RESEARCH ARTICLE

Genome Wide DNA Methylation Profiles Provide Clues to the Origin and Pathogenesis of Germ Cell Tumors

Martin A. Rijlaarsdam1, David M. J. Tax2, Ad J. M. Gillis1, Lambert C. J. Dorssers1, Devin C. Koestler3, Jeroen de Ridder2, Leendert H. J. Looijenga1 *

1 Department of Pathology, Erasmus MC Cancer Institute—University Medical Center Rotterdam, Rotterdam, The Netherlands, 2 Faculty of Electrical Engineering, Mathematics and Computer Science Intelligent Systems—Delft Bioinformatics Lab, Technical University of Delft, Delft, The Netherlands, 3 Department of Biostatistics, University of Kansas Medical Center, Kansas City, Kansas, United States of America

* l.looijenga@erasusmc.nl

Abstract

The cell of origin of the five subtypes (I-V) of germ cell tumors (GCTs) 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 (type I-IV) and four cell lines were profiled (Illumina’s HumanMethylation450BeadChip). Data was pre-processed controlling for cross hybridization, SNPs, 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), seminomas/dysgerminomas (type II), spermatocytic seminomas (type III) and teratomas/dermoid cysts (type I/IV). The seminomas, dysgerminomas and spermatocytic seminomas were globally hypomethylated, in line with previous reports and their demethylated precursor. Differential methylation and imprinting status between subtypes reflected their presumed cell of origin. Ovarian type I teratomas and dermoid cysts showed (partial) sex specific uniparental maternal imprinting. The spermatocytic seminomas and spermatocytic seminomas were globally hypomethylated, in line with previous reports and their demethylated precursor. Differential methylation and imprinting status between subtypes reflected their presumed cell of origin. Ovarian type I teratomas and dermoid cysts showed (partial) sex specific uniparental maternal imprinting. The spermatocytic seminomas showed uniparental paternal imprinting while testicular teratomas exhibited partial imprinting erasure. Somatic imprinting in type II GCTs might indicate a cell of origin after global demethylation but before imprinting erasure. This is earlier than previously described, but agrees with the totipotent/embryonic stem cell like potential of type II GCTs and their rare extra-gonadal localization. The results support the common origin of the type I teratomas and show strong similarity between ovarian type I teratomas and dermoid cysts. In conclusion, we identified specific and global methylation differences between GCT subtypes, providing insight into their developmental timing and underlying developmental biology. Data and extended annotation are deposited at GEO (GSE58538 and 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 (Fig 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].

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] (8.38/100,000 Dutch population. Dutch Cancer Registration (IKNL), www.cijfersoverkanker.nl). Risk factors have been thoroughly investigated and are integrated in a genenvironmental 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 (Fig 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
Fig 1. Tumor types/samples and cell lines analyzed and schematic visualization of genomic functional categories of interest. (A) GCT subtypes in the hypothesized context of normal germ cell development as proposed in earlier studies (grey box). 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) Samples included in this study. Abbreviations match Fig 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
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 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 (Fig 1A and 1B). 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 were investigated, starting from global methylation profiles, followed by functional enrichment analysis. Probes were functionally classified according to their relation to genes: transcription regulating (200 or 1500bp upstream of the TSS & 5'UTR) or gene coding (exon 1, gene body and 3'UTR). Probes covering micro-RNA (MIR) coding regions, CpG islands and/or transposon elements (LINE/SINE) were classified separately as were imprinting associated genes. For a detailed explanation, please see Fig 1C and the Materials and Methods section (section: (Additional) annotation 450K array). After functional enrichment analysis, specific differentially methylated probes were identified (DMPs). Probes represent individual CpG sites. Also, differentially methylated regions (DMRs) containing multiple adjacent probes were identified. Finally, imprinting status was evaluated. Please note that differential methylation indicates a statistically significant difference after correction for multiple testing, unless specifically stated otherwise. Differential methylation of $\Delta M > 0.9$ was considered relevant, in line with the recommendations of Du et al [41]. For details about the statistical procedures, please see the materials and methods section (analysis protocol). Abbreviations are explained in (the legend of) Fig 1.
SS and SE/DG show global hypomethylation when compared to EC/mNS and TE

