5 Results

5.1 Generation and use of testis-specific cDNA microarrays

We generated mouse testis-specific microarrays from several cDNA libraries with a total of 13,440 clones. After amplification and purification, 11,490 of the clones yielded single-band PCR products and were selected for spotting (Fig. 8). An aliquot of each PCR product was stored for sequencing. An additional 96 PCR products for testis control genes were spotted, and a total of 318 negative (no template) controls. Results from PCR and sequencing were tracked in a database.

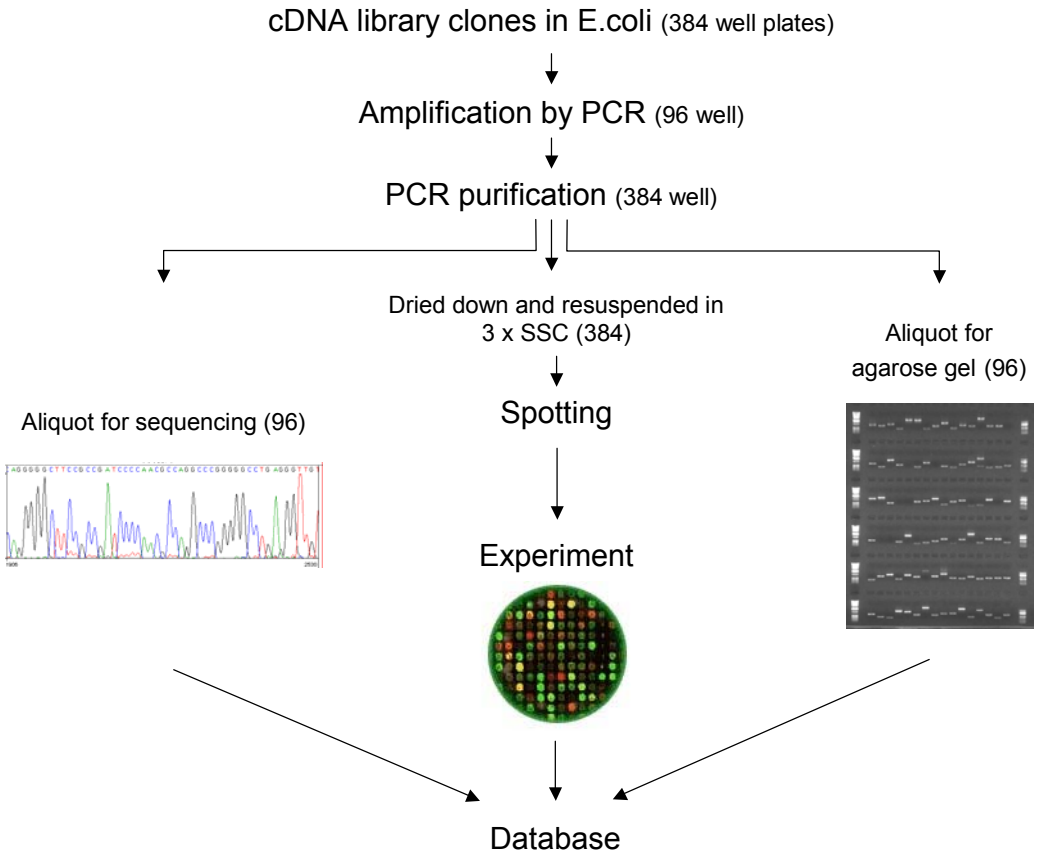

Figure 8. Generation of cDNA microarrays and data acquisition
Libraries of cDNA clones were amplified by PCR. Purified PCR products were used for DNA sequencing, agarose gel electrophoresis, and for spotting on glass slides. Sequence data, gel data, and the data from all experiments were entered into a database that allowed efficient analysis of the experiments.

The mouse testis cDNA microarrays were initially used with RNA isolated from cultured rat germ cells. The rat was the animal used in our laboratory for testicular cell cultures, and species differences between rat and mouse were not expected to lead to major hybridization differences on the microarrays. Populations of lam_B germ cells from 20-day-old rats, which consist of mostly undifferentiated spermatogonia, were compared to populations of lam_{NB} germ cells, which mainly consist of spermatocytes.

In one representative experiment, performed with a dye-reversal control, 2695 spots on the microarray showed a signal in at least one of the samples, 1242 of which showed regulation of more than 2-fold (Fig. 9).

