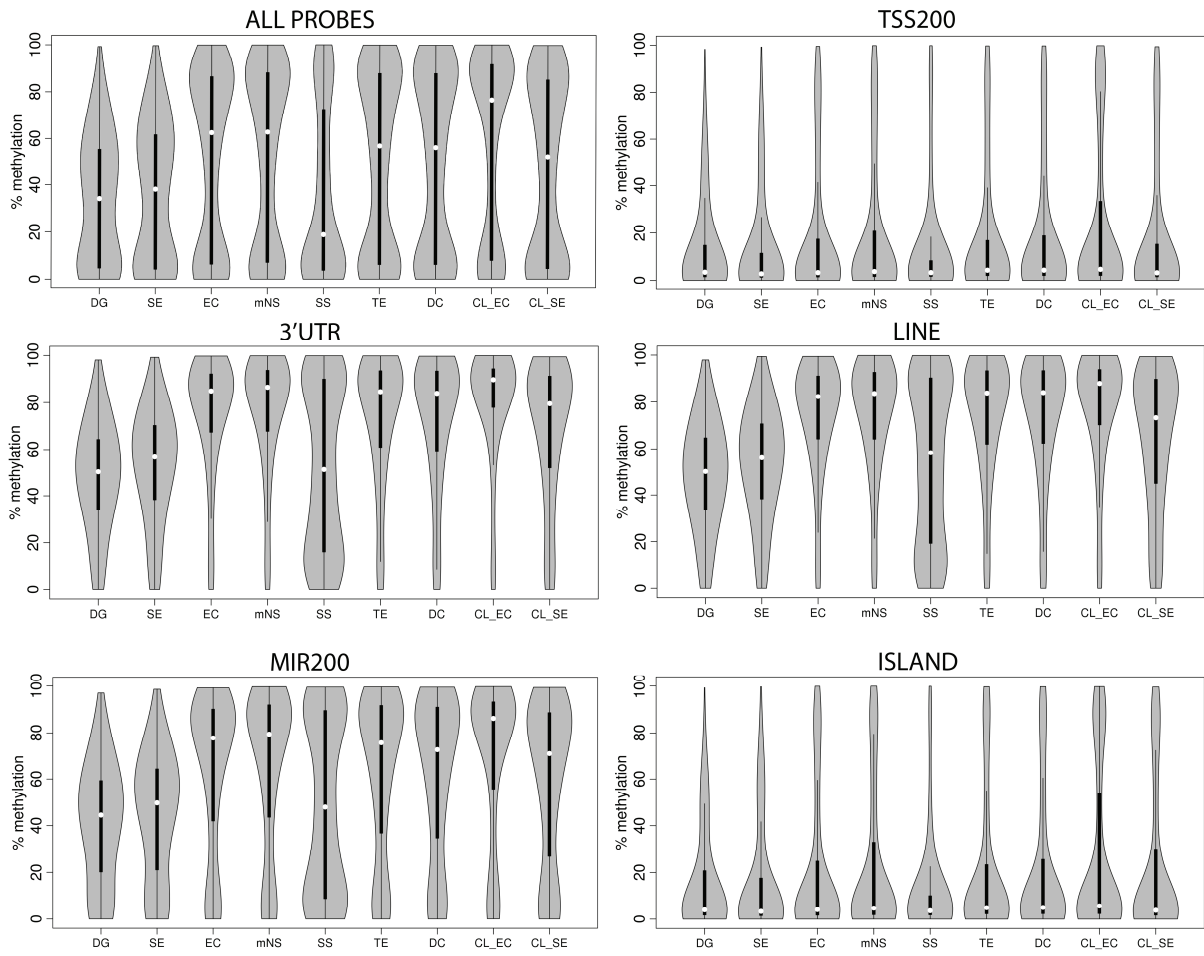


B. Principal Component Analysis
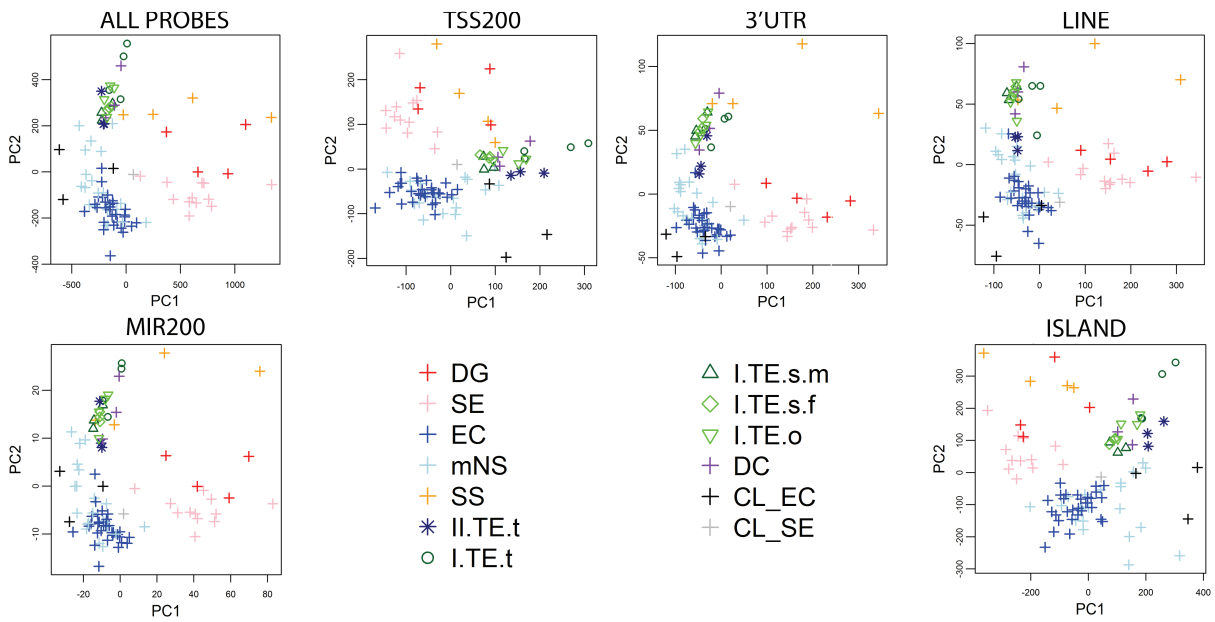
concentration of probes showing a low percentage of methylation and few probes showing a high methylation percentage. Hypermethylated configurations contain a large concentration of probes showing a high percentage of methylation and few probes showing a low methylation percentage. Hypomethylation was also shown in DG and SE samples albeit to a lesser extent, as can be observed from the mode at 50-60% methylation (Figure. 2A). The SE group showed consistent hypomethylation (Figure S2B, p.2), in contrast to a study of Nettersheim et al who showed separate groups of hypo- and hypermethylated SE in a larger sample series [41]. In contrast to the SE and DG samples, the EC, mNS, TE and DC samples consistently showed a bimodal pattern with one mode around 10% and one around 90% (Figure 2A). This bimodal pattern is also observed in three EC cell lines and a single SE cell line (Fig. 2A, CL_SE & CL_EC). In line with previous reports [14, 42], the EC cell lines were more methylated than the SE cell TCam-2 (Figure. 2A). The transcription regulatory region upstream of the TSS (TSSAssociated, TSS200) was generally hypomethylated in all tumor types as were regions annotated as first exon, 5'UTR and CpG islands. The gene body, 3'-UTR, micro-RNAs and LINE/SINE elements were generally hypermethylated except in SS, which showed a bimodal pattern (Figures 2A, S2A). At these sites, SE/DG showed a median methylation level of 50% in line with the maximal methylation of their global profile and previous reports [20, 43]. Hypermethylation of LINE/SINE elements NS and hypomethylation (Figure 2A) in SE was in line with a recent genome wide study [20] but contrasted with a targeted study that showed hypomethylation of 3 specific repetitive elements in both SE and NS [44].

GCT subtypes can be distinguished based on their methylation profile

Principal component analysis (PCA) showed robust separation of homogeneous clusters of EC/mNS, SE/DG, TE/DC and SS samples when all probes were considered (Figure 2B, S2A). In line with the larger inter-sample variation (Figure S2B), SE/DG and SS were more scattered in the PCA plot. Some mNS, which consist partly of differentiated tissue, showed a tendency towards the differentiated TE/DC group. The TE and DC showed an indistinguishable global methylation profile. Similar observations were made when subsets of probes were considered that were annotated to specific functional genomic regions (Figure 2B, S2A).

Figure 3. **Functional enrichment of DMPs.** DMPs were classified according to their functional genomic location (Figure 1C). Statistical over- and underrepresentation of probes in certain categories provides clues to differences between GCT subtypes in regarding function of methylation. Results are shown for four pairwise (A vs B) comparisons of histological subtypes: **(A)** SE/DG versus EC/mNS; **(B)** SE/DG vs type I TE; **(C)** EC/MNS vs type I TE and **(D)** SE/DG vs SS. **(LEFT)** The number (n) of DMPs identified in either the DMP[A-B] (hypermethylated in A, green) or DMP[A-B] (hypermethylated in B, red) group. **(MIDDLE/RIGHT)** Functional enrichment in the DMP[A-B] and DMP[A-B] group respectively. X-axis: positive numbers indicate a significant overrepresentation of DMPs in a functional category compared to non-DMPs while negative numbers indicate a significant underrepresentation. Depicted is the log2 ratio of (1) the % of either DMP group assigned to a category and (2) the % of non-DMPs assigned to that category. Only significant enrichments are depicted (2-sided Fisher's Exact test, see Methods section for Bonferroni corrected α threshold). DMPs[SE/DGvsSS].IMPR_P1500 showed significant underrepresentation, but could not be plotted on log scale (0 probes in DMP group). Details of calculations and raw counts and percentages are presented in Table S2. Y-axis: functional categories as specified in Figure 1C. →
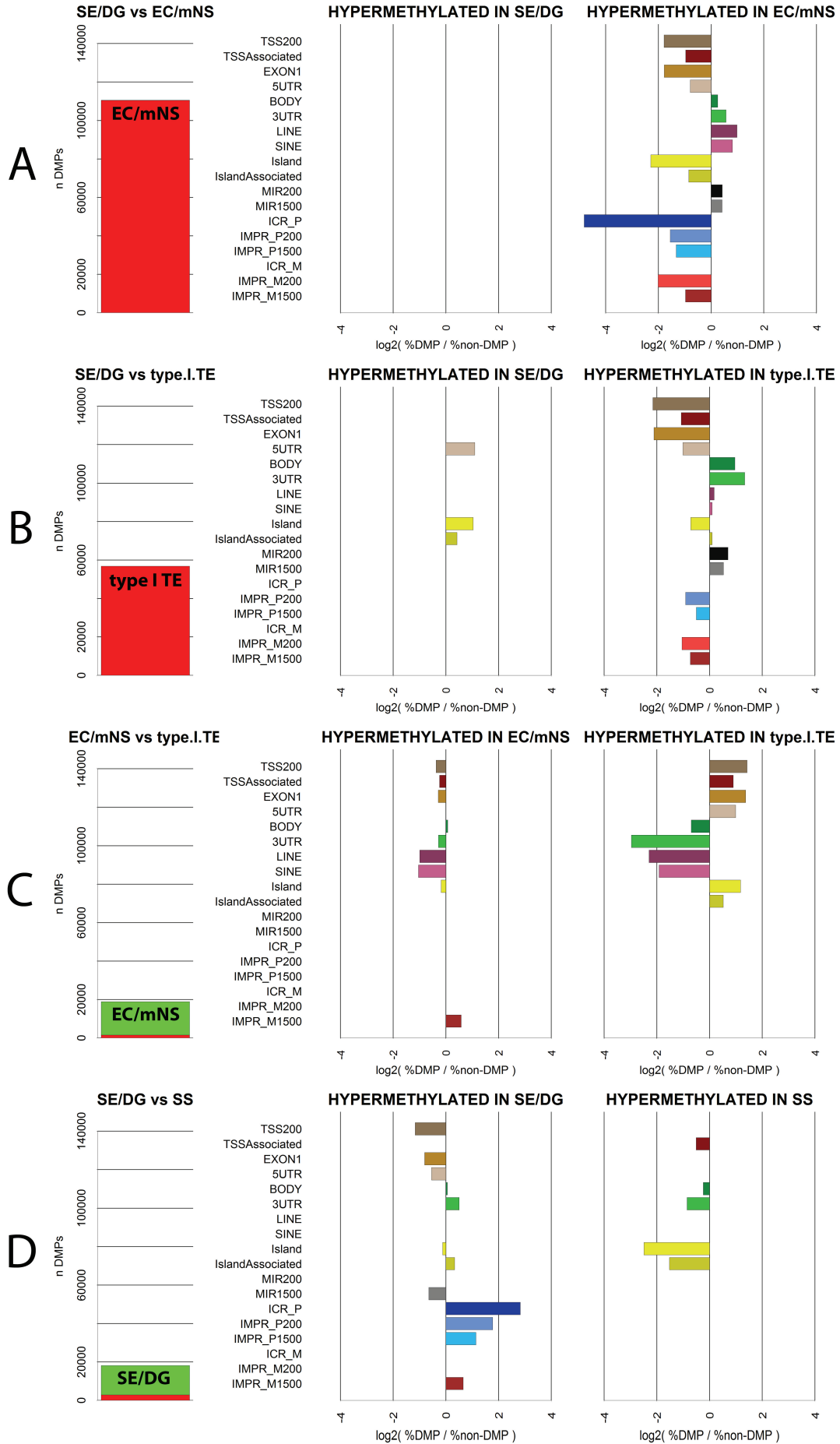
Zooming in: GCT subtype specific methylation patterns

To further pinpoint differences between pairs of GCT subtypes, DMPs were identified (Table 1, S1), tested for functional and chromosomal enrichment (Figure 3, S3; Table 1, S2) and grouped into DMRs (Figure 4, S4; Table 1, S3, File S1). SE + DG and EC + mNS subtypes were merged because of high similarity of the observed methylation profile (Figure 2A, 2B, S3A), in line with literature regarding their similar origin [45]. Recurrent DMRs were identified as genes occurring more than once within or between comparisons, which may indicate regions of importance (Table S3, n=149). (Differential) methylation of GCT cell lines (4136 DMRs between the cell lines: File S2) showed little similarity to their in vivo counterparts (Figure 2,6,S2) but matched with a biologically relevant DMR previously validated in these cell lines using bisulfite sequencing in [46] (microRNA-371/2/3 cluster, Table 2). 719 gene symbols intersected between tumor and cell line DMRs (Table S3). The major differences between the subgroups of GCT will be summarized hereafter.

Comparing SE/DG, EC/mNS and type I TE

Regardless of their presumed common origin, EC/mNS and SE/DG show vastly different methylation profiles. The relative hypermethylation in EC/mNS versus SE/DG was concentrated in regions not involved in transcription regulation (Figure 3A). This points to a global difference in methylation status rather than differential methylation of specific regulatory elements. This also holds for the hypermethylation of TE when compared to SE/DG (Figure 3B). The 61 DMPs hypermethylated in SE/DG relative to TE were concentrated at three specific genes: *NCOR2*, *ALOX12* and *ECEL1P2* (Table 1, S3, Figure S4A).

DMPs between TE and EC/mNS indicate a more methylated profile of the EC/mNS group (Figure 3C). Moreover, the majority of the probes hypermethylated in TE were located on the X chromosome and can therefore be traced back to hemi-methylation of chromosome X in females (TE=male/female, EC/mNS=male only) (Table 1, Figure S3B). DMRs included many genes involved in male gametogenesis like *DMRT3* (Figure 4A). The EC marker *SOX2* [17, 47] was present as one of the only 15 hypermethylated autosomal DMRs in TE (Figure 4B). These DMRs presumably relate to the cell of origin as well as to the sex of the patient (Figure S4B, Table 1, S3).

Type III (SS) versus type II seminomatous GCT (SE/DG)

The general, hypomethylation in SS (DMP[**SE/DG**-ss]) is enriched for regions associated with paternal expression (Figure 3D). DMRs hypermethylated in SE/DG predominantly included recurrent DMRs and DMRs within genes associated with germ cell and testis development (Table 1, S3). The promoter of *POU5F1* was relatively hypomethylated in SS, while it is a marker for the stem cell component of type II GCTs and not expressed in SS [17, 45, 48] (Figure 4C, Table 2). DMRs hypermethylated in SS also included genes associated with male germ cell determination, fertility and GCTs, enforcing the epigenetic relation between GCT cells and their cell of origin (Table 1, S3).

Table 1. Pairwise comparison of GCT subtypes. This table concisely summarizes the results of the search for differentially methylated (DM) probes (P) and differentially methylated regions (R) between pairs (A and B) GCT subtypes. Briefly, the number of DMPs and DMRs is shown separately for probes hypermethylated in A or B. The subtype in which the probes are hypermethylated is indicated in **bold and underlined**. Also, a brief interpretation of the genomic function of the DMPs is provided. For the DMRs the associated genes are discussed in the context of GCTs. (**Abbreviations**) ↓ significantly underrepresented; ↑ significantly overrepresented; % DMPs is calculated relative to the total number of valid probes (Materials and Methods section). tr=transcription regulation associated regions (TSS200/TSSAssociated/5'UTR/EXON1); non-tr=non transcription regulation associated gene coding regions (GENE.BODY/3'UTR). The other functional categories are depicted in Figure 1C. [global] = global methylation difference between subtypes; no distinguishable potential subtype specific differentially methylated regulatory elements. (**Associated sources**) Statistical procedures are described in the Materials and methods section. The overall methylation pattern of each histological subtype is visualized in Figure 2. Functional enrichment of DMPs is visualized in Figure 3. Details of enrichment calculations and raw counts and percentages are presented in Table S2. Enrichment of chromosomes is depicted more detailed in Figure S3B. DMRs, recurrent tumor DMR and DMPs are listed in Table S3 and Table S1 respectively. DMRs are visualized in Figures 4 and S4 and File S1, S2.

| A. Seminomatous (SE/DG) versus non-seminomatous (EC/mNS) GCTs	
DMProbes[**SE/DG**-EC/MNS]	DMProbes[SE/DG-**EC/mNS**]
6	110,462 (25%)
[no enrichment]	↓ in tr and ↑ in non-tr, LINE/SINE suggested global difference in methylation status rather than differential methylation of specific regulatory elements. CpG islands were ↓. miRs regions were weakly ↑. ICRs were ↓, suggesting no difference in imprinting status.
DMRegions[**SE/DG**-EC/MNS]	DMRegions[SE/DG-**EC/mNS**]
0	*[global]*
-	*[global]*
B. Seminomatous type II (SE/DG) versus type I (TE)	
DMProbes[**SE/DG**-TE]	DMProbes[SE/DG-**TE**]
61	56,764 (13%)
significantly overrepresented on chromosome 12 (13/61) and preferentially located in the 5'UTR of genes and CpG islands	↓on tr, CpG islands and ICRs and ↑of non-trs, transposons and miRs (≈DMPs[SE/DG-**EC/mNS**]).
DMRegions[**SE/DG**-TE]	DMRegions[SE/DG-**TE**]
3	*[global]*
NCOR2. (*SMRT*, silencing mediator for retinoic acid and thyroid hormone receptors) nuclear receptor co-repressor on 12q24.31 involved in mouse spermatogenesis [49] and vitamin D metabolism in GCTs [50]. *ALOX12.* lipoxygenase family [51], has not implicated in GC(T) biology. *ECEL1P2.* increased methylation upon aging [52] → although not implicated in normal/aberrant germ cell development, this might explain the hypomethylation in pediatric type I TE as compared to adult type II SE/DG. All three genes were also DMR hotspots.	*[global]*

C. Non-seminomatous type II (EC/mNS) versus type I (TE)	
DMProbes[**EC/mNS**-TE]	DMProbes[EC/MNS-**TE**]
17,407	1,520
Enrichments were weak, the strongest being ↓ in transposons.	≈80% on X chromosome → differences in sex (TE=male+female, EC/mNS=male). ↑ in tr and ↓ in non-tr and transposon elements suggests differential methylation of specific regulatory elements. ↑ in CpG islands.
DMProbes[**EC/mNS**-TE]	DMRegions[EC/MNS-**TE**]
580 (all autosomal)	128 (15 autosomal)
e.g. _DMRT3:_ implicated in testis development and male sex determination [45, 53] _MOV10L1:_ which has been implicated in human male infertility [54] and germ cell maturation in mice [55]. _DDR2:_ crucial for spermatogenesis in mice [56] ICR_P _WT1_ was also present.	e.g. _SOX2:_ EC marker (see: Table 1). _IRX5:_ germ cell migration in Xenopus laevis embryos; DMR ≈1kb downstream of its 3'UTR [57]. _MSX1:_ progression of germ cells into meiosis, leading to germ cell maturation arrest in mutant embryos; DMR at its 3'UTR [58]. Hypomethylation in type II GCTs [59].
D. Seminomatous type II (SE/DG) versus type III (SS)	
DMProbes[**SE/DG**-SS]	DMProbes[SE/DG-**SS**]
15,340	2,830
↑ in non-tr and ↓in the tr. ↑ in ICR_P/IMPR_P200/1500 in line with paternal cell of origin of SS.	↓ at non-tr and CpG islands.
DMProbes[**SE/DG**-SS]	DMRegions[SE/DG-**SS**]
559	30
e.g. hotspot genes like _NCOR2 , ALOX12, ECL1P2, MSX1_ (see above). _IRS2:_ associated with male germ cell and testis development [60]. _POU5F1:_ SE/DG/EC marker (Table 1). _TEX14:_ associated with known high risk GCT SNP [29].	_SERPINE1_ (plasminogen activator inhibitor 1, PAI-1): hypomethylated in GCT except in SS. PAI-1 SNPs have been associated with poor prognosis in GCTs [61]. The plasminogen activator system has been implicated in human infertility [62]. _MOG:_ hypermethylated in SS, knockdown causes male germ cell differentiation in _mog_ deficient C. Elegans [63, 64].

