

# Pediatric Malignant Germ Cell Tumors Show Characteristic Transcriptome Profiles

Roger D. Palmer,<sup>1,2</sup> Nuno L. Barbosa-Morais,<sup>3</sup> Emma L. Gooding,<sup>1</sup> Balaji Muralidhar,<sup>1</sup> Claire M. Thornton,<sup>4</sup> Mark R. Pett,<sup>1</sup> Ian Roberts,<sup>1</sup> Dominik T. Schneider,<sup>5</sup> Natalie Thorne,<sup>3</sup> Simon Tavaré,<sup>3</sup> James C. Nicholson,<sup>2</sup> and Nicholas Coleman<sup>1</sup> on behalf of Children's Cancer and Leukaemia Group

<sup>1</sup>MRC Cancer Cell Unit, Hutchison/MRC Research Center; <sup>2</sup>Department of Paediatric Oncology, Addenbrooke's Hospital; <sup>3</sup>Computational Biology Group, Department of Oncology, University of Cambridge, Cambridge Research Institute, Cambridge, United Kingdom; <sup>4</sup>Department of Pathology, Institute of Clinical Science, Royal Group of Hospitals, Belfast, United Kingdom; and <sup>5</sup>Clinic of Pediatrics, Dortmund, Germany

## Abstract

**Malignant germ cell tumors (GCT) of childhood are rare and heterogeneous neoplasms thought to arise from primordial germ cells. They vary substantially in their natural history and show important clinical differences from their adult counterparts. To address the biological basis for these observations, we have undertaken a comprehensive analysis of global gene expression patterns in pediatric malignant GCTs and compared these findings with published data on adult testicular GCTs (TGCT). Our study included 27 primary tumors and assessed the principal malignant histologic types of pediatric GCT, yolk sac tumor (YST;  $n = 18$ ), and seminoma ( $n = 9$ ). Analysis of Affymetrix U133A GeneChip data was performed using the statistical software environment *R*, including gene set enrichment analysis, with cross-validation at the RNA and protein level. Unsupervised analysis showed complete separation of YSTs and seminomas by global gene expression profiles and identified a robust set of 657 discriminatory transcripts. There was no segregation of tumors of the same histology arising at different sites or at different ages within the pediatric range. In contrast, there was segregation of pediatric malignant GCTs and adult malignant TGCTs, most notably for the YSTs. The pediatric seminomas were significantly enriched for genes associated with the self-renewing pluripotent phenotype, whereas the pediatric YSTs were significantly enriched for genes associated with a differentiation and proliferation phenotype. We conclude that histologic type is the key discriminator in pediatric malignant GCTs and that the observed clinical differences between malignant GCTs of children and adults are mirrored by significant differences in global gene expression. [Cancer Res 2008;68(11):4239–47]**

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Accession numbers: The entire data set discussed is compliant to the Minimum Information About a Microarray Gene Experiment (MIAME; see checklist in Supplementary Information) and is deposited with the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/projects/geo/>), under accession number GSE 10615. The data analysis can be reproduced completely by following the information in the Supplementary Materials and Methods.

**Requests for reprints:** Roger D. Palmer, MRC Cancer Cell Unit, Hutchison/MRC Research Center, Box 197, Hills Road, Cambridge, CB2 0XZ, United Kingdom. Phone: 44-1223-763279; Fax: 44-1223-763284; E-mail: [rdp@hutchison-mrc.cam.ac.uk](mailto:rdp@hutchison-mrc.cam.ac.uk).

©2008 American Association for Cancer Research.  
doi:10.1158/0008-5472.CAN-07-5560

## Introduction

Germ cell tumors (GCT) are a rare and complex group of heterogeneous neoplasms that comprise both benign and malignant histologies (1). Despite their heterogeneity, they are all presumed to arise from totipotent primordial germ cells (PGC). Benign GCTs are called teratomas. Of the malignant GCT histologies, seminomatous tumors (testicular seminomas, ovarian dysgerminomas, and extragonadal germinomas) recapitulate the undifferentiated and pluripotent PGC phenotype, whereas nonseminomatous tumors [e.g., yolk sac tumors (YST)] display lineage specific differentiation. A mixture of histologic subtypes may be present within any single GCT, suggesting a close interrelationship between the different histologies and the cell of origin.

Many epidemiologic and clinical differences exist between GCTs arising in adulthood and those arising in childhood (defined in United Kingdom as 0–16 years of age), with tumors in adolescents showing a degree of overlap with GCTs in both age groups (1). Because of these differences, GCTs of identical histology are treated differently depending on patient age and country of residence. Despite overall success in managing children with GCTs, there are subgroups of patients in whom the prognosis is less favorable (1, 2). Moreover, adoption of adult derived chemotherapeutic schedules in the treatment of pediatric malignant GCTs has resulted in significant toxicity and long-term morbidity (3), which may be avoidable (2). It is therefore essential to obtain a greater understanding of the biology of pediatric malignant GCTs, not least so that children can be given the most appropriate treatment.

Our understanding of pediatric GCTs is limited, and investigations to date have largely focused on genetic and epigenetic changes (4–6). A recently proposed classification attempted to separate GCTs based upon present epidemiologic and biological evidence (7). In this system, GCTs of children with ages <5 years (typically teratomas or YSTs) are classified as type I tumors. The remaining GCTs of childhood, and most GCTs of adulthood, are deemed type II. An invariable feature of the type II tumors is reported to be gain of 12p, which occurs infrequently in type I tumors and, when present, is restricted to 12p13 (7).

Global gene expression studies have only been performed in adult testicular GCTs (TGCT). These suggested that expression profiles can distinguish malignant TGCTs from other adult malignancies and can separate TGCTs of different histologic types (8–12). These studies included cases with pure seminoma and YST histology. However, because none of the cases were from children, it is not known how representative these are of the biology of

**Table 1.** Clinicopathologic data for the 27 pediatric malignant GCTs analyzed

Sample	Histology	Tumor composition	Site	Age (y)	Gender	Stage	Genomic 12p gain
1	Seminoma (Dysgerminoma)	Pure	Ovary	10	Female	1	No
2	Seminoma (Dysgerminoma)	Pure	Ovary	12	Female	1	Yes
3	Seminoma (Dysgerminoma)	Pure	Ovary	12	Female	1	Yes
4	Seminoma (Dysgerminoma)	Pure	Ovary	12	Female	1	No
5	Seminoma (Dysgerminoma)	Pure	Ovary	13	Female	1	Yes
6	Seminoma (Dysgerminoma)	Pure	Ovary	13	Female	3	Yes
7	Seminoma (Germinoma)	Pure	Brain	10	Female	1	No
8	Seminoma (Germinoma)	Pure	Brain	16	Male	1	Yes
9	Seminoma (Dysgerminoma)	Within teratoma	Ovary	12	Female	2	No
10	YST	Pure	Testis	0	Male	1	No
11	YST	Pure	Testis	0	Male	1	No
12	YST	Pure	Testis	1	Male	1	No
13	YST	Pure	Testis	1	Male	1	Yes
14	YST	Pure	Testis	1	Male	1	No
15	YST	Pure	Testis	2	Male	1	Yes
16	YST	Pure	Testis	4	Male	1	Yes
17	YST	Pure	Ovary	0	Female	2	No
18	YST	Pure	Ovary*	9	Female	2	No
19	YST	Pure	Ovary	12	Female	1	No
20	YST	Pure	Ovary	12	Female	1	No
21	YST	Pure	Ovary	13	Female	2	Yes
22	YST	Pure	Ovary	13	Female	3	No
23	YST	Pure	Ovary	14	Female	1	Yes
24	YST	Pure	Ovary	14	Female	4	No
25	YST	Within teratoma	SCT	2	Female	4	Yes
26	YST	Within teratoma	SCT	3	Female	4	No
27	YST	Within teratoma	Brain	12	Male	1	Yes

NOTE: The malignant GCTs called seminomas in this study were ovarian dysgerminomas and extragonadal germinomas. No testicular seminomas were available, as these are extremely rare in the pediatric age range. 12p gain was determined by metaphase CGH, as reported previously (6). For all tumors arising within a teratoma, the sample was confirmed to be completely or predominantly (>90%) composed of the single malignant element on initial frozen section analysis.

