

# Frequent Nonrandom Activation of Germ-Line Genes in Human Cancer

Michael Koslowski,<sup>1</sup> Carolin Bell,<sup>1</sup> Gerhard Seitz,<sup>2</sup> Hans-Anton Lehr,<sup>3</sup> Klaus Roemer,<sup>4</sup> Horst Müntefering,<sup>5</sup> Christoph Huber,<sup>1</sup> Ugur Sahin,<sup>1</sup> and Özlem Türeci<sup>1</sup>

<sup>1</sup>Department of Internal Medicine III, Johannes-Gutenberg University, Mainz; <sup>2</sup>Institute of Pathology, Bamberg Hospital, Bamberg; <sup>3</sup>Institute of Pathology, Johannes-Gutenberg University, Mainz; <sup>4</sup>Institute of Virology, University of Saarland, Homburg/Saar; and <sup>5</sup>Institute of Pediatric Pathology, Johannes-Gutenberg University, Mainz, Germany

## ABSTRACT

The growing class of cancer/germ-line genes is characterized by a unique expression pattern with transcription restricted to germ cells and cancer cells. It is not known which fraction of germ-line genes is ectopically activated in tumor cells and whether this fraction displays common features as compared with strictly germ-line genes remaining silent in cancer. Using an unbiased genome-wide scanning approach, representative samples of both cancer/germ-line genes as well as strictly germ-line-specific genes were determined. Comparative analysis disclosed highly significant diametric characteristics for these two categories of genes with regard to sex specificity, developmental stage of physiological expression during gametogenesis, chromosomal localization, and epigenetic regulation of expression. Our findings provide class predictors for germ cell-specific gene activation in cancer. The identification of highly congruent expression patterns in cancer and in DNA methyltransferase-deficient cells suggests an underlying common epigenetic mechanism for activation of germ-line genes in cancer.

## INTRODUCTION

Cloning strategies for the systematic identification of cancer-derived gene products recognized specifically by the immune system of the tumor-bearing host disclosed a multitude of novel tumor antigens during the last decade (1–3). The primary intention was to find the antigens of spontaneously occurring immunorecognition of tumors as potential targets for immunotherapeutical interventions. However, this did not only open new avenues for specific cancer immunotherapy (4–6) but also provided unexpected insights into cancer-related genotypic and phenotypic alterations (7, 8).

One of the intriguing observations was that germ cell-specific genes, which are undergoing stringent transcriptional repression in adult somatic tissues, might become ectopically activated in cancer cells. Profiling of human tissues by reverse transcription-PCR (RT-PCR) fails to detect these gene products in any normal tissue except for testis but in various frequencies in a broad spectrum of human tumors (9, 10). As the first monoclonal antibodies against some of these genes became available, germ cells were identified as the expressing cell population within testicular tissue (11, 12).

Although these so-called cancer/testis or cancer/germ-line (CG) antigens are used as vaccines in clinical trials (13), they are still enigmatic. The growing number of CG genes being found by different methods implies a general mechanism of activation. Which fraction of germ-line genes escapes silencing in tumors and what distinguishes them from those germ-line genes remaining stringently repressed throughout malignant transformation remains to be defined.

Recently, we reported a powerful strategy based on genome-wide data mining, which allows the systematic extraction of germ-line-specific genes out of the entirety of full-length genes in the public

domain. Assessing this unbiased and representative sample of the germ-line-specific transcriptome for aberrant expression in tumors, we discovered novel CG genes (14). Moreover, this provided us with an authenticated negative list of strictly germ cell-specific genes not activated in cancer.

We report here that by comparative analysis of these two sets we unraveled highly significant characteristics distinguishing robustly CG genes from strictly germ-line-specific genes. Our findings demonstrate that ectopic gene activation in cancer does not occur randomly and suggest an underlying common mechanism for their activation.

## MATERIALS AND METHODS

**Tissues and Cell Lines.** Adult tumor and normal tissue samples were obtained in an anonymous fashion from the tumor tissue bank of the program project (SFB432) supported by the Deutsche Forschungsgemeinschaft, derived from surplus tissue from the routine surgical pathology service. Fetal testis (24 gestational weeks) and fetal ovary (18 and 20 gestational weeks) from abortions were stored at  $-80^{\circ}\text{C}$  until use. Human brain total RNA was purchased from Clontech. To induce DNA demethylation, phytohemagglutinin-activated peripheral blood mononuclear cells were cultured in RPMI 1640 + 10% FCS supplemented with 2 or 10  $\mu\text{mol/L}$  5'-aza-2-deoxycytidine (Sigma) for 72 h. HCT 116 cell lines knocked out for either *DNMT1* or *DNMT3b* as well as the double knockout cell line were kindly provided by Bert Vogelstein.

**ENTREZ Keyword Search.** As previously described (14), GenBank was searched for genes annotated to be specifically expressed in testicular tissue using the ENTREZ Search and Retrieval System at National Center for Biotechnology Information.<sup>6</sup> Nucleotide sequence files were generated by phrase searching for “testis-specific,” “germ-specific,” “spermatogonia-specific,” “spermatocyte-specific,” or “sperm-specific,” respectively. The search was restricted to a particular subset of data by setting limits to “*Homo sapiens*” for organism and “mRNA” for molecule, excluding all expressed sequence tags (ESTs), genomic sequences, and partial coding sequences. Because the same gene may be represented in GenBank with different names the sequence homology-searching program BLASTN<sup>7</sup> was run sequentially for each nucleotide sequence against all of the human nucleotide sequences. The homology stringency was set high to detect identical rather than homologous sequences without erroneous loss of hits based on sequencing errors.