Fig 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 S2A Fig. SS showed global hypomethylation (Fig 2A), i.e. a large 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 (Fig 2A). The SE group showed consistent hypomethylation (S2B Fig, page 2), in contrast to study of Nettersheim et al who showed separate groups of hypo- and hypermethylated SE in a larger sample series [42]. In contrast to the SE and DG samples, the EC and partly differentiated mNS, type I TE and DC samples consistently showed a bimodal pattern with one mode around 10% and one around 90% (Fig 2A and Fig 1: relation between subtypes). This bimodal pattern was also observed in three EC cell lines and a single SE cell line (Fig 2A, CL_SE & CL_EC). In line with previous reports [14,43], the EC cell lines were more methylated than the SE cell TCam-2 (Fig 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 show a bimodal pattern (Fig 2A and S2A Fig). 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 [20,44]. Hypermethylation of LINE/SINE elements NS and hypomethylation (Fig 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 [45].

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 (Fig 2B and S2A Fig). In line with the larger inter-sample variation (S2B Fig), 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 type I 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 (Figs 2B and S2A).

Zooming in: GCT subtype specific methylation patterns

To further pinpoint differences between pairs of GCT subtypes, DMPs were identified (Table 1, S1 Table), tested for functional and chromosomal enrichment (Fig 3, S3 Fig, Table 1 and S2 Table) and grouped into DMRs (Fig 4, S4 Fig, Table 1, S3 Table, GSE58538: File S1). SE + DG and EC + mNS (including type II pure TE) subtypes were merged because of high similarity of the observed methylation profile (Figs 2A and 2B, S3A Fig), in line with literature regarding their similar origin [46] and their close relation in the current WHO classification [16,18]. Recurrent DMRs were identified as genes occurring more than once within or between comparisons, which may indicate regions of importance (S3 Table, n = 149).

(Differential) methylation of GCT cell lines (4136 DMRs between the cell lines: GSE58538: File S2) showed little similarity to their in vivo counterparts (Figs 2 and 6, S2 Fig). The cell line analysis did however identify a biologically relevant DMR previously validated in these cell lines.
Fig 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 (Fig 1C) is visualized using the distribution of the methylation percentage $\beta$ 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 ($\beta$) of all probes in all samples in a certain category. The boxplot indicates the interquartile range (black bars) and
using bisulfite sequencing in [63] (microRNA-371/2/3 cluster, Table 2). 719 gene symbols intersected between tumor and cell line DMRs (S3 Table). 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 showed vastly different methylation profiles. The relative hypermethylation in EC/mNS versus SE/DG was concentrated in regions not involved in transcription regulation (Fig 3A). This pointed to a global difference in methylation status rather than differential methylation of specific regulatory elements. This also held for the hypermethylation of type I TE when compared to SE/DG (Fig 3B). The 61 DMPs hypermethylated in SE/DG relative to type I TE were concentrated at three specific genes: **NCOA2**, **ALOX12** and **ECEL1P2** (Table 1, S3 Table, S4A Fig).

DMPs between type I TE and EC/mNS indicated a more methylated profile of the EC/mNS group (Fig 3C). Moreover, the majority of the probes hypermethylated in type I 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, S3B Fig). DMRs included many genes involved in male gametogenesis like **DMRT3** (Fig 4A). The EC marker **SOX2** [17,64] was present as one of the only 15 hypermethylated autosomal DMRs in type I TE (Fig 4B). These DMRs presumably relate to the cell of origin as well as to the sex of the patient (S4B Fig, Table 1 and S3 Table).

**Type III (SS) versus type II seminomatous GCT (SE/DG).** The general, probes significantly hypomethylated in SS as compared to SE/DG were enriched for regions associated with paternal expression (Fig 3D). DMRs hypermethylated in SE/DG predominantly included recurrent DMRs and DMRs within genes associated with germ cell and testis development (Table 1 and S3 Table). 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 [17,46,65] (Fig 4C, discussed in 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 and S3 Table).

**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 Fig 5 and S5A Table). 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] (Fig 4D) and **BAX1**, which also contained a DMR[SE/DG-SS] (all SNP related genes: S5B Table).

**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 (Fig 1C and S4 Table) in the tumor groups probes covering regions...
Table 1. Pairwise comparison of GCT subtypes.