Abbreviation: SCT, sacrococcygeal teratoma.

\*Contralateral streak ovary noted perioperatively.

pediatric malignant GCTs. More recently, limited expression data from cancer-focused arrays of 588 genes were published on GCTs that included five malignant cases from children with ages  $\leq 16$  years (13). This suggested that ovarian YSTs of adolescence (and adulthood) differed from ovarian dysgerminomas based on overexpression of genes in the Wnt/ $\beta$ -catenin pathway, a determinant of differentiation in embryonal carcinoma cell development *in vitro* and in adult malignant TGCTs (14, 15).

We have performed global gene expression analysis on 27 pediatric malignant GCTs. Such large tumor sets could be studied as samples were made available from the entire U.K. bank of pediatric GCTs (as well as cases from Germany). Our primary aim was to compare the two principal histologic types, seminomatous tumors (hereafter termed seminoma, irrespective of site of origin) and YSTs. All samples had previously been analyzed by metaphase comparative genomic hybridization (CGH) to identify genomic copy number imbalances (5). We interrogated our results, including a comparison with published adult TGCT transcriptome data (8) obtained using the same microarray platform, to identify differences in gene expression signature related to clinicopathologic variables, such as histologic type, genomic imbalances, site of origin, and patient age. Here, we present the first comprehensive

analysis of the transcriptome of malignant pediatric GCTs, which informs on the pathogenesis of these unique tumors.

## Materials and Methods

**Tumor samples and nomenclature.** Banked frozen tissue was provided by the Children's Cancer and Leukaemia Group (CCLG) from patients managed on U.K. pediatric treatment schedules (i.e., children of ages  $\leq 16$  y). Twenty-two samples were from U.K. patients treated on the extracranial GC8901 study, with three additional specimens from the intracranial study. We also obtained two unselected specimens from the German GPOH-MAKEI archive. Our study received Multicenter Research Ethics Committee approval (ref 02/4/071).

By the tumor-node metastasis staging system (1), 18 tumors were stage 1, four tumors were stage 2, two tumors were stage 3, and three tumors were stage 4. Two children died (one from disease, one unrelated to disease) and four relapsed (three of whom had stage 1 disease and received no primary chemotherapy), with follow-up ranging from 12 to 144 mo in disease-free cases. In each case, a histopathologic diagnosis had been provided after CCLG expert panel review. All specimens underwent frozen section analysis, with independent histologic review by three separate histopathologists, which confirmed the CCLG diagnosis. All cases used were completely or predominantly (>90%) composed of a single malignant element.

The histopathologic nomenclature used in this paper is based on the WHO criteria (16, 17). In particular, the term seminoma refers to all tumors with seminomatous histology, regardless of site (i.e., testicular seminoma, ovarian dysgerminoma, and extragonadal germinoma).

**RNA isolation and quality control.** Total RNA was extracted from tumor sections using the PolyTron 2100 homogenizer and TRIzol (Invitrogen Life Technologies). Glycoblu (Ambion, Inc.) was added to facilitate RNA visualization. Resultant total RNA was column cleaned using the RNeasy kit (Qiagen), with RNA quantity determined by spectrophotometry ( $A_{260}$ ) and RNA quality using the Bioanalyzer 2100 system (Agilent Technologies). The presence of mRNA was confirmed by gel electrophoresis of reverse transcription-PCR (RT-PCR) products for housekeeping genes *TBP* and *ACTB*.

**RNA amplification and microarray hybridization.** Double-stranded cDNA was synthesized from column cleaned total RNA using SuperScript II (Invitrogen Life Technologies), using the (dT)<sub>24</sub>-T7 promoter primer. Biotin-labeled cRNA was generated by Bioarray *in vitro* transcription (Enzo) and fragmented by metal-induced hydrolysis (Affymetrix). Fifteen micrograms of each probe was hybridized, washed, stained, and scanned by MRC GeneService using standard Affymetrix procedures with the GeneChip 3000 Scanner (Affymetrix). We used the U133A GeneChip (Affymetrix), composed of 22,283 probe sets corresponding to 13,042 gene names and 13,390 LocusLinks.

**Analysis and validation of microarray data.** Preprocessing of microarray data (including background correction and normalization) was performed applying the robust multiarray (RMA) method (18), included in the Bioconductor package *affy* for the statistical software environment *R* (19). Following the MIAME guidelines (20), raw (.CEL) data files have been deposited in the GEO repository (ref. 21; GEO accession number GSE 10615). The *B* statistic (22) and prediction analysis of microarrays (PAM;

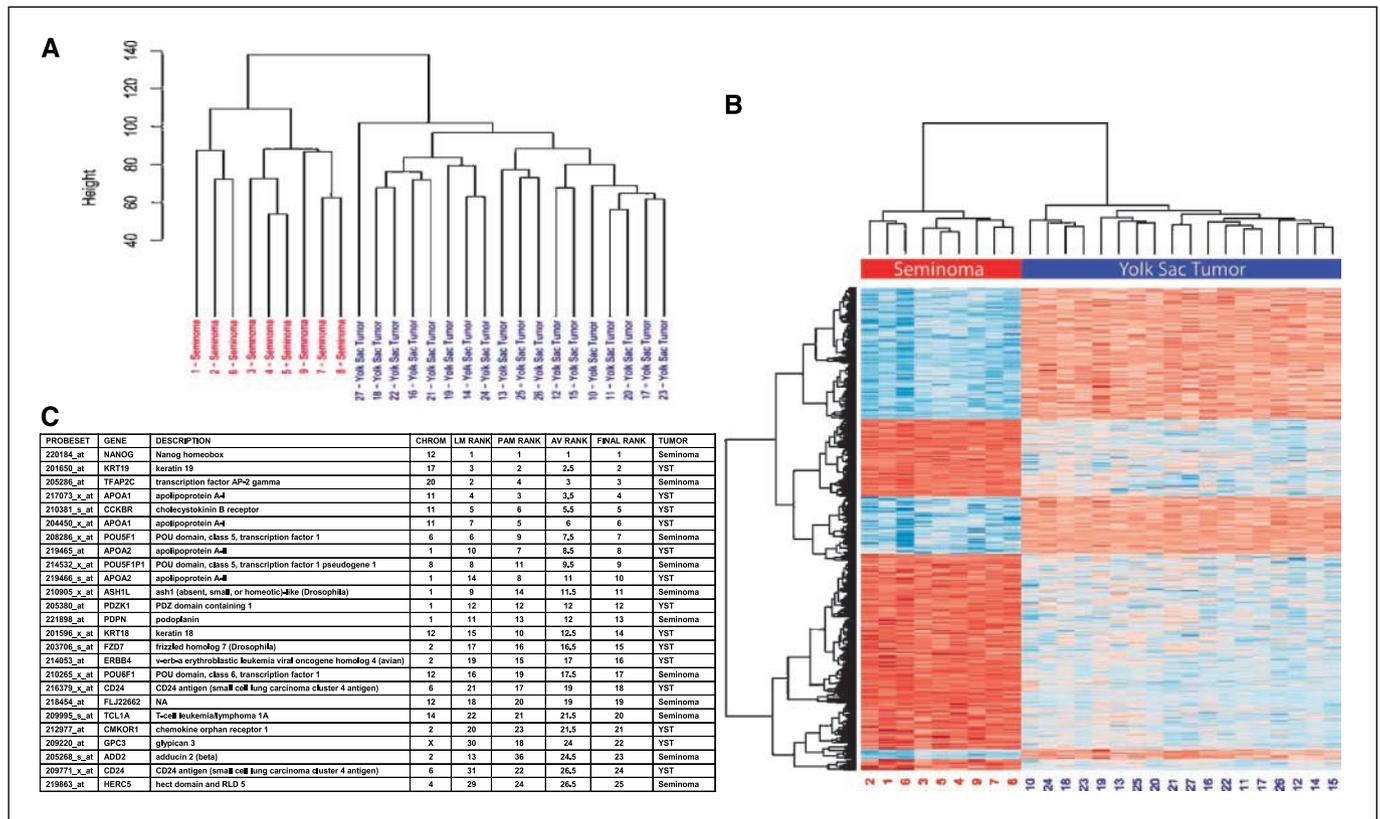
ref. 23) methods were used for identification of significantly differentially expressed genes and gene ranking, using Bioconductor packages *limma* and *pamr*, respectively. Gene set enrichment analysis (GSEA; ref. 24) was performed with the Bioconductor package *Category*.