The search criterion relies on the annotation provided by the submitting authors. Such uncurated annotation-based data needs additional authentication. As a second filter, electronic Northern was performed for all clones obtained by keyword search. Briefly, BLAST search of DNA sequences of interest against EST database at National Center for Biotechnology Information was performed.<sup>8</sup> The source libraries of ESTs homologous to genes of interest allow making inferences on the respective gene's tissue distribution. Thereby, genes were identified that are highly homologous to EST from testis-derived libraries but not to ESTs from nontesticular normal tissues, except placenta, ovary, and fetal tissues. It was taken into consideration that several cDNA libraries in the public domain are not properly annotated (15).<sup>9</sup> Sixty-five of the 150 sequences matched to multiple ESTs from normal nontesticular tissues and were therefore excluded from additional analysis.

Received 4/5/04; revised 6/10/04; accepted 7/7/04.

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.

**Note:** U. Sahin and Ö. Türeci contributed equally to this work.

**Requests for reprints:** Özlem Türeci, Department of Internal Medicine III, Johannes Gutenberg University, Obere Zahlbacherstr. 63, D-55131 Mainz, Germany. Phone: 49-6131-3933488; Fax: 49-6131-3933364; E-mail: tureci@mail.uni-mainz.de.

©2004 American Association for Cancer Research.

<sup>6</sup> Internet address: <http://www.ncbi.nlm.nih.gov/Entrez>.

<sup>7</sup> Internet address: <http://ncbi.nlm.nih.gov/blast>.

<sup>8</sup> Internet address: <http://www.ncbi.nlm.nih.gov/BLAST>.

<sup>9</sup> Internet address: <http://www.fau.edu/cmhb/publications/cancergenes6.htm>.

Table 1 Characteristics of germ cell-specific genes

Gene	Chromosome	Sex specificity	Developmental stage	Induction by DAC	Activation in DNMT1/3b k.o.
<b>Germ-line genes</b>					
DAZ	Yq11	M	Premeiotic	–	–
DMRT1	9p24	M/F	Premeiotic	–	nd
VASA	5q11	M/F	Premeiotic	+	nd
BOLL	2q33	M/F	Meiosis I	+	–
BPESC1	3q23	M/F	Meiosis I	–	nd
ADAM30	1p13	M	Meiosis II/Postmeiotic	–	nd
ADAM18	8p22	M	Meiosis II/Postmeiotic	–	nd
ADAM3A	8p21	M	Meiosis II/Postmeiotic	–	nd
AKAP4	Xp11	M	Meiosis II/Postmeiotic	–	nd
CAST	5q14	M	Meiosis II/Postmeiotic	–	nd
CST8	20q11	M	Meiosis II/Postmeiotic	–	–
CXORF2	Xq28	M	Meiosis II/Postmeiotic	+	nd
CYLC2	9q31	M	Meiosis II/Postmeiotic	–	–
MCSP	1q21	M	Meiosis II/Postmeiotic	–	nd
NYD-SP26	4q13	M	Meiosis II/Postmeiotic	–	nd
NYD-SP27	12p12	M	Meiosis II/Postmeiotic	–	nd
ODF1	8q22	M	Meiosis II/Postmeiotic	+	–
PP2Bs1	9q22	M	Meiosis II/Postmeiotic	–	nd
PRM1,2	16p13	M	Meiosis II/Postmeiotic	–	–
SPACA1	Xq27	M	Meiosis II/Postmeiotic	–	nd
SPAM1	7q31	M	Meiosis II/Postmeiotic	–	nd
TCPI0	6q27	M	Meiosis II/Postmeiotic	–	–
TETM4	11q12	M	Meiosis II/Postmeiotic	–	nd
TNP2	16p13	M	Meiosis II/Postmeiotic	–	nd
TSG	19p13	M	Meiosis II/Postmeiotic	–	nd
TSKS	19q13	M	Meiosis II/Postmeiotic	–	–
UBQLN3	11p15	M	Meiosis II/Postmeiotic	–	nd
VCX	Xp22	M	Meiosis II/Postmeiotic	–	–
<b>Cancer germ-line genes</b>					
BAGE	4p16	M	Premeiotic	+	+
GAGE2,3,4	Xp11	M/F	Premeiotic	+	+
HOM-Tes-85	Xq24	M/F	Premeiotic	+	nd
MAGEA2,3,4	Xq28	M/F	Premeiotic	+	+
MAGEB1,4	Xq21	M/F	Premeiotic	+	nd
MAGEC1	Xq26	M/F	Premeiotic	+	nd
MORC	3q13	M/F	Premeiotic	+	+
SAGE	Xq28	M/F	Premeiotic	+	nd
SPANXC	Xq27	M/F	Premeiotic	+	nd
TPTE	21p11	M/F	Premeiotic	+	+
TSPY	Yp11	M	Premeiotic	+	nd
XAGE1	Xp21	M/F	Premeiotic	+	+
BRDT	1p31	M/F	Meiosis I	+	+
SPO11	20q13	M/F	Meiosis I	–	–
SYCP1	1p13	M/F	Meiosis I	+	+
ADAM2	8p11	M	Meiosis II/Postmeiotic	–	–
LDHC	11p15	M	Meiosis II/Postmeiotic	–	–
NY-ESO-1	Xq28	M	Meiosis II/Postmeiotic	+	nd
SGY-1	19q13	M	Meiosis II/Postmeiotic	–	nd
TPX1	6p21	M	Meiosis II/Postmeiotic	–	nd