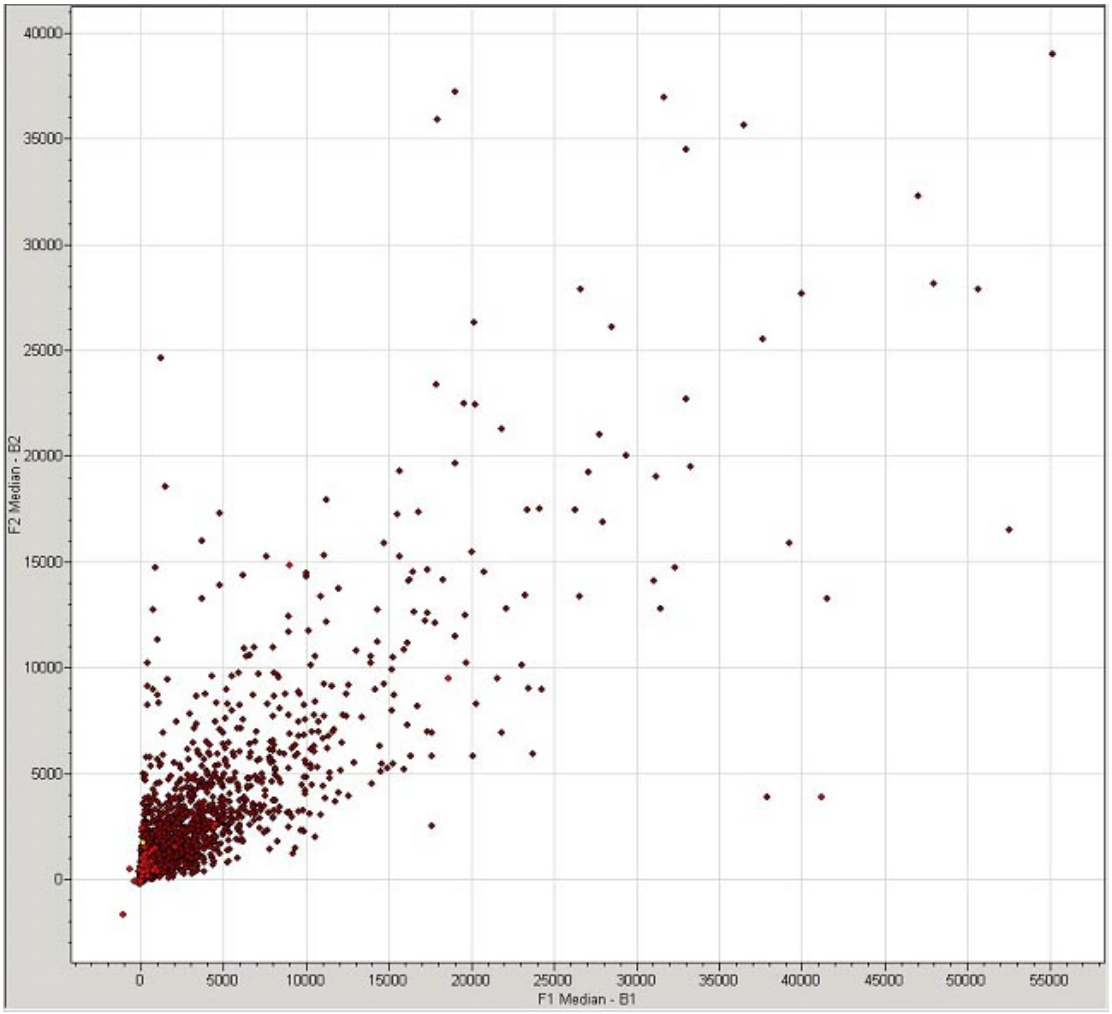

Figure 9. Gene expression changes between rat lam_B and lam_{NB} germ cells
Scatter plot for a spotted microarray experiment comparing lam_B vs lam_{NB} germ cells. Fluorescence intensities measured for both dyes are plotted for each gene (horizontal axis: lam_B, vertical axis: lam_{NB}). The plot demonstrates the large differences in gene expression between the two samples.

Ninety-five genes were expressed at least 5-fold higher in the lam_B population, 144 were expressed at least 5-fold higher in the lam_{NB} population. The list of lam_B-enriched genes contained interesting genes like *cRet*, *RagD*, and *Egr4*. Genes found to be enriched in lam_{NB} cells were Ldhc, *Tpx1*, *Tlbp*, *Spa17*, and others (Table 2). Several of the genes found in lam_{NB} cells were already known from the literature and therefore served as good internal controls for the experiment.

Table 2. Gene expression comparison of rat lam_B and lam_{NB} cells with cDNA microarrays

Gene name	Image clone id	lam_B Cy5	lam_{NB} Cy3	lam_{NB} Cy5	lam_B Cy3	Ratio lam_B/lam_{NB}
C-ret proto-oncogene	4063105	2822	66	124	530	**23.2**
Early growth response 4 (Egr4)	4063010	19963	952	1327	19428	**17.8**
Inosine monophosphate dehydrogenase	4062733	5523	202	127	860	**17.0**
Rag D protein	3664197	1620	57	217	1126	**16.6**
Suppressor of cytokine signaling 3	4062246	7346	427	644	8072	**14.8**
Proteasome regulator PA28 β	4063341	5318	283	873	5199	**12.4**
Myosin heavy chain	3664378	3970	270	433	1634	**9.2**
MHC class III region RD gene	2798286	3209	336	133	1134	**9.0**
Homolog to HYPOTHETICAL 75.9 kD protein	3663867	5937	530	1867	6294	**7.3**
MARCKS-like protein (Mlp)	3664061	2987	323	2638	13679	**7.2**
Glycerol kinase-like protein 2 (Gkrs2)	3664806	68	818	23363	398	**-35.3**
Lactate dehydrogenase-C (Ldhc)	3665170	955	24881	8504	191	**-35.4**
Testis lipid binding protein (Tlbp)	1747425	988	37579	61454	1703	**-37.3**
Sperm autoantigenic protein 17 (Spa17)	3664212	204	10026	15007	521	**-39.4**
Testis lipid binding protein (Tlbp)	1747399	128	4685	31332	618	**-43.9**
Four and a half LIM domains 4 (Fhl4)	3663687	508	15029	52650	893	**-44.2**
Testis-specific protein (Tpx1)	3665088	410	14109	10110	179	**-45.5**
Repressor of GATA	3664689	220	5916	39522	548	**-49.6**
Testis lipid binding protein (Tlbp)	1746654	116	3181	3316	38	**-57.1**
RIKEN testis clone, unclassifiable	3664749	104	4891	16668	241	**-58.4**
Testis specific gene 1 (Tpx1)	3664444	176	8189	22050	203	**-77.0**

Listed are known genes that were detected as differentially expressed between rat lam_B and lam_{NB} cells. The expression data is based on two microarray experiments, the dyes used for labeling the samples were reversed for the separate microarrays. Cy5 and Cy3 expression values were determined after a subtraction of the local background signal. Ratios shown are the averages of the two experiments.

While these experiments yielded interesting preliminary data, there were significant problems in many areas:

1) Not all library clones yielded PCR products during the initial clone amplification.
2) Many clones showed only weak signals on the microarrays (Fig. 9 indicates the different expression levels in a typical cDNA microarray experiment)
3) Some PCR products were extremely difficult to sequence. It was also very tedious and error-prone to hand-select only those clones that were found to be regulated in an experiment. On the other hand, a global sequencing approach would have been too expensive.
4) We experienced multiple problems during the spotting process, which included spotting interruptions and variable spot sizes.
5) Peeling of the poly-L-lysine coat on the microarrays was a common problem, probably caused by the spotting pins.
6) Multiple times we observed uneven hybridization, leading to areas on the microarray without or with low signal.
7) Finding the correct scanner settings for each microarray to compensate for uneven incorporation proved to be difficult at times, and repeated scanning led to photobleaching of the Cy5-dye.
8) Slide-to-slide variability was very high, requiring a large number of repetitions in order to get reliable results (Table 2 shows the signal differences observed between two experiments).