**Real-time quantitative RT-PCR validation.** Real-time single-step RT-PCR (Qiagen) was performed on 2.5 ng DNase-treated (TURBO DNase-free, Ambion, Inc.), column cleaned total RNA from nine of the tumors. QuantiTect validated primer sets (Qiagen) for nine differentially expressed genes (*HERC5*, *APOA1*, *POU5F1*, *FOXA1*, *CDK2AP1*, *GPC3*, *CTSL2*, *MANIC1*, and *FLJ22662*) were selected and normalized to three housekeeping genes [*glyceraldehyde-3-phosphate dehydrogenase (GAPDH)*, *ACTB*, and *TBP*] using the Opticon DNA Engine (MJ Research, Inc.) and QuantiTect SYBR Green RT-PCR kit (Qiagen). Expression relative to 2.5 ng pooled normal human RNA (FirstChoice Human Total RNA Survey Panel, Ambion, Inc.) was calculated, using the Pfaffl equation, once primer efficiencies had been determined (25).

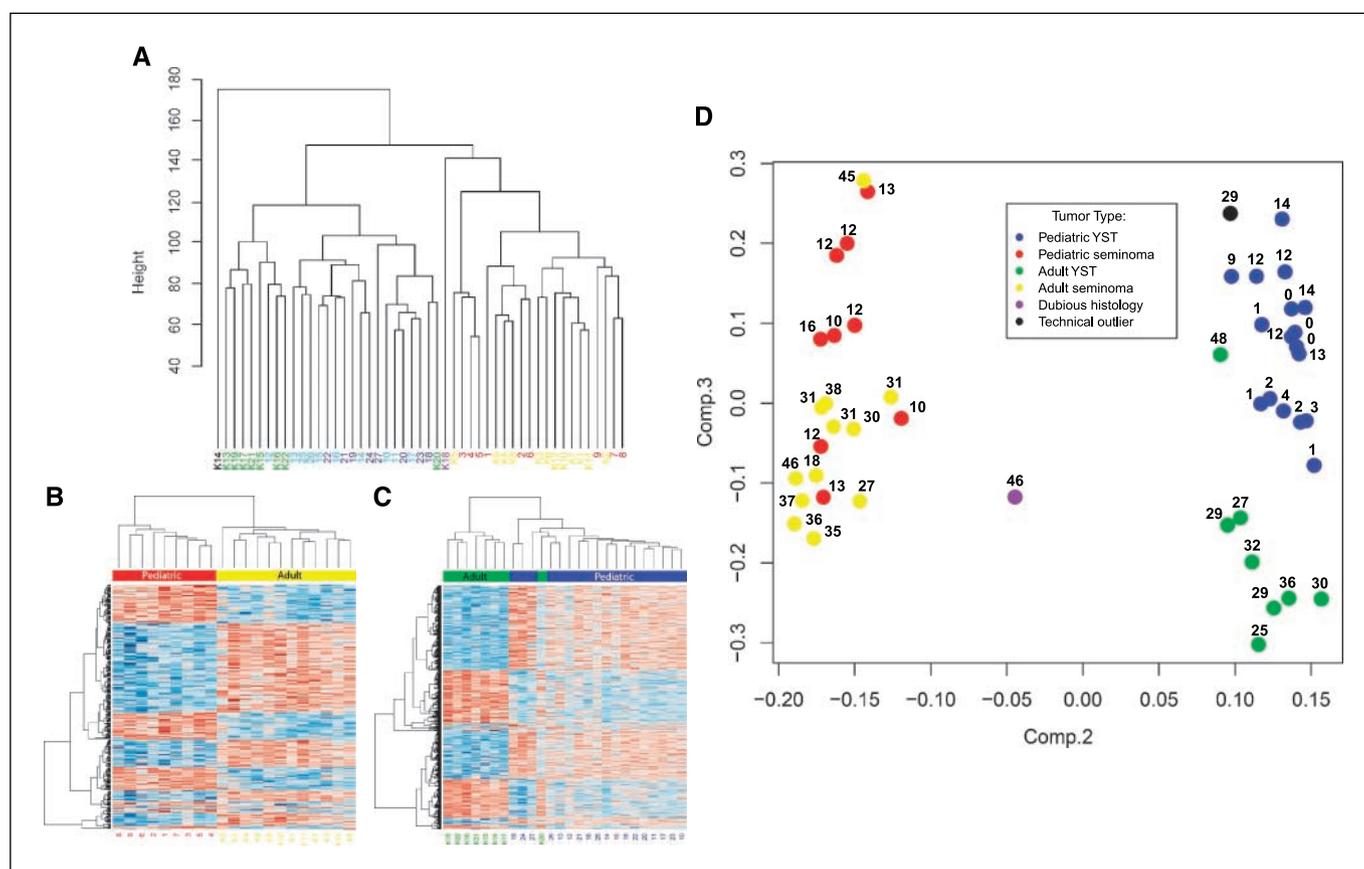
**Immunohistochemistry.** Validation of expression changes at the protein level was performed using immunohistochemistry, as previously described (26). We stained 5  $\mu$ m sections from a series of 10 unrelated paraffin-embedded cases of pediatric malignant GCTs, provided by the Addenbrooke's Hospital Tissue Bank. The primary antibodies were mouse monoclonal anti-AP2 $\gamma$  (dilution, 1:400), anti-TNNC1 (dilution, 1:250), anti-KRT19 (dilution, 1:300), and rabbit monoclonal anti-MAP4K1 (dilution, 1:150; all Abcam plc).

Results

**Characteristics of pediatric GCTs examined.** In total, 27 primary malignant GCTs from 27 different patients were analyzed (Table 1). There were 18 female and 9 male patients. Nine tumors



**Figure 1.** Analysis of pediatric GCT gene expression data. *A*, unsupervised cluster dendrogram for the 27 pediatric malignant GCTs after RMA preprocessing. *B*, heatmap for all 657 selected differentially expressed probes in the 27 GCTs (case numbers given at the base of the heatmap). In the heatmap, red represents relative overexpression and blue represents underexpression. *C*, top 25 discriminatory transcripts, with the average ranking (*AV RANK*) of the discriminatory probe sets obtained from Bayesian linear model (*LM*) and PAM analysis. *CHROM*, chromosomal location of the probe; *TUMOR*, histologic type of malignant GCT in which the gene in question is relatively overexpressed.



**Figure 2.** Comparison of gene expression in pediatric malignant GCTs versus adult malignant TGCTs. Tumors with a pure (i.e., not mixed) histologic diagnosis (GEO accession number GSE 3218; ref. 8) are compared with the 27 pediatric malignant GCTs. Cases from Korkola et al. are given the prefix K. The key is applicable to all panels. A, the unsupervised cluster dendrogram. B, comparison of the seminomas. C, comparison of the YSTs. D, PCA. The numbers above each dot refer to the age of the respective patient (in y).

were seminomas (six dysgerminomas, one dysgerminoma arising within an ovarian teratoma, and two intracranial germinomas), and 18 were YSTs (eight ovarian, seven testicular, two sacrococcygeal, and one intracranial). Twenty-two of the primary tumors were gonadal (15 ovarian and 7 testicular), and five were extragonadal. There was one potential case of gonadal dysgenesis (case 18), wherein a contralateral streak ovary was noted intraoperatively.

Seventeen tumors occurred in children with ages  $\geq 5$  years (14 ovarian, 3 intracranial) and included all nine seminoma cases. Ten tumors occurred in children with ages  $< 5$  years at diagnosis (seven testicular, two sacrococcygeal, one ovarian), and all of these were YSTs. This is not surprising, as malignant GCTs of other histologic types are exceedingly rare in this younger age group.