NOTE. Germ-line genes and CG genes were sorted according to sex specificity (F = female, M = male), gametogenic stage of expression initiation, transcriptional activation in somatic cells upon genomic hypomethylation induced either chemically by treatment with DAC or by deletion of DNA methyltransferases DNMT1 and DNMT3b. For breakdown in this table, members of a family of closely related genes, e.g., MAGE-A2, MAGE-A3, and MAGE-A4, were summarized.

Abbreviations: DAC, 5'-aza-2-deoxycytidine; nd, not determined.

**cDNA xProfiler.** The cDNA xProfiler tool of the Cancer Genome Anatomy Project at National Center for Biotechnology Information<sup>10</sup> compares gene expression between two pools (A and B) of cDNA libraries where each pool can be either a single library or several libraries. For a gene to be present in a library pool, there must be at least one EST sequence found in the UniGene cluster for that gene.

The search options for Pool A and Pool B were set to “*Homo sapiens*” for Organism and “All EST libraries” for Library Group to search all cDNA libraries in dbEST. The minimum number of sequences per library was set to “1000” in order not to lower the significance of the results. The search option for tissue preparation and library protocol was adjusted to default settings “all.” All cDNA libraries prepared from normal testis tissue matching the search option settings were assigned to Pool A, excluding mixed tissue libraries. For Pool B, all cDNA libraries prepared from normal tissues, except testis, ovary, and whole body fetus were selected. The Expression Profiler Set-Up was carefully reviewed for correct library selection in Pool A and Pool B, respectively. Because most of the libraries available for this purpose are not normalized and numbers of sequenced clones are not representative enough, there is a stochastically bias for false positives with this tool. PubMed was accessed for

additional confinement, disclosing published data proving expression in non-testicular normal organs for 66 of the 154 obtained sequences.

**RNA Isolation and RT-PCR.** For those genes predicted to be germ cell-specific by the above mentioned bioinformatical tools, tissue distribution was assessed by RT-PCR. First expression analysis was performed in all types of normal tissues available (spleen, thymus, mammary gland, liver, ovary, prostate, lymph node, uterus, kidney, thyroid, small intestine, colon, adrenal gland, esophagus, lung, skin, activated phytohemagglutinin-activated peripheral blood mononuclear cells, brain, and testis). Only those genes with confirmed silencing in nongonadal normal tissues were subjected to expression profiling in a panel of >100 different tumor samples (17 lung cancer, 20 breast cancer, 20 colon cancer, 16 melanoma, 8 prostate cancer, 7 renal cell carcinoma, 7 ovarian cancer, 6 cervical cancer, and 4 thyroid cancer) and 14 cancer cell lines (8 melanoma and 6 lung cancer).

Total cellular RNA was extracted from frozen tissue specimens using RNeasy Mini Kit (Qiagen), primed with a dT<sub>18</sub> oligonucleotide and reverse transcribed with Superscript II (Invitrogen-Life Technologies, Inc.) according to the manufacturer’s instructions. Integrity of the obtained cDNA was tested by amplification of p53 transcripts in a 30-cycle PCR (sense, 5'-CGT-GAGCGCTTCGATGTTCCG-3'; antisense, 5'-CCTAACCAGCTGCCA-ACTGTAG-3'; annealing temperature 67°C).

<sup>10</sup> Internet address: <http://cgap.nci.nih.gov/Tissues/xProfiler>.

PCR primers were designed and used for specific amplification from first-strand cDNA stocks. To exclude false positive PCR products due to contaminating genomic DNA in the RNA preparation or existence of pseudogenes, the individual gene-specific primer sets were designed to span exon/intron boundaries and quality controlled by PCR reactions using either DNA or not reverse-transcribed RNA as template. Twenty-nine genes were excluded from analysis because they had either processed pseudogenes or were intronless. For PCR analysis of individual gene transcripts, 0.5  $\mu$ L of first-strand cDNA were amplified with transcript-specific oligonucleotides (Invitrogen) using 1 unit of HotStarTaq DNA Polymerase (Qiagen) in a 30- $\mu$ L reaction according to the manufacturer's instructions. The primer sequences and the respective annealing temperatures are given in supplementary Table 1. In each experiment, a template-free negative control and testis as positive control were included. In case of discordant RT-PCR data controversial to studies published by other groups, we designed a second independent primer set and confirmed our results.