<table>
<thead>
<tr>
<th>A. Seminomatous (SE/DG) versus non-seminomatous (EC/mNS) GCTs</th>
<th>B. Seminomatous type II (SE/DG) versus type I (TE)</th>
<th>C. Non-seminomatous type II (EC/mNS) versus type I (TE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DMProbes[SE/DG-ec/mNS]</strong></td>
<td><strong>DMProbes[se/bo-EC/mNS]</strong></td>
<td><strong>DMProbes[EC/mNS-TE]</strong></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[no enrichment]</td>
<td>[no enrichment]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DMRegions[SE/DG-ec/mNS]</strong></td>
<td><strong>DMRegions[se/bo-EC/mNS]</strong></td>
<td><strong>DMRegions[ec/mNS-TE]</strong></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[global]</td>
<td>[global]</td>
<td>[global]</td>
</tr>
</tbody>
</table>

**A. Seminomatous (SE/DG) versus non-seminomatous (EC/mNS) GCTs**

- **DMProbes[SE/DG-ec/mNS]:**
  - 6 probes
  - [no enrichment]

- **DMRegions[SE/DG-ec/mNS]:**
  - 0 regions
  - [global]

**B. Seminomatous type II (SE/DG) versus type I (TE)**

- **DMProbes[SE/DG-TE]:**
  - 61 probes
  - Significantly overrepresented on chromosome 12 (13/61) and preferentially located in the 5'UTR of genes and CpG islands

- **DMRegions[SE/DG-TE]:**
  - 3 regions
  - **NCOR2:** (SMRT, silencing mediator for retinoic acid and thyroid hormone receptors) nuclear receptor co-repressor on 12q24.31 involved in mouse spermatogenesis [47] and vitamin D metabolism in GCTs [48].
  - **ALOX12:** lipoxygenase family [49], has not been implicated in GC(T) biology. **EC/MPS2:** increased methylation upon aging [50] — 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.

- **DMProbes[EC/mNS-TE]:**
  - 17,407 probes
  - Enrichments were weak, the strongest being ↑ in transposons.

- **DMRegions[EC/mNS-TE]:**
  - 580 regions (all autosomal)
  - e.g. **DMRT3:** implicated in testis development and male sex determination [46,51]
  - **MOVT1:** which has been implicated in human male infertility [52] and germ cell maturation in mice [53].
  - **DDR2:** crucial for spermatogenesis in mice [54] ICR_P WT1 was also present.

**C. Non-seminomatous type II (EC/mNS) versus type I (TE)**

- **DMProbes[EC/mNS-TE]:**
  - 1,520 probes
  - ≈80% on X chromosome - differences in sex (TE = male+female, EC/mNS = male).

- **DMRegions[EC/mNS-TE]:**
  - 128 regions (15 autosomal)
  - e.g. **SOX2:** EC marker (see: Table 1). **IRX5:** germ cell migration in Xenopus laevis embryos; DMR =1kb downstream of its 3'UTR [55].

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 Fig 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 Fig 2. Functional enrichment of DMPs is visualized in Fig 3. Details of enrichment calculations and raw counts and percentages are presented in Table S2. Enrichment of chromosomes is depicted more detailed in S3B Fig. DMRs, recurrent tumor DMR and DMPs are listed in S3 and S1 Tables respectively. DMRs are visualized in Fig 4, S4 Fig and GSE58538: Files S1 and S2.

PLOS ONE | DOI:10.1371/journal.pone.0122146 April 10, 2015 8/3
that are regulating paternally expressed genes (ICR_P) showed somatic methylation in type I and II GCTs with a trend towards hypermethylation in DC (Fig 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 seemed 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 (S2A Fig, pages 15 and 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 (Fig 6B). IMPR_M200/IMPR_P1500 probes showed hypomethylation similar to non-imprinted genes in all groups (S2A Fig, pages 18 and 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 (Fig 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 (S2B Fig) with the exception of type I TE and DC (Fig 6D and S2B Fig). In addition to the earlier analysis, where the type II TEs were grouped with the mNS and the type I TEs were assessed as one class, TEs were also investigated individually, grouped according to sex and anatomical site, in line with sex specific imprinting occurring during fetal/germ cell development (Fig 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 type I 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.

Validated ICRs (S4 Table) 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 ICRs 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. (Fig 7A and 7B, S6 Fig).