**Gene expression signatures.** Although seminomas and YSTs are presumed to arise from the same progenitor cell, the histologic differences between them were mirrored in their gene expression profiles. Unsupervised analysis of the entire probe set data showed complete separation of the histologies in the cluster dendrogram (Fig. 1A). Principal component analysis (PCA) showed that component 2 was responsible for the segregation by histologic type. We analyzed the data using two techniques (Bayesian linear model and PAM) to derive a final list of transcripts that robustly discriminated between the two tumor types. This list incorporated only transcripts shown to be significantly differentially expressed by both analysis techniques and, for which, there was no overlap in expression levels between the two histologic categories, irrespective

of site of origin. The minimum number of probe sets that were defined, using these strict criteria, as discriminating between seminomas and YSTs, was 657 (538 genes), with 266 overrepresented in seminomas relative to YSTs and 391 overrepresented in YSTs relative to seminomas (Fig. 1B). The top 25 transcripts are shown in Fig. 1C. The top-ranked gene ontology processes (27) for the 657 discriminatory transcripts included FZD signaling, Wnt receptor signaling, lipid metabolism, phosphatidylinositol binding, and protease/enzyme inhibitor activity (data not shown).

Clustering algorithms and PCA showed that there was no segregation of tumors of the same histology arising at different sites or at different ages within the pediatric range. When we compared our findings to published data obtained using the same platform from pure histology, malignant TGCTs of adults (10 YST, 12 seminoma; ref. 8; GEO accession number GSE 3218), we observed segregation of pediatric and adult tumors, most conspicuously for the YSTs (Fig. 2A–C). Principal component 2 again separated the samples on the basis of histology, although we note that one adult YST from the previous study (our reference K14) was technically poor and should be removed (see Supplementary Materials and Methods). Another adult YST (K18) did not clearly separate with the other adult YSTs on the basis of principal component 2, instead taking an intermediate position that we have also seen in preliminary work to be occupied by control tissues and teratomas (data not shown). This raises the possibility that the histologic diagnosis for case K18 was incorrect or that inadequate

amounts of the malignant element were present. Principal component 3 seemed to separate the tumors (especially the YSTs) on the basis of age group (pediatric versus adult), with the exception of one adult YST (K20), which intermingled with the pediatric cases (Fig. 2D).

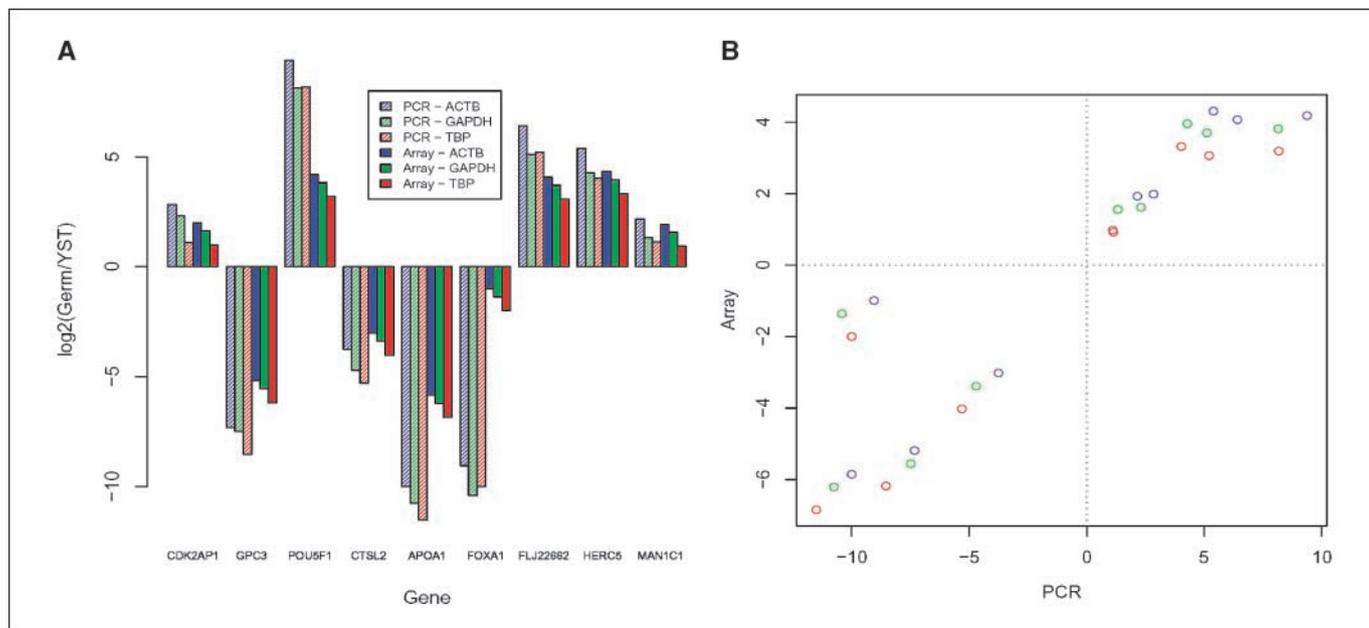
For seminomas, 380 transcripts discriminated between pediatric and adult tumors with a Bayesian linear model  $B$  statistic of  $>0$  (Fig. 2B), whereas 1,198 transcripts discriminated YSTs after removal of the two uncertain cases (K18 and K20; Fig. 2C). Sample expression plots for genes significantly overexpressed in pediatric seminomas (*CDC42BPA*) or adult seminomas (*RAC1*, *RPS4Y1*) and in pediatric YSTs (*LARPI*, *BMP4*) or adult YSTs (*CAPN7*) are shown in Supplementary Fig. S1. Of particular note, childhood testicular YSTs segregated away from the adult testicular YSTs (Fig. 2D), suggesting that site of origin, at least for the YSTs, was not responsible for the differences between the pediatric and adult cases. Similar direct comparison was not possible for the seminomas, as none of the childhood cases were testicular.

To confirm the separation of pediatric YSTs and adult testicular YSTs, all comparable data for the pediatric ( $n = 18$ ) and adult ( $n = 8$ ) YSTs were selected. This YST data were RMA normalized independently, and unsupervised clustering again showed segregation into two principal groups (Supplementary Fig. S2A). There were slight changes in the relationships between the YSTs compared with those observed in the overall analysis, with discrimination between pediatric and adult YSTs being improved by this approach (Supplementary Fig. S2A and B). The one adult YST (K20) still intermingled with the pediatric YSTs. Furthermore, there was clustering of all pediatric YSTs, irrespective of site (i.e., including testicular, ovarian, and extragonadal cases) and age (i.e., including cases from patients of ages  $<5$  and  $\geq 5$  years; Supplementary Fig. S2B).