## RESULTS AND DISCUSSION

We predicted a germ cell-specific set of genes expressed in testis by genome-wide data mining of public sequence databases (Fig. 1). For this purpose, hierarchical keyword search of GenBank, which had previously been successful in a prototype version (14), was combined with digital cDNA library subtraction to mitigate biases exerted by either of these procedures alone. We focused on known full-length genes and open reading frames, excluding ESTs and unclassified partial sequences to reduce complexity. Thus,  $\sim$ 12,000 gene equivalents (GenBank release March 2001), which according to recent estimations may cover one third of the human genome, were intro-

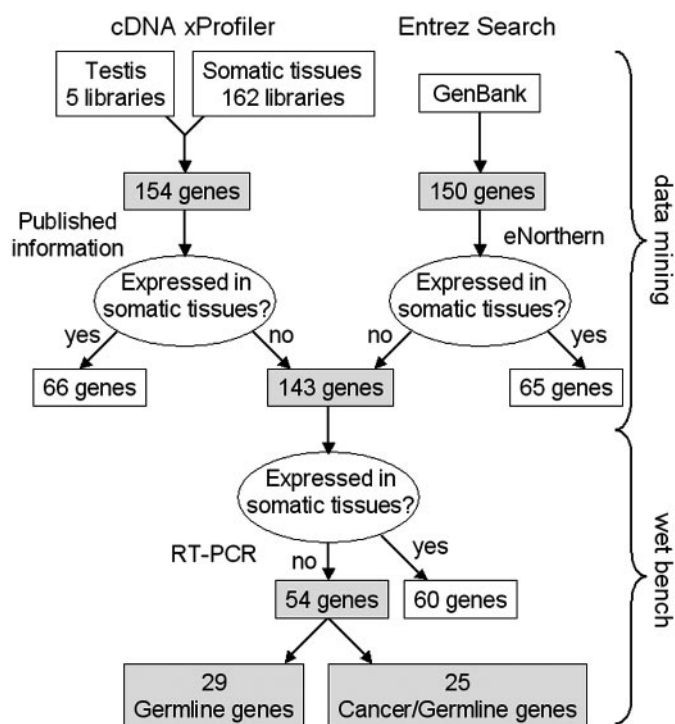


Fig. 1. Strategy for the identification of germ cell-specific genes. Data mining was conducted by using cDNA xProfiler and ENTREZ keyword search as described in Materials and Methods. The combination of these data mining tools predicted a total of 268 testicular genes as being germ cell-specific (36 genes were picked up by both queries). Supporting Table 2A provides a list of all genes assessed along this analysis process. This group of genes was additionally validated by electronic Northern and literature search as described in Materials and Methods, reducing the number of predicted germ cell-specific genes to 143. A total of 114 of these genes was subjected to RT-PCR analysis in testis and in a large panel of normal somatic tissues confirming germ-cell specificity for 54 of them. The remaining 29 genes (supplementary Table 2B) were excluded from analysis because they had either processed pseudogenes or were intronless and, as such, incompatible with the primer design strategy (supplementary Table 2C).

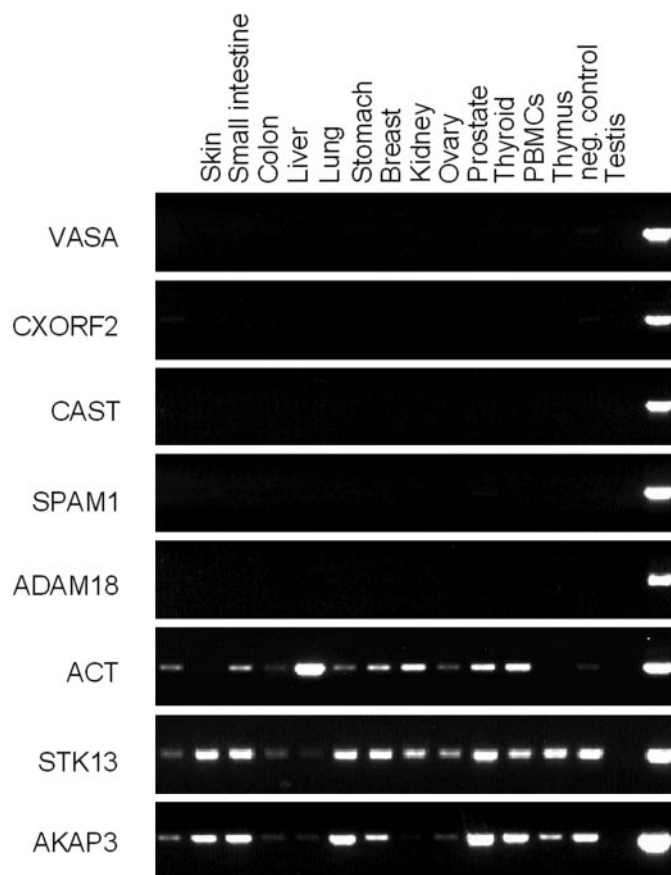


Fig. 2. RT-PCR analysis of genes in different tissues. Expression of genes was tested in a large panel of normal somatic tissues, as well as testis, to extract those with authentically germ cell-specific expression silenced in nontesticular normal tissues. As exemplified here for *ACT*, *STK13*, and *AKAP3*, a number of genes previously annotated as germ cell or testis specific were found to be expressed in multiple normal tissues and were therefore excluded from additional analysis.

duced into the outlined data mining approach. This data mining approach yielded 143 genes with predicted germ cell-specific expression. A total of 114 of these genes of interest was subjected to wet bench experiments for validation (supplementary Table 2, B and C). Only 54 genes could be confirmed as germ cell-specific genes by RT-PCR expression analysis in a broad panel of normal tissue specimen. The remaining 60 GOI were in fact frequently expressed in nontesticular tissues as well (Fig. 2) and thus excluded from additional analysis.