Figure 4. Methylation profile at GCT subtype specific differentially methylated regions (DMRs). Visualization of the methylation percentage at specific loci is used to zoom in on a predefined region and investigate local methylation differences between GCT subtypes. **(A)** _DMRT3,_ **(B)** _SOX2,_ **(C)** _POU5F1_ (_OCT3/4_), **(D)** _TEX14._ (Visualizations) From top to bottom the following is depicted: (1) Four-color heat map indicating methylation % for each individual probe in the depicted region. For the sample groups specified on the left the median methylation % is shown. (2) Position of all probes in the region of interest (ROI) is annotated as black rectangles. (3) HMM segments are displayed as grey boxes spanning the segment's width and grouped per state. Numbers indicate the state of each (group of) segment(s). (5) GC% was obtained from the UCSC genome browser database (gc5Base table). (6) Transcripts overlapping with the ROI are plotted at the bottom. Plot generated using the Gviz package. If red rectangles are indicated, they indicate DMPs in the visualized DMR. Abbreviations of histological subtypes are explained in Figure 1A and for specific subtypes of TE in the lgend of Figure 2. CL indicates cell lines. →

A. DMRT3

B. SOX2

C. POU5F1 (OCT3/4)

D. TEX14
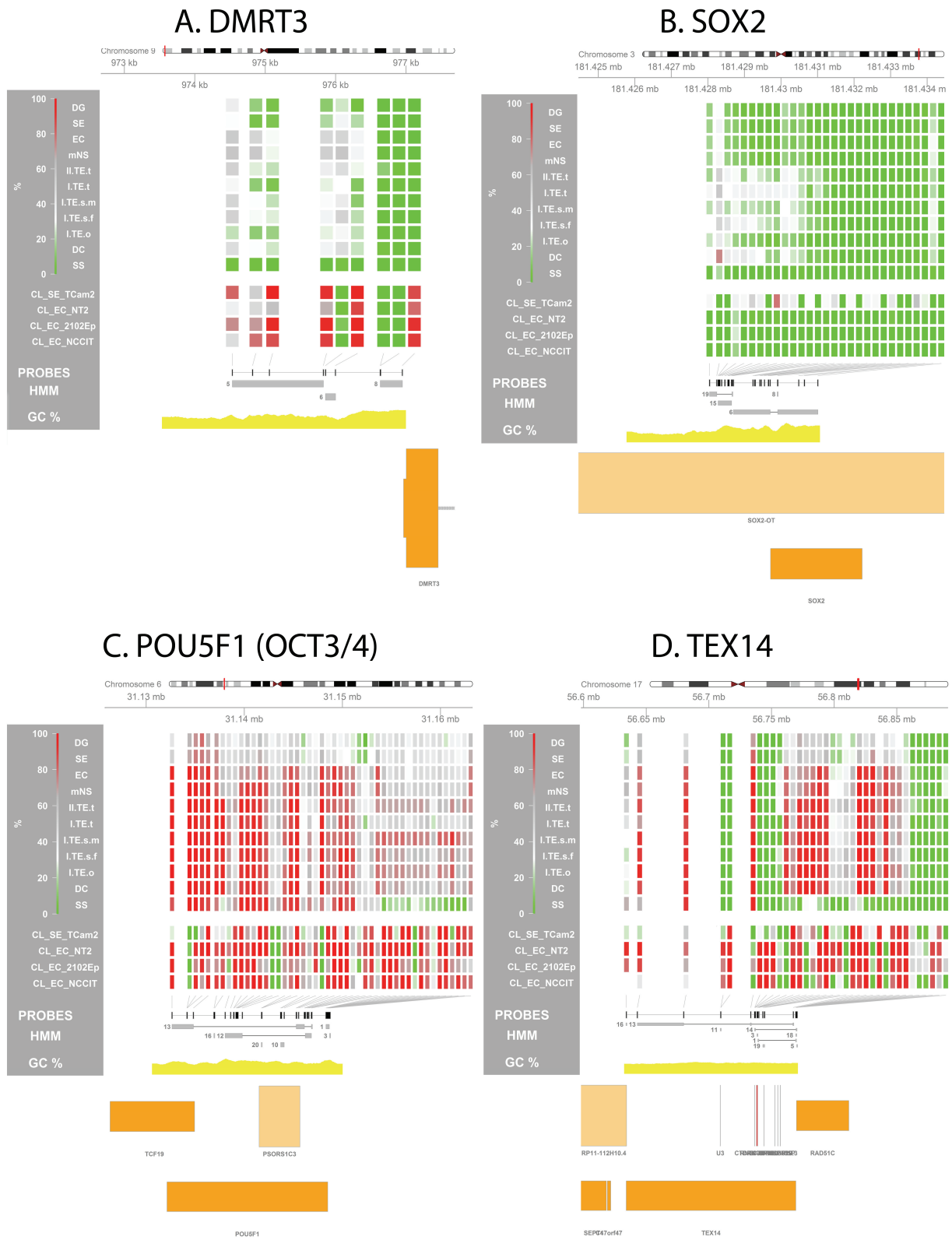
Specific GCT associated genes

A number of genes has been associated with (methylation in) GCTs, both regarding pathogenesis and diagnosis. Table 2 summarizes the literature for these genes and combines this with the methylation data from this study, e.g. overlap with DMRs and methylation profile of these genes (see also Figure 5, S5A). A recent meta-analysis of GCT GWAS studies identified 19 SNPs associated with 13 genes [29]. For most genes their methylation profile was non discriminative between the GCT subtypes, the exceptions being *TEX14* which was also independently identified as a DMR[SE/DG-SS] (Figure 4D) and *BAX1*, which also contained a DMR[SE/DG-SS] (all SNP related genes: Figure S5B).

A. AR

B. miR-371-2-3
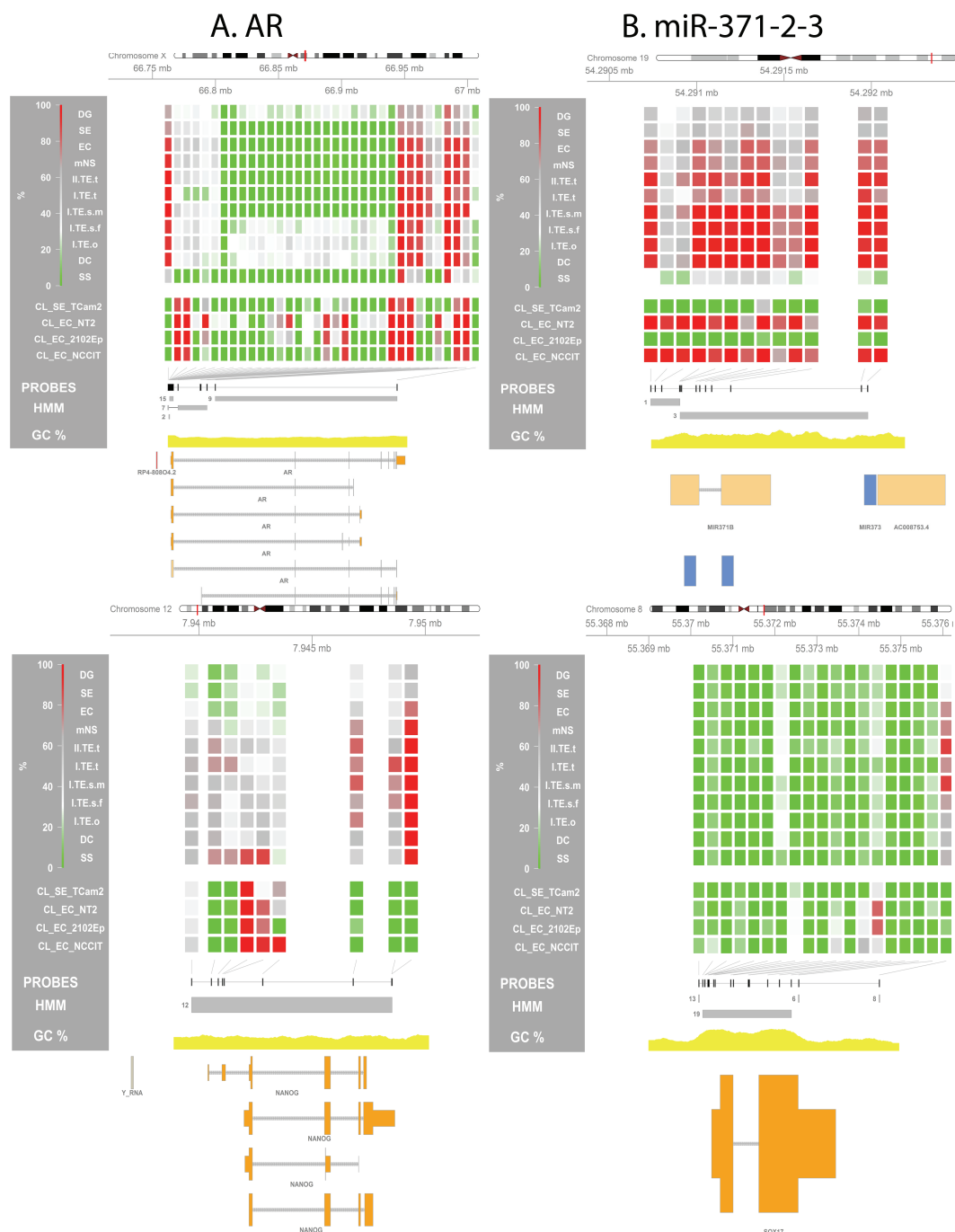
← Figure 5. Methylation profile of GCT specific genes and regions of interest (ROIs). Visualization of the methylation percentage at specific loci is used to zoom in on a predefined region and investigate local methylation differences between GCT subtypes. **(A)** *AR*, **(B)** *miR-371-2-3*, **(C)** *NANOG*, **(D)** *SOX17*. The visualization is described in the legend of Figure 4 and the genes are reviewed in Table 2. Abbreviations of histological subtypes are explained in Figure 1A and for specific subtypes of TE in the legend of Figure 2. CL indicates cell lines.

Table 2. GCT (methylation) associated genes. Genomic locations and strand were retrieved from genecards.com/UCSC. Detailed visualizations of the methylation status of these genes is presented in Figures 5 and S5A. DMRs in the cell lines are presented in File S2. ICRs are visualized in Figure 7 and S6.

Gene & region	Description
APC chr5 (+) 112,043,195 – 112,181,936	**GCT link:** A single study with small sample size (n=10) showed increase methylation in most YST as compared to germ cells in normal testis. Expression was high in germ cells and low in most YSTs [65]. **Findings:** 2102EP showed mild but significant relative hypermethylation compared to the other cell lines, but for all tumor groups *APC* was consistently hypomethylated.
AR chr X (+) 66,763,874 – 66,950,461	**GCT link:** Androgen receptor methylation can be used as a readout for X inactivation in non-germ cells. *AR* was methylated in differentiated NS, but unmethylated in a proportion of ECs and all SE & SS. This supports the hypothesis that methylation does not occur in the germ cell lineage [66]. **Findings:** the promoter region of the *AR* was completely deprived of methylation in all male tumors while a certain amount of methylation (ca. 50%) was present in the female samples. *AR* contained a DMR only in the CL where it was relatively methylated in NT2 as compared to all other cell lines (Figure 5A).
CTA genes	**GCT link:** Cancer Testis Antigens (CTA) are primarily unmethylated in SE. *MAGEA1/3* are predominantly methylated in NS while *SYCP1* is unmethylated in NS [67]. *MAGEA1*: chrX:152,481,522-152,486,116(-), *MAGEA3*: chrX:151,934,652-151,938,240(-), *SYCP1*: chr1:115,397,424-115,537,991(+) **Findings:** Methylation differences in these genes were not remarkable except for differential hypomethylation of TCam-2 and NCCIT compared to the other cell lines. The TSS associated regions of *MAGEA1* and *SYCP1* were consistently hypomethylated in SS.
GATA4 chr 8 (+) 11,534,468 – 11,617,511	**GCT link:** previously identified DMR between TCam-2 and NCCIT, promoter region hypermethylated in TCam-2 [68]. **Findings:** The *GATA4* promoter was differentially hypomethylated in all CL_EC as compared to EC_SE, but this was exactly opposite in the SE and EC tumor samples. Testicular TEs, like EC/mNS samples showed relative hypermethylation while sacral/ovarian TEs, DCs and SS showed relative hypomethylation like the SE/DG samples.
HIC1 chr17 (+) 1,957,448 – 1,962,981	**GCT link:** 55% of the GCT show methylation of this area which shows frequent loss of heterozygosity in somatic adult cancers. 5AZA treatment strongly induced *HIC1* expression in non-GCT CLs [69]. *HIC1* promoter methylation has been implicated in treatment resistance in GCTs [70]. **Findings:** HIC1 was showed predominantly hypomethylation in all GCT subtypes even though a weak DMR[<u>EC/mNS</u>-TE] was identified. Of the cell lines, only 2102EP showed differential hypermethylation.
KIT & KITL KIT: chr4 (+) 55,524,085 – 55,606,881	**GCT link:** KIT and KITL regulate primordial germ cell development and homing to the gonad [71-75]. In the embryonic phase the guidance of KIT+ primordial germ cells from the hind gut epithelium to the gonads depends strongly on KITL mediated chemo attraction [74, 76-78]. In the postnatal testis KIT-KITL signaling takes place via paracrine signaling in the

| Gene & region	Description
KITL: chr 12 (-) 88,886,570 – 88,974,628	germline stem cell niche and is crucial for spermatogenesis from the spre-matogonial stage onwards [72, 74, 79, 80]. More mature mouse spermatids and spermatozoa express a c-terminal truncated form of KIT transcribed from an intronic promotor [81]. Mechanistically, constitutive paracrine / autocrine activation of KIT/KITL signaling is implicated to be a crucial initiating event for the malignant transformation of maturation arrested germ cell progenitors [17, 19, 22]. In the early stages, KITL positivity is a hallmark of maturation arrested germ cells, CIS and intratubular SE [17, 82-84]. Progression into invasive SE is also strongly related to KIT/KITL signaling while much less association with the NS phenotype has been shown [79, 85-88]. Activating *KIT* mutations are identified in ca 13-60% of the SE (rare in NS) and result in constitutive kinase activity because of ligand independent dimerization and phosphorylation [89-92]. Recent GWAS studies identified susceptibility loci for GCTs close to, within or directly related to GCTs [29, 93-104]. No information about *KIT* or *KITL* methylation in tumors was presented in literature although *KITL* promoter methylation was significantly lower in blood of these patients [105] and SNPs in KITL combined with aberrations in cAMP regulation were suggested to contribute to tumor risk in these patients [104]. **Findings:** *KIT* (Figure S6A) and *KITL* (Figure S6B) were not differentially methylated between any of the tumor groups or cell lines.
miR-371/2/3 chr19 (+) *(371)* 54,290,929 – 54,290,995 *(372)* 54,290,995 – 54,291,210 *(373)* 54,291,959 – 54,292,027	**GCT link:** The *micro-371-2-3* cluster is expressed in the stem cell component of GCT [106] and is a potential diagnostic serum marker for GCT [107]. Upstream of the TSS of this cluster a DMR has been identified between TCam-2 and NCCIT [68]. Differential methylation in GCT cell lines has been validated using pyrosequencing and the methylation level showed significant and strong inverse correlation with the expression of miR-373 (Spearman's ρ -0.90, p=0.037) [46]. **Findings:** The *miR-371-2-3* cluster was hypomethylated in TCam-2 (CL_SE) and 2102EP and hypermethylated in NT2 and NCCIT (Figure 5B). However, with the exception of SS the tumors showed hypermethylation of this region, despite known expression in the stem cell components of type II tumors [46, 106].
NANOG chr12 (+) 7,940,390 – 7,948,655	**GCT link:** Specific marker for the all stem cell components of GCTs [17]. RA treatment of NT2 cells also increased methylation here [108]. Analogous to this CpG sites in the *NANOG* promoter (0-306 bp upstream of the TSS) were found hypomethylated in spermatogonia and hypermethylated in sperm [109]. **Findings:** The *NANOG* promoter region showed a trend towards relative hypomethylation in the undifferentiated stem cell components of the type II tumors as compared to all other (more differentiated) GCT subtypes including the type II TE and mNS (intermediate status). However, the number of probes+consistency of the difference lacked significance (Figure 5C).
POU5F1 (*OCT3/4*) chr6 (-) 31,132,114 – 31,148,508	**GCT link:** Specific marker for the all stem cell components of GCTs [17, 48, 110]. *OCT3/4* transcription is regulated by methylation of conserved regions up to 2.6kb upstream of the TSS. Another study also showed that little increase of methylation at specific sites upstream of *OCT3/4* strongly inhibited expression [108, 111, 112]. Differentiation of NT2 after retinoic acid treatment resulted in increased methylation and loss of expression [108]. **Findings:** A promoter DMR[**SE/DG**-ss] was identified despite the fact the SE/DG express the OCT3/4 protein and SS do not [17, 45, 48] (Figure 4C). However, probes located close to its transcription start site are generally methylated between 20 and 40% in OCT3/4 positive tumors (SE/EC) which results in unmethylated alleles primed for expression. Moreover, the promoter region of *OCT3/4* showed a non-significant trend towards lower methylation levels

Gene & region	Description
	in SE/DG and EC/mNS when compared to the differentiated tumors (TE). Most importantly however, regulation of *OCT3/4* expression is (also) crucially influenced by specific sites more upstream (ca. 2.6 kb) and a set of distant enhancer [111, 112]. Also, we previously showed that even though high promoter methylation is generally associated with low expression, this is not always the case [68].
PRSS21 chr16 (+) 2,867,164 – 2,876,305	**GCT link:** *TESTISIN* (PRSS21) is a proposed tumor suppressor gene in TGCT regulated by methylation of a 385bp long CpG rich island [113] and CpG sites close to the TSS [114]. **Findings:** Al GCT subtypes except SS (DMR[<u>SE/DG</u>-SS]) showed hypermethylation of *PRSS21*.
RUNX3 chr1 (-) 25,226,002 – 25,291,612	**GCT link:** 90% of the infantile YSTs (type I) showed methylation of *RUNX3* while methylation was only rarely observed in the adult GCTs [59, 115, 116]. **Findings:** The promoter region of *RUNX3* was consistently hypomentylated, progressing to hemimethylation on larger distances from the TSS (except SS). *RUNX3* only showed differential methylation between the cell lines, most consistently showing hypomethylation in NCCIT and hypermethylation in 2102EP.
SOX17 chr8 (+) 55,370,495 – 55,373,456	**GCT link:** Discriminative marker between EC (+) and SE (-) [17, 47]. **Findings:** *SOX17* was consistently hypomethylated in all tumor groups and cell lines (Figure 5D).
SOX2 chr3 (+) 181,429,712 – 181,432,224	**GCT link:** Discriminative marker between EC (+) and SE (-) [17, 47]. Previously identified DMR upstream of TSS between (≈50%) TCam-2 and (≈0%) NCCIT [68] ≈1kb upstream of the *SOX2* TSS. The region directly upstream of the *SOX2* TSS has consistently been found hypomethylated in both cell lines [68, 117]. TCam-2 has been shown to differentiate and become *SOX2* positive after extra-gonadal injection in mice [118]. **Findings:** A region ≈1 kb upstream of the *SOX2* TSS was differentially hypomethylated in all CL_ECs as compared to TCam-2 (File S2). EC and SE tumor samples showed consistent hypomethylation of the region -154 – -2283bp upstream of the *SOX2* TSS in contrast to the TE samples which showed higher levels of methylation (DMR[EC/MNS-**<u>TE</u>**], Figure 4B).
TFAP2C (*AP-2γ*) chr20 (+) 55,204,358 – 55,214,339 *TFAP2A* (*AP-2α*) chr 6 (-) 10,393,419 – 10,419,892	**GCT link:** *AP-2γ* is crucial for progression of PGCs into the germ line [119]. It is a known germ cell marker, abundantly expressed in CIS and SE, and heterogeneously expressed in NS and somatic tumors [119, 120]. *AP-2γ* expression is induced by estrogens [121]. Epigenetically, ChIP-seq analysis targeting activating histone marks showed strong enrichment of *AP-2α* and *AP-2γ* motifs in the SE-like cell line TCam-2 [68]. **Findings:** *TFAP2A* showed mostly hypomethylation in all tumor groups and cell lines. Only NCCIT was showed significantly increased methylation at the gene coding region compared to the other cell lines (File S2). All TE samples showed a non-significant block of hemimethylated probes close to the TSS of *TFAP2A*. *TFAP2C* was consistently hypomethylated in all tumor groups and cell lines.

Gene & region	Description
XIST chrX (-) 73,040,486 – 73,072,588	**GCT link:** *XIST* is completely methylated in male somatic cells, in contrast to female somatic cells. Testicular GCTs show hypomethylation of the 5' end of *XIST* which, have been suggested for TGCT diagnostics [122] but has so far not been validated. SE/NS/SS showed *XIST* expression (X inactivation) [66]. **Findings:** *XIST* showed no significant differential methylation in the comparison of the tumor groups or cell lines. Female gonadal tumors, SE and SS showed a trend towards less methylation as compared to the strongly methylated profile of the non-seminomatous tumors and male type I TE.
ICR_P: *SNURF/SNRPN* chr15 25,199,934 – 25,200,343	**GCT link:** *SNURF/SNRPN* has been described to show derivation from somatic imprinting in type II GCTs (non-quantitative, not necessarily indicating erasure) [123]. Low, but not absent methylation in non-allele-specific analysis [124]. Schneider and colleagues showed absence of the methylated band in bisulfite restriction analysis in 9 dysgerminomas [125]. **Findings:** In this dataset, this *SNURF/SNRPN* (controlling paternal expression) was only covered by a single probe (Figure S6). This very limited evidence suggests somatic imprinting in the type II tumors and sacral TE and uniparental status in the other subtypes: loss of imprinting in the I.TE.m.t and complete methylation in the ovarian tumors (DC, I.TE.f.o).
ICR_P: *MEST* chr7 130,130,740 – 130,133,111	**GCT link:** The *MEST* ICR regulates paternal expression, is already erased in fetal spermatogonia and remains so during male germ cell development [126]. **Findings:** The imprinting during germ cell development is reflected in our findings: (1) hypomethylation in the testicular type I TE and SS, (2) somatic imprinting in the type II tumors, (3) somatic-high imprinting in the ovarian and sacral TE, (4) high methylation in DC (Figure 7A).
ICR_M: *H19-IGF2* chr11 2,020,834 – 2,023,499	**GCT link:** *H19* (M expressed) and *IGF2* (P expressed) are inversely controlled by this ICR upstream of *H19* [124]. In mice oocytes are erased at *H19* before meiosis while bialelic methylation occurs before the gonocyte stage in males [127]. In humans *H19* is erased in fetal spermatogonia, but becomes fully methylated before meisosis (spematogonia) [126]. *H19* erasure fis unctionally illustrated in [128] and related to pluripotency markers (SOX2 and OCT3/4) in germ cell development in [129]. Previous studies using have suggested low methylation of the *H19-IFG2* ICR in a variable, but generally high percentage of the type II GCTs. This has generally been interpreted as imprinting erasure. Somatic imprinting has been shown in non-gonadal TE and mimicking of female germ cells has been seen in ovarian TE. Most studies investigated imprinting indirectly using allele specific expression limiting the sample sizes because of the mandatory presence of SNPs in this analysis to be informative [125, 130-132]. But a number of studies inquired the DNA methylation status directly using bisulfite restriction analysis, identifying consistent demethylation of one allele and variable methylation of the other in allele specific analysis and low, but not absent methylation in non-specific analysis [124, 133]. Low-somatic imprinting in DG was also shown by Amatruda *et al.* in a high throughput approach [20]. **Findings:** The SS in our series show complete methylation at 1 of the two *H19/IFG2* sites indicating a paternal committed origin. The sacral TEs exhibit mainly a somatic pattern, presumably indicating a pre-erasure origin. The gonadal I TE/DC show the lowest level of methylation presumably representing (partial) erasure (I.TE.m.t, TE) or complete maternal imprinting (I.TE.f.o, DC). Type II GCTs were found to consistently show somatic imprinting (Figure 7B; 2 regions from literature: Table S4).

Imprinting status and X chromosome reactivation

As reviewed in the introduction, gradual and tightly controlled establishment of uniparental imprinting and X chromosome reactivation (female only) has been demonstrated in developing germ cells which is at least partly mirrored in their malignant counterparts. Regarding imprinting controlled regions (Figure 1C, Table S4) in the tumor groups probes covering regions that are regulating paternally expressed genes (ICR_P) showed somatic methylation in type I and II GCTs with a trend towards hypermethylation in DC (Figure 6A). SS and the cell lines showed hypomethylation of ICR_Ps, a distinction also visible in the PCA plots. In IMPR_P200/1500 the pattern of the ICR_P probes seems to be pooled with a set of unmethylated probes (type I, II, IV GCT) presumably indicating contamination by non-imprinting related regions and hence not informative for imprinting status (Figure S2A, p.15,16). A somatic methylation state was shown for ICR_M except in the SS (bimodal) and the CL_SE (hypomethylated); a difference corroborated by the separation of these groups in the PCA plot (Figure 6B). IMPR_M200/IMPR_P1500 probes showed hypomethylation similar to non-imprinted genes in all groups (Figure S2A, p.18,19). No reactivation of chromosome X was seen in GCTs from female patients, which is reflected by the consistent 50% median methylation of the X chromosome in these cases (Figure 6C). The cell lines did not reflect the imprinting status of their *in vivo* counterpart, warranting caution when using the cell lines as a GCT model system in methylation based experiments.