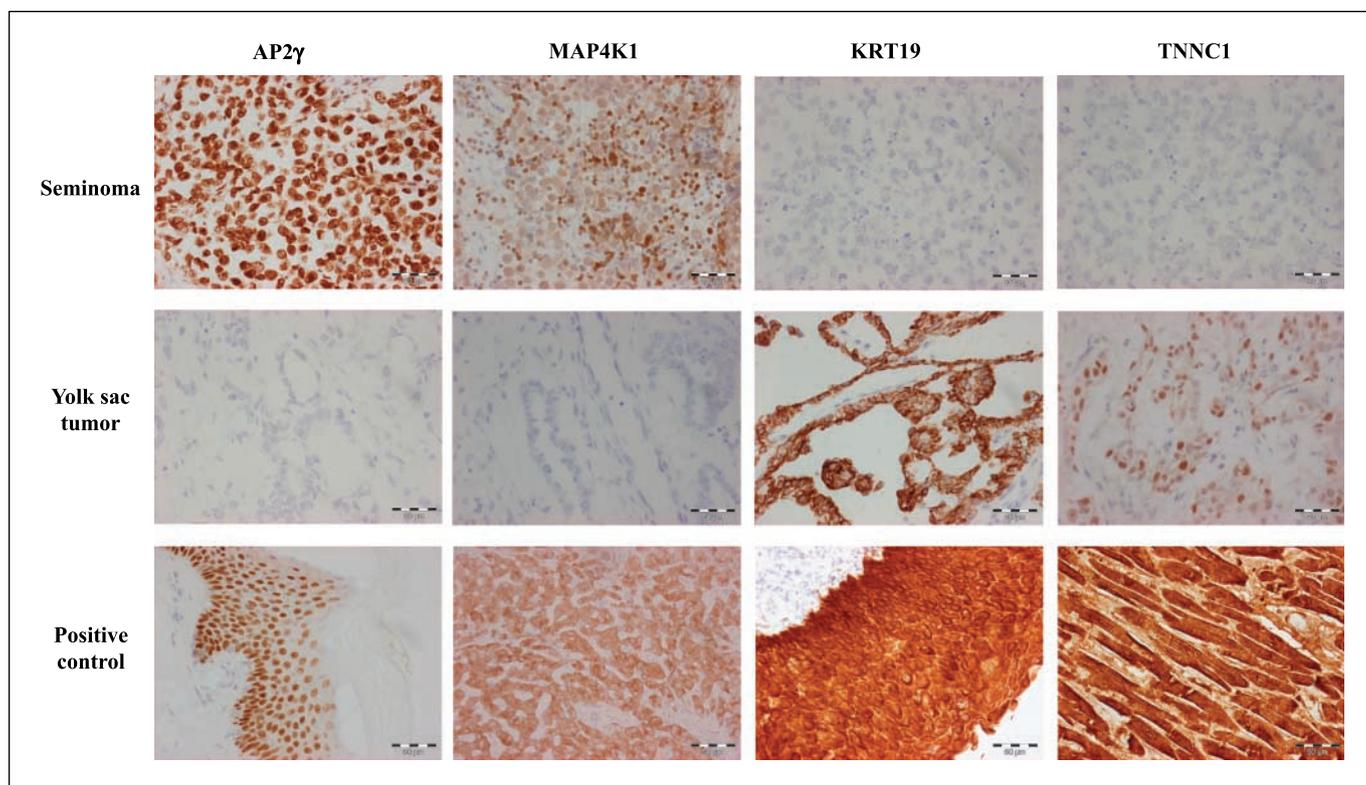
**Validation of microarray data.** Array data were confirmed by real-time quantitative RT-PCR analysis of nine genes (five overexpressed in YSTs and four overexpressed in seminomas), using total RNA from nine samples (five YSTs and four seminomas; Fig. 3). Additionally, overexpression of the proteins encoded by four of these genes (two overexpressed in each histologic grouping) was confirmed by immunohistochemistry using an unrelated series of formalin-fixed, paraffin-embedded pediatric GCTs (five seminomas, five YSTs; Fig. 4).

**Chromosome locations of differentially expressed genes.** GSEA (24) was performed for all 46 chromosomal arms using the entire gene list and reperformed for transcripts where the  $B$  statistic was  $>0$  (3,670 probe sets) to identify any additional enrichment. Seminomas were significantly more likely to express transcripts on 15q, 12p, 8q, and 1p, whereas YSTs were significantly more likely to express transcripts on 11q, 11p, 3p, 5q, 10p, and 13q (Supplementary Fig. S3). Similarly, analysis of transcript location by age revealed that malignant GCTs from children of ages  $<5$  years were enriched for transcripts on 11q, 3p, 11p, and 13q, whereas malignant GCTs from children of ages  $\geq 5$  years were enriched for transcripts on 19q, 2p, 1p, 15q, and 12p (Supplementary Fig. S4A). Because none of the cases in patients of ages  $<5$  years were seminomas, the age-related analysis was repeated for the YSTs only. The entire gene list was used in this analysis, as only 61 probe sets that separated YSTs on the basis of age had a  $B$  statistic of  $>0$ . YSTs in children of ages 5 to 16 years were enriched for transcripts on 19, Xp, Xq, 2q, 7q, and 22q, whereas YSTs in children of ages  $<5$  years were enriched for transcripts on 12q, 2p, 3p, and 18p (Supplementary Fig. S4B).

Of our final list of 538 discriminatory genes, 19 reside on chromosome 12p. Of these, 16 were overexpressed in seminomas, including the highly discriminatory genes *FLJ22662*, *Nanog*, and



**Figure 3.** Real-time PCR validation of microarray data for selected gene transcripts. *A*, expression data relative to each of the three housekeeping genes (*ACTB*, *GAPDH*, and *TBP*), as determined by PCR or from microarray data. *B*, comparison of expression levels determined by PCR to expression levels determined by microarray. Where expression levels were determined in seminomas, they are given a positive value, and where determined in YSTs, they are given a negative value. *FOXA1* is the only transcript noted to differ from the linear association between microarray and PCR, although both methods show that the gene is overexpressed in YSTs relative to seminomas. The discrepancy is attributable to the fact that 4 of the 11 *FOXA1* probes on the Affymetrix U133A microarray not do correspond to any sequence match within the *FOXA1* annotation.



**Figure 4.** Immunohistochemical validation of expression of four proteins. Consistent with transcript levels, AP-2 $\gamma$  and MAP4K1 were relatively overexpressed in seminomas, whereas KRT19 and TNNC1 were relatively overexpressed in YSTs. The positive control tissues were (from left to right) normal epidermis, normal hepatocytes, normal cervical squamous epithelium, and normal cardiac muscle.

*Cl2orf35*, whereas three were overexpressed in YSTs. The only chromosomal locations of overexpressed genes that were exclusive to any particular histology were 10p (eight genes), 18p (eight genes), and Y (one gene) for YSTs and 21p (three genes) for seminomas. In tumors showing 12p copy number gain by metaphase CGH (compared with tumors without 12p gain; Table 1; ref. 5), there was enrichment for genes on 12p, as well as on 12q and 1p (Supplementary Fig. S5).

**Genes overexpressed in pediatric seminomas.** The 266 transcripts overexpressed in pediatric seminomas relative to pediatric YSTs included the known pluripotency genes *Nanog* (rank 1), *POU5F1* (*OCT3/4*; ranks 3 and 4), and *UTF* (rank 31); *TFAP2C* (*ERF1*; rank 2), an estrogen induced transcription factor that may stimulate or repress gene transcription and have crucial, but distinct, functions in embryologic development and malignant transformation (28); *ASH1L* (*YY1API*; rank 5), a putative transcription factor that may be involved in chromatin remodeling and cell-cell adhesion (29); and *PDPN* (*Aggrus*; rank 6), a trans-membrane glycoprotein involved in cell-cell adhesion and migration (30). We observed that an increased number of zinc finger genes were overexpressed in seminomas, consistent with previous findings (31).

By GSEA of the differentially expressed transcripts, wherein the *B* statistic was >0, and of the entire gene set, the most significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in seminomas included genetic information processing (notably ribosome genes), cellular processes (hematopoietic cell lineage, natural killer cell mediated, antigen processing, T-cell and B-cell receptor signaling), and signal transduction pathways (Fig. 5

and Supplementary Fig. S6). The latter included Toll-like receptor and phosphatidylinositol signaling, as well as the Janus-activated kinase (JAK)/STAT pathway (including *PIAS3* and *PIK3CD*) and the canonical Wnt/ $\beta$ -catenin pathway (*FZD10*, *MAP4K1*, *WNT2B*; Supplementary Fig. S7).

**Genes overexpressed in pediatric YSTs.** The 391 transcripts overexpressed in pediatric YSTs relative to pediatric seminomas included the known tumor marker  $\alpha$ -fetoprotein (AFP), as well as differentiation genes, such as *KRT19* (rank 1) and *KRT8* (rank 8). Other overexpressed transcripts included *APOA1* (ranks 2 and 4) and *APOA2* (rank 5 and 6), which have a role in lipid transport (32); *CCKBR* (rank 3), potentially involved in tumor growth (including leukemia, pancreatic, and gastrointestinal tumors) in response to gastrin via the phosphatidylinositol-calcium second messenger system (33); and *PDZK1* (rank 7), involved in ion transport (notably chloride via the CFTR), lipid and peptide handling, and potentially, therefore, peptide-like drug resistance (34).