Of the 54 GOI with experimentally confirmed germ-cell specific expression, 19 gene products turned out to be known and previously characterized CG genes. Expression of the remaining 35 GOI was analyzed by RT-PCR in a panel of cDNAs from  $>$ 100 different tumor samples. Six genes qualified as novel CG genes in that they were expressed in varying frequencies in different tumor entities. These genes were described in detail elsewhere (14). Twenty-nine GOI were not expressed in tumors and therefore were classified as germ-line genes.

Seeking to understand why some but not all germ cell-specific genes are ectopically activated in cancer, we additionally characterized the 27 germ-line genes and 25 CG genes (Table 1).

First, we tested for sex specificity assessing their expression in human spermatogenesis and oogenesis. In oogenesis, meiosis is initiated from week 11 after fertilization (16), but oocytes may remain arrested in first meiotic prophase for  $>$ 40 years (refs. 17, 18; Fig. 3A). Therefore, we included fetal ovary in our study to cover premeiotic

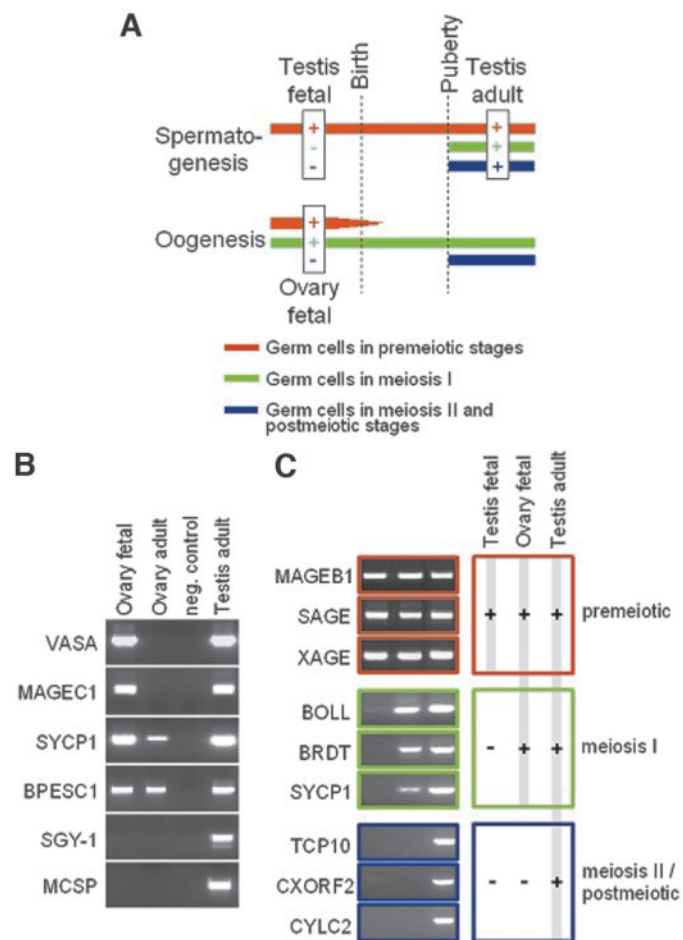


Fig. 3. Assessment of sex-specificity and allocation of transcriptional activation of genes to different developmental stages of gametogenesis. A. Male and female gametogenesis display major differences in the timing of developmental stages. In view of these developmental differences, sex-specificity of germ-line-specific genes was dissected by analyzing expression in adult testis (representing all stages of spermatogenesis), fetal ovary (representing early stages of oogenesis until meiosis I), and adult ovary (representing meiotic stages of oocyte development). B. With inclusion of fetal ovary in this study and use of quality controlled adult ovarian tissue, several germ cell-specific genes previously reported as male specific (e.g., *MAGE* family members) were detected in female germ cells. C. Expression analysis of genes in fetal testis, fetal ovary and adult testis were used for categorization of genes in premeiotic, early meiotic and meiotic/postmeiotic stages.

gene expression as well. The study of adult ovary tissue is affected by the lack of germ cells because of the death of >99% of the oogonia and oocytes between the fifth month of gestation and the time of puberty. To ensure representation of germ cells in the bulk ovary specimen, we chose *SYCP1* (19) as a surrogate marker for these cells and restricted our investigations to *SYCP1*-positive samples of adult ovary tissue. Surprisingly, by doing so, we detected expression of several CG genes in ovarian tissue samples, which had previously been reported not to be expressed in ovary (Fig. 3B; Table 1).

Next, we addressed the question in which stage of gametogenesis these genes were activated. Studies with pure populations of human gametogenic cells in different developmental stages are hampered by availability of specimen and technical feasibility. Therefore, we resorted to using bulk tissue but exploiting the differences in timing between male and female gametogenesis (Fig. 3A). In contrast to oogonia, spermatogonia enter meiosis not before puberty and after a long phase of mitotic arrest. Furthermore, although fetal testis harbors only premeiotic germ cells (20), adult testis contains all stages of differentiation from spermatogonia to postmeiotic spermatids. We deduced that genes belonging to different stages of germ cell devel-

opment should have distinct but predictable expression patterns in fetal testis, fetal ovary, and adult testis (Fig. 3A). Analysis of the expression of genes that have been reported to be associated with the different gametogenic stages (*DMRT1* and *DAZ* for premeiotic stage, synaptonemal complex proteins *SYCP1* and *Spo11* for meiosis, and spermatid proteins *TNP2*, *ODF1*, and *SPACA1* for postmeiotic stage) confirmed the credibility of this approach to distinguish the distinct developmental stages (data not shown).