In summary, ovarian type I TE and DC showed partial sex specific uniparental 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 (Fig 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 + type I TE and hypomethylated SS + SE/DG clusters, in line with
Fig 3. Functional enrichment of DMPs. DMPs were classified according to their functional genomic location (Fig 1C). Statistical over- and underrepresentation of probes in certain categories provides clues to differences between GCT subtypes in regarding function of methylation. Enrichment was assessed by comparing the number of probes in a functional category in a subset of DMPs with the that in the total dataset (Fisher’s Exact test, see Materials & Methods section). Results are shown for four pairwise (A vs B).
previous reports [14,43,117,135] (Fig 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 (Fig 2B). Hypermethylation in EC/mNS and type I 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 type I 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 (Figs3, 4 and 5, Tables1 and 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 (Figs6 and 7). The local and global methylation difference observed between GCTs could be matched to physiological germ cell development, but did not match with their respective cell line models (Fig 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 (Fig 6D). This (partial) mimicking of female germ cells in ovarian type I TE is in line with in several studies [20,131,132]). 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] (Fig 8).

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 [136]. Earlier studies have suggested separated NS-CIS and a SE-CIS types
Fig 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. 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
but the lack of methylation in CIS combined with absence of SOX2 (EC marker) expression \cite{64,136,137} 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 \cite{14,43,117,135,136}. EC and mNS show a (de novo) methylated profile (Fig 2A). This is in line with the previously reviewed increased methylation in the transition of CIS into NS \cite{13,14,43,138}, possibly illustrating reversal to a hypermethylated ES like state \cite{7,16,139–142} 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 (Fig 6, S6 Fig and S4 Table) was in line with an earlier report \cite{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 (Fig 6A and 6B) situates the cell of origin of the type II tumors possibly earlier than previously described \cite{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 \cite{16,139–142} (Fig 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-spermatogonial state with regard to their cell of origin \cite{36–39,46}. Earlier epigenetic data showed a heterogeneous profile of histone modification and methylation profiles, not corresponding with a pre-spermatogonial origin \cite{143}. Our limited series of SS show a consistent pattern of distinct hypomethylation and loss of imprinting at the paternally expressed ICRs (ICR_M: heterogeneous \( \geq 50\% \), Fig 2B). This matches with a cell of origin between the gonocyte and spermatogonial 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 \cite{16} (Figs 2B, 6 and 8).

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 (Fig 8). Methylation profiles allowed for separation of clusters of NS, SE/DG, SS and TE/DC, largely in line with the current WHO classification. 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 cell of origin as did imprinting status. However, somatic imprinting in type II GCT might indicate a cell of origin after global demethylation but before imprinting erasure. This is earlier than previously described, but agrees with the totipotent/embryonic stem cell like potential of type II GCTs and their 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. However, the limited samples size and
### Table 2. GCT (methylation) associated genes.

<table>
<thead>
<tr>
<th>Gene &amp; region</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APC</strong>, chr5 (+), 112,043,195–112,181,936</td>
<td><strong>GCT link:</strong> 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 [86]. <strong>Findings:</strong> 2102EP showed mild but significant relative hypermethylation compared to the other cell lines, but for all tumor groups APC was consistently hypomethylated.</td>
</tr>
<tr>
<td><strong>AR</strong>, chr X (+), 66,763,874–66,950,461</td>
<td><strong>GCT link:</strong> 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 &amp; SS. This supports the hypothesis that methylation does not occur in the germ cell lineage [87]. <strong>Findings:</strong> 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 (Fig 5A).</td>
</tr>
<tr>
<td><strong>CTA genes</strong></td>
<td><strong>GCT link:</strong> Cancer Testis Antigens (CTA) are primarily unmethylated in SE. MAGeA1/3 are predominantly methylated in NS while SYCP1 is unmethylated in NS [88]. MAGeA1: chrX:152,481,522–152,486,116(-), MAGeA3: chrX:151,934,652–151,938,240(-), SYCP1: chr1:115,397,424–115,537,991(+). <strong>Findings:</strong> Methylation differences in these genes were not remarkable except for differential hypermethylation of TCam-2 and NCCIT compared to the other cell line. The TSS associated regions of MAGEA1 and SYCP1 were consistently hypomethylated in SS.</td>
</tr>
<tr>
<td><strong>GATA4</strong>, chr 8 (+), 11,534,468–11,617,511</td>
<td><strong>GCT link:</strong> previously identified DMR between TCam-2 and NCCIT, promoter region hypermethylated in TCam-2 [89]. <strong>Findings:</strong> 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 hypomethylation while sacral/ovarian TE, DCs and SS showed relative hypomethylation like the SE/DG samples.</td>
</tr>
<tr>
<td><strong>HIC1</strong>, chr17 (+), 1,957,448–1,962,981</td>
<td><strong>GCT link:</strong> 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 [70]. HIC1 promoter methylation has been implicated in treatment resistance in GCTs [71]. <strong>Findings:</strong> HIC1 was showed predominantly hypomethylated in all GCT subtypes even though a weak DMR [EC/mNS-re] was identified. Of the cell lines, only 2102EP showed differential hypermethylation.</td>
</tr>
<tr>
<td><strong>KIT</strong>, chr4 (+), 55,524,085–55,606,881 / <strong>KITL</strong>, chr 12 (-), 88,886,570–88,974,628</td>
<td><strong>GCT link:</strong> KIT and KITL regulate primordial germ cell development and homing to the gonad [72–76]. In the embryonic phase the guidance of KIT+ primordial germ cells from the hind gut epithelium to the gonads depends strongly on KITL mediated chemotactic attraction [73,77–79]. In the postnatal testis KIT+KITL signaling takes place via paracrine signaling in the germine stem cell niche and is crucial for spermatogenesis from the spermatogonial stage onwards [73,75,80,81]. More mature mouse spermatids and spermatozoa express a c-terminal truncated form of KIT transcribed from an intronic promoter [82]. 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 [17,83–85]. Progression into invasive SE is also strongly related to KIT/KITL signaling while much less association with the NS phenotype has been shown [80,86–89]. 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 [90–93]. Recent GWAS studies identified susceptibility loci for GCTs close to, within or directly related to GCTs [29,94–105]. 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 [106] and SNPs in KITL combined with aberrations in cAMP regulation were suggested to contribute to tumor risk in these patients [105]. <strong>Findings:</strong> KIT (S6A Fig) and KITL (S6B Fig) were not differentially methylated between any of the tumor groups or cell lines.</td>
</tr>
</tbody>
</table>