By GSEA, the most significantly enriched KEGG pathways (Fig. 5 and Supplementary Fig. S6) included cellular communication (cell communication, focal adhesion, adherens junction, and tight junction) and signaling pathways [e.g., extracellular matrix receptor interaction and transforming growth factor- $\beta$  (TGF- $\beta$ )/BMP signaling]. Other signaling pathways with overrepresentation in the YSTs included calcium signaling, the canonical (*FZD2*, *FZD4*, *FZD5*, *FZD7*, *SFRP1*, *ILK*, *CDH2*, *TCF7L1*, *TCF7L2*) and noncanonical Wnt/ $\beta$ -catenin pathways including members of the planar cell polarity (PCP) pathway, notably *WNT11*, *DAAMI*, *RHOA*, and *ROR2*, and the Wnt/calcium release pathway, notably *CAMKII*; Supplementary Fig. S6.

**Discussion**

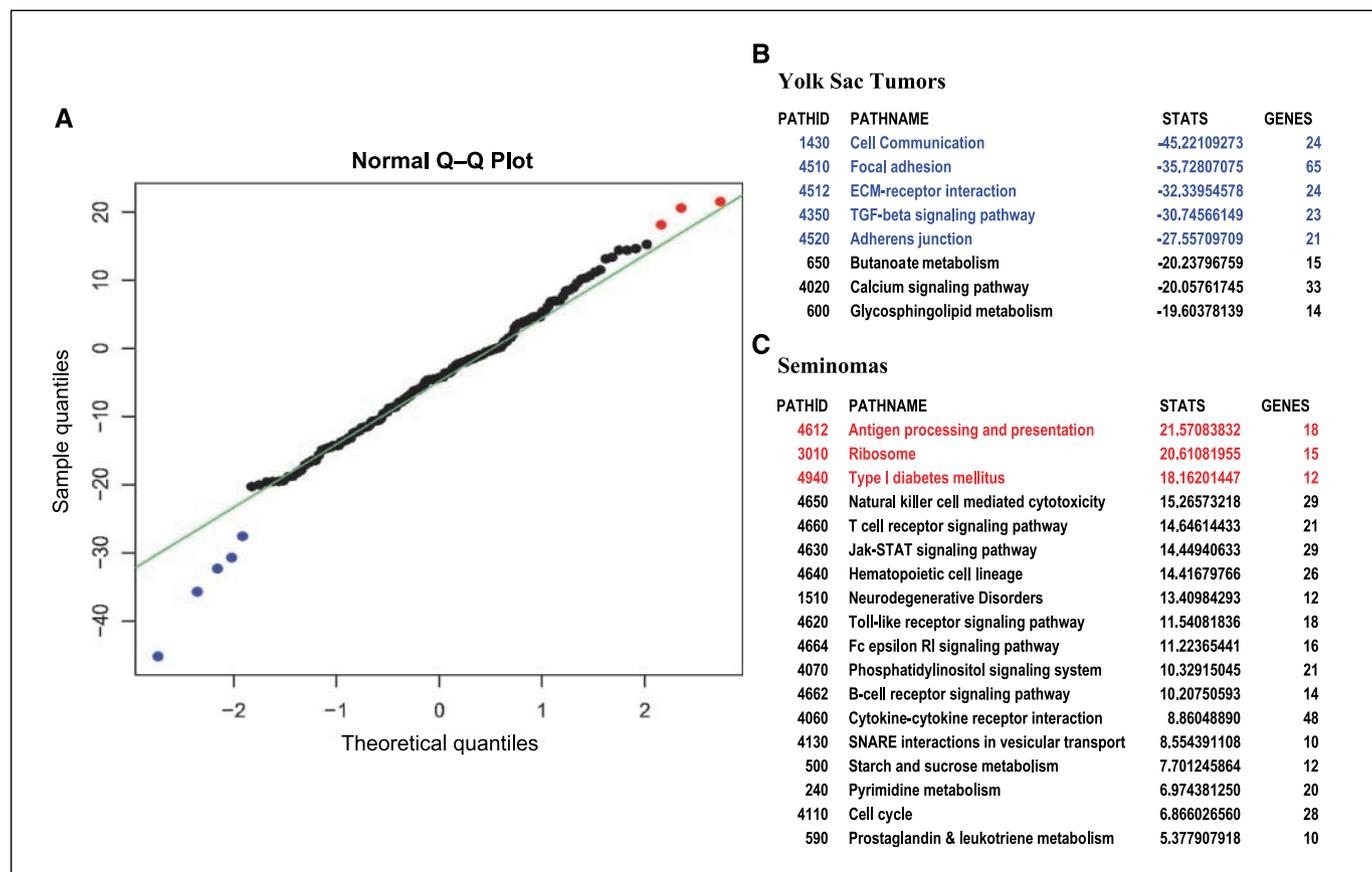
Our data indicate that the key discriminating factor between malignant GCTs arising across the pediatric age range, irrespective of the site of origin, is the histologic diagnosis. We have determined a set of genes responsible for the divergence of YSTs and seminomas that is highly robust, thereby permitting interpretation of function and downstream hypothesis testing with confidence. Adult TGCTs are similarly discriminated by histology, but their expression profiles are far from identical to those of pediatric GCTs (8–11).

Several top-ranked genes that discriminate on the basis of histology in adult malignant TGCTs were associated with the histologic equivalent in the dataset for pediatric malignant GCTs. For instance, top-ranked YST genes in adult TGCT series (8–11) included *GPC3*, *BMP2*, and *VTN*, which were respectively ranked 15, 28, and 36 by the Bayesian linear model in the pediatric YST dataset. Similarly, top-ranked seminoma genes in adult TGCT series included *Nanog*, *POU5F1*, and *PDPN*, which ranked 1, 3, and 6 in the pediatric seminomas (Fig. 1C and Supplementary Fig. S1). Such consistency did not apply to all genes, however. For example, *C5* ranked in the top 10 adult YST genes (11), but was lowly ranked (rank 166) in pediatric YSTs. Moreover, highly ranked pediatric YST genes, such as *ERBB4* (rank 10) and *CD24* (rank 12), did not appear in adult YST discriminatory gene lists.

These observations are reflected in the unsupervised clustering dendrogram (Fig. 2A) and PCA plot (Fig. 2D), which show that malignant GCTs most resemble tumors of the same histologic type, regardless of age, but that there is also separation by age category (i.e., pediatric or adult) within the seminomas and, especially, the YSTs (Fig. 2D). Indeed, we were able to derive gene lists that discriminate between adult and pediatric tumors within each histologic group (Fig. 2B and C).

As expected, immunohistochemical markers currently used for the histopathologic diagnosis of GCTs were represented in our gene expression lists. For example, diagnosis of seminoma commonly involves immunohistochemical detection of OCT3/4 (35) and KIT (CD117; ref. 36), both of which are overexpressed in the seminomas in our dataset (ranks 6 and 1857, respectively, in the pediatric YST versus seminoma list). Diagnosis of YSTs frequently involves immunohistochemical detection of AFP and GPC3 (37), which respectively ranked 125 and 30 in the discriminatory list. These observations support our own immunohistochemical validation data, as presented in Fig. 4.

Our data provide support for clinical evidence, suggesting that pediatric and adult malignant GCTs differ biologically. The pediatric seminomas in our study were all from ovarian or extragonadal sites, and we could not exclude the possibility that differences from adult testicular seminomas may, at least in part, be site related. Our



**Figure 5.** KEGG pathways significantly enriched in genes differentially expressed in YSTs and seminomas. The pathways were determined from differentially expressed transcripts identified using the Bayesian linear model ( $B$  statistic  $> 0$ ). A, the quantile-quantile (Q-Q) plot of the actual (sample) values and expected values (based on a theoretical normal distribution) of the statistical score for the 191 KEGG pathways. Enrichment toward seminomas is shown as positive values, whereas enrichment toward YSTs is shown as negative values. The green line passes through the first and third quartiles. The most deviant top score pathways are highlighted in red for seminomas and blue for YSTs. The statistical score and number of genes in each top-ranked pathway are given in B (YSTs) and C (seminomas).

comparison of YSTs alone shows that pediatric and adult cases cluster separately in unsupervised analysis and that all pediatric YSTs group together regardless of tumor site or patient age (Supplementary Fig. S2). In particular, testicular YSTs of adults and children were separated on clustering analysis and PCA. As expected for pediatric YSTs, the testicular cases occurred exclusively in children of ages <5 years. However, it is noteworthy that one adult testicular YST clustered with the pediatric cases in unsupervised analysis. The clinical and biological behavior of such adult cases will be of interest. In view of the lack of published global gene expression data for adult ovarian or extragonadal malignant GCTs, it is not possible to determine the timing of the potential biological shift between childhood and adult-type patterns, although this might occur during puberty and be hormone driven. As CGH profiles in childhood ovarian GCT have been reported to resemble those in adults (38), further study of adolescent tumors is warranted. Moreover, despite having access to the entire United Kingdom archive of pediatric malignant GCTs, plus additional German cases, only five extragonadal tumors were available, and investigation of more examples of such tumors is required.