Remarkably, we found that the majority of CG genes are activated in the early premeiotic stages of gametogenesis, whereas expression of germ-line genes is induced in late meiotic and postmeiotic stages of gametogenesis (Fig. 3C; Table 1).

DNA methylation is involved in gene regulation and constitutes an efficient mechanism of gene silencing (21–23). Furthermore, genomic demethylation has been documented to lead to the expression of several germ cell-specific genes in tumors (24–26). To assess the role of DNA methylation in the transcriptional control of the genes under investigation, we tested their induction of expression in somatic cells upon treatment with the methylation-inhibiting drug 5'-aza-2-deoxycytidine. Expression of the majority of the CG genes (80%), but only a small fraction of germ-line genes (18%) could be induced that way (Fig. 4A; Table 1). To substantiate this finding and to gain information on the DNA methyltransferases involved, we studied expression of selected genes in an isogenic set of human colorectal cancer cell lines in which the gene for either one of the DNA methyltransferases *DNMT1* or *DNMT3b*, or both genes together, had been disrupted by targeted homologous recombination. Expression of most of the premeiotically active, germ cell-specific genes was induced in the *DNMT1/3b* double-knockout cells but not in either of the single knockouts (Fig. 4B; Table 1). Our findings are in accordance with a recent study demonstrating a cooperation between DNA methyltransferases 1 and 3b in gene silencing (27).

Gene location of tissue-specific or coregulated genes may cluster within the genome. A particular role in gametogenesis as well as in ectopic activation in tumors has been assigned to the sex chromosomes (28–30). Our studies revealed that indeed CG genes cluster on the X chromosome, whereas germ-line genes are more evenly distrib-

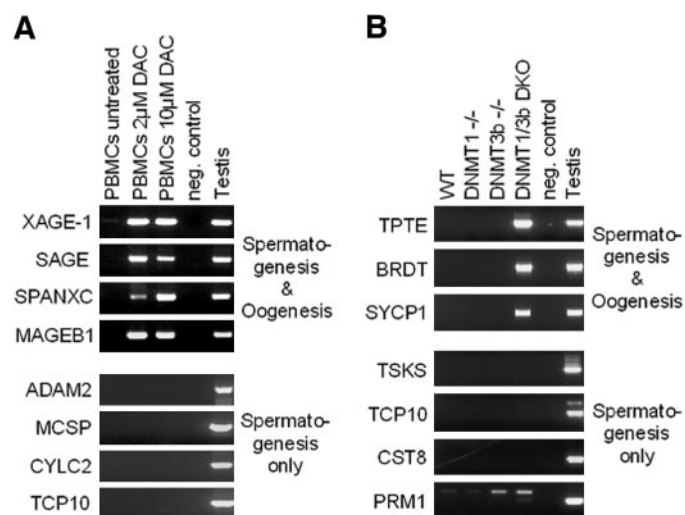


Fig. 4. Effect of genomic demethylation on transcriptional regulation of germ cell genes. A. Peripheral blood mononuclear cells (PBMCs) were cultured for 72 h either with or without the methylation-inhibiting drug 5'-aza-2-deoxycytidine (DAC) and were subsequently tested for the expression of germ cell-specific genes. B. Expression of genes in parental HCT 116 cell line was compared with expression in variants deleted for either *DNMT1* (*DNMT1*-), *DNMT3b* (*DNMT3b*-), or both methyltransferases (*DKO*). The larger band in *PRM1* (protamin)-PCR results from amplification of genomic DNA containing a small intron.

uted throughout the genome (Table 1). Furthermore and remarkably, X-chromosomal localization was observed in particular for those CG genes, which are premeiotically expressed in both sexes and are inducible by demethylation (Fig. 5).

A summary of all of the data obtained (Fig. 6; Table 1) disclosed that the analyzed features significantly cluster diametrically into the predefined categories of germ-line genes and CG genes.

Although the genes studied here definitely do not represent the complete specific transcriptome of gametogenesis, they constitute an unbiased and representative random sample out of the entire germ cell-specific transcriptional program, therefore, allowing for general inferences. The highly significant and partly unexpected correlations we disclosed in this sample (Table 1) have several implications.

First, our systematic and unbiased approach disclosed that ectopic expression of germ cell-specific genes in cancer does not represent occasional events of dysregulation. In fact, nearly half of the genes analyzed were expressed in cancer cells.

Second, we defined criteria associated with those germ-line-specific genes susceptible to ectopic activation in cancer. Our findings document the previously speculated predominant role of the X chromosome (13) on which more than half of the CG genes in our set are localized. Furthermore, affirming recent data (13, 31), genomic hypomethylation is necessary and sufficient for CG gene expression. *DNMT1* together with *DNMT3b* activity are obviously essentially involved in the stringent transcriptional repression of CG genes in somatic cells. In contrast, germ-line genes seem to underlie a more complex regulation, presumably involving specific transcription factors and/or chromatin-remodeling mechanisms (32). In fact, it has been proposed that DNA methylation could be the primary mechanism for the selective expression of tissue-specific genes with CpG-rich promoters (21, 33), whereas those with CpG-poor promoters depend solely on the presence of tissue-specific transcription factors (23). A preliminary survey of CG gene promoters suggests that most of them indeed have CpG-rich promoters. The relative independence of CG genes from individual transcription factors may be the reason for their frequent ectopic activation in tumors.