Methylation status of ICR_Ps and ICR_Ms was similar between individual samples of the same histology (Figure S2B) with the exception of type I TE and DC (Figure 6D, S2B). TEs were investigated individually, grouped according to sex and anatomical site, in line with sex specific imprinting occurring during fetal/germ cell development (Figure 6D). The genome-wide methylation pattern was similar for all TEs. No reactivation of chromosome X was seen in the GCTs from female patients. Sacral type I TEs showed somatic imprinting patterns both in males and females. In line with sex specific imprinting, ICR_P sites in testicular type I TEs were relatively hypomethylated compared to sacral TEs. In contrast, ovarian type I TEs showed a tendency towards hypermethylation. Of note, testicular type I TE also showed a trend towards hypomethylation in ICR_M (only 18 probes). On the other hand, the expected inverse pattern of ICR_P was seen in the ovarian TEs at the ICR_M sites. A pattern similar to ovarian TE was observed in the individual DC samples: heterogeneity and gradual deviation from biparental imprinting towards uniparental maternal imprinting. Two out of three type II TEs showed a somatic imprinting pattern of both ICR_P and ICR_M.

A. ICR_P

B. ICR_M

C. X-chromosome

D. TE - individual samples - all probes

ICR_P

ICR_M
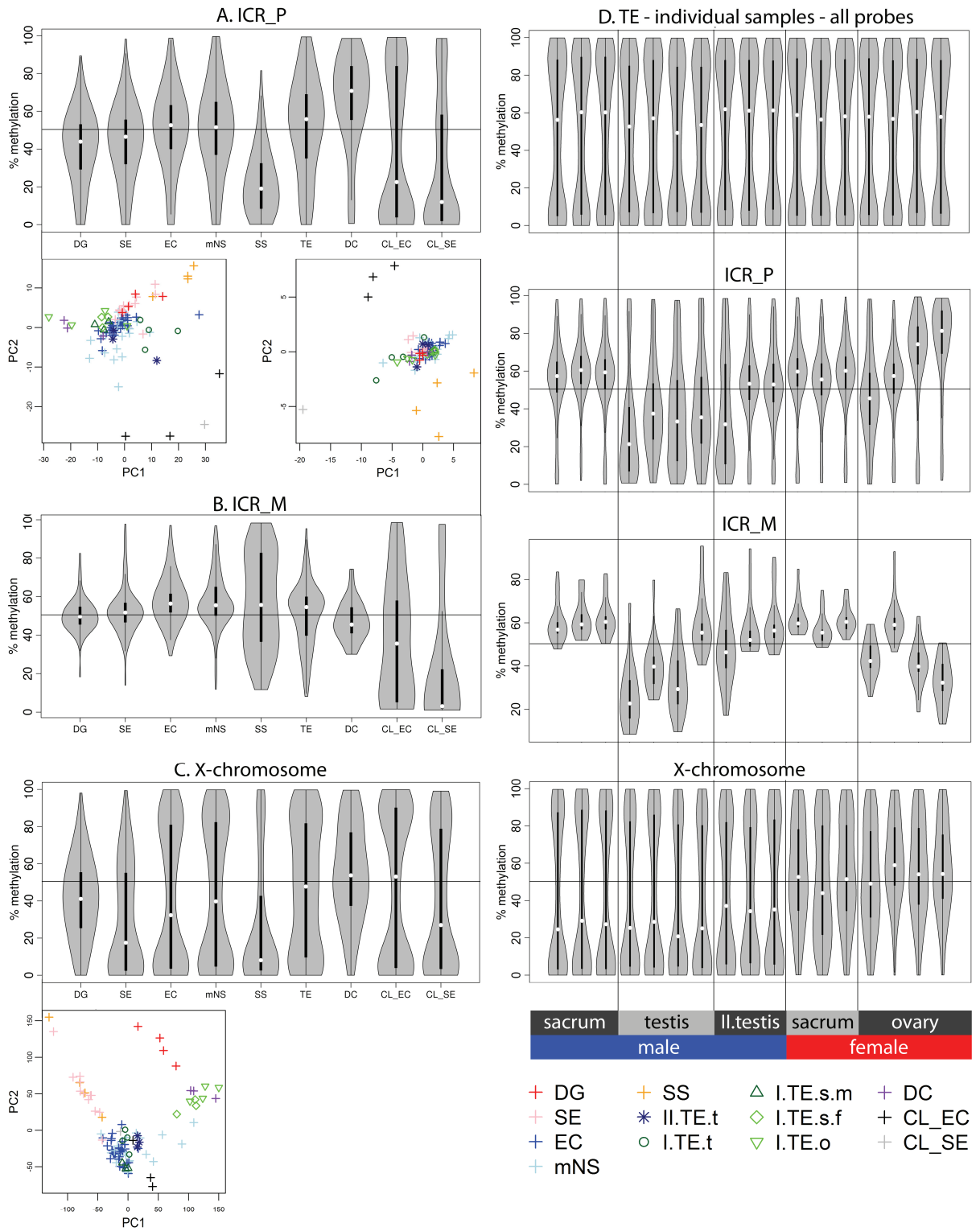
X-chromosome

← Figure 6. Methylation of imprinting control regions and the X chromosome. Analogous to Figure 2 the differences in methylation status between histological GCT subtypes is illustrated by two methods. Firstly, the methylation pattern is visualized using the distribution of the methylation percentage β. Next, the discriminatory power of the methylation pattern for each individual sample is shown using principal component analysis. Please see the legend of Figure 2 for a more detailed description of the visualizations. **(A)** All probes associated with paternally expressed genes (ICR_P). **(B)** All probes associated with maternally expressed genes (ICR_M). **(C)** All probes located on the X chromosome. **(D)** Distribution of methylation in individual TE samples ordered by sex and localization. To compare type I and II TE the n=3 type II pure TEs from the mNS were included in this visualization. Methylation levels of all probes, and probes associated with ICRs (P/M) and probes on the X chromosome are subsequently shown.
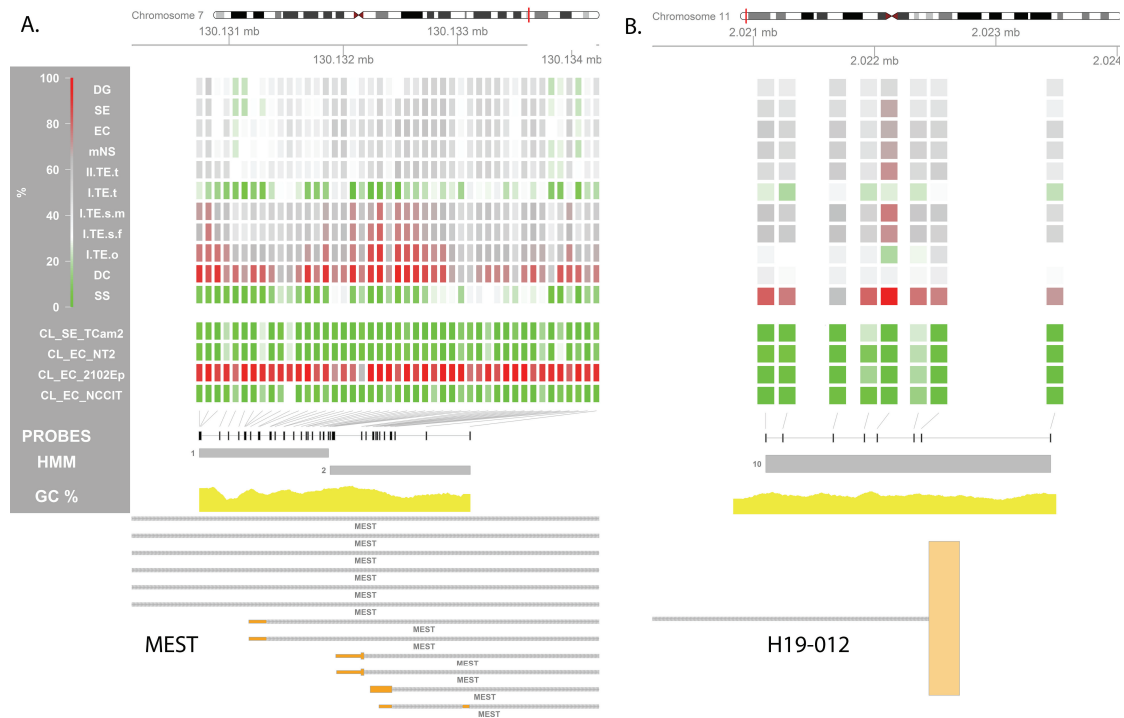

Figure 7. Methylation status of imprinting control regions. Visualization of the methylation percentage at specific loci is used to zoom in on a predefined region and investigate local imprinting differences between GCT subtypes. Two illustrative regions are depicted. **(A)** ICR_P: *MEST*. **(B)** ICR_M: *H19-IGF2*. The overlapping *H19* transcript is an aberrant, long alternative transcript (H19-012, ENST00000428066). This ICR regulates *H19* and *IGF2* expression and lies upstream all other transcripts of *H19*. The other ICRs are visualized in Figure S6 and listed in Table S4. The visualization is described in the legend of Figure 4. Abbreviations of histological subtypes are explained in Figure 1A and for specific subtypes of TE in the legend of Figure 2. CL indicates cell lines.

Validated ICRs (Table S4) were also studied individually. After merging overlapping validated ICRs from literature, 28 unique ICRs remained of which 21 were covered by the 450K array (4 ICR_M, 16 ICR_P, 1 unknown). ICRs controlling the expression of *H19/IGF2*, *SNURF/SRPN* and *MEST* have been studied in GCTs previously (review & results in Table 2). In the ICR_Ps which constitute the majority of the validated ICRs, the dominating pattern is: (1) somatic methylation in the type II tumors (2) hypomethylation in the type I testicular TEs and SS and (3) a trend towards hypermethylation in DC and ovarian TE. (Figure 7AB,S6).

In summary, ovarian type I TE and DC showed partial sex specific uniparental maternal imprinting, inverse of the uniparental paternal imprinting of SS. Testicular type I TE shows a trend towards erasure and type II GCTS (SE/DG/EC/mNS) showed somatic imprinting status.

Discussion

This study provides a detailed overview of the differences in global and local methylation status between type I-IV GCTs (Figure 1) and relates it to their cell of origin during normal germ cell development. Normal germ cell maturation includes complete de- and subsequent remethylation. Establishment of sex specific uniparental imprinting is physiological as is reactivation of chromosome X in female gametes. The largest methylation differences were detected between the hypermethylated EC/mNS + TE and hypomethylated SS + SE/DG groups, in line with previous reports [14, 42, 116, 134] (Figure 2A). However, the methylation profiles also allowed for a more detailed separation of EC/mNS, SE/DG, TE/DC and SS clusters, which is in line with the differentiation status of the tumors and their cell of origin. This distinction was also apparent when specific functional genomic regions were evaluated (Figure 2B). Hypermethylation in EC/mNS and TE is concentrated at non-transcription related regions when compared to SE/DG, pointing to a global difference in methylation status rather than differential methylation of specific regulatory elements. Moreover, EC/mNS is somewhat more methylated than TE and shows specific differences at transcription regulating genomic regions including genes implicated in male germ cell development. Regarding type III tumors, differential hypomethylation in SS relative to SE/DG is enriched for paternally expressed imprinting associated regions and DMRs cover male germ cell related genes (Figure 3, 4, 5, Table 1, 2). In addition, marked differences in imprinting status were observed. Ovarian type I TE and DC showed partial uniparental maternal imprinting, inverse of the uniparental paternal imprinting of SS. Testicular type I TE shows a trend towards imprinting erasure and type II GCTS (SE/DG/EC/mNS) showed somatic imprinting status (Figure 6, 7). The local and global methylation difference observed between GCTs could be matched to physiological germ cell development (Figure 8).

Limited knowledge exists about the progenitor of type I tumors. The absence of CIS and clinically different presentation (pediatric, frequently extra-gonadal, fully differentiated histology: TE/YST) sets them apart from the type II tumors [16-18]. Their bimodal global methylation status could a pattern generally observed in normal differentiated tissues and in very early germ cell progenitors (pre-migration. Historically type I and II tumors are also thought to be different with regard to their imprinting status. Imprinting status in these tumors was earlier shown to be somatic (biparental) or partially erased in case of the type I tumors and erased in case of the type II GCTs [16]. This positions the progenitor cell of type I tumors before imprinting erasure in the gonad. Indeed biparental (somatic) imprinting status in extra-gonadal TE was confirmed in this study and by Amatruda and colleagues [20]. There is a trend towards imprinting erasure in testicular type I TE. Ovarian type I TE show a trend towards completely maternal imprinting, but starting from a biparental status (50%), not showing any evidence of prior complete erasure (Figure 6D). This (partial) mimicking of female germ cells in ovarian TE is in line with in several studies [20, 125,

132]). However, the non-erased imprinting status, inactivated X chromosome and generally methylated state fits with the cell of origin at the very early PGC stage, which is then blocked in physiological complete demethylation, erasure and X reactivation and, when subjected to a gonadal micro-environment, shows partial erasure/uniparental imprinting [16-18] (Figure 8).
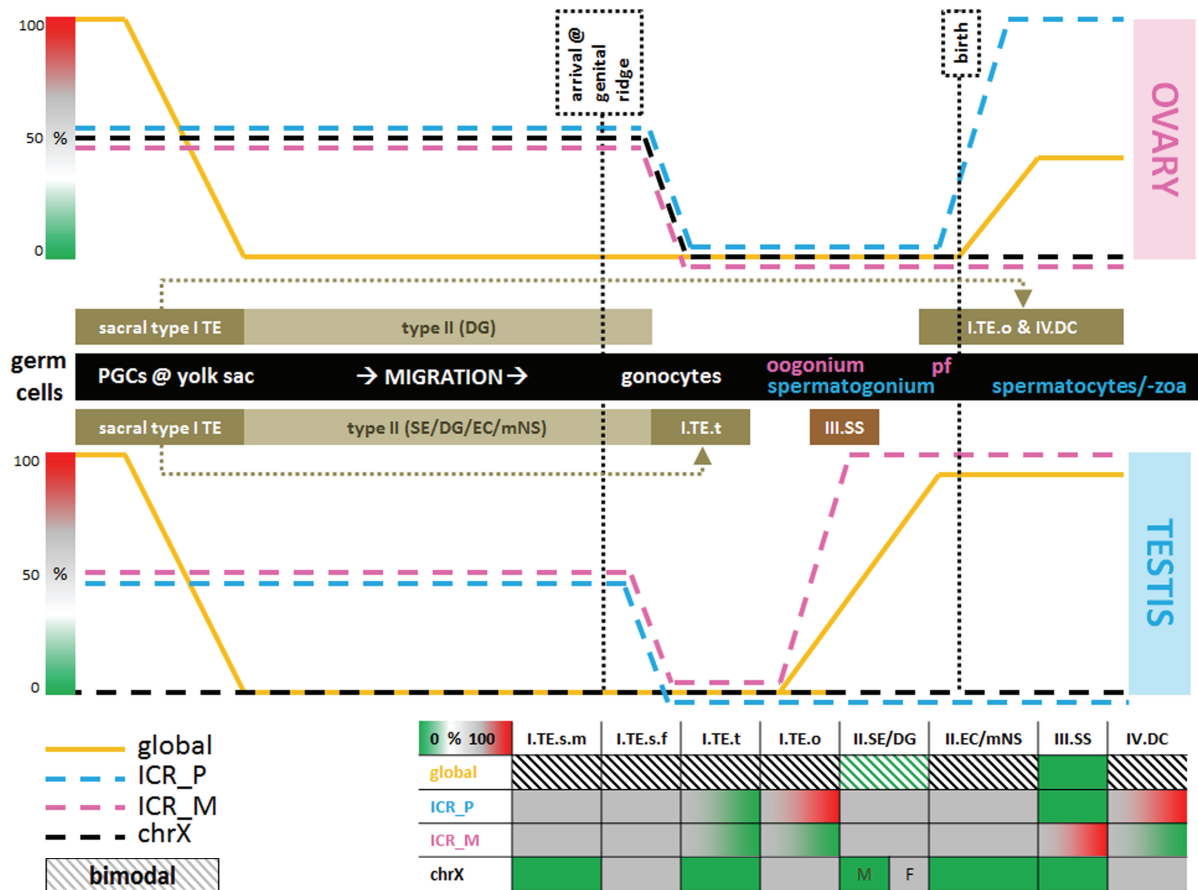

Figure 8. GCT methylation status in context of methylation during germ cell development. The top and bottom line charts depict normal germ cell development in female and male respectively (stages specified in the middle black bar). Methylation status during normal germ cell development is depicted for the global genome, ICRs and chromosome X (see Discussion). Putative cells of origin of the various types of GCTs are indicated in the brown boxes. ICR_P/M=ICR regulating paternally/maternally expressed genes. Bimodal indicates a methylation pattern peaking 0 and 100% with the exception of SE/DG (between 0 and ≈50). The table (bottom) provides a summary of the results, mainly Figures 2 and 6. Abbreviations: pf=primordial follicle. Type I tumors are indicated with their type (I), sex (m=male, f=female) and location (s=sacral, t=testis, o=ovary). Other GCT subtypes are indicated with their type (I, II, IV) and the abbreviation of each histological class, which are explained in the main text. Gradient bars indicate percentages of methylation (0→100%, green-white-grey-red) analogous to the gradient used in the other figures.

Most data is available on the epigenetic constitution of the type II tumors, as reviewed before [13, 21]. A strongly hypomethylated state was recently shown for all CIS, the common precursor of SE and EC [135]. Earlier studies have suggested separated NS-CIS and a SE-CIS types [134], but the lack of methylation in CIS combined with absence of SOX2 (EC marker) expression [47, 135, 136] increases the likelihood of a single precursor and progression into SE or NS. The CIS-like state is evident in the hypomethylated profile of SE/DG as shown in this article and previous research [14,

42, 116, 134, 135]. EC and mNS show a (*de novo*) methylated profile (Figure 2A). This is in line with the previously reviewed increased methylation in the transition of CIS into NS [13, 14, 42, 137], possibly illustrating reversal to a hypermethylated ES like state [7, 16, 138-141] or a bimodal methylation state normally present in differentiated tissues as shown in the differentiated NS. The consistent somatic imprinting pattern in general and at specific ICRs (Figure 6, S6, Table S4) was in line with an earlier report [20] but contrasted with targeted studies suggesting erased imprinting status at specific ICRs in these tumors using mainly indirect methods (allele specific expression analysis) and or non-quantitative methylation analysis (bisulfite specific restriction enzymes) (for review Table 2). The hypomethylated progenitor and somatic imprinting pattern (Figure 6A,B) situates the cell of origin of the type II tumors possibly earlier than previously described [16]: after global demethylation but before imprinting erasement, which is also in line with the occurrence of extra-gonadal type II GCTs (brain, anterior mediastinum) and their totipotent, embryonic stem cell like potential [16, 138-141] (Figure 8).

The other GCT subtypes are historically hypothesized to originate from more mature germ cell progenitors. Their marker profile has placed the type III tumors at the pre-spermatogonium state with regard to their cell of origin [36-39, 45]. Earlier epigenetic data showed a heterogeneous profile of histone modification and methylation profiles, not corresponding with a pre-spermatogonial origin [142]. Our limited series of SS show a consistent pattern of distinct hypomethylation and loss of imprinting at the paternally expressed ICRs (ICR_M: heterogeneous ≥ 50%, Figure 2B). This matches with a cell of origin between the gonocyte and spermatogonium stage, after establishment of uniparental imprinting but before initiation of *de novo* methylation. The type IV tumors (DC) show a pattern comparable to other differentiated tissues (ovarian type I TE) and show a general trend towards uniparental maternal imprinting but not starting from a completely erased state, potentially placing their cell of origin and pathogenesis parallel to the type I ovarian TE and not as a separate entity originating from a completely maternally imprinted an differentiated female germ cell as described before [16] (Figure 6,8,2B).

In conclusion this exploratory study of genome wide methylation profiles of GCT subtypes identified specific and global methylation differences, providing novel insight into the developmental timing and underlying biology of the various subtypes of GCTs and their (embryonic) cells of origin. However interpret the function of differential methylation between GCT subtypes, targeted validation the findings using matched expression data or careful evaluation of the effects of methylation in cell line models of GCTs is a crucial next step, even though validation of a biological relevant and representative DMR in *microRNA-371/2/3* (Table 2) showed excellent match with the results of bisulfite sequencing. The in-depth review of related literature and extensive accompanying online data (supplementary and on GEO) serve as a hypothesis generating source for future research.

Materials and Methods

Samples

Patient samples

Use of tissue samples remaining after diagnosis for scientific reasons was approved by Medical Ethical Committee (MEC) of the Erasmus MC Rotterdam (The Netherlands), permission 02.981. This included the permission to use the secondary tissue without further consent. Samples were used according to the "Code for Proper Secondary Use of Human Tissue in The Netherlands[b]" developed by the Dutch Federation of Medical Scientific Societies (FMWV[c] (Version 2002, update 2011)). An overview of the samples in this study is presented in Figure 1A,B.

Cell lines

Four cell lines were included (Figure 1B), all modelling type II GCTs. Cell lines derived from EC (CL_EC) include NT2[143-147], NCCIT [144, 148] and 2102EP[143-147]. TCam-2 closely resembles SE (CL_SE) [149-151]. TCam-2 was grown in RPMI1640 (#61870-010, Thermo Fisher Scientific / Life Technologies, Carlsbad, CA, USA). 2102EP and NCCIT were grown in DMEM/F12 (#12634-010, Thermo Fisher Scientific / Life Technologies). NT2 was grown in DMEM high glucose (#31966-021, Thermo Fisher Scientific / Life Technologies). All cell lines were cultured in T75 cm^2 flasks to 75-90% confluence, each with the addition of 10% Fetal Calf Serum (#CH30160.03, FCS, GE Healthcare Life Sciences, HyClone Laboratories, Utah, USA) and 1/100 Penicillin/Streptomycin (#15140, Thermo Fisher Scientific / Life Technologies).

Methylation profiling

DNA was isolated as described in [109]. The GCT material used contained > 75% tumor cells. Bisulfite conversion and methylation detection was performed using Illumina's HumanMethylation450 BeadChip (450K array) and exported as described in [68]. This array does not distinguish between DNA methylation variants like 5mC and 5hmC [152].

---

[b] http://www.federa.org/codes-conduct
[c] http://www.federa.org/

Data analysis

Data (pre-)processing

Further processing was carried out in R using the LUMI[d] package [153] according to [154, 155]. In the raw data, no structural differences in quality or batch effects were observed. Poorly performing probes (detection p<0.01 in > 95% of the samples), cross hybridizing probes and probes with a SNP at or within 10 bp of the target CpG (allele frequency >= 0.05) were excluded [155]. As a result 44,540 probes were discarded, leaving 437,881 valid, methylation related probes for processing and analysis. Finally, color adjustment, quantile normalization and BMIQ-based correction for probe type bias (Infinium I vs II) were performed [153, 154, 156]. Data processing resulted in two quantifications of a CpG site's methylation status: the methylation percentage β and an associated M-value which (logit$_2$(β)). M-values were used for statistical computations because of a more favorable tradeoff between true positive rate and detection rate [157]. All data is available via GEO (GSE58538[e]).