Differences between adult and pediatric malignant GCTs are supported by the genomic differences between the two groups. Isochromosome 12p is present in 80% of adult TGCTs, and gain on 12p is invariably present in the remaining cases (39). Adult malignant GCTs at other sites (e.g., ovarian and intracranial tumors) seem to show near-identical genetic aberrations (38, 40). However, 12p gain is found with a lower incidence in adolescence and is much less common in infancy (<5 years of age; refs. 4, 5). We observed enrichment for transcripts on 12p in GCTs of children of ages  $\geq 5$  years compared with children of ages <5 years (Supplementary Figs. S3–S5). However, this can be attributed to increased expression of 12p genes in seminomas compared with YSTs, as there was no enrichment for 12p genes in YSTs of children of ages  $\geq 5$  years compared with YSTs of children of ages <5 years (Supplementary Fig. S4). Taken together, our findings do not support the claim that 12p gain has a ubiquitous role in the biology of malignant GCTs in children of ages  $\geq 5$  years. Indeed, our overall expression data does not clearly support the proposed separation of pediatric malignant GCTs into type I and type II tumors (7) and would also be consistent with an age-related cutoff, as used in previous studies (4).

When compared with YSTs, seminomatous tumors largely recapitulated the features of self-renewing, pluripotent human embryonic stem (hES) cells, which by microarray analysis are characterized by activation of signaling pathways, including the Wnt/ $\beta$ -catenin, fibroblast growth factor, and leukemia inhibitory factor (JAK/STAT) pathways (41), together with up-regulation of cell cycle control genes, DNA repair enzymes, zinc finger-specific transcription repressors, and ribosomal proteins and down-regulation of proapoptotic genes (42). Whereas such similarities have previously been described for adult seminomas (11, 15), we now show that this also applies to pediatric seminomas, irrespective of patient age (at least within the range of 10–16 years) and tumor site.

Relative to seminomas, YSTs overexpressed genes activated during differentiation, including early markers of hES cell differentiation, such as *GATA3* and *GATA6*; keratin genes, such as *KRT19*, *KRT18* and *KRT8*; and markers of extraembryonic endodermal differentiation, such as *VTN*, *BMP2*, and *FOXA2*. There was also evidence of overexpression of genes in the Wnt/ $\beta$ -catenin pathway, in concordance with previous studies of nonseminomatous malignant GCTs (13, 14). In comparison to the seminomas, we

observed overexpression in YSTs of genes from both canonical and noncanonical (PCP and Wnt/calcium release) Wnt/ $\beta$ -catenin pathways, as well as overexpression of *RYK*, which could potentially cause Wnt activation independently of *FZD* or act as a coreceptor (43). Whereas the Wnt/ $\beta$ -catenin pathway is essential to maintain the pluripotency phenotype in hES cells (41, 44) and seminomatous tumors (8, 12), overexpression of more Wnt pathway genes in YSTs may lead to differentiation, as well as deregulated growth. Our study is the first to show that overexpression of the Wnt/ $\beta$ -catenin pathway genes occurs in YSTs across the pediatric age spectrum. YSTs also showed enrichment for cell-to-cell interaction and other signaling pathways (Fig. 5), with overexpression of genes in the TGF- $\beta$ /BMP pathway almost exclusively seen in YSTs. Whereas tight junction genes, notably *GJA1* and *CLDN6*, have previously been reported to be important in defining the stem cell phenotype (44), we observed a significant enrichment for overexpression of tight junction, focal adhesion, and adherens junction genes in YSTs, consistent with the differentiated state of the tumor cells.

From a clinical perspective, the discriminatory genes that we have identified may be of value in the diagnosis or differential diagnosis of GCTs, particularly in small biopsy samples. We did not look at associations with clinical outcome in pediatric malignant GCTs, as these tumors are largely curable with platinum-based combination chemotherapy. Our data provide biological evidence supporting clinical differences between pediatric and adult malignant GCTs. Resistance to cisplatin chemotherapy is extremely rare in pediatric malignant GCTs, unlike in adults (45). Moreover, whereas less toxic carboplatin chemotherapy is associated with an inferior outcome in adult malignant TGCTs (46), this has not been observed for pediatric malignant GCTs. Our evidence of differences in global gene expression supports the view that chemotherapeutic treatments can be different between children and adults without adversely affecting outcome (1, 2).

Several of our findings also have direct clinical implications. Overexpression of *CD99* (*MIC2*) in YSTs suggests its use as a specific marker of stromal cells (47) when investigating gonadal tumors may be inappropriate. *TSPY*, a possible marker of invasive potential in adult TGCTs (48), was only overexpressed in three pediatric GCTs, all from male patients, raising the possibility of a gender bias to *TSPY* expression levels (Supplementary Fig. S8). Cardiac troponin C (*TNNC1*), which we showed (and confirmed by immunohistochemistry; Fig. 4) to be overexpressed in YSTs, may be a useful adjunct to AFP in the diagnosis and monitoring of GCTs in childhood, particularly during investigation of sacrococcygeal teratomas in infants, where serum AFP levels are typically raised (49). Moreover, as cardiac troponin is released from cells, there may be potential to use it as a serum marker for YST diagnosis in children. Finally, YSTs showed overexpression of both *ERBB4* (*HER4*; rank 16) and *ERBB2* (*HER2*; rank 1952), raising the possibility of heterodimerization of the proteins, which has been described in childhood medulloblastomas (50) and may represent a cell surface therapeutic target.

In conclusion, we have shown that pediatric malignant GCTs differ at the gene expression level on the basis of histology and also differ from their adult TGCT histologic counterparts. Consequently, an understanding of the biology of pediatric malignant GCTs cannot be inferred from adult GCT studies alone. Our findings also support the use of different therapeutic strategies in pediatric compared with adult malignant GCTs, thereby reducing the life-long morbidity associated with treatment in childhood without affecting overall survival (2).

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Acknowledgments

Received 9/20/2007; revised 2/22/2008; accepted 3/26/2008.

**Grant support:** Medical Research Council, Cancer Research UK, CLIC Sargent, Parthenon Trust, and Addenbrooke's Charities. D.T. Schneider was supported by a

Max-Eder grant from German Cancer Aid. S. Tavaré is a Royal Society Wolfson Merit Award holder.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank John Brown and Pam Stacey for technical assistance; the Children's Cancer and Leukaemia Group, especially Dr Juliet Hale and Professor Richard Grundy, for their support; and Dr. James Korkola and Dr. Jane Houldsworth of Memorial Sloan-Kettering Cancer Center for providing the clinicopathologic data relating to the adult TGCTs.