Third, for several CG genes, expression in ovary has been denied based on RT-PCR data. However, because bulk adult ovary tissue is stroma-rich but poor of germ cells, RT-PCR regularly fails even in detection of established oogenesis-related transcripts due to sampling reasons. Therefore, we included *SYCP1* as a surrogate marker for the presence of germ cells in the respective tissue specimen. Remarkably,

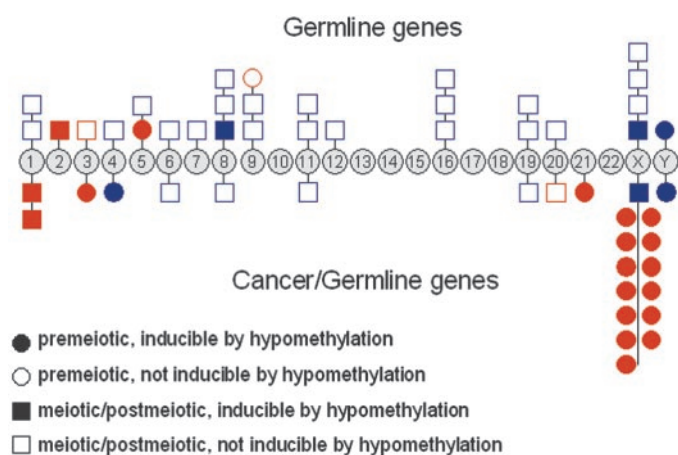


Fig. 5. Chromosomal mapping of the genes analyzed. Distribution of the 54 genes throughout the genome broken down into germ-line genes and CG genes. Blue circles and boxes represent genes restricted to spermatogenesis, and red circles and boxes show genes expressed during both spermatogenesis and oogenesis.

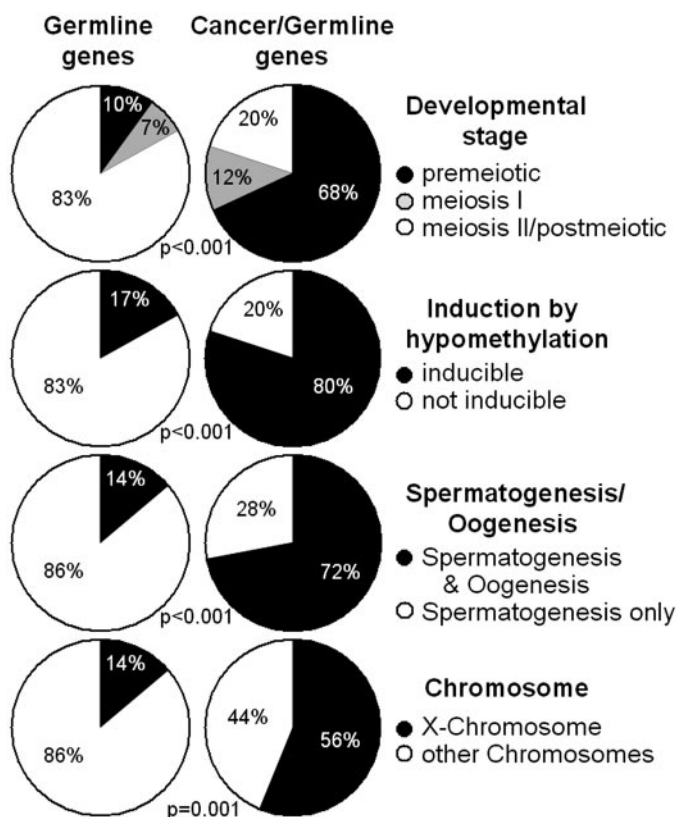


Fig. 6. Correlation of investigated criteria with the germ-line gene and CG gene expression patterns. Distribution of germ-line genes and CG genes related to developmental stage of expression, induction by genomic hypomethylation, sex-specificity, and chromosomal localization. Significances were calculated using Fisher's exact test and  $\chi^2$  test, respectively.

the overwhelming majority of CG genes (72%) are expressed in gametogenesis of both sexes, whereas, in contrast, most of the germ cell-specific transcripts resisting ectopic activation in cancer (86%) are restricted to spermatogenesis and thus to male gametogenic cells.

In summary, our findings suggest that a considerable number of germ-line-specific genes undergo ectopic activation in cancer. This is significantly correlated with a set of specific features, which we have defined in this study. We provide evidence that, in particular, those genes activated in the premeiotic stages of germ cell development relying predominantly on DNA methylation for transcriptional regulation are prone to such aberrant expression. An underlying common epigenetic molecular alteration may be the reason for their induction in cancer, which is the subject of our ongoing work. Moreover, our study provides class predictors to distinguish between CG and germ-line genes. Compilation of these class predictors within an appropriate data mining script will allow an even more targeted and tailor-made prediction and identification of new CG genes.