We were able to alleviate some of these problems. Switching to UltraGaps slides (Corning) eliminated the issue with peeling coats, and direct labeling protocols promised to guarantee even incorporation. However, the persistence of most of the problems and the rapid improvements and price drop of the Affymetrix GeneChips led us to abandon spotted microarrays and use Affymetrix arrays. We had initially assumed that the greatest advantage of using custom tissue-specific microarrays was the larger proportion of tissue-specific genes, but the Affymetrix chips turned out to contain a surprisingly high number of testis-specific genes. Furthermore, the Affymetrix GeneChips would allow us to study the gene expression of a larger number of genes with greater sensitivity and much higher reproducibility.

5.2 Gene expression analysis of mouse spermatogenesis

5.2.1 Gene expression patterns

For the first few days after birth, the seminiferous tubules in the mouse testis principally contain gonocytes, spermatogonia and Sertoli cells. They are rapidly out-numbered by differentiating germ cells as the animal ages. Spermatocytes appear around days 11-14 and spermatids after about day 21 of age in the mouse (Fig. 10).

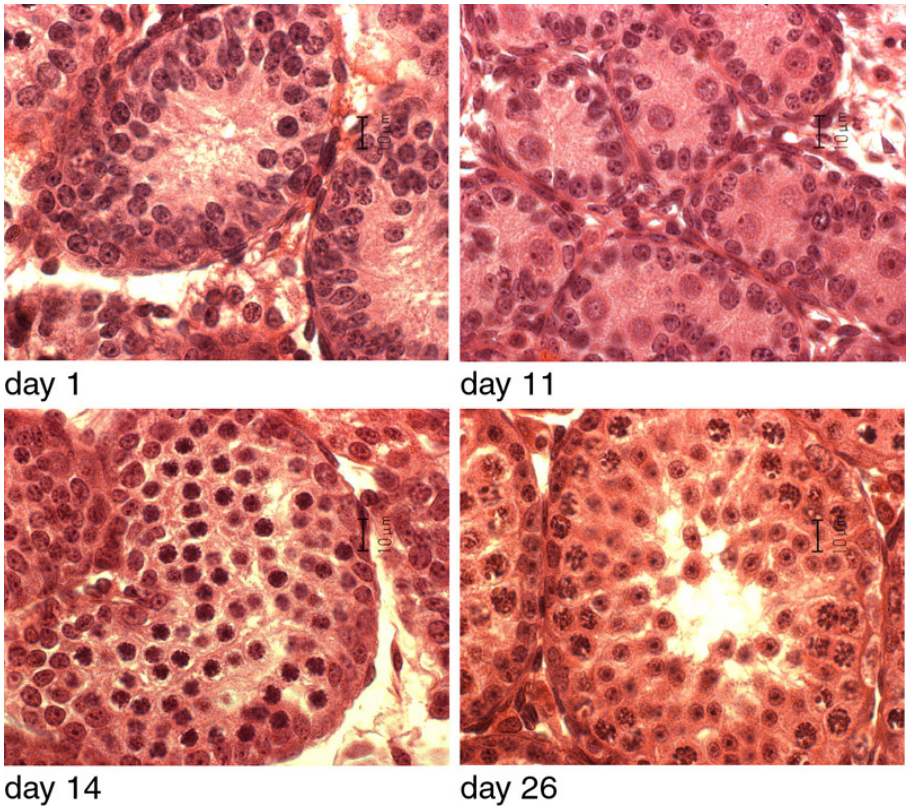

Figure 10. Tissue sections of mouse testes at different days of postnatal development
Hematoxylin/eosin staining of testis sections from mice at day 1, 11, 14, and 26 demonstrate germ cell differentiation at each age. At day 1, only Sertoli cells and gonocytes (large cells) are visible. At day 11, the tubules look more organized, and late stage spermatogonia are dominant. At day 14, pachytene spermatocytes dominate the testis, and at day 26 spermatocytes and round spermatids are visible.

Testes from C57BL/6 mice were collected at various days ranging from day 1 *post partum* to adult. The mRNA expression profiles were determined using the Affymetrix Mouse U74v2 oligonucleotide array set. RNA labeling for subsequent microarray analysis was performed on three 1-day-old animals, two four-day-olds, two eight-day-olds, two 11-day-olds, and one animal each for days 14, 18, 21, 26, 29, and adult (Fig. 11). The

complete data set can be queried in the NCBI Gene Expression Omnibus (GEO) repository (www.ncbi.nlm.nih.gov/geo), GEO accession number GSE640.