## References

- Palmer RD, Nicholson JC, Hale JP. Management of germ cell tumours in childhood. *Curr Paediatr* 2003;13:213–20.
- Mann JR, Raafat F, Robinson K, et al. The United Kingdom Children's Cancer Study Group's second germ cell tumor study: carboplatin, etoposide, and bleomycin are effective treatment for children with malignant extracranial germ cell tumors, with acceptable toxicity. *J Clin Oncol* 2000;18:3809–18.
- Marina N, Chang KW, Malogolowkin M, et al. Amifostine does not protect against the ototoxicity of high-dose cisplatin combined with etoposide and bleomycin in pediatric germ-cell tumors: a Children's Oncology Group study. *Cancer* 2005;104:841–7.
- Schneider DT, Schuster AE, Fritsch MK, et al. Genetic analysis of childhood germ cell tumors with comparative genomic hybridization. *Klin Padiatr* 2001;213:204–11.
- Palmer RD, Foster NA, Vowler SL, et al. Malignant germ cell tumours of childhood: new associations of genomic imbalance. *Br J Cancer* 2007;96:667–76.
- Schneider DT, Schuster AE, Fritsch MK, et al. Multi-point imprinting analysis indicates a common precursor cell for gonadal and nongonadal pediatric germ cell tumors. *Cancer Res* 2001;61:7268–76.
- Oosterhuis JW, Looijenga LH. Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer* 2005;5:210–22.
- Korkola JE, Houldsworth J, Chadalavada RS, et al. Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with *in vivo* differentiation of human male germ cell tumors. *Cancer Res* 2006;66:820–7.
- Juric D, Sale S, Hromas RA, et al. Gene expression profiling differentiates germ cell tumors from other cancers and defines subtype-specific signatures. *Proc Natl Acad Sci U S A* 2005;102:17763–8.
- Sugimura J, Foster RS, Cummings OW, et al. Gene expression profiling of early- and late-relapse nonseminomatous germ cell tumor and primitive neuroectodermal tumor of the testis. *Clin Cancer Res* 2004;10:2368–78.
- Korkola JE, Houldsworth J, Dobrzynski D, et al. Gene expression-based classification of nonseminomatous male germ cell tumors. *Oncogene* 2005;24:5101–7.
- Skotheim RI, Monni O, Mousset S, et al. New insights into testicular germ cell tumorigenesis from gene expression profiling. *Cancer Res* 2002;62:2359–64.
- Fritsch MK, Schneider DT, Schuster AE, Murdoch FE, Perlman EJ. Activation of Wnt/ $\beta$ -catenin signaling in distinct histologic subtypes of human germ cell tumors. *Pediatr Dev Pathol* 2006;9:115–31.
- Walsh J, Andrews PW. Expression of Wnt and Notch pathway genes in a pluripotent human embryonic carcinoma cell line and embryonic stem cell. *APMIS* 2003;111:197–210; discussion-1.
- Sperger JM, Chen X, Draper JS, et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A* 2003;100:13350–5.
- Calaminus G, Patte K. Germ Cell Tumors in Children and Adolescents. In: Agarwal BR, Perilongo G, Rogers P, Strahlendorf C, Eden OB, editors. *Education Book*, 37th meeting of SIOP. Eindhoven (the Netherlands): International Society of Paediatric Oncology; 2005.
- Eble JN, Sauter G, Epstein JI, Sesterhenn IA, editors. *World Health Organization Classification of Tumours. Pathology and genetics of tumours of the urinary system and male genital organs*. Lyon: IARC Press; 2004.
- Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–64.
- Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71.
- Barrett T, Suzek TO, Troup DB, et al. NCBI GEO: mining millions of expression profiles-database and tools. *Nucleic Acids Res* 2005;33:D562–6.
- Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology [electronic resource]* 2004;3:Article3.
- Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002;99:6567–72.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- Herdman MT, Pett MR, Roberts I, et al. Interferon- $\beta$  treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants. *Carcinogenesis* 2006;27:2341–53.
- Scott IS, Morris LS, Bird K, et al. A novel immunohistochemical method to estimate cell-cycle phase distribution in archival tissue: implications for the prediction of outcome in colorectal cancer. *J Pathol* 2003;201:187–97.
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–9.
- Hoei-Hansen CE, Nielsen JE, Almstrup K, et al. Transcription factor AP-2 $\gamma$  is a developmentally regulated marker of testicular carcinoma *in situ* and germ cell tumors. *Clin Cancer Res* 2004;10:8521–30.
- Nakamura T, Blechman J, Tada S, et al. huASH1 protein, a putative transcription factor encoded by a human homologue of the *Drosophila* ash1 gene, localizes to both nuclei and cell-cell tight junctions. *Proc Natl Acad Sci U S A* 2000;97:7284–9.
- Sonne SB, Herlihy AS, Hoei-Hansen CE, et al. Identity of M2A (D2-40) antigen and gp36 (Aggrus, T1A-2, podoplanin) in human developing testis, testicular carcinoma *in situ* and germ-cell tumours. *Virchows Arch* 2006;449:200–6.
- Port M, Schmelz HU, Stockinger M, et al. Gene expression profiling in seminoma and nonseminoma. *J Clin Oncol* 2005;23:58–69.
- Mastroianni A, Bellati C, Facchetti G, Oldani S, Franzini C, Berrino F. Increased plasma HDL-cholesterol and apo A-I in breast cancer patients undergoing adjuvant tamoxifen therapy. *Clin Biochem* 2000;33:513–6.
- Stubbs M, Khan K, Wickremasinghe RG, Ganeshguru K, Caplin ME. CCK2 gastrin receptor as a potential target for therapy in leukaemia cell lines. *Oncol Rep* 2005;14:1055–8.
- Rossmann H, Jacob P, Baisch S, et al. The CFTR associated protein CAP70 interacts with the apical Cl<sup>-</sup>/HCO<sub>3</sub> exchanger DRA in rabbit small intestinal mucosa. *Biochemistry* 2005;44:4777–87.
- Jones TD, Ulbright TM, Eble JN, Baldrige LA, Cheng L. OCT4 staining in testicular tumors: a sensitive and specific marker for seminoma and embryonal carcinoma. *Am J Surg Pathol* 2004;28:935–40.
- Izquierdo MA, Van der Valk P, Van Ark-Otte J, et al. Differential expression of the c-kit proto-oncogene in germ cell tumours. *J Pathol* 1995;177:253–8.
- Zynger DL, Dimov ND, Luan C, Teh BT, Yang XJ. Glypican 3: a novel marker in testicular germ cell tumors. *Am J Surg Pathol* 2006;30:1570–5.
- Kragerud SM, Szymanska J, Abeler VM, et al. DNA copy number changes in malignant ovarian germ cell tumors. *Cancer Res* 2000;60:3025–30.
- Chaganti RS, Houldsworth J. Genetics and biology of adult human male germ cell tumors. *Cancer Res* 2000;60:1475–82.
- Rickert CH, Simon R, Bergmann M, Dockhorn-Dworniczak B, Paulus W. Comparative genomic hybridization in pineal germ cell tumors. *J Neuropathol Exp Neurol* 2000;59:815–21.
- Rao M. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev Biol* 2004;275:269–86.
- Bhattacharya B, Miura T, Brandenberger R, et al. Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* 2004;103:2956–64.
- Cadigan KM, Liu YI. Wnt signaling: complexity at the surface. *J Cell Sci* 2006;119:395–402.
- Skottman H, Mikkola M, Lundin K, et al. Gene expression signatures of seven individual human embryonic stem cell lines. *Stem Cells* 2005;23:1343–56.
- Schneider DT, Wessalowski R, Calaminus G, et al. Treatment of recurrent malignant sacrococcygeal germ cell tumors: analysis of 22 patients registered in the German protocols MAKEI 83/86, 89, and 96. *J Clin Oncol* 2001;19:1951–60.
- Horwich A, Shipley J, Huddart R. Testicular germ-cell cancer. *Lancet* 2006;367:754–65.
- Hoei-Hansen CE, Sehested A, Juhler M, et al. New evidence for the origin of intracranial germ cell tumours from primordial germ cells: expression of pluripotency and cell differentiation markers. *J Pathol* 2006;209:25–33.
- Kersemakers AM, Honecker F, Stoop H, et al. Identification of germ cells at risk for neoplastic transformation in gonadoblastoma: an immunohistochemical study for OCT3/4 and TSPY. *Hum Pathol* 2005;36:512–21.
- Blohm ME, Vesterling-Horner D, Calaminus G, Gobel U.  $\alpha$ 1-fetoprotein (AFP) reference values in infants up to 2 years of age. *Pediatr Hematol Oncol* 1998;15:135–42.
- Gilbertson RJ, Perry RH, Kelly PJ, Pearson AD, Lunec J. Prognostic significance of HER2 and HER4 coexpression in childhood medulloblastoma. *Cancer Res* 1997;57:3272–80.