(Continued)
Table 2. (Continued)

<table>
<thead>
<tr>
<th>Gene &amp; region</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>miR-371/2/3</strong>, chr19 (+): [371] 54,290,929–54,290,995 / [372] 54,290,995–54,291,210 / [373] 54,291,959–54,292,027</td>
<td>GCT link: The micro-371-2-3 cluster is expressed in the stem cell component of GCT [107] and is a potential diagnostic serum marker for GCT [108]. Upstream of the TSS of this cluster a DMR has been identified between TCam-2 and NCCIT [69]. 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 $\rho$ -0.90, $p = 0.037$) [63]. <strong>Findings:</strong> The miR-371-2-3 cluster was hypomethylated in TCam-2 (CL_SE) and 2102EP and hypermethylated in NT2 and NCCIT (Fig 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 [63,107].</td>
</tr>
<tr>
<td><strong>NANOG</strong>, chr12 (+), 7,940,390–7,948,655</td>
<td>GCT link: Specific marker for the all stem cell components of GCTs [17]. RA treatment of NT2 cells also increased methylation here [109]. 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 [110]. <strong>Findings:</strong> 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 and consistency of the difference lacked significance (Fig 5C).</td>
</tr>
<tr>
<td><strong>POUSF1 (OCT3/4)</strong>, chr6 (-), 31,132,114–31,148,508</td>
<td>GCT link: Specific marker for the all stem cell components of GCTs [17,65,111]. 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 [109,112,113]. Differentiation of NT2 after retinoic acid treatment resulted in increased methylation and loss of expression [109]. <strong>Findings:</strong> A promoter DMR[SE/EC]-ss was identified despite the fact the SE/DG express the OCT3/4 protein and SS do not [17,46,65] (Fig 4C). However, probes located close to its transcription start site are generally unmethylated 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 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 [112,113]. Also, we previously showed that even though high promoter methylation is generally associated with low expression, this is not always the case [69].</td>
</tr>
<tr>
<td><strong>PRSS21</strong>, chr16 (+), 2,867,164–2,876,305</td>
<td>GCT link: TESTISIN (PRSS21) is a proposed tumor suppressor gene in TGCT regulated by methylation of a 385bp long CpG rich island [114] and CpG sites close to the TSS [115]. <strong>Findings:</strong> Al GCT subtypes except SS (DMR[SE/EC]-ss) showed hypermethylation of PRSS21.</td>
</tr>
<tr>
<td><strong>RUNX3</strong>, chr1 (+), 25,226,002–25,291,612</td>
<td>GCT link: 90% of the infantile YSTs (type I) showed methylation of RUNX3 while methylation was only rarely observed in the adult GCTs [57,116,117]. <strong>Findings:</strong> The promoter region of RUNX3 was consistently hypomethylated, 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.</td>
</tr>
<tr>
<td><strong>SOX17</strong>, chr8 (+), 55,370,495–55,373,456</td>
<td>GCT link: Discriminative marker between EC (+) and SE (-) [17,64]. <strong>Findings:</strong> SOX17 was consistently hypomethylated in all tumor groups and cell lines (Fig 5D).</td>
</tr>
<tr>
<td><strong>SOX2</strong>, chr3 (+), 181,429,712–181,432,224</td>
<td>GCT link: Discriminative marker between EC (+) and SE (-) [17,64]. Previously identified DMR upstream of TSS between ($\approx$50%) TCam-2 and ($\approx$0%) NCCIT [89] ($\approx$1kb upstream of the SOX2 TSS. The region directly upstream of the SOX2 TSS has consistently been found hypomethylated in both cell lines [89,118]. TCam-2 has been shown to differentiate and become SOX2 positive after extra-gonadal injection in mice [119]. <strong>Findings:</strong> A region $\approx$1 kb upstream of the SOX2 TSS was differentially hypomethylated in all CL_ECs as compared to TCam-2 (GSE58538: File S2). EC and SE tumor samples showed consistent hypermethylation of the region -154 --2283bp upstream of the SOX2 TSS in contrast to the TE samples which showed higher levels of methylation (DMR[cc/mNS-TE], Fig 4B).</td>
</tr>
</tbody>
</table>