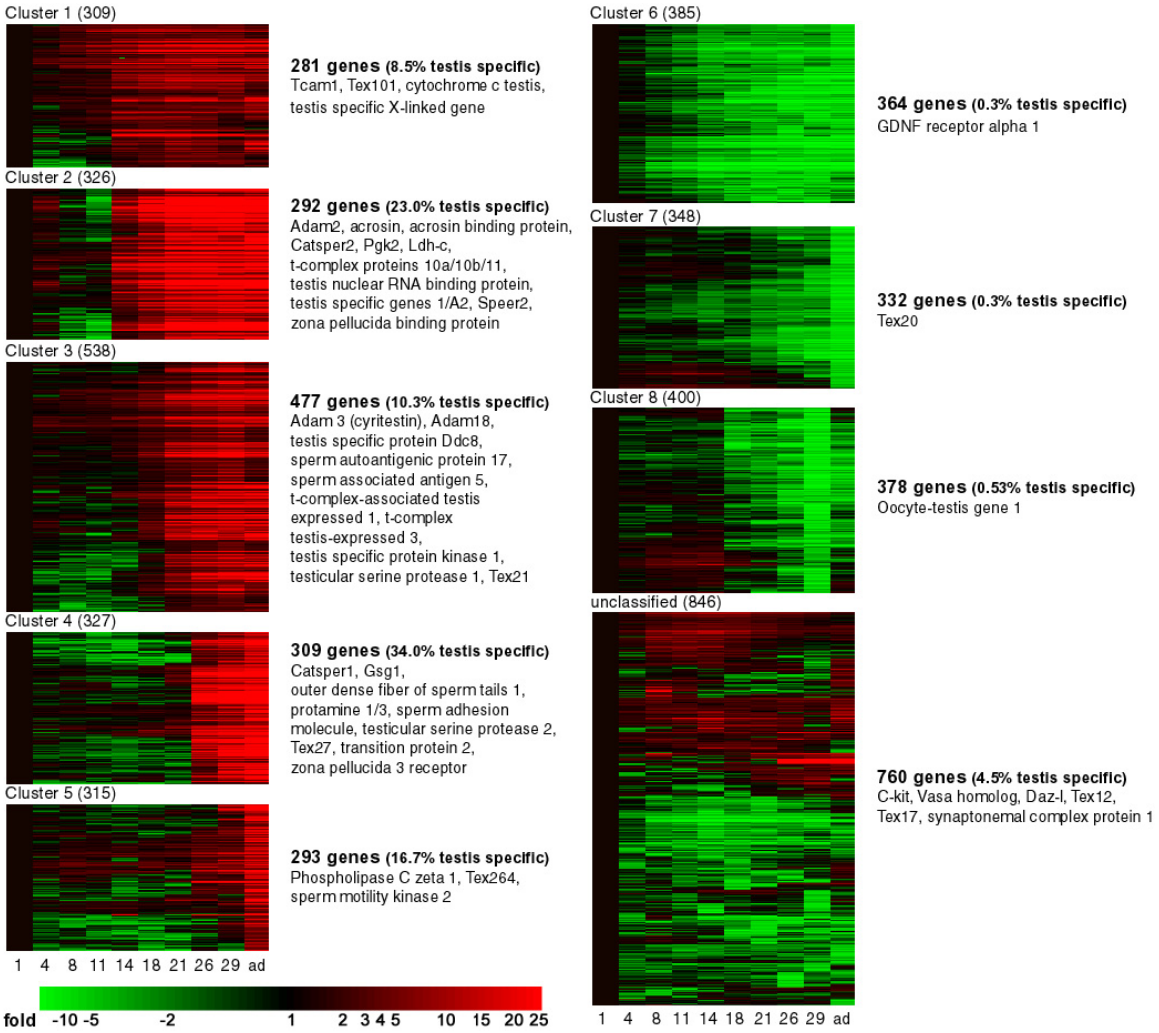

Figure 11. Gene expression clusters during mouse spermatogenesis
The gene expression heat maps show the eight clusters of transcripts that change coincident with the formation of meiotic and post-meiotic male germ cells and the unclassified genes. Each row of the heat map represents a gene, and each column represents a time point in development (as labeled at the bottom). The color saturation represents differences in gene expression compared with day 1. Red indicates an increase in gene expression, whereas green indicates a decrease. The genes listed next to the heat maps are examples of known genes found in the respective clusters.

A total of 21,374 transcripts were identified as expressed in the developing mouse testis (see 4.2.3 for the definition of ‘present’ and ‘absent’ calls). To more clearly characterize a germ cell population of transcripts, we separately screened for genes found in a cultured testis somatic cell preparation containing principally Sertoli cells. We also determined the genes expressed in an interstitial cell population. The found to be expressed

in these two somatic cell populations were subtracted from those for whole testis to yield a germ cell-enriched population. Three thousand seven hundred ninety-four transcripts (3486 different genes) were changed at least 3-fold compared to day one of age, and of these, 2245 represented uncharacterized genes. Because some genes are represented by multiple oligonucleotide probe sets on the Affymetrix arrays, the numbers of transcripts reported represent a slightly smaller number of different gene products. Transcript expression profiles were clustered into 8 distinct patterns; in 5 clusters the expression of genes substantially increased coincident with the appearance of spermatocytes or spermatids. The pattern of gene expression in the other three clusters was highest during times when the spermatogonial population dominated the testis (see Fig. 11).

While the transcript profiles of male germ cells early in development are critical to an understanding of male germline stem cells and the early steps of commitment to differentiation, we first concentrated on the five clusters where transcript expression markedly increased late in development. If these transcripts were expressed by the germ cells, these would be the best candidate genes for a role in fertilization. The average patterns of expression for the genes found in clusters one to five are shown in Fig. 12. Meiosis begins in the mouse at about day 11, and elevations in transcript expression are evident starting on or after this day. It should be noted that the relative levels to which transcript expression increased was considerably different between each of the clusters, with genes in clusters two (expression starting in pachytene spermatocytes) and cluster four (expression in spermatids) showing the highest degree of induction.

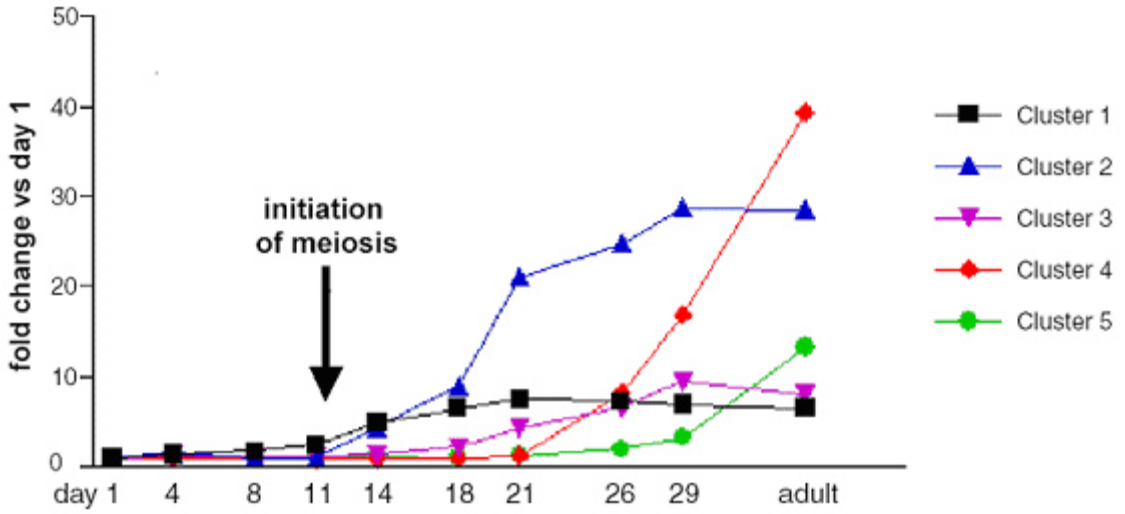

Figure 12. Testis developmental profiles for clusters 1–5
The relative gene transcript abundance in testis from mice at day 4, 8, 11, 14, 18, 21, 26, and 29 and adult compared with testis at day 1 of age. The values are the mean relative abundance (compared with day 1) for all transcripts found in each of the clusters. Note the marked increase in relative transcript abundance for cluster 4 coincident with formation of spermatids.