(Additional) annotation 450K array

The 450K annotation manifest[f] (v1.2) as supplied by Illumina contains a number of functional genomic classes like a probe's association with CpG islands, gene coding regions, etc. The manifest was extended with (additional) functional genomic classes, based on the GRch37/hg19 assembly. Briefly, probes close to small nuclear RNAs and microRNAs from snoRNABase[g] and miRBase[h] were identified, as were probes within repeats defined by RepeatMasker[i] (source: UCSC[j]). Probes close to the transcription start site (TSS) of imprinted genes were also identified (geneimprint.com / igc.otago.ac.nz). Known imprinting control regions (ICR) were retrieved from WAMIDEX[k] and igc.otago.ac.nz. Imprinting is indicated using the expressed allele. Illumina probe classes were extended with a number of merged categories. Where applicable, the upstream (-) and downstream (+) margins reported in this manuscript are analogous to the Illumina annotation (-1500+0; -200+0). The eighteen functional categories of primary interest to this manuscript are illustrated in Figure 1C. The extended annotation including its documentation is available at GEO (GPL18809[l]).

---

[d] http://www.bioconductor.org/packages/release/bioc/html/lumi.html
[e] http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58538
[f] http://support.illumina.com/downloads/humanmethylation450_15017482_v12.ilmn
[g] https://www-snorna.biotoul.fr/
[h] http://www.mirbase.org/
[i] http://www.repeatmasker.org/
[j] http://genome.ucsc.edu/
[k] https://atlas.genetics.kcl.ac.uk/
[l] http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL18809

Analysis protocol

Below, the subsequent steps of the data analysis are described. More details are presented in Figure S1. Depending on the context, "feature" can refer to a probe or a segment. All results are based on the GRch37/hg19 assembly.

Global methylation

Violin plots were created per histological subtype using all (global) or functional subsets of 450K probes. Violin plots ([vioplot package](http://cran.r-project.org/web/packages/vioplot/index.html)[m]) integrate the benefits of a boxplot and a kernel density plot. Two-dimensional principal component analysis (PCA) was applied and bootstrap-validated to assess how well the methylation values of (subsets of) the probes separated the GCC subtypes.

Defining genomic segments and discriminative methylation states

To detect regions of interest rather than only selecting individual differentiating probes (CpG sites) a HMM was trained on the tumor samples. Without a priori information about tumor type, the HMM combines adjacent probes into segments and assigns these segments to k mutually exclusive states, each with distinct methylation profiles over all tumor samples. k=20 was used as the likelihood of the model saturated around this number of states (Figure S1, p.11). In total, 133,730 segments were identified. The median methylation value (M or β) of all probes in a segment or state was taken as methylation proxy. As a proof of concept, Figure S1 (p. 17) shows clear separation of male and female samples based on state 15 which almost exclusively contains probes on the X chromosome. The result of the HMM is included in the GEO submission of the data ([GSE58538](#)) and its properties/procedures are summarized in Figure S1.

Differentiating features (probes or segments)

Features showing low variability over all samples were excluded before formally testing for differential methylation ($\sigma_{M,probes}$<0.8, n=77,154/437,881 (17,62%) & $\sigma_{M,segments}$<0.6, n=13,229/133,730 (9,89%), Figure S1, p.8). A Mann Whitney U test was applied to each feature, comparing the distribution of M values between two histological subtypes. If significant ($p<0.05$, Benjamini-Hochberg corrected [158]), the subtype specificity was validated in 100 stratified bootstrap samples. If the feature proved to be significant in ≥95% of the validation samples and showed a difference in median M values > |0.9| between the pair of histological subtypes it was considered potentially discriminating. The value of 0.9 was chosen as the mean of the cut-off range recommended by Du and coworkers (0.4-1.4) [157]. Although a less stringent setting might result in a higher detection rate, it will considerably reduce the true positive rate [157]. The sign of the difference in median M value was used to assign a relative methylation status (hyper/hypo) in either of the two subtypes under pairwise consideration.

---

[m] http://cran.r-project.org/web/packages/vioplot/index.html

Differentiating HMM states

To identify non-adjacent regions that showed similar patterns of methylation a logistic LASSO[n] regression model was fitted on the M values of the HMM states (glmnet packgage[o]) [159]. Coefficients > 0 were selected from the most regularized regression model within 1 standard deviation of the model with minimal cross validation error. A 10 fold cross validated λ was used. Features included in the selected state(s) and showing a difference in median M values > |0.9| (see above) between the pair of histological subtypes compared were considered potentially discriminating. The sign of the difference in median M value was used to assign a relative methylation status (hyper-/hypomethylated) in each of the subtypes.

Final selection of differentially methylated probes (DMPs)

Features of interest were identified in the intersection of (1) all probes in discriminating states, (2) all probes in discriminating segments and (3) all individually discriminating probes (Figure S1 (p. 3/5), sections 2.3.3.3 & 2.3.3.4). This was done separately for probes that showed relative hypermethylation in either of the subtypes under pairwise comparison. This way, two groups of differentially methylated probes (DMPs) were identified, showing relative hypermethylation in one subtype and relative hypomethylation in the other.

Functional enrichment

The sets of DMPs were subjected to enrichment analysis for 18 functional categories (Figure 1C) using a two sided Fisher's Exact test. Analogously, association with chromosome and state was tested. $p < 0.05/(18+24+20) = 0.00080645161$ was considered significant, hence retaining a Bonferroni corrected Type I error rate of 5% (18 functional categories, 24 chromosomes, 20 states).

Differentially methylated regions (DMRs)

Regions with ≥ 5 adjacent DMPs and a maximal inter-DMP distance ≤ 1 kb were identified as DMRs between the tumor groups. Annotations were retrieved for DMRs including flanking regions of 20% of the length of each DMR.

Analysis of the cell lines

Cell lines were compared to tumor samples in the evaluation of specific regions of interest in the tumor samples and with regard to their global methylation profile. Moreover, they were analyzed using the DMRforPairs[p] package to identify specific DMR in these unique samples ([160], using the default settings except min_dM=0.9, see above). For the NCCIT and TCam-2 cell lines this analysis matches the one performed in [68].

---

[n] http://statweb.stanford.edu/~tibs/lasso.html
[o] http://cran.r-project.org/web/packages/glmnet/index.html
[p] http://www.bioconductor.org/packages/release/bioc/html/DMRforPairs.html

Software

Analyses were performed in R 3.1.0/Bioconductor 2.14 (Windows 7 x64) and 2.15.2/Bioconductor 2.11 (Redhat Linux x64). Additional R packages: parallel[q] & gplots[r]. The HMM was trained using Matlab R2012a (64 bits) using the MOGHMM[s] toolbox.

Acknowledgements

Supporting Information

Figure S1. Supplementary methods. A flowchart / detailed description of the analysis protocol, the motivation for the variance filtering, the properties of the Hidden Markov Model (HMM) and HMM state 15 discriminating male and female samples are presented subsequently.

Figure S2. Methylation patterns in GCT subtypes and cell lines. (A) All categories and validation. In addition to the selected categories presented in Figure 2 this figure contains all 18 functional categories presented in Figure 1C and includes the primary PCA as well as an its validation (i.e. robustness of the result). PCA was performed on the total dataset (left) and validated using stratified bootstrapping (middle: training, right: validation). Please see the legend of figure 2 for an explanation of the density and PC plots/legends. **(B)** Global methylation patterns in individual samples. X-axis indicates arbitrary sample ID. The sex of the patient from which the sample originates is indicated in blue (male) or red (female). Density plots are explained in the legend of Figure 2. Distributions are shown for all probes individual per sample. The ICR_P and ICR_M categories are presented separately to facilitate the discussion about imprinting. The red dashed line indicates somatic imprinting (50%). Please note that details on the TE group are presented in the main text (Figure 6D) and that this category is therefore omitted here. This also holds for the n=3 type II pure TE included in the mNS group.

Figure S3. Enrichment of differentially methylated probes (DMPs) for chromosomal position and HMM state. (A) Merged GCT subtypes in pairwise comparisons. The SE+DG and EC+mNS categories were merged because of high similarity in biological classification and methylation profile. Despite their similarities, the DC and type I TE because they belong to different histological classes. **(B)** Association between DMPs and chromosome / HMM state. Stacked bar charts indicate the fraction of probes in a subset (DMP[A-B], DMP[A-B], non-DMP) that is mapped to a specific chromosome or assigned to a specific state. Grey indicates the non-DMPs and red and green indicated the DMPs hypermethylated in the subtype with the matching color in the figure (alternating green/white=A, alternating red/white=B). *=significant over-/underrepresentation of DMPs relative to the non-DMP subset (tested per chromosome/state, 2-sided Fisher's exact test, see Methods for Bonferroni corrected α threshold). In the right bottom of each figure the coefficients of the LASSO regression model are depicted. These roughly match the strongest over- and underrepresentations identified by the Fisher's Exact tests on the states. The LASSO selected states are marked orange in the table indicating the significant associations between each state and either DMP group.

Figure S4. Methylation profile at GCT subtype specific differentially methylated regions (DMRs) - continued. This figure depicts the DMRs discussed in the main text in addition to those already visualized in Figure 4. Legend to the visualizations is provided in the legend of Figure 4.

Figure S5. Methylation status of (A) GCT specific genes and (B) genes with SNPs significantly associated with GCTs. This figure depicts the genes discussed in the main text and Table 2 in addition to those already visualized in Figure 5. Legend to the visualizations is provided in the legend of Figure 4. Genes are annotated 1.5kb upstream of their TSS and 1.5kb downstream of their transcription termination site.

Figure S6. Methylation status of known imprinting control regions (ICRs). ICRs identified as described in the materials and methods sections were checked for coverage on the 450K array. 21/28 unique ICRs were covered by one or more probes. These were visualized here (overview: Table S4). *H19_IGF2* regions: the overlapping transcript is an aberrant, long alternative transcript (*H19-012*, ENST00000428066). These ICRs regulates *H19* and *IGF2* expression and lie upstream all other transcripts of H19. Legend to the visualizations is provided in the legend of Figure 4.

Table S1. List of DMPs resulting from pairwise comparison of GCT subtypes.

Table S2. Counts, percentages, log scores and statistical test results for enrichment in functional genomic categories. (A) SE/DG vs EC/mNS; (B) SE/DG vs type I TE; (C) EC/mNS vs type I TE; (D) SE/DG vs SS. Rows indicate the functional categories. Columns indicate the number of probes in the non-DMP and both subtype specific DMP sets. Next, the fraction (%) of this count relative to all non-DMPs or either set of DMPs is presented. The log-scores are calculated as log2(%DMP/%non-DMPs) and visually presented in Figure 3 for those categories showing significant over-/underrepresentation. Significance of the enrichment was evaluated using a two-sided Fisher Exact test with a Bonferroni corrected α threshold as specified in the Materials & Methods section.

Table S3. DMRs between tumor groups List of DMRs for each pair of GCT subtypes. (Recurrent tumor DMRs) Gene symbols that occurred in more than one DMR; either irrespective of DMR subset (n.total.occurences) or in multiple independent DMR subsets (n.dmr.lists) (Overlap tumor and CL DMRs) Gene symbols involved in DMRs identified between both the tumor groups and the cell lines. The second column indicates in which tumor comparisons the gene symbol was involved in a DMR.

Table S4. Merged known ICRs from literature with sources. Also see Figure S6 for a visual representation of the methylation status at the ICRs if covered by the 450K array (21/28).

File S1. Differentially methylated regions between tumors classes as discussed in Table 1. All figures. Please investigate using the genomic position as search term. (ZIP, included in GEO submission GSE58538)

File S2. Differentially methylated regions between GCT cell lines. DMRforPairs output. Significant results only. Please start from the html file. (ZIP, included in GEO submission GSE58538)

Supplementary information will be made available upon publication.

References

1.   Messerschmidt DM: **Should I stay or should I go: protection and maintenance of DNA methylation at imprinted genes.** *Epigenetics : official journal of the DNA Methylation Society* 2012, 7:969-975.
2.   Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W: **Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.** *Nature* 2010, 463:1101-U1126.
3.   Reik W, Dean W, Walter J: **Epigenetic reprogramming in mammalian development.** *Science* 2001, 293:1089-1093.
4.   Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W: **The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.** *Mol Cell* 2012, 48:849-862.
5.   Seisenberger S, Peat JR, Reik W: **Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells.** *Curr Opin Cell Biol* 2013, 25:281-288.
6.   Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W, et al: **Programming and inheritance of parental DNA methylomes in mammals.** *Cell* 2014, 157:979-991.
7.   Matsui Y, Mochizuki K: **A current view of the epigenome in mouse primordial germ cells.** *Mol Reprod Dev* 2014, 81:160-170.
8.   Saitou M, Kagiwada S, Kurimoto K: **Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells.** *Development* 2012, 139:15-31.
9.   Payer B, Lee JT, Namekawa SH: **X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells.** *Hum Genet* 2011, 130:265-280.
10.   Ohhata T, Wutz A: **Reactivation of the inactive X chromosome in development and reprogramming.** *Cellular and molecular life sciences : CMLS* 2013, 70:2443-2461.
11.   Deuve JL, Avner P: **The coupling of X-chromosome inactivation to pluripotency.** *Annual review of cell and developmental biology* 2011, 27:611-629.
12.   de Napoles M, Nesterova T, Brockdorff N: **Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells.** *PloS one* 2007, 2:e860.
13.   Kristensen DG, Skakkebaek NE, Rajpert-De Meyts E, Almstrup K: **Epigenetic features of testicular germ cell tumours in relation to epigenetic characteristics of foetal germ cells.** *The International journal of developmental biology* 2013, 57:309-317.
14.   Wermann H, Stoop H, Gillis AJ, Honecker F, van Gurp RJ, Ammerpohl O, Richter J, Oosterhuis JW, Bokemeyer C, Looijenga LH: **Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance.** *J Pathol* 2010, 221:433-442.
15.   Almstrup K, Nielsen JE, Mlynarska O, Jansen MT, Jorgensen A, Skakkebaek NE, Rajpert-De Meyts E: **Carcinoma in situ testis displays permissive chromatin modifications similar to immature foetal germ cells.** *Br J Cancer* 2010, 103:1269-1276.
16.   Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5:210-222.
17.   Looijenga LH: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J Pathol* 2009, 218:146-162.
18.   Woodward PJ, Heidenreich A, Looijenga LHJ, al. e: **Testicular germ cell tumors.** In *World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs.* Edited by Eble JN, Sauter G, Epstein JI, Sesterhann IA, eds. Lyon: IARC Press; 2004: 17-278
19.   Rijlaarsdam MA, Looijenga LH: **An oncofetal and developmental perspective on testicular germ cell cancer.** *Seminars in cancer biology* 2014.
20.   Amatruda JF, Ross JA, Christensen B, Fustino NJ, Chen KS, Hooten AJ, Nelson H, Kuriger JK, Rakheja D, Frazier AL, Poynter JN: **DNA methylation analysis reveals distinct methylation signatures in pediatric germ cell tumors.** *BMC cancer* 2013, 13:313.
21.   Van Der Zwan YG, Stoop H, Rossello F, White SJ, Looijenga LH: **Role of epigenetics in the etiology of germ cell cancer.** *The International journal of developmental biology* 2013, 57:299-308.
22.   Boublikova L, Buchler T, Stary J, Abrahamova J, Trka J: **Molecular biology of testicular germ cell tumors: unique features awaiting clinical application.** *Critical reviews in oncology/hematology* 2014, 89:366-385.

23. Gobel U, Schneider DT, Calaminus G, Haas RJ, Schmidt P, Harms D: **Germ-cell tumors in childhood and adolescence. GPOH MAKEI and the MAHO study groups.** *Ann Oncol* 2000, 11:263-271.

24. Rescorla FJ: **Pediatric germ cell tumors.** *Seminars in surgical oncology* 1999, 16:144-158.

25. Sievers S, Alemazkour K, Zahn S, Perlman EJ, Gillis AJ, Looijenga LH, Gobel U, Schneider DT: **IGF2/H19 imprinting analysis of human germ cell tumors (GCTs) using the methylation-sensitive single-nucleotide primer extension method reflects the origin of GCTs in different stages of primordial germ cell development.** *Genes Chromosomes Cancer* 2005, 44:256-264.

26. Horwich A, Shipley J, Huddart R: **Testicular germ-cell cancer.** *Lancet* 2006, 367:754-765.

27. Looijenga LH, Van Agthoven T, Biermann K: **Development of malignant germ cells - the genvironmental hypothesis.** *The International journal of developmental biology* 2013, 57:241-253.

28. Kratz CP, Mai PL, Greene MH: **Familial testicular germ cell tumours.** *Best Pract Res Clin Endocrinol Metab* 2010, 24:503-513.

29. Chung CC, Kanetsky PA, Wang Z, Hildebrandt MA, Koster R, Skotheim RI, Kratz CP, Turnbull C, Cortessis VK, Bakken AC, et al: **Meta-analysis identifies four new loci associated with testicular germ cell tumor.** *Nat Genet* 2013.

30. Dieckmann KP, Pichlmeier U: **Clinical epidemiology of testicular germ cell tumors.** *World journal of urology* 2004, 22:2-14.

31. Cools M, Drop SL, Wolffenbuttel KP, Oosterhuis JW, Looijenga LH: **Germ cell tumors in the intersex gonad: old paths, new directions, moving frontiers.** *Endocr Rev* 2006, 27:468-484.

32. Czene K, Lichtenstein P, Hemminki K: **Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database.** *Int J Cancer* 2002, 99:260-266.

33. Oosterhuis JW, Stoop H, Honecker F, Looijenga LH: **Why human extragonadal germ cell tumours occur in the midline of the body: old concepts, new perspectives.** *International journal of andrology* 2007, 30:256-263; discussion 263-254.

34. Fan R, Ulbright TM: **Does intratubular germ cell neoplasia, unclassified type exist in prepubertal, cryptorchid testes?** *Fetal and pediatric pathology* 2012, 31:21-24.

35. Leman ES, Gonzalgo ML: **Prognostic features and markers for testicular cancer management.** *Indian journal of urology : IJU : journal of the Urological Society of India* 2010, 26:76-81.

36. Looijenga LH: **Spermatocytic seminoma: toward further understanding of pathogenesis.** *J Pathol* 2011, 224:431-433.

37. Lim J, Goriely A, Turner GD, Ewen KA, Jacobsen GK, Graem N, Wilkie AO, Rajpert-De Meyts E: **OCT2, SSX and SAGE1 reveal the phenotypic heterogeneity of spermatocytic seminoma reflecting distinct subpopulations of spermatogonia.** *J Pathol* 2011, 224:473-483.

38. Verdorfer I, Rogatsch H, Tzankov A, Steiner H, Mikuz G: **Molecular cytogenetic analysis of human spermatocytic seminomas.** *J Pathol* 2004, 204:277-281.

39. Rajpert-De Meyts E, Jacobsen GK, Bartkova J, Aubry F, Samson M, Bartek J, Skakkebaek NE: **The immunohistochemical expression pattern of Chk2, p53, p19INK4d, MAGE-A4 and other selected antigens provides new evidence for the premeiotic origin of spermatocytic seminoma.** *Histopathology* 2003, 42:217-226.

40. Chen YT, Chiu R, Lee P, Beneck D, Jin B, Old LJ: **Chromosome X-encoded cancer/testis antigens show distinctive expression patterns in developing gonads and in testicular seminoma.** *Hum Reprod* 2011, 26:3232-3243.

41. Nettersheim D, Heukamp LC, Fronhoffs F, Grewe MJ, Haas N, Waha A, Honecker F, Waha A, Kristiansen G, Schorle H: **Analysis of TET Expression/Activity and 5mC Oxidation during Normal and Malignant Germ Cell Development.** *PloS one* 2013, 8:e82881.

42. Netto GJ, Nakai Y, Nakayama M, Jadallah S, Toubaji A, Nonomura N, Albadine R, Hicks JL, Epstein JI, Yegnasubramanian S, et al: **Global DNA hypomethylation in intratubular germ cell neoplasia and seminoma, but not in nonseminomatous male germ cell tumors.** *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2008, 21:1337-1344.

43. Jeyapalan JN, Noor DA, Lee SH, Tan CL, Appleby VA, Kilday JP, Palmer RD, Schwalbe EC, Clifford SC, Walker DA, et al: **Methylator phenotype of malignant germ cell tumours in children identifies strong candidates for chemotherapy resistance.** *Br J Cancer* 2011, 105:575-585.

44. Ushida H, Kawakami T, Minami K, Chano T, Okabe H, Okada Y, Okamoto K: **Methylation profile of DNA repetitive elements in human testicular germ cell tumor.** *Molecular carcinogenesis* 2012, 51:711-722.

45. Looijenga LH, Hersmus R, Gillis AJ, Pfundt R, Stoop HJ, van Gurp RJ, Veltman J, Beverloo HB, van Drunen E, van Kessel AG, 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**:**290-302.

46. Eini R, Dorssers LC, Looijenga LH: **Role of stem cell proteins and microRNAs in embryogenesis and germ cell cancer.** *The International journal of developmental biology* 2013, 57:319-332.

47. de Jong J, Stoop H, Gillis AJ, van Gurp RJ, van de Geijn GJ, Boer M, Hersmus R, Saunders PT, Anderson RA, Oosterhuis JW, Looijenga LH: **Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.** *J Pathol* 2008, 215:21-30.

48. Rijlaarsdam MA, van Herk HA, Gillis AJ, Stoop H, Jenster G, Martens J, van Leenders GJ, Dinjens W, Hoogland AM, Timmermans M, Looijenga LH: **Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours.** *Br J Cancer* 2011, 105**:**854-863.

49. Wu J, Bao J, Wang L, Hu Y, Xu C: **MicroRNA-184 downregulates nuclear receptor corepressor 2 in mouse spermatogenesis.** *BMC developmental biology* 2011, 11**:**64.

50. Bremmer F, Thelen P, Pottek T, Behnes CL, Radzun HJ, Schweyer S: **Expression and function of the vitamin D receptor in malignant germ cell tumour of the testis.** *Anticancer research* 2012, 32:341-349.

51. Funk CD, Funk LB, FitzGerald GA, Samuelsson B: **Characterization of human 12-lipoxygenase genes.** *Proceedings of the National Academy of Sciences of the United States of America* 1992, 89:3962-3966.

52. Salpea P, Russanova VR, Hirai TH, Sourlingas TG, Sekeri-Pataryas KE, Romero R, Epstein J, Howard BH: **Postnatal development- and age-related changes in DNA-methylation patterns in the human genome.** *Nucleic Acids Res* 2012, 40:6477-6494.

53. Matson CK, Murphy MW, Sarver AL, Griswold MD, Bardwell VJ, Zarkower D: **DMRT1 prevents female reprogramming in the postnatal mammalian testis.** *Nature* 2011, 476:101-104.

54. Sarkardeh H, Totonchi M, Asadpour O, Sadighi Gilani MA, Zamani Esteki M, Almadani N, Borjian Boroujeni P, Gourabi H: **Association of MOV10L1 gene polymorphisms and male infertility in azoospermic men with complete maturation arrest.** *Journal of assisted reproduction and genetics* 2014.

55. Frost RJ, Hamra FK, Richardson JA, Qi X, Bassel-Duby R, Olson EN: **MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107**:**11847-11852.

56. Kano K, Kitamura A, Matsuwaki T, Morimatsu M, Naito K: **Discoidin domain receptor 2 (DDR2) is required for maintenance of spermatogenesis in male mice.** *Mol Reprod Dev* 2010, 77:29-37.

57. Bonnard C, Strobl AC, Shboul M, Lee H, Merriman B, Nelson SF, Ababneh OH, Uz E, Guran T, Kayserili H, et al: **Mutations in IRX5 impair craniofacial development and germ cell migration via SDF1.** *Nat Genet* 2012, 44**:**709-713.