(Continued)
Table 2. (Continued)

<table>
<thead>
<tr>
<th>Gene &amp; region</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TFAP2C (AP-2y)</strong>, chr20 (+), 55,204,358–55,214,339 / <strong>TFAP2A (AP-2a)</strong>, chr 6 (-), 10,393,419–10,419,892</td>
<td>GCT link: AP-2y is crucial for progression of PGCs into the germ line [120]. It is a known germ cell marker, abundantly expressed in CIS and SE, and heterogeneously expressed in NS and somatic tumors [120,121]. AP-2y expression is induced by estrogens [122]. 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 [69]. <strong>Findings:</strong> 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 (GSE58538: 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.</td>
</tr>
<tr>
<td><strong>XIST</strong>, chrX (-), 73,040,486–73,072,588</td>
<td>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 [123] but has so far not been validated. SE/NS/SS showed XIST expression (X inactivation) [67]. <strong>Findings:</strong> 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.</td>
</tr>
<tr>
<td><strong>ICR_M: H19-IGF2</strong>, chr11, 2,020,834–2,023,499</td>
<td>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 bialeric methylation occurs before the gonocyte stage in males [125]. In humans H19 is erased in fetal spermatogonia, but becomes fully methylated before meiosis (spermatagonia) [126]. H19 erasure fits uncially illustrated in [127] and related to pluripotency markers (SOX2 and OCT3/4) in germ cell development in [128]. Previous studies using have suggested low methylation of the H19-IGF2 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 [129–132]. 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 and coworkers in a high throughput approach [20]. <strong>Findings:</strong> The SS in our series show complete methylation at 1 of the two H19/IGF2 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 (Fig 7B; 2 regions from literature: S4 Table).</td>
</tr>
<tr>
<td><strong>ICR_P: SNURF/SNRPN</strong>, chr15, 25,199,934–25,200,343</td>
<td>GCT link: SNURF/SNRPN has been described to show derivation from somatic imprinting in type II GCTs (non-quantitative, not necessarily indicating erasure) [134]. 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 [131]. <strong>Findings:</strong> In this dataset, this SNURF/SNRPN (controlling paternal expression) was only covered by a single probe (S6 Fig). 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).</td>
</tr>
<tr>
<td><strong>ICR_P: MEST</strong>, chr7, 130,130,740–130,133,111</td>
<td>GCT link: The MEST ICR regulates paternal expression, is already erased in fetal spermatogonias and remains so during male germ cell development [128]. <strong>Findings:</strong> 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 (Fig 7A).</td>
</tr>
</tbody>
</table>

Genomic locations and strand were retrieved from genecards.com/UCSC. Detailed visualizations of the methylation status of these genes is presented in Fig 5 and S5A Fig. DMRs in the cell lines are presented in GSE58538: File S2. ICRs are visualized in Fig 7 and S6 Fig.

doi:10.1371/journal.pone.0122146.t002
Fig 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. The genes are reviewed in Table 2. (A) AR, (B) miR-371-2-3, (C) NANOG, (D) SOX17. (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.
conflicting results with some of the current literature warrants careful interpretation of the results and validation in a larger/extended dataset. Moreover, to 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” developed by the Dutch Federation of Medical Scientific Societies (FMWV (Version 2002, update 2011)). An overview of the samples in this study is presented in Fig 1A and 1B. Samples were collected when submitted to the pathology department and stored in liquid nitrogen.