The expression patterns of various transcripts are consistent with the behavior of these genes described in the literature (Fig. 11 lists several known genes for each cluster). Examples of well-characterized genes include lactate dehydrogenase C and phosphoglycerate kinase 2, which are known to be initially transcribed during meiosis (Wheat et al., 1977; McCarrey et al., 1992); as expected then, they appear in cluster 2. Likewise, protamine 1 and transition protein 2 are initially expressed in spermatids (Mali et al., 1989), and therefore as expected appear in cluster 4. An interesting aspect of the analysis is that transcription of CatSper 1 (cluster 4) and CatSper 2 (cluster 2), which may form subunits of the same channel, is not initiated at the same time during spermatogenesis.

5.2.2 Analyzing tissue specificity

Of the 1652 genes found in the five clusters, where gene expression increased relative to day one, about 17.8% were represented by transcripts that may be testis-specific based on the Unigene database (two or more sequence entries; no tissue other than testis) (Table 3, Supplemental Tables 1 and 2). We required at least two sequence entries in our analysis to help ensure that an EST sequence entered was real. The addition of transcripts present in one or two other tissues would increase the percentage to 28.9%. In contrast, in the other 3 clusters, only 0.4% of the genes appeared as testis-specific transcripts, suggesting that only a few of the genes in male germline stem cells or spermatogonia are unique to the testis. The known or named testis-specific genes are listed in Supplemental Table 1.

Table 3. The number of genes in each of the 8 clusters that show apparent testis-specific expression

Cluster	total	Named genes	Genes with unknown/ unassigned function	Apparent testis-specific [Unigene database][a]	Only 1 other tissue	Only 2 other tissues
1	323	106	217	24 (7.4%)	18 (5.6%)	9 (2.8%)
2	308	116	192	68 (22.1%)	32 (10.4%)	19 (6.4%)
3	538	196	342	51 (9.5%)	36 (6.0%)	7 (1.3%)
4	312	115	197	105 (33.7%)	29 (9.3%)	13 (4.2%)
5	317	127	190	49 (15.5%)	18 (5.7%)	6 (1.9%)
6	480	197	283	1 (0.2%)	1 (0.4%)	2 (0.4%)
7	432	155	277	1 (0.2%)	0	0
8	571	161	410	3 (0.5%)	2 (0.4%)	1 (0.2%)
Unclassified	910[b]	328	582	36 (4.1%)	15 (1.7%)	3 (0.3%)

[a]Expression data are based on the tissue distribution information in the Unigene database.
[b]The 910 gene transcripts in the unclassified category did not show a pattern of expression that fit the other 8 clusters.

To determine whether transcripts reported only for the testis actually reflected testis-specific expression, 50 genes with unassigned function (principally from cluster four) were examined for expression in spleen, liver, testis, brain, ovary, thymus, kidney, heart, lung, and embryo. Of the 50 transcripts, 46 yielded signals in the testis only (see Supplemental Table 2). Thus, it is predicted that more than 90% of the transcripts now entered in the Unigene database as testis-only (at least two sequence entries), in fact, are testis-specific.

To then determine whether all such genes were also germ cell-specific (likely given the subtraction against testicular somatic cells), we performed *in situ* hybridization on 11 (randomly chosen) genes with unassigned function from clusters one to five. All transcripts were expressed specifically by spermatocytes or spermatids, strongly suggesting that all of the testis-specific genes identified above are germ cell-specific (Fig. 13, Table 4).