58. Le Bouffant R, Souquet B, Duval N, Duquenne C, Herve R, Frydman N, Robert B, Habert R, Livera G: **Msx1 and Msx2 promote meiosis initiation.** *Development* 2011, 138:5393-5402.

59. Lind GE, Skotheim RI, Fraga MF, Abeler VM, Esteller M, Lothe RA: **Novel epigenetically deregulated genes in testicular cancer include homeobox genes and SCGB3A1 (HIN-1).** *J Pathol* 2006, 210:441-449.

60. Griffeth RJ, Carretero J, Burks DJ: **Insulin receptor substrate 2 is required for testicular development.** *PloS one* 2013, 8:e62103.

61. de Haas EC, Zwart N, Meijer C, Suurmeijer AJ, Meijer K, Guchelaar HJ, Hoekstra HJ, van Leeuwen FE, Sleijfer DT, Boezen HM, Gietema JA: **Association of PAI-1 gene polymorphism with survival and chemotherapy-related vascular toxicity in testicular cancer.** *Cancer* 2010, 116:5628-5636.

62. Ebisch IM, Steegers-Theunissen RP, Sweep FC, Zielhuis GA, Geurts-Moespot A, Thomas CM: **Possible role of the plasminogen activation system in human subfertility.** *Fertility and sterility* 2007, 87:619-626.

63. Graham PL, Kimble J: **The mog-1 gene is required for the switch from spermatogenesis to oogenesis in Caenorhabditis elegans.** *Genetics* 1993, 133:919-931.

64. Graham PL, Schedl T, Kimble J: **More mog genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of Caenorhabditis elegans.** *Developmental genetics* 1993, 14:471-484.

65. Kato N, Shibuya H, Fukase M, Tamura G, Motoyama T: **Involvement of adenomatous polyposis coli (APC) gene in testicular yolk sac tumor of infants.** *Human pathology* 2006, 37:48-53.

66.   Looijenga LH, Gillis AJ, van Gurp RJ, Verkerk AJ, Oosterhuis JW: **X inactivation in human testicular tumors. XIST expression and androgen receptor methylation status.** *The American journal of pathology* 1997, 151**:**581-590.

67.   Zhang C, Kawakami T, Okada Y, Okamoto K: **Distinctive epigenetic phenotype of cancer testis antigen genes among seminomatous and nonseminomatous testicular germ-cell tumors.** *Genes Chromosomes Cancer* 2005, 43**:**104-112.

68.   van der Zwan YG, Rijlaarsdam MA, Rossello FJ, Notini AJ, de Boer S, Watkins DN, Gillis AJ, Dorssers LC, White SJ, Looijenga LH: **Seminoma and embryonal carcinoma footprints identified by analysis of integrated genome-wide epigenetic and expression profiles of germ cell cancer cell lines.** *PloS one* 2014, 9**:**e98330.

69.   Rathi A, Virmani AK, Harada K, Timmons CF, Miyajima K, Hay RJ, Mastrangelo D, Maitra A, Tomlinson GE, Gazdar AF: **Aberrant methylation of the HIC1 promoter is a frequent event in specific pediatric neoplasms.** *Clin Cancer Res* 2003, 9**:**3674-3678.

70.   Koul S, McKiernan JM, Narayan G, Houldsworth J, Bacik J, Dobrzynski DL, Assaad AM, Mansukhani M, Reuter VE, Bosl GJ, et al: **Role of promoter hypermethylation in Cisplatin treatment response of male germ cell tumors.** *Mol Cancer* 2004, 3**:**16.

71.   Yabuta Y, Kurimoto K, Ohinata Y, Seki Y, Saitou M: **Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling.** *Biol Reprod* 2006, 75**:**705-716.

72.   Lennartsson J, Ronnstrand L: **Stem cell factor receptor/c-Kit: from basic science to clinical implications.** *Physiological reviews* 2012, 92**:**1619-1649.

73.   McLaren A: **Primordial germ cells in the mouse.** *Dev Biol* 2003, 262**:**1-15.

74.   Mithraprabhu S, Loveland KL: **Control of KIT signalling in male germ cells: what can we learn from other systems?** *Reproduction* 2009, 138**:**743-757.

75.   Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M, Clark AT: **The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation.** *Nat Cell Biol* 2013, 15**:**113-122.

76.   Mamsen LS, Brochner CB, Byskov AG, Mollgard K: **The migration and loss of human primordial germ stem cells from the hind gut epithelium towards the gonadal ridge.** *The International journal of developmental biology* 2012, 56**:**771-778.

77.   Farini D, La Sala G, Tedesco M, De Felici M: **Chemoattractant action and molecular signaling pathways of Kit ligand on mouse primordial germ cells.** *Dev Biol* 2007, 306**:**572-583.

78.   Runyan C, Schaible K, Molyneaux K, Wang Z, Levin L, Wylie C: **Steel factor controls midline cell death of primordial germ cells and is essential for their normal proliferation and migration.** *Development* 2006, 133**:**4861-4869.

79.   Devouassoux-Shisheboran M, Mauduit C, Tabone E, Droz JP, Benahmed M: **Growth regulatory factors and signalling proteins in testicular germ cell tumours.** *APMIS* 2003, 111**:**212-224; discussion 224.

80.   Oatley JM, Brinster RL: **The germline stem cell niche unit in mammalian testes.** *Physiological reviews* 2012, 92**:**577-595.

81.   Rossi P, Dolci S, Sette C, Geremia R: **Molecular mechanisms utilized by alternative c-kit gene products in the control of spermatogonial proliferation and sperm-mediated egg activation.** *Andrologia* 2003, 35**:**71-78.

82.   Honecker F, Stoop H, de Krijger RR, Chris Lau YF, Bokemeyer C, Looijenga LH: **Pathobiological implications of the expression of markers of testicular carcinoma in situ by fetal germ cells.** *J Pathol* 2004, 203**:**849-857.

83.   Rajpert-de Meyts E, Hoei-Hansen CE: **From gonocytes to testicular cancer: the role of impaired gonadal development.** *Ann N Y Acad Sci* 2007, 1120**:**168-180.

84.   Rorth M, Rajpert-De Meyts E, Andersson L, Dieckmann KP, Fossa SD, Grigor KM, Hendry WF, Herr HW, Looijenga LH, Oosterhuis JW, Skakkebaek NE: **Carcinoma in situ in the testis.** *Scand J Urol Nephrol Suppl* 2000**:**166-186.

85.   Goddard NC, McIntyre A, Summersgill B, Gilbert D, Kitazawa S, Shipley J: **KIT and RAS signalling pathways in testicular germ cell tumours: new data and a review of the literature.** *International journal of andrology* 2007, 30**:**337-348; discussion 349.

86.   Izquierdo MA, Van der Valk P, Van Ark-Otte J, Rubio G, Germa-Lluch JR, Ueda R, Scheper RJ, Takahashi T, Giaccone G: **Differential expression of the c-kit proto-oncogene in germ cell tumours.** *J Pathol* 1995, 177:253-258.

87.   Leroy X, Augusto D, Leteurtre E, Gosselin B: **CD30 and CD117 (c-kit) used in combination are useful for distinguishing embryonal carcinoma from seminoma.** *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2002, 50:283-285.

88.   Strohmeyer T, Reese D, Press M, Ackermann R, Hartmann M, Slamon D: **Expression of the c-kit proto-oncogene and its ligand stem cell factor (SCF) in normal and malignant human testicular tissue.** *J Urol* 1995, 153:511-515.

89.   McIntyre A, Summersgill B, Grygalewicz B, Gillis AJ, Stoop J, van Gurp RJ, Dennis N, Fisher C, Huddart R, Cooper C, et al: **Amplification and overexpression of the KIT gene is associated with progression in the seminoma subtype of testicular germ cell tumors of adolescents and adults.** *Cancer Res* 2005, 65:8085-8089.

90.   Mol CD, Dougan DR, Schneider TR, Skene RJ, Kraus ML, Scheibe DN, Snell GP, Zou H, Sang BC, Wilson KP: **Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase.** *J Biol Chem* 2004, 279:31655-31663.

91.   Agarwal S, Kazi JU, Ronnstrand L: **Phosphorylation of the activation loop tyrosine 823 in c-Kit is crucial for cell survival and proliferation.** *J Biol Chem* 2013, 288:22460-22468.

92.   Fukushima S, Otsuka A, Suzuki T, Yanagisawa T, Mishima K, Mukasa A, Saito N, Kumabe T, Kanamori M, Tominaga T, et al: **Mutually exclusive mutations of KIT and RAS are associated with KIT mRNA expression and chromosomal instability in primary intracranial pure germinomas.** *Acta neuropathologica* 2014.

93.   Kanetsky PA, Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, Letrero R, Ciosek SL, Doody DR, Smith LM, Weaver J, et al: **Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer.** *Nat Genet* 2009, 41:811-815.

94.   Kanetsky PA, Mitra N, Vardhanabhuti S, Vaughn DJ, Li M, Ciosek SL, Letrero R, D'Andrea K, Vaddi M, Doody DR, et al: **A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility.** *Hum Mol Genet* 2011, 20:3109-3117.

95.   Kratz CP, Han SS, Rosenberg PS, Berndt SI, Burdett L, Yeager M, Korde LA, Mai PL, Pfeiffer R, Greene MH: **Variants in or near KITLG, BAK1, DMRT1, and TERT-CLPTM1L predispose to familial testicular germ cell tumour.** *Journal of medical genetics* 2011, 48:473-476.

96.   Kratz CP, Greene MH, Bratslavsky G, Shi J: **A stratified genetic risk assessment for testicular cancer.** *International journal of andrology* 2011, 34:e98-102.

97.   Rapley EA, Hockley S, Warren W, Johnson L, Huddart R, Crockford G, Forman D, Leahy MG, Oliver DT, Tucker K, et al: **Somatic mutations of KIT in familial testicular germ cell tumours.** *Br J Cancer* 2004, 90:2397-2401.

98.   Rapley EA, Turnbull C, Al Olama AA, Dermitzakis ET, Linger R, Huddart RA, Renwick A, Hughes D, Hines S, Seal S, et al: **A genome-wide association study of testicular germ cell tumor.** *Nat Genet* 2009, 41:807-810.

99.   Ruark E, Seal S, McDonald H, Zhang F, Elliot A, Lau K, Perdeaux E, Rapley E, Eeles R, Peto J, et al: **Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14.** *Nat Genet* 2013.

100.   Schumacher FR, Wang Z, Skotheim RI, Koster R, Chung CC, Hildebrandt MA, Kratz CP, Bakken AC, Bishop DT, Cook MB, et al: **Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23.** *Hum Mol Genet* 2013, 22:2748-2753.

101.   Turnbull C, Rahman N: **Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour.** *International journal of andrology* 2011, 34:e86-96; discussion e96-87.

102.   Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, Ricketts M, Linger R, Nsengimana J, Deloukas P, et al: **Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer.** *Nat Genet* 2010, 42:604-607.

103.   Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, Davies B, Peterse EF, Sacilotto N, Walker GJ, et al: **A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection.** *Cell* 2013, 155:410-422.

104.   Azevedo MF, Horvath A, Bornstein ER, Almeida MQ, Xekouki P, Faucz FR, Gourgari E, Nadella K, Remmers EF, Quezado M, et al: **Cyclic AMP and c-KIT signaling in familial testicular germ cell tumor predisposition.** *J Clin Endocrinol Metab* 2013.

105.   Mirabello L, Kratz CP, Savage SA, Greene MH: **Promoter methylation of candidate genes associated with familial testicular cancer.** *International journal of molecular epidemiology and genetics* 2012, 3:213-227.

106.   Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, et al: **A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.** *Cell* 2006, 124:1169-1181.

107.   Gillis AJ, Rijlaarsdam MA, Eini R, Dorssers LC, Biermann K, Murray MJ, Nicholson JC, Coleman N, Dieckmann KP, Belge G, et al: **Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle.** *Molecular oncology* 2013, 7:1083-1092.

108.   Deb-Rinker P, Ly D, Jezierski A, Sikorska M, Walker PR: **Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation.** *J Biol Chem* 2005, 280:6257-6260.

109.   Nettersheim D, Biermann K, Gillis AJ, Steger K, Looijenga LH, Schorle H: **NANOG promoter methylation and expression correlation during normal and malignant human germ cell development.** *Epigenetics : official journal of the DNA Methylation Society* 2011, 6:114-122.

110.   Looijenga LH, Stoop H, de Leeuw HP, de Gouveia Brazao CA, Gillis AJ, van Roozendaal KE, van Zoelen EJ, Weber RF, Wolffenbuttel KP, van Dekken H, et al: **POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.** *Cancer Res* 2003, 63:2244-2250.

111.   De Jong J, Weeda S, Gillis AJ, Oosterhuis JW, Looijenga LH: **Differential methylation of the OCT3/4 upstream region in primary human testicular germ cell tumors.** *Oncology reports* 2007, 18:127-132.

112.   Nordhoff V, Hubner K, Bauer A, Orlova I, Malapetsa A, Scholer HR: **Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences.** *Mamm Genome* 2001, 12:309-317.

113.   Manton KJ, Douglas ML, Netzel-Arnett S, Fitzpatrick DR, Nicol DL, Boyd AW, Clements JA, Antalis TM: **Hypermethylation of the 5' CpG island of the gene encoding the serine protease Testisin promotes its loss in testicular tumorigenesis.** *Br J Cancer* 2005, 92:760-769.

114.   Kempkensteffen C, Christoph F, Weikert S, Krause H, Kollermann J, Schostak M, Miller K, Schrader M: **Epigenetic silencing of the putative tumor suppressor gene testisin in testicular germ cell tumors.** *Journal of cancer research and clinical oncology* 2006, 132:765-770.

115.   Kato N, Tamura G, Fukase M, Shibuya H, Motoyama T: **Hypermethylation of the RUNX3 gene promoter in testicular yolk sac tumor of infants.** *The American journal of pathology* 2003, 163:387-391.

116.   Lind GE, Skotheim RI, Lothe RA: **The epigenome of testicular germ cell tumors.** *APMIS* 2007, 115:1147-1160.

117.   Nettersheim D, Gillis A, Biermann K, Looijenga LH, Schorle H: **The seminoma cell line TCam-2 is sensitive to HDAC inhibitor depsipeptide but tolerates various other chemotherapeutic drugs and loss of NANOG expression.** *Genes Chromosomes Cancer* 2011, 50:1033-1042.

118.   Nettersheim D, Westernstroer B, Haas N, Leinhaas A, Brustle O, Schlatt S, Schorle H: **Establishment of a versatile seminoma model indicates cellular plasticity of germ cell tumor cells.** *Genes Chromosomes Cancer* 2012, 51:717-726.

119.   Weber S, Eckert D, Nettersheim D, Gillis AJ, Schafer S, Kuckenberg P, Ehlermann J, Werling U, Biermann K, Looijenga LH, Schorle H: **Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance.** *Biol Reprod* 2010, 82:214-223.

120.   Hoei-Hansen CE, Nielsen JE, Almstrup K, Sonne SB, Graem N, Skakkebaek NE, Leffers H, Rajpert-De Meyts E: **Transcription factor AP-2gamma is a developmentally regulated marker of testicular carcinoma in situ and germ cell tumors.** *Clin Cancer Res* 2004, 10:8521-8530.

121.   Orso F, Cottone E, Hasleton MD, Ibbitt JC, Sismondi P, Hurst HC, De Bortoli M: **Activator protein-2gamma (AP-2gamma) expression is specifically induced by oestrogens through binding of the oestrogen receptor to a canonical element within the 5'-untranslated region.** *The Biochemical journal* 2004, 377:429-438.

122.   Kawakami T, Okamoto K, Ogawa O, Okada Y: **XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer.** *Lancet* 2004, 363:40-42.

123.   Bussey KJ, Lawce HJ, Himoe E, Shu XO, Heerema NA, Perlman EJ, Olson SB, Magenis RE: **SNRPN methylation patterns in germ cell tumors as a reflection of primordial germ cell development.** *Genes Chromosomes Cancer* 2001, 32:342-352.

124.   Furukawa S, Haruta M, Arai Y, Honda S, Ohshima J, Sugawara W, Kageyama Y, Higashi Y, Nishida K, Tsunematsu Y, et al: **Yolk sac tumor but not seminoma or teratoma is associated with abnormal epigenetic reprogramming pathway and shows frequent hypermethylation of various tumor suppressor genes.** *Cancer science* 2009, 100:698-708.

125. Schneider DT, Schuster AE, Fritsch MK, Hu J, Olson T, Lauer S, Gobel U, Perlman EJ: **Multipoint imprinting analysis indicates a common precursor cell for gonadal and nongonadal pediatric germ cell tumors.** *Cancer Res* 2001, 61:7268-7276.

126. Kerjean A, Dupont JM, Vasseur C, Le Tessier D, Cuisset L, Paldi A, Jouannet P, Jeanpierre M: **Establishment of the paternal methylation imprint of the human H19 and MEST/PEG1 genes during spermatogenesis.** *Hum Mol Genet* 2000, 9:2183-2187.

127. Ueda T, Abe K, Miura A, Yuzuriha M, Zubair M, Noguchi M, Niwa K, Kawase Y, Kono T, Matsuda Y, et al: **The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development.** *Genes to cells : devoted to molecular & cellular mechanisms* 2000, 5:649-659.

128. Kawasaki Y, Lee J, Matsuzawa A, Kohda T, Kaneko-Ishino T, Ishino F: **Active DNA demethylation is required for complete imprint erasure in primordial germ cells.** *Scientific reports* 2014, 4:3658.

129. Zimmerman DL, Boddy CS, Schoenherr CS: **Oct4/Sox2 binding sites contribute to maintaining hypomethylation of the maternal igf2/h19 imprinting control region.** *PloS one* 2013, 8:e81962.

130. van Gurp RJ, Oosterhuis JW, Kalscheuer V, Mariman EC, Looijenga LH: **Biallelic expression of the H19 and IGF2 genes in human testicular germ cell tumors.** *J Natl Cancer Inst* 1994, 86:1070-1075.

131. Verkerk AJ, Ariel I, Dekker MC, Schneider T, van Gurp RJ, de Groot N, Gillis AJ, Oosterhuis JW, Hochberg AA, Looijenga LH: **Unique expression patterns of H19 in human testicular cancers of different etiology.** *Oncogene* 1997, 14:95-107.

132. Ross JA, Schmidt PT, Perentesis JP, Davies SM: **Genomic imprinting of H19 and insulin-like growth factor-2 in pediatric germ cell tumors.** *Cancer* 1999, 85:1389-1394.

133. Kawakami T, Zhang C, Okada Y, Okamoto K: **Erasure of methylation imprint at the promoter and CTCF-binding site upstream of H19 in human testicular germ cell tumors of adolescents indicate their fetal germ cell origin.** *Oncogene* 2006, 25:3225-3236.

134. Smiraglia DJ, Szymanska J, Kraggerud SM, Lothe RA, Peltomaki P, Plass C: **Distinct epigenetic phenotypes in seminomatous and nonseminomatous testicular germ cell tumors.** *Oncogene* 2002, 21:3909-3916.

135. Kristensen DG, Nielsen JE, Jorgensen A, Skakkebaek NE, Rajpert-De Meyts E, Almstrup K: **Evidence that active demethylation mechanisms maintain the genome of carcinoma in situ cells hypomethylated in the adult testis.** *Br J Cancer* 2014, 110:668-678.

136. Sonne SB, Perrett RM, Nielsen JE, Baxter MA, Kristensen DM, Leffers H, Hanley NA, Rajpert-De-Meyts E: **Analysis of SOX2 expression in developing human testis and germ cell neoplasia.** *The International journal of developmental biology* 2010, 54:755-760.

137. Kristensen DM, Sonne SB, Ottesen AM, Perrett RM, Nielsen JE, Almstrup K, Skakkebaek NE, Leffers H, Rajpert-De Meyts E: **Origin of pluripotent germ cell tumours: the role of microenvironment during embryonic development.** *Mol Cell Endocrinol* 2008, 288:111-118.

138. Honecker F, Stoop H, Mayer F, Bokemeyer C, Castrillon DH, Lau YF, Looijenga LH, Oosterhuis JW: **Germ cell lineage differentiation in non-seminomatous germ cell tumours.** *J Pathol* 2006, 208:395-400.

139. Yang S, Lin G, Deng L, Lu GX: **Tumourigenic characteristics of embryonal carcinoma cells as a model for studying tumour progression of human embryonic stem cells.** *Cell proliferation* 2012, 45:299-310.

140. Josephson R, Ording CJ, Liu Y, Shin S, Lakshmipathy U, Toumadje A, Love B, Chesnut JD, Andrews PW, Rao MS, Auerbach JM: **Qualification of embryonal carcinoma 2102Ep as a reference for human embryonic stem cell research.** *Stem Cells* 2007, 25:437-446.

141. Andrews PW, Matin MM, Bahrami AR, Damjanov I, Gokhale P, Draper JS: **Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin.** *Biochem Soc Trans* 2005, 33:1526-1530.

142. Kristensen DG, Mlynarska O, Nielsen JE, Jacobsen GK, Rajpert-De Meyts E, Almstrup K: **Heterogeneity of chromatin modifications in testicular spermatocytic seminoma point toward an epigenetically unstable phenotype.** *Cancer genetics* 2012, 205:425-431.

143. Andrews PW, Bronson DL, Benham F, Strickland S, Knowles BB: **A comparative study of eight cell lines derived from human testicular teratocarcinoma.** *Int J Cancer* 1980, 26:269-280.

144. Andrews PW, Casper J, Damjanov I, Duggan-Keen M, Giwercman A, Hata J, von Keitz A, Looijenga LH, Millan JL, Oosterhuis JW, et al: **Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours.** *Int J Cancer* 1996, 66:806-816.

145. Wang N, Trend B, Bronson DL, Fraley EE: **Nonrandom abnormalities in chromosome 1 in human testicular cancers.** *Cancer Res* 1980, 40:796-802.

146. Fogh J, Trempe G: **New Human Tumor Cell Lines.** In *Human Tumor Cells in Vitro.* Edited by Fogh J: Springer US; 1975: 115-159

147. Fogh J: **Cultivation, characterization, and identification of human tumor cells with emphasis on kidney, testis, and bladder tumors.** *National Cancer Institute monograph* 1978:5-9.

148. Teshima S, Shimosato Y, Hirohashi S, Tome Y, Hayashi I, Kanazawa H, Kakizoe T: **Four new human germ cell tumor cell lines.** *Laboratory investigation; a journal of technical methods and pathology* 1988, 59:328-336.

149. Eckert D, Nettersheim D, Heukamp LC, Kitazawa S, Biermann K, Schorle H: **TCam-2 but not JKT-1 cells resemble seminoma in cell culture.** *Cell Tissue Res* 2008, 331:529-538.

150. de Jong J, Stoop H, Gillis AJ, Hersmus R, van Gurp RJ, van de Geijn GJ, van Drunen E, Beverloo HB, Schneider DT, Sherlock JK, et al: **Further characterization of the first seminoma cell line TCam-2.** *Genes Chromosomes Cancer* 2008, 47:185-196.

151. Mizuno Y, Gotoh A, Kamidono S, Kitazawa S: **[Establishment and characterization of a new human testicular germ cell tumor cell line (TCam-2)].** *Nihon Hinyokika Gakkai zasshi The japanese journal of urology* 1993, 84:1211-1218.

152. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y: **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.** *Science* 2011, 333:1300-1303.

153. Du P, Kibbe WA, Lin SM: **lumi: a pipeline for processing Illumina microarray.** *Bioinformatics* 2008, 24:1547-1548.

154. Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerstrom-Billai F, Jagodic M, Sundberg CJ, Ekstrom TJ, Teschendorff AE, Tegner J, Gomez-Cabrero D: **An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform.** *Epigenetics : official journal of the DNA Methylation Society* 2013, 8:333-346.

155. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R: **Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.** *Epigenetics : official journal of the DNA Methylation Society* 2013, 8:203-209.

156. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S: **A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.** *Bioinformatics* 2013, 29:189-196.

157. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM: **Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.** *BMC Bioinformatics* 2010, 11:587.

158. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B (Methodological)* 1995, 57:289-300.

159. Tibshirani: **Regression Shrinkage and Selection via the Lasso.** *J R Statist Soc* 1996, 58:267-288.

160. Rijlaarsdam MA, van der Zwan YG, Dorssers LC, Looijenga LH: **DMRforPairs: identifying differentially methylated regions between unique samples using array based methylation profiles.** *BMC Bioinformatics* 2014, 15:141.

CHAPTER 10

General Discussion

This thesis presents results on functional genomics in germ cell tumors (GCTs) and possible clinical applications of the unique onco-fetal properties of these germ cell derived tumors [1-3]. Analyzing various types of "omics" datasets comprises an important part of this work. This discussion will therefore focus on the onco-fetal properties of GCTs as well as the paradigm shift towards open data and the required multidisciplinary approach to the integrated analysis of "omics" data.

(Embryonic) miRs and OCT3/4: onco-fetal GCC hallmarks

OCT3/4

As reviewed in **chapter 2**, type II GCTs, also called germ cell cancer (GCC) show strong similarities to developing (early fetal) germ cells, i.e. primordial germ cells (PGCs) or gonocytes [3]. Moreover, one of the stem cell components (embryonal carcinoma EC) closely represents embryonic stem cells. A mainstream marker for GCC is OCT3/4 [2, 4-8]. OCT3/4 is an evolutionary conserved core regulator of pluripotency [9, 10] and one of the four Yamanaka factors required to create induced pluripotent stem cells [11, 12].

Chapter 3 demonstrates the specificity of the OCT3/4-A isoform for GCC [7]. In our isoform specific analysis, the OCT3/4-B and B1 isoforms were less specific for GCC at the mRNA level. Isoform specific OCT3/4 mRNA levels showed a discrepancy with protein levels. OCT3/4-B and B1 have been associated with other functions than pluripotency. In fact, OCT3/4-B is related to stress response and OCT3/4-B1 has been associated with pluripotency, induction of apoptosis and cell cycle deregulation. Non-isoform specific detection of OCT3/4 mRNA and protein levels might account for the reports of OCT3/4 in non-GCC malignancies as reviewed by us [7]. Recently non-specific OCT3/4 was also detected in bladder cancer [13]. Because OCT3/4 isoforms differ only in their N-terminal region, antibodies targeting the C-terminal region [14] will not be isoform specific.

Overall, our findings and other studies emphasize the role of alterative transcripts and post-transcriptional/translational regulation in the execution of OT3/4 function, e.g. by pseudogene expression [15, 16], miR regulation [17, 18] and phosphorylation [19]. Therefore, the use of isoform specific primers and monoclonal antibodies against the N terminus of the OCT3/4 protein is of crucial importance to sensitively detect the pluripotency related isoform OCT3/4-A. Moreover, OCT3/4 is post-translationally regulated. AKT mediated phosphorylation of OCT3/4 regulates OCT3/4 stability, nuclear localization and transcriptional activity. Via a positive feedback loop OCT3/4 also stimulates AKT1 expression [19]. AKT levels are in turn crucial for maintaining PGC identity [20] and regulate the migration of PGCs during embryogenesis [21]. We are therefore currently investigating phospho-OCT3/4 as a potential marker of pluripotency/GCTs, using specific antibodies against phosphorylated OCT3/4.

(Embryonic) miRs

Analogous to OCT3/4-A expression, the micro-RNA (miR) profile of GCC reflects their pluripotent embryonic origin. The progenitor and stem-cell components of GCC express a specific set of miRs that are physiologically only expressed in early embryonic (stem) cells, specifically the miR-371-2-3 and miR-302abcd/367 clusters. [22-24]}. The role of these and other miRs in maintaining the onco-fetal phenotype of these tumors is reviewed in **chapter 2**. In this review, miRs are shown to be core players in facilitating the crossover between pluripotency, cell cycle regulation and therapy sensitivity [3]. Relevant miRs include the embryonic miRs mentioned above as well as miR-34a and miR-145. Recently, a study by Siemens and colleagues also showed a relation between TP35 mediated miR-34 expression and repression of the dependence receptor KIT and downstream functions (chemosensitivity, migration, stemness) [25]. This is of specific interest because KIT signaling is crucial during germ cell migration as well as in GCC pathogenesis [3].

The detection of miR-mRNA interactions like miR-34/KIT is often hampered by the large number of predicted targets for each miR. In **Chapter 4** a tool (miMsg) is presented to identify promising patterns of matched miR and mRNA expression from genome wide profiles. miMsg is implemented in Matlab and is freely available online[a]. In data from seminoma (SE), EC and type III GCTs (spermatocytic seminomas), miMsg identified a number of potentially GCT related interactions. These showed an enrichment of the miR-23ab/24//27 paralog cluster, which is significantly lower expressed in SE as compared to EC [26]. A significant amount of the selected miR-mRNA interactions has also been verified in human embryonic stem (ES) cells [26, 27]. The miR-23ab/24//27 clusters and their mRNA targets have been implicated in several (patho)biological processes, but have not yet been functionally associated with GCTs [26].

The potential clinical applicability of embryonic miRs in GCC is explored in **chapters 5 & 6** [22, 23, 28-34]. For GCT diagnosis and follow-up, a number of serum markers is available, including α-fetoprotein (AFP), human chorionic gonadotropin (hCG), and to a lesser extend lactate dehydrogenase (LDH) [35-38]. About 80% of all non-seminomas (NS) and 20% of all SE show increased levels of these markers. AFP and hCG are mainly related to the presence of a YS or CH component respectively. They therefore show little sensitivity for the stem cell components SE and EC. False positivity might be caused by liver disease (AFP) and chemotherapy (hCG), hampering specificity for GCTs. Finally, positivity for either marker can change during disease progression. Therefore, better biomarkers are of interest, particularly markers with a high sensitivity for SE and EC. **Chapter 5** describes the TSmiR test, applying a para-magnetic bead based approach to isolate miRs from serum and then quantify the levels of the embryonic miRs using a stringently quality controlled pipeline [30]. This miR panel showed an overall sensitivity of 98% for detecting GCCs, clearly outperforming the traditional serum markers AFP/hCG (36%/57%, sensitivity$_{AFP}$=3%/45%;

---

[a] www.martinrijlaarsdam.nl/mimsg

sensitivity$_{hCG}$=62%/66%, SE/NS). In line with the GCC specificity of these markers, serum levels returned to baseline after orchidectomy (stage-I disease). Moreover, there was a trend toward higher miR levels in patients with metastasis. These results were very recently validated in an independent study [34]. Regarding false positivity, embryonic miRs are normally not expressed in somatic tissues, but miR-371-2-3 has been shown to be expressed by a subgroup of thyroid adenomas with specific chromosomal rearrangements [39]. **Chapter 6** builds on the targeted observations from chapter 5 and validates miR-371 and 372 as GCC specific in a high throughput serum miR profile covering ≈750 miRs. In addition, potential novel serum miR biomarkers are identified which show high levels in serum from GCC cases as compared to controls. These most prominently included miR-511 and, less outstanding, miR-26b, 769, 23a, 106b, 365, 598, 340 and let-7a. miR-511, a presumed tumor suppressor miR in adenocarcinoma with a role in tumor immune response [40-42], might be of interest based on its relatively high levels in serum from GCC patients. Further studies are required to inquire the function of this miR in GCC pathogenesis. Moreover, miR-mRNA interactions involving miR-23a have been shown to be differentiating between GCT subtypes [26] (see discussion of miMsg above). Despite the limited difference in miR levels between GCC cases and controls, miR-23a might therefore also be an interesting target for future exploration.

Taken together, these results illustrate the relevance of studying the function of miRs in disease pathogenesis, specifically in GCC. The data predominantly illustrate the biological importance of miRs in maintaining the onco-fetal identity of GCC and the related clinical opportunities. Based on the proof of concept presented by TSmiR, our group received a grant from the Dutch Cancer Society (grant number: 13-6001) to further investigate and validate the clinical application of serum miRs in the biology and diagnostics/follow-up of GCC. In a national and international collaboration the results of TSmiR will be validated and extended in a prospective cohort including long term follow-up. To sensitively quantify the low miR-levels in serum, optimization of the measurement protocols and adequate normalization is crucial here. Along the way, results from high throughput studies like chapter 6, should be taken into account as discussed above.

Epigenetics in GCTs: an (embryonic) blueprint

During embryonic development, the progenitors of mature germ cells migrate from the yolk sac, via the hindgut to the genital ridge where they further mature. During this migration and at the genital ridge these PGCs / gonocytes undergo a characteristic epigenetic "reset" which is reflected in their malignant counterparts: GCTs (for review: **chapters 2 and 9,** core literature: [43-50]). Focusing on DNA CpG methylation, genome-wide methylation is erased early during migration and re-established after arrival at the gonads. Uniparental genomic imprinting patterns are established via total imprinting erasure and in females the inactivated X chromosome is (re)activated.

Tools to compare groups of samples with regard to their methylation status are broadly available [51]. In **chapter 7** a tool (DMRforPairs) is presented which allows comparison unique samples [52]. DMRforPairs is implemented in R and freely available as a Bioconductor package[b]. In **chapter 8 and 9**, DMRforPairs was successfully applied to GCT cell line data. **Chapter 8** illustrates the similarity of GCC cell lines to their PGC / gonocyte ancestor by integrating genome-wide histone modification, methylation and gene expression data [53]. A seminoma-like (TCam-2) and an extragonadal embryonal carcinoma cell line (NCCIT) were compared. Germ cell marker specificity was shown not only in the mRNA data, but also at the epigenetic level. The results support the hypothesis that both cell types originate from embryonic germ cells arrested in early gonadal development. Subtle difference in the (integrated) epigenetic and expression profiles indicated TCam-2 to exhibit a more germ cell-like profile, whereas NCCIT showed a more pluripotent phenotype. **Chapter 9** continues by investigating the genome wide methylation profiles of 91 GCTs and four representative cell lines, providing novel insight into the developmental timing and underlying biology of the various subtypes of GCTs and their (embryonic) cells of origin. Cell lines showed little correlation with the actual tumor samples, warranting care when extrapolating methylation findings from cell lines to *in vivo* tumors. Different GCT subtypes could be clearly distinguished based on their global methylation profile. The methylation status of specific functional regions like imprinting control regions was also highly informative in identifying specific GCT subtypes. The observed methylation profiles matched with specific stages of germ cell development. Some specific issues are of particular interest for future studies. The imprinting status in the GCCs was previously thought to be erased (**review: chapter 9**) but was found to be somatic (50%) in our data. Therefore, the specific imprinting control regions evaluated in **chapter 9** are primary targets for validation studies. The non-erased status identified in **chapter 9** changes the proposed developmental stage of the precursor cell of GCC. These findings do however fit nicely with the totipotent potential of GCC, as demonstrated in mice [1]. Moreover, the "location dependent" imprinting pattern of type I TE is of particular interest. The general pattern of imprinting erasure in our small series of testicular

---

[b] http://bioconductor.org/packages/release/bioc/html/DMRforPairs.html

type I TE might indicate a role for the micro-environment in determining the epigenetic landscape of these cells. Furthermore, the strong epigenetic likeness of dermoid cysts (type IV GCTs) and ovarian type I TE should be investigated further in a larger dataset, potentially warranting re-evaluation of the current GCT classification. Finally, the methylation status of the X chromosomes in the GCTs represented a normal diploid situation (50% methylation in females or 0% in males) while specific GCT subclasses are generally considered aneuploid (tri-/tetraploid) [1, 54, 55], showing intact X inactivation [56]. These new insights need further evaluation to understand their impact on GCT pathogenesis.

In summary, these studies use novel tools and existing methods were applied to elucidate part of the epigenetic blueprint of GCTs and illustrate their close, onco-fetal relation to (fetal) developing germ cells. The datasets and results, including regions differentially methylated between different GCT subtypes, are publically available online, intending to serve as a hypothesis generating source for future targeted experiments.

From ball to bytes and back again

Successful exploration of the (integrated) functional genomics underlying disease faces two major challenges: the paradigm of open science / data and the required multidisciplinary approach.

Crowd science

Despite significant advances, generating omics data is still expensive and analyzing it takes a considerable amount of time and manpower. It is therefore logical to expect some return on investment before throwing the data to the online wolves. At this moment, most online databases like ArrayExpress[c] or GEO[d] provide free access to the published data. Data submission to an online repository is often mandatory for publication. Researchers will generally not submit unpublished data to such a public data warehouse. Indeed, all datasets discussed in this thesis were made available online in a standardized format only upon publication. This also goes for miMsg[e] and DMRforPairs[f], the bioinformatics tools discussed in **chapters 4 and 7**. However, omics data grows old very fast and publication of such data in a static form alongside a publication does not necessarily stimulate expansion of the scientific debate. This suggests that scientific and economic revenue could be increased by providing early access to datasets and ongoing experiments/analysis in an interactive environment in context of the research aims [57]. In this context, crowd funding models are an interesting emerging approach to research and

---

[c] http://www.ebi.ac.uk/arrayexpress/
[d] http://www.ncbi.nlm.nih.gov/geo/
[e] http://martinrijlaarsdam.nl/mimsg
[f] http://bioconductor.org/packages/release/bioc/html/DMRforPairs.html

development [58, 59]. While becoming more mainstream at the edges of technological development and the performing arts, specific alternatives aiming at science are still very limited in size. For example, experiment.com includes 139 registered projects registered opposed to >65,000 on a general crowd funding site like kickstarter.com. In the end, crowd funding might just not be directly applicable to science. Crowd science might be a better term, suggesting a business model that supersedes the financial focus of classical crowd funding. By contributing to projects, people gain access to data, analysis and results in context of the research aims as they become available. An added advantage of crowd science is that relevance of a project is judged by all contributors together, increasing efficiency. Successful application of these novel approaches requires the business model of biomedical research to change. A paradigm shift from an "inwards" a posteriori disclosure towards "outwards" shared research and development is indicated in addition to legal, organizational and infrastructural issues that need to be addressed. These long-term changes will require significant investment and diligence of all involved.

Lost in translation

In the analysis of omics data the difference between data and information is essential. Purely economically speaking data costs money and information generates revenue. A multidisciplinary team is crucial to convert (integrated) omics data into biologically or, even better, clinically relevant information [60]. This frequently means learning to communicate in new languages and mixing of different scientific cultures and work processes [57]. It would be fruitful to emphasize this approach in all levels of scientific education. Only when sufficient synthesis is accomplished, a successful translation from bytes to bench and eventually to the patient's bed can occur, contributing to personalized medicine by better understanding the etiology and pathogenesis of disease (Figure 1). First and foremost, all people involved need to understand the general ideas behind the clinical targets, biological hypothesis and technical pitfalls. An important issue in this context is superficial transfer of knowledge to other disciplines which might lead to unjustified expectations or even complete misinterpretation of the aims. An extreme example is that just asking for "relevant" genes might be interpreted completely different by each person/discipline involved. With regard to the expectations, too much is often expected from omics data. Omics datasets are large and subject to technical errors and bias because of the complex way they are generated and processed. They are intended to be evaluated globally, zooming in on relevant patterns as you go. Directly focusing on subsets of data might subject the analysis to bias (e.g. known genes of interest). Depending on the hypothesis, this is not necessarily wrong, but needs to be taken into consideration when evaluating the results. Another potential pitfall is a lack of evaluation during design and analysis. All participants need to be involved from the start and the team needs to frequently evaluate (1) the status of the analysis, (2) the available results and (3) the general course of the research. It is important for all participants to continually loop back to the original research aims: *from balls to bytes and back again.*
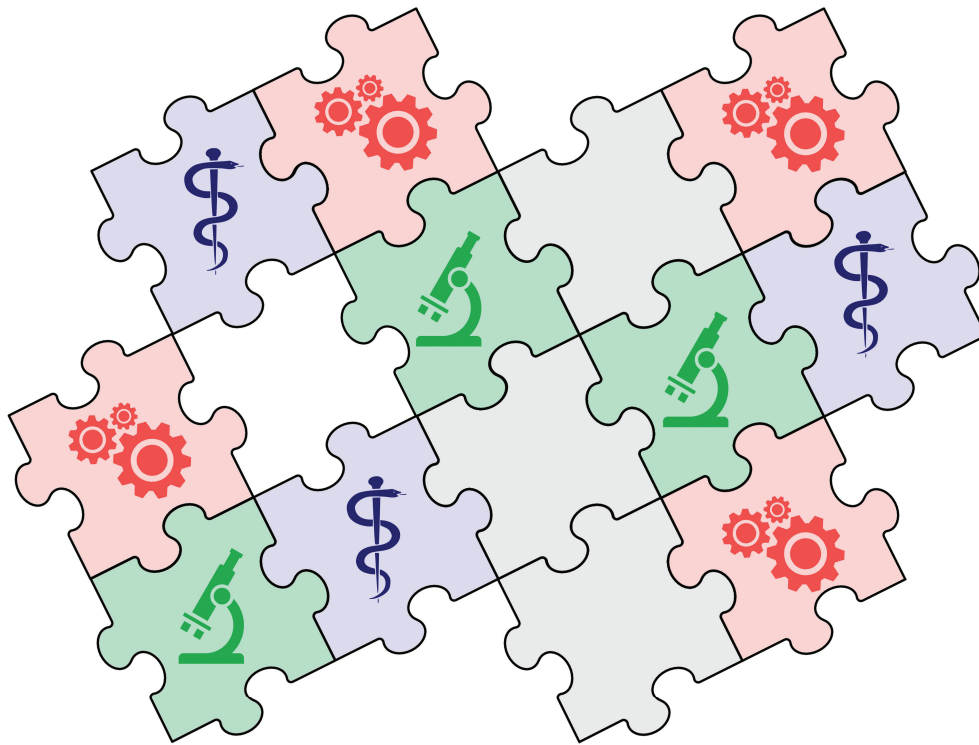
Figure 1. A multidisciplinary approach is essential for the successful exploration of functional genomics. A high level of understanding should be created between team members. Teams are composed of specialized members from specific backgrounds in biology (green), medicine (blue), technology (red) or other disciplines (grey) and all contribute their part. Only when the position of each piece is correctly identified and all pieces are combined, the full picture emerges.

Concluding remarks

To conclude, this thesis illuminates functional genomic, onco-fetal properties of GCTs from a developmental point of view, mainly by investigating (integrated) omics datasets using existing and newly developed computational tools. **Chapters 3, 5 and 6** zoom in on the role of post-transcriptional maintenance of the early embryonic and pluripotent phenotype of SE and EC and provide a potential clinical application of these unique properties of GCTs. These observations need to be validated in prospective studies, which are already initiated. **Chapters 8 and 9** focus on the epigenetic constitution of GCTs and representative cell lines, providing insight into their developmental timing and underlying biology. These results provide hypotheses that can be validated in future targeted experiments. **Chapters 4 and 7** are devoted to novel exploratory bioinformatics tools related to miR-mRNA interactions and detection of differentially methylated regions. There tools were also applied to GCT data. In all studies, close collaboration with other disciplines including biologists, clinicians and engineers from various specific backgrounds has contributed greatly to the research process and the results. In conclusion the results presented in this thesis provide insight into epigenetic and (post-)transcriptional regulation in GCT pathogenesis. The research focuses on clinical applications of the findings and contributes to the emergence of new hypothesis that can be experimentally and clinically validated in the future.

References

1. Oosterhuis JW, Looijenga LH: **Testicular germ-cell tumours in a broader perspective.** *Nat Rev Cancer* 2005, 5:210-222.
2. Looijenga LH: **Human testicular (non)seminomatous germ cell tumours: the clinical implications of recent pathobiological insights.** *J Pathol* 2009, 218:146-162.
3. Rijlaarsdam MA, Looijenga LH: **An oncofetal and developmental perspective on testicular germ cell cancer.** *Seminars in cancer biology* 2014.
4. Looijenga LH, Stoop H, de Leeuw HP, de Gouveia Brazao CA, Gillis AJ, van Roozendaal KE, van Zoelen EJ, Weber RF, Wolffenbuttel KP, van Dekken H, et al: **POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors.** *Cancer Res* 2003, 63:2244-2250.
5. de Jong J, Looijenga LH: **Stem cell marker OCT3/4 in tumor biology and germ cell tumor diagnostics: history and future.** *Critical reviews in oncogenesis* 2006, 12:171-203.
6. Cheng L, Sung MT, Cossu-Rocca P, Jones TD, MacLennan GT, De Jong J, Lopez-Beltran A, Montironi R, Looijenga LH: **OCT4: biological functions and clinical applications as a marker of germ cell neoplasia.** *J Pathol* 2007, 211:1-9.
7. Rijlaarsdam MA, van Herk HA, Gillis AJ, Stoop H, Jenster G, Martens J, van Leenders GJ, Dinjens W, Hoogland AM, Timmermans M, Looijenga LH: **Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours.** *Br J Cancer* 2011, 105:854-863.
8. Cheng CJ, Wu YC, Shu JA, Ling TY, Kuo HC, Wu JY, Chang EE, Chang SC, Huang YH: **Aberrant expression and distribution of the OCT-4 transcription factor in seminomas.** *Journal of biomedical science* 2007, 14:797-807.
9. Frankenberg SR, Frank D, Harland R, Johnson AD, Nichols J, Niwa H, Scholer HR, Tanaka E, Wylie C, Brickman JM: **The POU-er of gene nomenclature.** *Development* 2014, 141:2921-2923.
10. Jerabek S, Merino F, Scholer HR, Cojocaru V: **OCT4: dynamic DNA binding pioneers stem cell pluripotency.** *Biochimica et biophysica acta* 2014, 1839:138-154.
11. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.** *Cell* 2006, 126:663-676.
12. Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA: **Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2.** *Nat Biotechnol* 2008, 26:1269-1275.
13. Hatefi N, Nouraee N, Parvin M, Ziaee SA, Mowla SJ: **Evaluating the expression of oct4 as a prognostic tumor marker in bladder cancer.** *Iranian journal of basic medical sciences* 2012, 15:1154-1161.
14. Arakawa T, Yoshimi T, Azuma M, Tachibana T: **Production of a monoclonal antibody specific for Pou5f1/Oct4.** *Monoclonal antibodies in immunodiagnosis and immunotherapy* 2013, 32:229-231.
15. Wang L, Guo ZY, Zhang R, Xin B, Chen R, Zhao J, Wang T, Wen WH, Jia LT, Yao LB, Yang AG: **Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma.** *Carcinogenesis* 2013, 34:1773-1781.
16. Panagopoulos I, Moller E, Collin A, Mertens F: **The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1.** *Oncology reports* 2008, 20:1029-1033.
17. Xu N, Papagiannakopoulos T, Pan G, Thomson JA, Kosik KS: **MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.** *Cell* 2009, 137:647-658.
18. Hu J, Guo H, Li H, Liu Y, Liu J, Chen L, Zhang J, Zhang N: **MiR-145 regulates epithelial to mesenchymal transition of breast cancer cells by targeting Oct4.** *PloS one* 2012, 7:e45965.
19. Lin Y, Yang Y, Li W, Chen Q, Li J, Pan X, Zhou L, Liu C, Chen C, He J, et al: **Reciprocal regulation of Akt and Oct4 promotes the self-renewal and survival of embryonal carcinoma cells.** *Mol Cell* 2012, 48:627-640.
20. Kimura T, Nakano T: **Induction of pluripotency in primordial germ cells.** *Histology and histopathology* 2011, 26:643-650.
21. Glover JD, Taylor L, Sherman A, Zeiger-Poli C, Sang HM, McGrew MJ: **A novel piggyBac transposon inducible expression system identifies a role for AKT signalling in primordial germ cell migration.** *PloS one* 2013, 8:e77222.

22. Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW, Sun Y, Chen C, Guenther S, Sherlock J, Veltman I, Baeten J, et al: **High-throughput microRNAome analysis in human germ cell tumours.** *J Pathol* 2007, 213:319-328.

23. Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, et al: **A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.** *Cell* 2006, 124:1169-1181.

24. Eini R, Dorssers LC, Looijenga LH: **Role of stem cell proteins and microRNAs in embryogenesis and germ cell cancer.** *The International journal of developmental biology* 2013, 57:319-332.

25. Siemens H, Jackstadt R, Kaller M, Hermeking H: **Repression of c-Kit by p53 is mediated by miR-34 and is associated with reduced chemoresistance, migration and stemness.** *Oncotarget* 2013, 4:1399-1415.

26. Rijlaarsdam MA, Rijlaarsdam DJ, Gillis AJ, Dorssers LC, Looijenga LH: **miMsg: a target enrichment algorithm for predicted miR-mRNA interactions based on relative ranking of matched expression data.** *Bioinformatics* 2013, 29:1638-1646.

27. Lipchina I, Elkabetz Y, Hafner M, Sheridan R, Mihailovic A, Tuschl T, Sander C, Studer L, Betel D: **Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response.** *Genes Dev* 2011, 25:2173-2186.

28. Belge G, Dieckmann KP, Spiekermann M, Balks T, Bullerdiek J: **Serum levels of microRNAs miR-371-3: a novel class of serum biomarkers for testicular germ cell tumors?** *Eur Urol* 2012, 61:1068-1069.

29. Dieckmann KP, Spiekermann M, Balks T, Flor I, Loning T, Bullerdiek J, Belge G: **MicroRNAs miR-371-3 in serum as diagnostic tools in the management of testicular germ cell tumours.** *Br J Cancer* 2012, 107:1754-1760.

30. Gillis AJ, Rijlaarsdam MA, Eini R, Dorssers LC, Biermann K, Murray MJ, Nicholson JC, Coleman N, Dieckmann KP, Belge G, et al: **Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle.** *Molecular oncology* 2013, 7:1083-1092.

31. Murray MJ, Coleman N: **Testicular cancer: a new generation of biomarkers for malignant germ cell tumours.** *Nat Rev Urol* 2012, 9:298-300.

32. Murray MJ, Halsall DJ, Hook CE, Williams DM, Nicholson JC, Coleman N: **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.

33. Bezan A, Gerger A, Pichler M: **MicroRNAs in testicular cancer: implications for pathogenesis, diagnosis, prognosis and therapy.** *Anticancer research* 2014, 34:2709-2713.

34. Syring I, Bartels J, Holdenrieder S, Kristiansen G, Muller SC, Ellinger J: **Circulating serum microRNA (miR-367-3p, miR-371a-3p, miR-372-3p, miR-373-3p) as biomarkers for patients with testicular germ cell cancers.** *J Urol* 2014.

35. von Eyben FE, Liu FJ, Amato RJ, Fritsche HA: **Lactate dehydrogenase isoenzyme 1 is the most important LD isoenzyme in patients with testicular germ cell tumor.** *Acta oncologica (Stockholm, Sweden)* 2000, 39:509-517.

36. Horwich A, Shipley J, Huddart R: **Testicular germ-cell cancer.** *Lancet* 2006, 367:754-765.

37. Sturgeon CM, Duffy MJ, Stenman UH, Lilja H, Brunner N, Chan DW, Babaian R, Bast RC, Jr., Dowell B, Esteva FJ, et al: **National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers.** *Clin Chem* 2008, 54:e11-79.

38. Salem M, Gilligan T: **Serum tumor markers and their utilization in the management of germ-cell tumors in adult males.** *Expert review of anticancer therapy* 2011, 11:1-4.

39. Rippe V, Dittberner L, Lorenz VN, Drieschner N, Nimzyk R, Sendt W, Junker K, Belge G, Bullerdiek J: **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:e9485.

40. Zhang HH, Pang M, Dong W, Xin JX, Li YJ, Zhang ZC, Yu L, Wang PY, Li BS, Xie SY: **miR-511 induces the apoptosis of radioresistant lung adenocarcinoma cells by triggering BAX.** *Oncology reports* 2014, 31:1473-1479.

41. Zhang C, Chi YL, Wang PY, Wang YQ, Zhang YX, Deng J, Lv CJ, Xie SY: **miR-511 and miR-1297 inhibit human lung adenocarcinoma cell proliferation by targeting oncogene TRIB2.** *PloS one* 2012, 7:e46090.

42. Squadrito ML, Pucci F, Magri L, Moi D, Gilfillan GD, Ranghetti A, Casazza A, Mazzone M, Lyle R, Naldini L, De Palma M: **miR-511-3p modulates genetic programs of tumor-associated macrophages.** *Cell reports* 2012, 1:141-154.

43. Matsui Y, Mochizuki K: **A current view of the epigenome in mouse primordial germ cells.** *Mol Reprod Dev* 2014, 81**:**160-170.

44. Wermann H, Stoop H, Gillis AJ, Honecker F, van Gurp RJ, Ammerpohl O, Richter J, Oosterhuis JW, Bokemeyer C, Looijenga LH: **Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance.** *J Pathol* 2010, 221**:**433-442.

45. Almstrup K, Nielsen JE, Mlynarska O, Jansen MT, Jorgensen A, Skakkebaek NE, Rajpert-De Meyts E: **Carcinoma in situ testis displays permissive chromatin modifications similar to immature foetal germ cells.** *Br J Cancer* 2010, 103**:**1269-1276.

46. Kristensen DG, Skakkebaek NE, Rajpert-De Meyts E, Almstrup K: **Epigenetic features of testicular germ cell tumours in relation to epigenetic characteristics of foetal germ cells.** *The International journal of developmental biology* 2013, 57**:**309-317.

47. Seisenberger S, Peat JR, Reik W: **Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells.** *Curr Opin Cell Biol* 2013, 25**:**281-288.

48. Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, Eggan K, Meissner A: **DNA methylation dynamics of the human preimplantation embryo.** *Nature* 2014, 511**:**611-615.

49. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J, et al: **The DNA methylation landscape of human early embryos.** *Nature* 2014, 511**:**606-610.

50. Nakaki F, Saitou M: **PRDM14: a unique regulator for pluripotency and epigenetic reprogramming.** *Trends in biochemical sciences* 2014, 39**:**289-298.

51. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, Marsit CJ, Houseman EA, Brown R: **Review of processing and analysis methods for DNA methylation array data.** *Br J Cancer* 2013, 109**:**1394-1402.

52. Rijlaarsdam MA, van der Zwan YG, Dorssers LC, Looijenga LH: **DMRforPairs: identifying differentially methylated regions between unique samples using array based methylation profiles.** *BMC Bioinformatics* 2014, 15**:**141.

53. van der Zwan YG, Rijlaarsdam MA, Rossello FJ, Notini AJ, de Boer S, Watkins DN, Gillis AJ, Dorssers LC, White SJ, Looijenga LH: **Seminoma and embryonal carcinoma footprints identified by analysis of integrated genome-wide epigenetic and expression profiles of germ cell cancer cell lines.** *PloS one* 2014, 9**:**e98330.

54. Woodward PJ, Heidenreich A, Looijenga LHJ, al. e: **Testicular germ cell tumors.** In *World Health Organization Classification of Tumours Pathology and Genetics of the Urinary System and Male Genital Organs*. Edited by Eble JN, Sauter G, Epstein JI, Sesterhann IA, eds. Lyon: IARC Press; 2004: 17-278

55. van Echten J, Oosterhuis JW, Looijenga LH, van de Pol M, Wiersema J, te Meerman GJ, Schaffordt Koops H, Sleijfer DT, de Jong B: **No recurrent structural abnormalities apart from i(12p) in primary germ cell tumors of the adult testis.** *Genes Chromosomes Cancer* 1995, 14**:**133-144.

56. Looijenga LH, Gillis AJ, van Gurp RJ, Verkerk AJ, Oosterhuis JW: **X inactivation in human testicular tumors. XIST expression and androgen receptor methylation status.** *The American journal of pathology* 1997, 151**:**581-590.

57. Bernstein R: **Shifting ground for big data researchers.** *Cell* 2014, 157**:**283-284.

58. Weigmann K: **Tapping the crowds for research funding. Crowdfunding, a common practice to support projects in the arts, music or gaming, has also attracted the attention of scientists.** *EMBO Rep* 2013, 14**:**1043-1046.

59. Cameron P, Corne DW, Mason CE, Rosenfeld J: **Crowdfunding genomics and bioinformatics.** *Genome Biol* 2013, 14**:**134.

60. Teiten MH, Grigorakaki C, Orlikova B, Diederich M: **OMICS, a multidisciplinary friendship.** *Cell Death Dis* 2012, 3**:**e267.

SUMMARY

Samenvatting

English

The work discussed in this thesis explains the role of the functional genome in germ cell tumor (GCT) pathogenesis by applying newly developed and existing computational methods to (genome-wide) functional genomic datasets. Specifically, epigenetic and (post-)transcriptional regulation in GCTs was studied to gain a deeper understanding of disease pathogenesis, also aiming at clinical application of the findings.

GCTs are a unique class of neoplasms originating from (fetal) developing germ cells. Five subtypes can be distinguished (I-V), which are related to physiological germ cell development. Type I GCTs, also called infantile or pediatric GCTs, are rare and generally benign. Type II GCTs, also called germ cell cancer (GCC), include a heterogeneous set of histological subtypes with clearly defined totipotent stem cell components. GCC accounts for 60% of all cancers in Caucasian males between the ages of 20 and 40. Type III, IV and V GCTs are generally benign. They originate from more mature germ cells. Generally speaking, the spectrum of GCT subtypes represents a unique class of neoplastic entities in relatively young patients. Moreover, GCTs show strong and traceable onco-fetal roots advantageous for clinical applications.

The functional genome includes a broad range of interacting regulatory features in a cell that determine its identity and behavior. These features are therefore highly relevant in the investigation of GCT pathogenesis and underlying biological processes. Functional genomics comprises the field of study targeted at integrating multiple (genome-wide) datasets, aiming to better understand these interactions. A general introduction into the field of GCTs and functional genomics is provided in **chapter 1** of this thesis. **Part one / chapter 2** gives a review of the onco-fetal roots of GCC. It illustrates that GCC is a fascinating model for studying onco-fetal processes like cell cycle control, pluripotency maintenance and KIT-KITLG signaling.

In the second part of this thesis onco-fetal proteins and micro-RNAs (miR) are investigated that can serve as diagnostic markers and help understand GCT pathogenesis. For example, the OCT3/4 protein (transcription factor) is both a mainstream pluripotency regulator and a specific marker for GCCs. **Chapter 3** underlines the specificity of the OCT3/4-A isoform for GCC showing at the same time that the other isoforms (B, B1) are not GCC specific. The findings illustrate the importance of using isoform specific primers and monoclonal antibodies against the N terminus of the OCT3/4 protein to specifically detect the pluripotency related isoform OCT3/4-A. Besides proteins like OCT3/4, specific miRs are important in GCC development/identity. miRs are short, non-coding RNAs that target mRNAs, primarily promoting their degradation and inhibiting translation. The detection of potentially functional miR-mRNA interactions is often hampered by the large number of predicted targets for each miR. Therefore, a tool (miMsg) is presented in **chapter 4** to identify promising patterns of matched miR and mRNA expression from genome wide profiles. In this chapter, miMsg is also successfully applied to GCT data. Functionally, miRs are core players in GCC, facilitating the crossover between pluripotency, cell cycle regulation and therapy sensitivity of as

reviewed in **chapter 2**. Previous studies have shown that the miR expression profile of GCC tissues reflects their pluripotent embryonic origin. **Chapters 5 & 6** explore the potential clinical applicability of (embryonic) miRs as serum biomarkers in GCC. **Chapter 5** describes the Targeted Serum miR (TSmiR) test, which uses the levels of the embryonic miR-371-2-3 and 302abc/367 clusters in serum as biomarkers for GCC. TSmiR achieved 98% sensitivity in detecting GCC in serum of patients versus controls, clearly outperforming classical serum markers like AFP and hCG. Prospective studies to validate and extend these findings are already initiated. **Chapter 6** builds on the targeted observations in **chapter 5** and validates miR-371 and 372 as GCC specific in a high throughput serum miR profile covering ≈750 miRs. In addition, potential novel serum miR biomarkers were identified, including miR-511 and, less prominent, miR-26b, 769, 23a, 106b, 365, 598, 340 and let-7a.

The work presented in the third part of this thesis investigates components of the epigenetic blueprint of GCTs, and subsequently illustrates their close onco-fetal relation to (fetal) developing germ cells as reviewed in **chapter 2**. During embryonic development, the progenitors of mature germ cells migrate from the yolk sac, via the hindgut to the genital ridge where they further mature. During this migration and at the genital ridge, these PGCs / gonocytes undergo a characteristic epigenetic "reset," including a reorganization of their methylation state. These changes are reflected in their malignant counterparts: GCTs. This process is extensively reviewed in **chapter 9**. Tools to compare the methylation status between groups of samples are widely available. **Chapter 7** presents a new tool (DMRforPairs) which allows comparison of unique samples in cases where large groups of samples are not available. DMRforPairs was successfully applied to GCT cell line data in **chapters 8 and 9**. **Chapter 8** illustrates the similarity of GCC cell lines to their PGC / gonocyte ancestor by integrating genome-wide histone modification, methylation and gene expression data. Subtle differences in the (integrated) epigenetic and expression profiles indicated the seminoma-like cell line TCam-2 to exhibit a more germ cell-like profile. The mediastinal embryonal carcinoma cell line NCCIT showed a more pluripotent phenotype. **Chapter 9** investigates the genome wide methylation profiles of 91 GCTs and four representative cell lines. Different GCT subtypes could be clearly distinguished based on their global methylation profile. The methylation status of specific functional regions like imprinting control regions was also highly informative in identifying specific GCT subtypes. The observed methylation profiles matched with specific stages of germ cell development, providing insight into the developmental timing and biology of the various GCT subtypes.

To conclude, the studies presented in this thesis illuminate onco-fetal properties of GCTs from a developmental point of view, mainly by investigating (integrated) functional genomic datasets using existing and newly developed computational tools. In all studies, close collaboration with other disciplines including biologists, clinicians and engineers from various backgrounds contributed greatly to the research process and the results. All data and results are freely available online. The results presented provide insight into epigenetic and (post-)transcriptional regulation in GCT pathogenesis, focus on clinical application of the findings and contribute to the emergence of new hypothesis that can be experimentally and clinically validated in the future.

Nederlands

De resultaten in dit proefschift zijn erop gericht om de rol van het functionele genoom in het ontstaan van kiemceltumoren (KT) beter te begrijpen. Nieuwe en bestaande analysemethoden werden hiervoor toegepast op functional genomic KT datasets. Met name epigenetische en post-transcriptionele regulatie in KT werden onderzocht om meer inzicht te verkrijgen in de pathogenese van KT, zo mogelijk resulterend in klinische toepassingen.

KT ontstaan uit de (foetale) voorlopers van uitgerijpte kiemcellen (geslachtscellen). Er zijn vijf KT subtypes (I-V) die allemaal gerelateerd zijn aan normale kiemcel ontwikkeling. Type I KT zijn zeldzaam, meestal goedaardig en worden ook wel infantiele KT genoemd. Type II KT wordt ook wel kiemcelkanker (KK) genoemd. KK betreft een heterogene groep van histologische subtypes. Van deze tumoren is veel bekend over hun (totipotente) stamcelcomponenten. Zestig procent van alle Kaukasische mannen met kanker tussen de 20 en 40 jaar heeft KK. De type III, IV en V KT zijn meestal goedaardig en komen voort uit verder ontwikkelde kiemcellen. KT vormen een unieke groep maligniteiten in jonge patiënten waarvan relatief veel bekend is over de onco-foetale oorspong. Dit biedt mogelijk kansen in de kliniek.

Het functionele genoom bevat een grote groep van interacterende regulatoire processen die de identiteit en het gedrag van een cell bepalen. Deze processen zijn cruciaal voor het begrijpen van de pathologie van KT en de onderliggende biologische processen. "Functional genomics" is het vakgebied wat zich bezighoudt met het integreren van datasets die deze processen beschrijven. In **hoofdstuk 1** worden KT en functional genomics verder geïntroduceerd. **Deel één / hoofdstuk 2** geeft een overzicht van de literatuur die de onco-foetale eigenschappen van KK beschijft. Daaruit blijkt dat KK een goed model is voor het onderzoek naar onco-foetale ontwikkelingsprocessen, met name cel cyclus controle, plupotentie en KIT-KITL gerelateerde processen.

Het tweede deel van dit proefschrift gaat over onco-foetale eiwitten en micro-RNAs (miRs) die inzicht geven in de ontwikkeling van KT. Deze eiwitten en miRs kunnen ook als markers voor KT dienen in de kliniek. OCT3/4 is een voorbeeld van zo'n eiwit (transcriptiefactor) die tegelijkertijd een bekende regulator van pluripotentie en een specifieke marker voor KK is. **Hoofdstuk 3** beschrijft de specificiteit van een isoform van OCT3/4 (OCT3/4-A) voor KK en laat zien dat andere isovormen (B en B1) niet specifiek zijn voor KK. Isovorm specifieke detectie van OCT3/4 mRNA en eiwit is daarom van groot belang als deze marker gebruikt wordt voor onderzoek en in de kliniek. Naast de aanwezigheid van zulke eiwitten laten KT ook een karakteristiek onco-foetaal micro-RNA (miR) expressieprofiel zien. miRs zijn korte stukjes RNA die de afbraak van mRNAs stimuleren of voorkomen dat mRNAs vertaald worden in eiwitten. Het identificeren van relevante miR-mRNA interacties is vaak lastig omdat er veel mRNA targets voorspeld worden voor iedere miR. In **hoofdstuk 4** wordt een tool beschreven (miMsg) die voor specifieke datasets de meest relevante interacties selecteert. miMsg wordt in dit hoofdstuk ook succesvol toegepast op KT data. **Hoofdstuk 2** beschrijft het belang van miRs in KT, met name met betrekking tot het integreren van

pluripotentie, cell cyclus controle en therapie gevoeligheid. **Hoofdstuk 5 en 6** onderzoeken de klinische toepassing van (embryonale) miRs als serum biomarkers voor KK. **Hoofdstuk 5** beschrijft de Targeted Serum miR (TSmiR) test die embryonale miR clusters miR-371-2-3 en miR302abc/367 detecteert in serum. Deze test identificeerde 98% van de KK correct, wat een veel hogere sensitiviteit impliceert dan de traditionele serum markers AFP en hCG. Prospectieve studies om deze resultaten te valideren zijn inmiddels gestart. **Hoofdstuk 6** gaat verder op de bevindingen van **hoofdstuk 5** en vergelijkt de niveaus van ≈750 miRs in serum van KK patienten en controles zonder KT. miR-371 en 372 blijken in deze studie wederom sensitief voor het maken van het onderscheid tussen KK en controles. Ook werd een aantal nieuwe potentiele KT serum markers geïdentificeerd: miR-511, miR-26b, 769, 23a, 106b, 365, 598, 340 en let-7a.

Het derde deel van dit proefschrift geeft inzicht in de epigenetische blauwdruk van KT en illustreert de directe relatie tussen KT en (foetale) ontwikkelende kiemcellen zoals eerder besproken in **hoofdstuk 2**. De karakteristieke epigenetische veranderingen die optreden tijdens kiemcelontwikkeling tonen mogelijk gelijkenis met de epigenetische status van KT. Dit geeft inzicht in het celtype waarvan de verschillende typen KT afstammen. Tijdens normale ontwikkeling migreren kiemcellen door het embryo. Vanaf de dooierzak migreren ze naar de "gential ridge" waar later de gonade gevormd wordt die zich uiteindelijk tot testis of ovarium ontwikkelt. Daar ontwikkelen de kiemcellen zich verder. Het ontwikkelproces van kiemcellen wordt gekenmerkt door een epigenetische "reset" die terug te zien is in de epigenetische staat van KT. Deze reset betreft ook de methylatiestatus van het DNA. Dit proces wordt uitgebreid besproken in **hoofdstuk 9**. Voor het vergelijken van de methylatie tussen groepen van weefsels zijn veel tools beschikbaar. **Hoofdstuk 7** beschrijft een nieuwe tool (DMRforPairs) die specifiek ontwikkelt is om unieke samples te vergelijken als er geen groepen beschikbaar zijn. DMRforPairs wordt toegepast op KT data in de **hoofdstukken 8 en 9**. **Hoofdstuk 8** beschrijft de vergelijking tussen de epigenetische status van KK cell lijnen en relateert de resultaten aan hun vroege kiemcel voorlopers. Deze studie integreert histone modificatie, methylatie en gen expressie data. Subtiele verschillen tussen de geïntegreerde (epigenetische) profielen laten zien dat de seminoom-achtige cellijn (TCam-2) meer op een kiemcel lijkt. De embryonaal carcinoom cellijn (NCCIT) heeft relatief meer pluripotente eigenschappen. **Hoofdstuk 9** beschrijft het methylatieprofiel van verschillende typen KT en vergelijkt dit met bekende resultaten van normaal ontwikkelende kiemcellen. KT subtypes konden duidelijk onderscheiden worden op basis van hun methylatie status. Ook specifieke regio's op het genoom zoals de imprinting control regio's waren informatief voor het voorspellen van het type KT. De methylatieprofielen konden worden teruggevoerd naar specifieke ontwikkelingsstadia van kiemcellen.

Concluderend beschrijft dit proefschift de onco-foetale eigenschappen van KT vanuit een ontwikkelingsbiologisch perspectief. Grootschalige functional genomic datasets werden geanalyseerd met nieuwe en bestaande methoden. Bij alle studies bleek de nauwe samenwerking tussen meerdere disciplines zoals biologen, clinici en ingenieurs uit specifieke vakgebieden cruciaal voor het onderzoeksproces en de resultaten. De resultaten zoals beschreven in dit proefschrift geven inzicht in de epigenetische en (post-)transcriptionele regulatie in KT pathogenese en zijn gericht op klinische toepassingen en het ontwikkelen van toetsbare nieuwe hypothesen.

ACKNOWLEDGEMENTS

Dankwoord

Science is teamwork. Bij iedere publicatie komt het weer terug: de lange lijst van co-auteurs die hebben bijgedragen aan het onderzoek. En bij presentaties is de laatste powerpoint sheet vaak niet groot genoeg om iedereen kwijt te kunnen die cruciaal was voor de resultaten. Geen enkel onderzoek is het werk van één persoon. En toch heeft, wonderlijk genoeg, een proefschrift meestal maar één enkele auteur...

Het staat voor mij als een paal boven water dat dit proefschift er nooit zou zijn geweest zonder alle onmisbare samenwerkingsverbanden, (technische) ondersteuning en opbouwende kritiek. Ik ben blij dat ik hier de ruimte en de tijd mag nemen om een grote groep mensen en organisaties meer algemeen of specifiek te bedanken. Dit is het enige niet peer-reviewed hoofdstuk van het proefschrift, dus het betreft mijn eigen ongezouten mening. Sta je er niet tussen en heb je toch het gevoel dat je voor mij onmisbaar bent geweest? Waarschijnlijk heb je gelijk. Laat het weten, dan maak ik het een keer goed.