## REFERENCES

- Boon T, Coulie PG, Van den Eynde B. Tumor antigens recognized by T cells. *Immunol Today* 1997;18(6):267–8.
- Boon T, Old LJ. Cancer Tumor antigens. *Curr Opin Immunol* 1997;9(5):681–3.
- Sahin U, Tureci O, Pfreundschuh M. Serological identification of human tumor antigens. *Curr Opin Immunol* 1997;9(5):709–16.
- Marchand M, van Baren N, Weynants P, et al. Tumor regressions observed in patients with metastatic melanoma treated with an antigenic peptide encoded by gene MAGE-3 and presented by HLA-A1. *Int J Cancer* 1999;80(2):219–30.
- Jager E, Jager D, Knuth A. Clinical cancer vaccine trials. *Curr Opin Immunol* 2002;14(2):178–82.
- Knuth A, Jager D, Jager E. Cancer immunotherapy in clinical oncology. *Cancer Chemother Pharmacol* 2000;46 (Suppl):46–51.
- Disis ML, Cheever MA. Oncogenic proteins as tumor antigens. *Curr Opin Immunol* 1996;8(5):637–42.

8. Wang RF, Wang X, Rosenberg SA. Identification of a novel major histocompatibility complex class II-restricted tumor antigen resulting from a chromosomal rearrangement recognized by CD4(+) T cells. *J Exp Med* 1999;189(10):1659–68.
9. Chen YT, Old LJ. Cancer-testis antigens: targets for cancer immunotherapy. *Cancer J Sci Am* 1999;5(1):16–7.
10. Old LJ. Cancer/testis (CT) antigens: a new link between gametogenesis and cancer. *Cancer Immunol* 2001;1(1):1–7.
11. Jungbluth AA, Busam KJ, Kolb D, et al. Expression of MAGE-antigens in normal tissues and cancer. *Int J Cancer* 2000;85(4):460–5.
12. Juretic A, Spagnoli GC, Schultz-Thater E, Sarcevic B. Cancer/testis tumour-associated antigens: immunohistochemical detection with monoclonal antibodies. *Lancet Oncol* 2003;4(2):104–9.
13. Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol Rev* 2002;188(1):22–32.
14. Koslowski M, Tureci O, Bell C, et al. Multiple splice variants of lactate dehydrogenase C selectively expressed in human cancer. *Cancer Res* 2002;62(22):6750–5.
15. Scheurle D, DeYoung MP, Binninger DM, Page H, Jahanzeb M, Narayanan R. Cancer gene discovery using digital differential display. *Cancer Res* 2000;60(15):4037–43.
16. Gondos B, Westergaard L, Byskov AG. Initiation of oogenesis in the human fetal ovary: ultrastructural and squash preparation study. *Am J Obstet Gynecol* 1986;155(1):189–95.
17. Goto T, Adjaye J, Rodeck CH, Monk M. Identification of genes expressed in human primordial germ cells at the time of entry of the female germ line into meiosis. *Mol Hum Reprod* 1999;5(9):851–60.
18. Motta PM, Makabe S. Elimination of germ cells during differentiation of the human ovary: an electron microscopic study. *Eur J Obstet Gynecol Reprod Biol* 1986;22(5–6):271–86.
19. Meuwissen RL, Meerts I, Hoovers JM, Leschot NJ, Heyting C. Human synaptonemal complex protein 1 (SCP1): isolation and characterization of the cDNA and chromosomal localization of the gene. *Genomics* 1997;39(3):377–84.
20. Gondos B, Hobel CJ. Ultrastructure of germ cell development in the human fetal testis. *Z Zellforsch Mikrosk Anat* 1971;119(1):1–20.
21. Eden S, Cedar H. Role of DNA methylation in the regulation of transcription. *Curr Opin Genet Dev* 1994;4(2):255–9.
22. Meehan R, Lewis J, Cross S, Nan X, Jeppesen P, Bird A. Transcriptional repression by methylation of CpG. *J Cell Sci Suppl* 1992;16:9–14.
23. Bird A. The essentials of DNA methylation. *Cell* 1992;70(1):5–8.
24. De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc Natl Acad Sci USA* 1996;93(14):7149–53.
25. De Smet C, Lurquin C, Lethe B, Martelange V, Boon T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* 1999;19(11):7327–35.
26. Honda T, Tamura G, Waki T, et al. Demethylation of MAGE promoters during gastric cancer progression. *Br J Cancer* 2004;90(4):838–43.
27. Rhee I, Bachman KE, Park BH, et al. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature (Lond.)* 2002;416(6880):552–6.
28. Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 2001;27(4):422–6.
29. Hurst LD, Randerson JP. An eXceptional chromosome. *Trends Genet* 1999;15(10):383–5.
30. Mintz A, Debinski W. Cancer genetics/epigenetics and the X chromosome: possible new links for malignant glioma pathogenesis and immune-based therapies. *Crit Rev Oncog* 2000;11(1):77–95.
31. Lorient A, Boon T, De Smet C. Five new human cancer-germline genes identified among 12 genes expressed in spermatogonia. *Int J Cancer* 2003;105(3):371–6.
32. Sassone-Corsi P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science (Wash. DC)* 2002;296(5576):2176–8.
33. Riggs AD, Jones PA. 5-Methylcytosine, gene regulation, and cancer. *Adv Cancer Res* 1983;40:1–30.