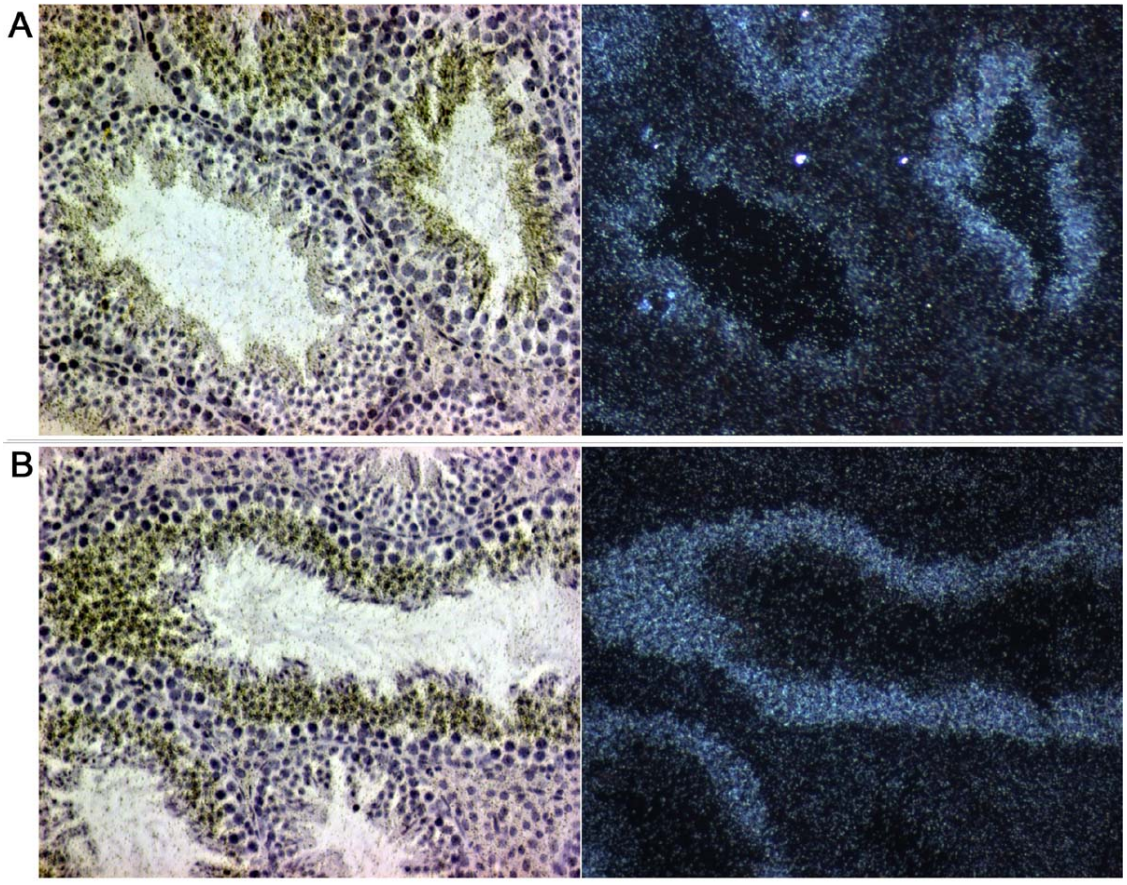

Figure 13. *In situ* hybridization of two testis-specific genes from cluster 4
Both genes are expressed exclusively in elongating spermatids. *Mm.46121* (A) is expressed in step 7-14 spermatids, and *Mm.45302* (B) is expressed in step 7-12 spermatids. Shown are hybridization results with antisense probes (*Left*: bright field image; *Right*: dark field image). Sense probes showed no signal.

Table 4. Meiotic and post-meiotic germ cells express the transcripts found in clusters one to five

Unigene identifier	Probe id.	Cluster	Germ cell stage of expression
Mm.56002	163837_at	One	Spermatocytes / spermatids
Mm.87419	168262_at	Three	Round spermatids
Mm.59283	129740_at	Three	Spermatids (steps 1-14)
Mm.45611	164171_at	Four	Spermatids (steps 1-12)
Mm.84522	167943_f_at	Four	Spermatids (steps 1-16)
Mm.46121	165889_f_at	Four	Spermatids (steps 7-14) (see Fig. 13A)
Mm.45306	165824_at	Four	Spermatids (steps 5-14)
Mm.46151	165900_f_at	Four	Spermatids (steps 5-14)
Mm.45302	165822_f_at	Four	Spermatids (steps 7-12) (see Fig. 13B)
Mm.56501	166073_at	Four	Spermatids (steps 7-14)
Mm.72970	166193_at	Five	Spermatids (steps 7-16)

All 11 genes have no assigned function. Expression was determined by *in situ* hybridization.

5.2.3 Mining the Unigene database for testis-specific genes

Intrigued by the large number of testis-specific genes in clusters 1-5, we examined the Unigene database to determine whether a significant number of other transcripts not represented on the arrays would also be entered as testis-specific. As of May 2003, there were 62,692 entries with a cluster size of at least two sequence entries. Again, we limited our survey to at least two entries to help ensure that a given EST entry represented a real transcript sequence. Of these, 2931 gene transcripts were reported only for the testis (seven percent with known functions, 68% with a predicted open reading frame but unknown or unassigned function, and 25% as ESTs). The question then arose whether these apparent testis-specific gene transcripts were also elevated during or after meiosis. Therefore, we arbitrarily chose 12 of the 2931 (the 12 were genes not represented on the microarrays). Quantitative PCR showed that for 11 of 12, expression was significantly elevated after day 11, again suggesting expression in spermatocytes or spermatids (Table 5).

Additionally, a screen of testis-specific genes subtracted by using the somatic cell populations suggests that less than 10% of all testis-specific genes on the microarray (independent of cluster) are found in Sertoli or interstitial cells. Thus, a vast majority of the 2931 gene sequences appear to represent meiotic or post-meiotic germ cell-specific transcripts. Subtraction of the low numbers of testis-specific genes that appear early in spermatogenesis, those confined to Sertoli or interstitial cells, and the approximate 10% of genes expressed in tissues outside the testis, leads to the conclusion that about 2375 of the 62,692 genes represented in Unigene (at least two sequence entries) are ex-

pressed specifically in meiotic and post-meiotic male germ cells. This represents about 3.8% of the genome. Although some EST entries will represent identical full-length transcripts, there is no apparent reason that the testis would have more such redundant entries than other tissues.

Table 5. Expression patterns of testis-specific genes identified in the Unigene database

Day	*Mm.140115*	*Mm.141471*	*Mm.158168*	*Mm.188680*	*Mm.23402*	*Mm.245982*
8	3.3 (2.4-4.5)	0.5 (0.2-1.4)	2.4 (0.5-10.7)	0.7 (0.5-0.9)	0.6 (0.4-0.9)	1.2 (0.6-2.4)
11	1.4 (0.3-7.6)	1.5 (1.3-1.8)	2.9 (2.4-3.4)	1.8 (1.6-2.1)	0.6 (0.4-0.9)	2.3 (1.5-3.5)
14	9.2 (1.8-47.5)	0.7 (0.5-0.9)	1.7 (0.7-4.4)	0.1 (0.1-0.2)	6.6 (5.5-8.0)	1.6 (1.2-2.0)
18	3.9 (2.1-7.0)	0.5 (0.2-1.0)	0.8 (0.7-0.8)	0.3 (0.1-0.7)	25.2 (24.4-26.1)	0.7 (0.6-0.9)
21	31.9 (18.8-53.9)	24.6 (22.8-26.5)	10.36 (5.2-20.8)	5.0 (3.4-7.3)	332.4 (317.0-348.6)	2.8 (2.7-2.9)
26	79.2 (73.8-84.9)	4.4 (3.5-5.7)	7.09 (3.6-13.8)	0.5 (0.4-0.7)	36.7 (32.5-41.5)	3.4 (2.5-4.7)
29	376.4 (333.6-424.6)	9.3 (8.1-10.6)	269.7 (237.7-306.0)	3.1 (2.3-4.2)	28.8 (23.4-35.5)	18.0 (12.8-25.4)