Om te beginnen, zonder fianciering geen promotietraject: **MRACE**, bedankt! Het onderzoek mocht ik uitvoeren op de **afdeling pathologie** binnen de context van wat tegenwoordig het **Erasmus MC Cancer Institute** heet. Het werk vond plaats op het **Laboratorium voor Experimentele Patho-Oncologie** (**LEPO**) onder leiding van **prof. Leendert H.J. Looijenga**. Op dit lab en het team wat er werkt wil graag later ingaan. Ik wil hier volstaan met de afdeling te bedanken voor de geboden faciliteiten. Het gepresenteede onderzoek was met name onmogelijk geweest zonder de unieke weefselbank van het LEPO lab / de afdeling pathologie. Een bijzonder woord van dank gaat dan ook uit naar alle patienten die hier aan hebben bijgedragen en hun familie.

Vervolgens, zonder verdediging geen promotie. In willekeurige volgorde, **prof.dr. Sjoerd Repping** (AMC), **prof.dr. Edwin Cuppen** (UMCU), **prof.dr. Guido Jenster** (Erasmus MC), veel dank dat u in mijn leescommissie wilde plaatsnemen. De grote commissie zoals die bij het ter perse gaan van dit proefschrift is vastgesteld wil ik alvast van harte bedanken voor hun komst en de getoonde interesse: **prof.dr. Robert Hofstra** (Erasmus MC), **prof.dr. Andre Uitterlinden** (Erasmus MC), **prof.dr. Lodewyk Wessels** (TU Delft / NKI-AVL), **dr. Jeroen de Ridder** (TU Delft) en **dr. Andreas Kremer** (Siemens Corporate Technology). Ik ben vereerd dit proefschrift te mogen verdedigen tegenover een vakinhoudelijk en geografisch zo diverse commissie.

Een kernbegrip bij onderzoek doen is samenwerking. Binnen en buiten de eigen werkgroep/instelling, maar ook binnen en buiten het eigen vakgebied. Het delen van kennis, materiaal en technieken speelde een belangrijke rol bij al het onderzoek in dit proefschrift. In hoofdstuk 3 zijn we samen met de **afdelingen urologie** en **interne geneeskunde** aan de slag gegaan om een panel aan weefsels samen te stellen waarin we de specificiteit van een marker konderen valideren. **Hilde, Guido, John, Arno, Winand, Marije en Mieke,** bedankt! The comparison of the epigenetic properties of germ cell cancer cell lines in chapter 8 would not have been

possible without the group of **Stefan White** at **Monash University** in Australia. **Fernando, Amanda, Suzan, Neil and Stefan**, thanks for showing me the magic of epigenetics. In chapter 5 a group of international experts brought together the samples and knowledge to investigate micro-RNAs as an emerging type of serum markers for germ cell cancer. **Matthew, James, Nicholas, Klaus-Peter, Gazanfer and Jörn** thanks for the collaboration. In chapter 6, this concept is further explored in a high thoughput approach. Like in chapter 5, this work was done in close collaboration with **Thermo Fisher**, formerly **Life Technologies**, formerly **Applied Biosystems.** Names change, but your team always remained highly involved. **Kathy**, **Nathalie**, **Tom**, **Sunali** and **Kathleen**, it was a pleasure working with you, sometimes at the edges of technological possibilities.

Samenwerken met technische universiteiten bleek buitengewoon zinvol en leuk. Het was uitdagend om een bruggen te slaan tussen de klinische, biologische en technische werelden en vanuit daar samen verder te denken over oplossingen. **David (Rijlaarsdam)**, mijn tweelingbroer en destijds promovendus aan de **TU Eindhoven**, bedankt voor je geduld als ik tijd nodig had voor de abstracte taal van de wiskunde. Het was bijzonder om samen met jou na te mogen denken over hoofdstuk 4. **Devin (University of Kansas)**, thank you for responding to my cry for help when I started working on the application of feature selection and classification methods to methylation profiles. **David (Tax)** en **Jeroen,** bedankt dat jullie mij welkom heetten op de **TU Delft** en mee wilden denken over het scheiden van het kaf en het koren in de methylatie data waar hoofdstuk 9 over gaat. Ik hoop dat dit het begin is van een langdurige samenwerking.

Binnen het Erasmus MC zit ons lab in het **Josphina Nefkens** gebouw waar één, twee of drie deuren verderop altijd wel iemand zit die mee wil denken over een project. De aanwezigheid van verschillende disciplines en de laagdrempeligheid van het contact heb ik erg gewaardeerd. Dit geldt ook voor de **afdeling Bioinformatica**. Al tijdens mijn master mocht ik gebruik maken van de faciliteiten van deze afdeling en dat heeft zich voortgezet tijdens mijn promotie. Het heeft zich ook vertaald in de implementatie van een gecentraliseerd opslag en backup systeem voor alle onderzoeksdata van ons lab. **Peter, Mario, Andrew** en **Anton**, bedankt / thank you. Mijn bijzondere dank gaat uit naar de systeembeheerders: **Mirjam**, **Ivo** en **Sylvia** bedankt voor jullie korte lijntjes en daadkracht.

Ik neem met de verdediging van dit proefschrift voor het moment afscheid van het Erasmus MC en de regio Rotterdam. Ik ben hier met veel plezier opgeleid tot basisarts en onderzoeker. Ik wil het **Erasmus MC, de Erasmus Universiteit** en alle docenten en collegas waar ik de afgelopen tien jaar van heb mogen leren en mee heb mogen werken daarvoor van harte bedanken. Daarnaast wil ik het **AMC en de gelieerde ziekenhuizen** bedanken voor het getoonde vertrouwen om vanaf 1 januari 2015 aan de slag te gaan met mij als AIOS interne geneeskunde.

Na een kleine duizend woorden over belangrijke afdelingen/organisaties en waardevolle samenwerkingsverbanden zijn we nu aangeland bij de bron van alle resultaten in dit proefschrift: het **team van het LEPO lab**. Als lezer is u het u misschien opgevallen dat deze directe collegas in het bovenstaande overzicht schitteren door hun afwezigheid. Dat is alleen om eindeloze herhalingen te voorkomen, bij iedere stap van het onderzoeks- en publicatieproces bleken zij weer onmisbaar. Daar komen ze dan...

Prof. Looijenga, Leendert, ik ken jou vanaf mijn eerste jaar als geneeskunde student waar je mijn tutor was. De vrijheid die je me toen gaf om in en uit te lopen in jullie lab was eigenlijk kenmerkend voor alles wat daarna kwam. Je laat alles schijnbaar op zijn beloop, waardoor creativiteit maximaal tot zijn recht komt. Ik heb je er echter nog nooit op kunnen betrappen dat je niet tot in detail op de hoogte was van wat er in een project speelde. Dank dat je me deze zelfstandigheid gegund hebt en dank voor je 24/7 bereikbaarheid en betrokkenheid. De cirkel was rond toen we vorig jaar samen tutor waren van een nieuwe groep eerstejaars geneeskundestudenten. Bedankt dat je mijn promotor wilde zijn. Ik vond het ook leuk om jou, **Katharina** en jullie gezin beter te leren kennen.

Ad, het is heel simpel: zonder jou was ik nooit gepromoveerd. We hebben vaak gelachen om mijn ehm... beperkte vaardigheden in het lab, maar ik ben blij dat we de afgelopen jaren zo'n goed team hebben kunnen vormen. Je hebt me veel geleerd over hoe de data tot stand komt in het lab. Je bent co-auteur bij vijf van mijn zeven niet-review artikelen en we zijn gedeeld eerste auteur van een zesde publicatie in dit proefschrift. Rekenen, visualiseren, hypothetiseren en publiceren heeft alleen zin met top-kwaliteit data. Jouw vaardigheden in het lab zijn dan ook onmisbaar voor al het werk wat daaruit is voortgekomen.

Wolter, jij was deze jaren mijn kamergenoot. Als je als kersverse basisarts aantreedt lijkt het misschien even slikken als je op een kamer zit met het ermitus hoofd van de afdeling. Jij was daar heel duidelijk over: geen fratsen, doe maar gewoon normaal. En gelukkig houden we van dezelfde muziek. Het was voor mij ook een voorrecht om soms mee te mogen kijken bij de diagnostiek, om met je te kunnen discussiëren over de biologie van kiemceltumoren en natuurlijk om te praten over het minimaal invasieve obductie project, samen met **Britt** en **Ivo**. Bedankt ook voor je interesse in Kirsten en mij, ook naast het werk. Ik had me geen betere kamergenoot kunnen wensen. Ik ben alleen benieuwd wanneer die kar met archiefspullen een keer leeg gaat ☺ .

Hans, samen met **Remko, Ad, Ton en Lambert** vorm jij de harde kern van LEPO. Jouw vaardigheden met immunohistochemie zijn inspirerend. Dank dat ik dit al helemaal in het begin van jou heb mogen leren.

Lambert, jij neemt nooit een blad voor je mond en dat is waardevol. Op ten minste vijf hoofdstukken uit dit proefschrift heb jij je biologische licht laten schijnen. Regelmatig zette je mij daarbij met beide benen terug op de grond. In de kern van je boodschap heb je altijd een goed punt. Dank voor de discussies.

Yvonne, voordat jij mijn pad kruiste wist ik niets van epigenetica. Op een gegeven moment had jij "wat data" waarbij ik je kon aanvullen op het gebied van de computational biology. Een paar weken later hadden we samen twee publicaties. Het was een intensieve samenwerking onder forse tijdsdruk maar ik vond het zo leuk dat ik er het tweede deel van mijn proefschrift maar aan gewijd heb. Dank voor de samenwerking en de gezelligheid.

Ton, jou leerde ik vooral beter kennen tijdens de serum micro-RNA projecten. Na toekennen van de KWF grant voor dit onderzoek heb jij dit stokje ten dele van Ad en mij overgenomen. Ik wens je veel succes en ben enorm benieuwd naar de resultaten.

Martine en **Berdine**, bedankt dat ik met jullie mee mocht kijken in de keuken van de KNO. Spannende resultaten waar ik graag meer van hoor ook al zit ik straks niet meer naast jullie. Vanuit de urologie mocht ik soms even wat proeven van de kliniek via **Katja** en **Lisette**: dank. **Katja**, ook veel dank voor je persoonlijke betrokkenheid. Ten slotte heb ik met veel plezier samengewerkt met alle **studenten** en **stagaires** die de afgelopen jaren op het lab hebben gewerkt.

Niet alles draait om werk gelukkig. Het koffie-groepje, een noodzakelijke ontsnapping aan de verschikkelijke automatenkoffie, zorgde voor regelmatige onderbrekingen. **Yvonne**, **Berdine**, **Martine** en **Remko**, bedankt voor de afleiding. Lunch met als vaste kern **Ad**, **Hans** en **Remko** was ook altijd een goed moment om even ergens anders over te praten dan onderzoek. En **Wil**, bedankt dat je soms zomaar met een casus of dia voor de deur stond met de "simpele" vraag: wat heeft de patient? Want van jou mocht ik geen moment vergeten dat ik toch in de eerste plaats een dokter ben. Ten slotte **Ingrid**, jij weet alles, ziet alles, je kent iedereen en je hebt altijd een oplossing met een glimlach, super!

De laatste alineas van dit dankwoord zijn natuurlijk voor de achterban, het vangnet, de boksballen. **(Schoon)familie** en **vrienden** vormen een lijst die te lang is om op te noemen, maar zijn voor mij een onmisbare steunpilaar geweest tijdens dit promotietraject. De beschrijvingen in deze en de volgende alineas zijn te kort en doen onvoldoende recht aan de betekenis van deze mensen voor mij, maar jullie weten wat ik bedoel. **Opa Jan** en **Oma Tan**, altijd geïnteresseerd en begripvol. **Mieneke** en **Jeroen**, nooit aflatend betrokken en ondersteunend waar nodig. **Mieneke**, veel dank voor jou correcties op de tekst van dit proefschrift. **Frits, Renske, Eva** en **Joost**: dank voor jullie interesse en dank dat ik de afgelopen jaren deel mocht worden van jullie gezin en dat jullie Kirsten aan mij toevertrouwd hebben.

Tijdens de verdediging heb ik het voorrecht dat mijn tweelingbroer en mijn zus mij bijstaan als paranymfen. Vroeger zouden jullie de verdediging van mij moeten overnemen als het nodig mocht blijken, maar vandaag staan jullie hier als symbool van het belang wat ik hecht aan ons gezin. **David**, zoals in bijna alles ging jij mij voor met promoveren. Ik was ook jouw paranymf en we hebben samen gepubliceerd. Uniek om al deze dingen samen te doen. Ik wil niet de kracht van stelling 8 afhalen, maar het is soms best fijn/handig om tweeling te zijn en een beetje op elkaar te lijken. **Anne Fleur**, je bent net als David een onmisbaar deel van mijn leven. Bedankt dat je je grote broer laat voorgaan met promoveren. En met goede reden: jullie geweldige zoonje en mijn neefje Seb. **Seb**, bedankt dat ik je moeder voor mijn verdediging mag lenen. **Lex** en **Sonja**, bedankt voor het lenen van jullie partners en bedankt voor alle gezelligheid en betrokkenheid.

De laatste alinea reserveer ik voor drie mensen: mijn ouders en mijn vrouw. Jaja, Kirsten, net op tijd getrouwd om dat op te kunnen schrijven ☺ . Lieve ouders, jullie steun aan ons alle drie en aan elkaar is altijd onvoorwaardelijk geweest. Dit is niet vanzelfsprekend, een goed voorbeeld en voor een belangrijke deel de reden dat ik hier (samen met David en Anne Fleur) sta. Bedankt. **Anne**, als vader en als persoon ben jij altijd enorm bescheiden, volgens mij niet altijd terecht. Nooit heb jij geneeskunde als optie voor mij genoemd, zelfs niet toen ik het zat was bij bedrijfseconomie en rechten. Maar je ging wel mee naar de cursus biologie (had ik nooit gehad) en liet me na mijn definitieve keuze voor geneeskunde het werk van een patholoog zien. Je hebt me altijd voorgehouden (als dokter) bescheiden te zijn, je grenzen te kennen en altijd verder te zoeken dan je eerste conclusie/diagnose ("don't jump into a conclusion"). Aan het onderzoek droeg jij ook bij door David en mij jaarlijks mee te nemen op een wandelvakantie: hoofdstuk 4 is daar grotendeels ontstaan; zonder computer, op kladpapier. **Henriëtte**, jij bent van de gezelligheid. Altijd een open house en altijd een luisterend oor, hoewel Anne dat sinds zijn pensioen ook behoorlijk heeft opgepakt. Je verstaat als geen ander de kunst om met honderd dingen tegelijk voor jan en alleman bezig te zijn, waarbij de (klein)kinderen en Anne altijd volop in jouw onverdeelde aandacht mogen staan. Het knappe en leerzame aan jou is dat je vaak uit die moordende schemas alleen maar energie lijkt te putten. Inspirerend. **Kirsten**, met jou mag ik dit proefschrift afsluiten en daarna in Amsterdam een nieuw leven beginnen. Het is niet altijd makkelijk om naast een work-a-holic als ik te leven die vergroeid lijkt met zijn telefoon. Maar zoals een goede vriend van me laatst zei: sinds je Kirsten kent ben je gelukkig wel wat relaxter geworden. Geef dus niet op, wat je doet heeft wel degelijk effect ☺ . Je bent mijn steun en toeverlaat, mijn thuisbasis en soms mijn symbolische boksbal bij stress (sorry!). Dit proefschrift is daarom ook van jou. Je helpt me steeds weer herinneren wat echt belangrijk is als ik met mijn hoofd in de cijfertjes zit. Dank je wel. Bij het formuleren van de stellingen en het ontwerpen van de cover bleek je een enorme hulp. Niet in de laatste plaats door aan te geven dat veel suggesties "veel te saai" of juist te druk waren. Ik ben benieuwd wat de toekomst ons gaat brengen, maar één ding weet ik zeker: ik heb er zin in!

Tot ziens. Het ga jullie goed. *Martin*

CURRICULUM VITAE

name	Martin Anne Rijlaarsdam	age & gender	31, male
telephone	+31 6 45 408 508	date & place of birth	December 31st, Leiden
e-mail	m.a.rijlaarsdam@gmail.com	nationality	Dutch
website	www.martinrijlaarsdam.nl	marital status	Married

EDUCATION

Erasmus University Rotterdam

2004 - 2011	**drs, MD in Medicine**
	Master thesis: "Comparing what cannot be compared" (Gerrit Jan Mulder thesis award); Development of algorithms for integrated analysis of copy number and expression data.
	Activities: Medical Student Association, various committees and member of the advisory board.
	Elective internship: internal medicine (oncology), Maasstad Hospital, Rotterdam.
	Senior internship: radiology, Erasmus MC, Rotterdam.
2005 - 2009	**MSc in Molecular Medicine** (research master)
	Thesis: "Transcriptomics of Germ Cell Tumors." Research aims:
	1. Validation of two computational models aiming at the detection of biologically relevant miR/mRNA interactions. Implementation of quality control.
	2. Functional studies in tumor cell lines investigating the function of miR-122a as a tumor marker in yolk sac tumors.
2002 - 2008	**BSc in Business Economics**
	Thesis: "Hypertext healthcare." Application of explicit and implicit knowledge and communication in patient-centered innovation.

Northgo College (high school)

| 1995 - 2002 | **HAVO** (cum laude) |
| | **VWO** (Natuur & Techniek, cum laude) |

EXPERIENCE

Erasmus MC Rotterdam, department of Pathology

2011 - 2014	**Ph.D. candidate** (promotor: prof.dr. L.H.J. Looijenga)
	Functional genomics of germ cell tumors. By applying (experimental) computational methods to (matched) genome-wide datasets the research aims to provide insight into epigenetic and (post-)transcriptional regulation in germ cell tumor pathogenesis. The research focuses on clinical applications of the findings and contributes to the emergence of new hypothesis that can be experimentally and clinically validated in the future. Please see my Ph.D. portfolio in this thesis for an overview of courses, presentations and teaching activities.
2008 - 2011	**Research student** (prof.dr. L.H.J. Looijenga)
	Integrated transcriptomics and comparative genomics in germ cell tumors.

Promeras

2011 – 2014	**Board member**
	Treasurer of the association of Ph.D. candidates

YOU@WWW

2009 - 2013	**Website design**
	Sole proprietorship

Erasmus MC Rotterdam, ICT department

2006 - 2008	**Elpado developer**
	Assisting the medical staff in the design of the interface of the Electronic Patient Record software of the Erasmus Medical Center (Elpado) at various departments, e.g. cosmetic/reconstructive surgery and physiotherapy.

Erasmus MC Rotterdam, department of Education (Medicine)

2006 - 2008	Supervising the financial planning and control concerning the budgets student foundations within the faculty of Medicine receive from Erasmus Medical Center.

Erasmus College & Holland Internet School

2005/6/7	**Remedial Teaching**
	Organizing and teaching an annual 4 day course in economy for HAVO & VWO students in preparation for their final exams (3x) as well as organizing a trial for e-learning (2006).

CED Group

2004/5	**CITO reports.**
	Writing of statistical reports about the performance of elementary schools.

Music school Warmond

2002 - 2004	**Mucis teacher**
	Clarinet

SKILLS & INTERESTS

Computer	Database design: *Microsoft Access, SQL, SPSS, Structured Data Entry* Statistical analysis: *SPSS, R, Matlab* Computational biology / modelling: *R, Matlab* Electronic patient record: *EPLADO (development & user), ChipSoft* Graphical design: *Adobe Photoshop & Illustrator* Website development: *WordPress, HTML, PHP* Office software: *Microsoft Office, VBA*
Languages	Ducth (native proficiency), English (full professional proficiency)
Interests	Fitness, scuba diving, classical music (including playing clarinet in ochestras)

LIST OF PUBLICATIONS

Published or in press

- Rijlaarsdam MA, van Herk HA, Gillis AJ, Stoop H, Jenster G, Martens J, van Leenders GJ, Dinjens W, Hoogland AM, Timmermans M, Looijenga LH: **Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours.** Br J Cancer 2011, 6:854-863.

- Gillis AJ*, Rijlaarsdam MA*, Eini R, Dorssers LC, Biermann K, Murray MJ, Nicholson JC, Coleman N, Dieckmann KP, Belge G, Bullerdiek J, Xu T, Bernard N, Looijenga LJ: **Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle.** Molecular oncology 2013, 7:1083-1092.

- Rijlaarsdam MA, Rijlaarsdam DJ, Gillis AJ, Dorssers LC, Looijenga LH: **miMsg: a target enrichment algorithm for predicted miR-mRNA interactions based on relative ranking of matched expression data.** Bioinformatics 2013, 29:1638-1646.

- Rijlaarsdam MA*, van der Zwan YG*, Dorssers LCJ, Looijenga LHJ: **DMRforPairs: identifying Differentially Methylated Regions between unique samples using array based methylation profiles.** BMC Bioinformatics 2014, 15:141.

- van der Zwan YG*, Rijlaarsdam MA*, Rossello FJ, Notini AJ, de Boer S, Watkins DN, Gillis AJ, Dorssers LC, White SJ^, Looijenga LH^: **Seminoma and Embryonal Carcinoma Footprints Identified by Analysis of Integrated Genome-Wide Epigenetic and Expression Profiles of Germ Cell Cancer Cell Lines.** PloS one 2014, 9:e98330.

- Rijlaarsdam MA, Looijenga LHJ: **An oncofetal and developmental perspective on testicular germ cell cancer.** Seminars in Cancer Biology 2014.

- Rijlaarsdam MA*, van Agthoven T*, Gillis AJM, Patel S, Hayashibara K, Lee KY, Looijenga LHJ: **Identification of known and novel germ cell cancer specific (embryonic) miRs in serum by high throughput profiling.** Andrology 2014. *in press*

Submitted

- Rijlaarsdam MA, Tax DJM, Gillis AJM, Dorssers LCJ, Koestler DC, de Ridder J, Looijenga LHJ: **Genome wide DNA methylation profiles provide clues to the origin and pathogenesis of germ cell tumors.**

- Oosterhuis JW, Stoop H, Rijlaarsdam MA, Smit V, Hersmus R, Looijenga LHJ: **Pediatric germ cell tumors presenting beyond childhood?**

*/^ shared first or last authors

Ph.D. PORTFOLIO

Summary of Ph.D. training

Department: Department of Pathology
Erasmus MC Cancer Institute – University Medical Center, Rotterdam, NL

Research School: Molecular Medicine Postgraduate School

PhD period: December 2011 – November 2014

Promotor: Prof.dr. L.H.J. Looijenga

Research skills	Year
Biostatistical Methods I: basic principles, NIHES	2012
Biostatistical Methods II: classical regression models	2013

In-depth courses	
Complete Genomics course, Complete Genomics	2011
Adobe Indesign CS5, MolMed	2011
Translational Oncology, MolMed	2012
UCSC Gene Browsing workshop, MolMed	2012
Introduction to clinical and fundamental Oncology, NVvO	2013
Federa scientific day (NextGen Sequencing)	2013

Presentations	
Work group (LEPO) meetings (6x)	2012-2014
Scientific meetings (JNI) (3x)	2012-2014
MolMed Day 2013 (1 oral presentation, 1 poster)	2013

Seminars and workshops (not presenting)	
Weekly LEPO workgroup meeting	2012-2014
Weekly JNI research meeting	2012-2014
Bi-weekly PALMs (translational research meetings)	2013-2014

Teaching activities	
Supervision of students, Junior Medschool	2013
Tutor for first year medical students	2013
Support of other research projects	2012-2014