**Cell lines.** Four cell lines were included (Fig 1B), all modelling type II GCTs. Cell lines derived from EC (CL_EC) include NT2 [144–148], NCCIT [145,149] and 2102EP [144–148]. TCam-2 closely resembles SE (CL_SE) [150–152]. 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 cm² 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 [110]. 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 [69]. This array does not distinguish between DNA methylation variants like 5mC and 5hmC [153].

Data analysis

**Data (pre-)processing.** Further processing was carried out in R using the LUMI package [154] according to [155,156]. 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 [156]. As a result 44,540 probes were discarded, leaving 437,881 valid, methylation related probes for processing and analysis. Finally, color
Fig 6. Methylation of imprinting control regions and the X chromosome. Analogous to Fig 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 \( \beta \). Next, the discriminatory power of the methylation pattern for each individual sample is shown using principal component analysis. (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
adjustment, quantile normalization and BMIQ-based correction for probe type bias (Infinium I vs II) were performed [154,155,157]. Data processing resulted in two quantifications of a CpG site's methylation status: the methylation percentage $\beta$ and an associated M-value which ($\logit_2(\beta)$). M-values were used for statistical computations because of a more favorable

---

**Fig 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 S6 Fig and listed in S4 Table. (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. Abbreviations of histological subtypes are explained in Fig 1A. Please note that 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. CL indicates cell lines.

doi:10.1371/journal.pone.0122146.g007
tradeoff between true positive rate and detection rate [41]. All data is available via GEO (GSE58538).

(Additional) annotation 450K array. The 450K annotation manifest (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 and miRBase were identified, as were probes within repeats defined by RepeatMasker (source: UCSC). 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) and their association with either paternal or maternal expression were retrieved from WAMIDEX 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 Fig 1C. The extended annotation including its documentation is available at GEO (GPL18809).

**Analysis protocol.** Below, the subsequent steps of the data analysis are described. More details are presented in S1 Fig. 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) integrate the benefits of a boxplot and a kernel density plot. Two-dimensional principal component analysis (PCA) was applied and validated using bootstrapping to assess how well the methylation values of (subsets of) the probes separated the histological subtypes. Formal statistical testing of the distribution of the methylation values identified (small) significant differences between almost all tumor classes (data not shown, Kruskal Wallis test followed by pairwise, Benjamini-Hochberg corrected Mann Whitney U tests). The PCA and violin plot based approach was preferred/used to identify the largest, most relevant differences.

**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 Hidden Markov Model (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 (S1 Fig, page 11). In total, 133,730 segments were identified. The median methylation value (M or $\beta$) of all probes in a segment or state was taken as methylation proxy. As a proof of concept, S1 Fig (page 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 S1 Fig.

**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%), S1 Fig, page 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 $\geq 95\%$ 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) [41]. Although a less stringent setting might result in a higher detection rate, it will considerably reduce the true positive rate [41]. 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.

**Differentiating Hidden Markov Model (HMM) states:** To identify non-adjacent regions that showed similar patterns of methylation a logistic LASSO regression model was fitted on the M values of the HMM states (glmnet package) [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 $\lambda$ 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 (S1 Fig (p. 3/5), also see above). 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 (Fig 1C) using a two sided Fisher’s Exact test. Analogously, association with chromosome and state was tested. p<0.05/(18+24+20) = 0.0080645161 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 \( \geq 5 \) adjacent DMPs and a maximal inter-DMP distance \( \leq 1 \) 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 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 [69].

**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 & gplots. The HMM was trained using Matlab R2012a (64 bits) using the MOGHMM toolbox.

**Supporting Information**

**S1 Fig. Supplementary methods.** A detailed flowchart and description of the analysis protocol is presented, followed by a visual illustration of the selection of differentially methylated features. The motivation for the threshold when filtering low variability probes is presented next. Finally, the properties of the Hidden Markov Model (HMM) are presented together with a detailed description of its construction. HMM state 15 is presented as a proof of concept, discriminating between male and female samples based on X-chromosomal localization of the large majority of the probes in this class.