Day	*Mm.247000*	*Mm.27027*	*Mm.46175*	*Mm.61213*	*Mm.66843*	*Mm.72404*
8	0.6 (0.2-1.5)	0.2 (0.1-0.4)	0.4 (0.3-0.4)	0.6 (0.6-0.6)	0.4 (0.3-0.4)	0.7 (0.7-0.7)
11	0.2 (0.2-0.2)	0.3 (0.2-0.3)	0.3 (0.3-0.4)	0.2 (0.2-0.3)	0.6 (0.4-1.0)	0.1 (0.1-0.4)
14	3.2 (2.7-3.9)	0.7 (0.7-0.8)	1.0 (0.8-1.1)	12.3 (11.9-12.8)	1.2 (1.1-1.5)	0.8 (0.7-0.8)
18	16.6 (15.5-17.7)	3.7 (3.0-4.4)	5.7 (5.5-5.9)	21.7 (20.3-23.1)	2.1 (1.6-2.7)	0.4 (0.4-0.5)
21	184.2 (179.0-189.6)	141.4 (132.6-150.9)	87.2 (84.2-90.3)	125.1 (120.6-129.8)	24.9 (22.7-27.1)	179.4 (166.5-193.4)
26	73.1 (68.6-77.8)	45.1 (43.7-46.4)	47.3 (45.2-49.4)	19.7 (19.0-20.5)	6.7 (6.3-7.1)	41.4 (39.9-43.0)
29	93.2 (87.0-99.9)	168.5 (156.5-181.3)	136.9 (128.0-146.5)	16.8 (15.6-17.9)	10.9 (10.2-11.7)	88.4 (82.7-94.6)

Genes with unknown or unassigned function from the Unigene database that have been reported only in the testis (at least 2 sequence entries) show significant increases in expression after day 11, coincident with the start of meiosis. Gene expression was detected using qRT-PCR. Shown are fold changes in transcript expression compared to day 1 of age. Ranges in parentheses were calculated using standard deviations (n=2).

5.3 The spermatogonial stem cell

5.3.1 Testicular cell cultures

When germ cell-enriched suspensions from cultures of GCS-GFP rat testicular cells were added to collagen-coated culture dishes, cells that failed to bind to collagen (col_{NB} cells) were highly enriched in GFP^+ cells. After transferring these freshly isolated germ cells to laminin-coated culture dishes, two populations of GFP^+ cells were obtained, those that bound to laminin (lam_B) and those that did not bind (lam_{NB}) (Fig. 14A).

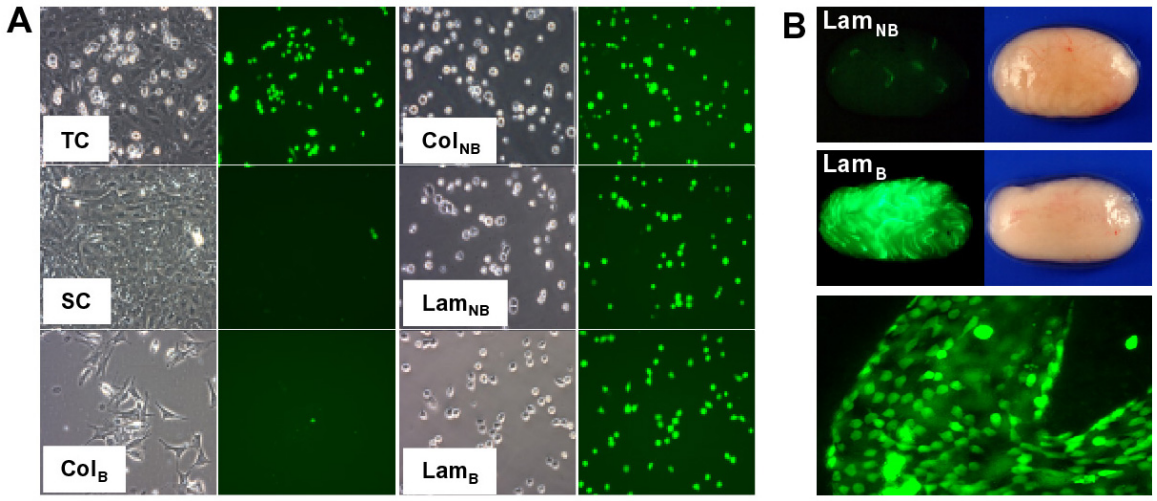

Figure 14. Testis cell cultures from GCS-GFP rats
(A) GFP^+ male germ cells (green fluorescence) purified by selection on collagen (col) and laminin (lam) matrices. *TC*, ~65-h culture of testis cells on plastic; *SC*, somatic cells on plastic after removal of the germ cell-enriched population from TC; *col*$_B$, cells in the germ cell-enriched population that bound to col; *col*$_{NB}$, cells in the germ cell enriched population that did not bind to col; *lam*$_{NB}$, cells from col$_{NB}$ that did not bind to lam; *lam*$_B$, cells from col$_{NB}$ that bound to lam.
(B) Spermatogonial stem cell activity is enriched in the lam$_B$ germ cell population initially isolated from GCS-GFP rats. Expression of GCS-GFP rat transgene in donor cells (green fluorescence) colonizing the seminiferous tubules of testes from WT recipient rats viewed 32 days after being transplanted with ~$5x10^4$ lam$_{NB}$ (*Upper Left*) or lam$_B$ (*Middle Left*) cells per testis. (*Middle & Upper Right*) Respective, bright-field microscopy images for panels at Left. (*Bottom*) High magnification image of *Middle Left* that shows spermatogonia from the lam$_B$ population colonizing a seminiferous tubule.

The cells at this initial time of selection on laminin are considered as day 0. Greater than 98% of the cells present in the lam_B and lam_{NB} fractions were GFP^+, and about 95% of these distributed to the lam_{NB} group. In contrast, cells that initially bound to plastic and collagen (col_B cells) were more than 98% GFP^-. When the originally selected lam_B and lam_{NB} GFP^+ germ cell populations were transferred to the testis of recipient males, os-

tensibly a measure of spermatogonial stem cell activity (Brinster and Zimmermann, 1994; Dobrinski et al., 1999; Nagano et al., 1999), only the lam_B fraction substantially colonized (Fig. 14B). This finding verified previous results where the lam_B fraction, which represented about 5% of the total germ cell population, was greatly enriched in spermatogonial stem cells (Hamra et al., 2002).

This conclusion was also supported by three additional observations: First, determination of the ploidy of cells showed that a majority of cells in the initially selected lam_{NB} fraction were 4N while most of the cells in the lam_B fraction were 2N (Fig. 15A). The diploid cells are most likely germ cells at a stage of development prior to meiosis, while the 4N cells probably represent cells proceeding through meiosis (Malkov et al., 1998).

Second, electron microscopy studies confirmed that the lam_B population was enriched with cells displaying distinct patterns of heterochromatin staining and organelles ascribed to type A spermatogonia (91±3%), with relatively few intermediate/type B spermatogonia or spermatocytes (4±1%), or somatic cells (5±1%) (Fig. 15B). In contrast, the lam_{NB} population was represented predominantly by spermatocytes plus intermediate/type B spermatogonia (89±6%) and contained a much smaller percentage of type A-like spermatogonia (11±6%) (see Fig. 15B).