(PDF)

**S2 Fig. A, Methylation patterns in GCT subtypes and cell lines—All categories and validation.** In addition to the selected categories presented in Fig 2 this figure contains all 18 functional categories presented in Fig 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). (Distribution plot of methylation percentage) Violin plots: grey areas indicate a kernel density plot of the methylation percentage (\( \beta \)) 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 Fig 1A and 1B. TE indicates type I TE only. (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 Fig 1A. CL indicates cell lines. Please note that in the 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. S2B Fig. Methylation patterns in GCT subtypes and cell lines—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.
Density plots are explained in the legend of Fig 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 (Fig 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. (Distribution plot of methylation percentage) Violin plots: grey areas indicate a kernel density plot of the methylation percentage ($\beta$) 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 Fig 1A and 1B. TE indicates type I TE only.

S3 Fig. A, Enrichment of differentially methylated probes (DMPs) for chromosomal position and HMM state—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. S3B Fig, Enrichment of differentially methylated probes (DMPs) for chromosomal position and HMM state—Association between DMPs and chromosome / HMM state. Stacked bar charts indicate the fraction of probes in a subset (DMP[A]=A, DMP [B]=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 $\alpha$ 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.

S4 Fig. A, Methylation profile at GCT subtype specific differentially methylated regions (DMRs)—continued—SE/DG versus type I TE. This figure depicts the DMRs between GCT subtypes discussed in the main text in addition to those already visualized in Fig 4. (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. Abbreviations of histological subtypes are explained in Fig 1A. Please note that 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. CL indicates cell lines. S4B Fig, Methylation profile at GCT subtype specific differentially methylated regions (DMRs)—continued—EC/mNS versus type I TE. This figure depicts the DMRs between GCT subtypes discussed in the main text in addition to those already visualized in Fig 4. (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. Abbreviations of histological subtypes are explained in Fig 1A. Please note that 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. CL indicates cell lines. S4C Fig. Methylation profile at GCT subtype specific differentially methylated regions (DMRs)—continued—SE/DG versus SS. This figure depicts the DMRs between GCT subtypes discussed in the main text in addition to those already visualized in Fig 4. (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. Abbreviations of histological subtypes are explained in Fig 1A. Please note that 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. CL indicates cell lines.

(PDF)

S5 Fig. A, Methylation status of GCT specific genes. This figure (together with S5B Fig) depicts the genes discussed in the main text and Table 2 in addition to those already visualized in Fig 5. Genes are annotated 1.5kb upstream of their TSS and 1.5kb downstream of their transcription termination site. (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. Abbreviations of histological subtypes are explained in Fig 1A. Please note that 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. CL indicates cell lines. S5B Fig. Methylation status of genes with SNPs significantly associated with GCTs. This figure (together with S5A Fig) depicts the genes discussed in the main text and Table 2 in addition to those already visualized in Fig 5. Genes are annotated 1.5kb upstream of their TSS and 1.5kb downstream of their transcription termination site. (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. Abbreviations of histological subtypes are explained in Fig 1A. Please note that 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. CL indicates cell lines.

(PDF)
S6 Fig. 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: S4 Table). 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. (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. Abbreviations of histological subtypes are explained in Fig 1A. Please note that 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. CL indicates cell lines.

S1 Table. List of DMPs resulting from pairwise comparison of GCT subtypes.

S2 Table. 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 Fig 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.

S3 Table. (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.

S4 Table. Merged known ICRs from literature with sources. Also see S6 Fig for a visual representation of the methylation status at the ICRs if covered by the 450K array (21/28).

Acknowledgments

The authors thank the Department of Bioinformatics, Erasmus MC, Rotterdam, for their support. They especially thank Ms Sylvia de Does and Mr Ivo Palli. Prof. dr. J.W. Oosterhuis, Department of Pathology, Erasmus MC, Rotterdam is kindly acknowledged for his critical reading of the manuscript.
Author Contributions
Conceived and designed the experiments: MR AG LD LL. Performed the experiments: AG. Analyzed the data: MR LD DT JR. Contributed reagents/materials/analysis tools: DT LD DK JR. Wrote the paper: MR DT AG LD DK JR LL.

References
1. Messerschmidt DM (2012) Should I stay or should I go: protection and maintenance of DNA methylation at imprinted genes. Epigenetics 7: 969–975. PMID: 22869105