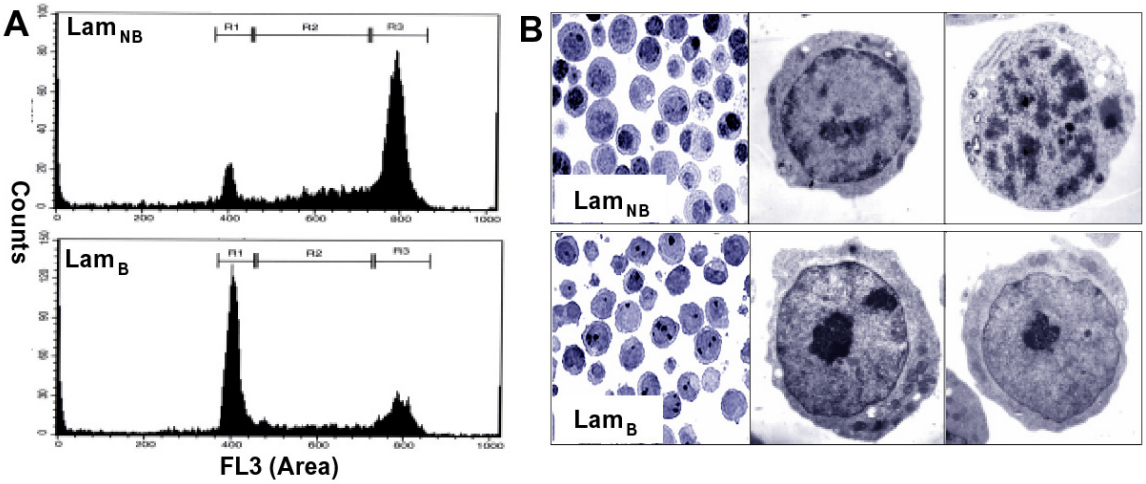

Figure 15. Analysis of laminin-binding and non-binding cells
(A) Ploidy of lam_{NB} and lam_B cells. The relative percentage DNA content/cell of lam_{NB} (*Top*) and lam_B (*Bottom*) cell populations from day 0 cultures were determined by FACS (Malkov et al., 1998) to be 11.5% 2N, 23.9% 3N, 63.8% 4N and 58.9% 2N, 19.7% 3N, 20.6% 4N, respectively. Values represent the mean number of counts under the illustrated regions (R1=2N, R2=3N, R3=4N) from duplicate experiments.
(B) Bright field microscopy images, 60x (*Left*) and electron micrographs, 5000x (*Middle and Right*) of lam_{NB} (*Top*) and lam_B cells (*Bottom*) from day 0 cultures. Note, cellular staining patterns characteristic for intermediate/type B spermatogonia (*Top Middle*), pachytene spermatocytes (*Top Right*) and type A spermatogonia (*Bottom Middle & Bottom Right*).

Third, the profile of transcripts expressed in the initially selected lam_B cells (Affymetrix microarrays) correlated strongly with transcripts expressed at early ages in the postnatal mouse testis. As shown in a Venn diagram, about 7% (906; 820) of the transcripts present in the originally selected lam_B population are absent in either day 0 somatic or day 0 lam_{NB} cultures of rats and mice (Fig. 16; see *Methods* for calculation of present or absent scores for microarrays). About 10% (rat) or 17% (mouse) of transcripts present in day 0 lam_{NB} cultures are absent in somatic or lam_B populations, and about 12% of transcripts present in cultures of mouse or rat testicular somatic cells are absent in either day 0 lam_B or lam_{NB} cultures (see Fig. 16). Thus, transcripts present exclusively within cultures of lam_B, lam_{NB} or somatic testis cells provided molecular markers for type A spermatogonia, spermatocytes or testicular somatic cells, respectively.

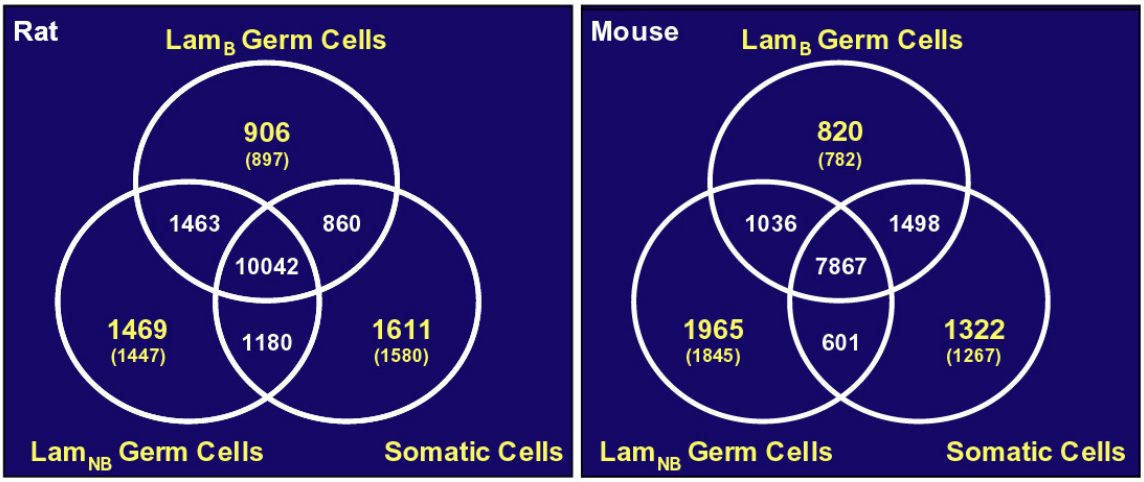

Figure 16. Overlap of gene expression in cultures of mouse and rat testis cells
Venn diagrams showing the distribution of transcripts that are exclusively expressed (yellow numbers) and mutually expressed (white numbers) by each culture population. Exclusively expressed transcripts are defined as transcripts detected as "present" within one culture population, but scored as "absent" within the other two culture populations of the same species. Mutually expressed transcripts were detected as present in more than one culture population of the same species. The numbers of different genes represented by exclusively expressed transcripts in each culture population are shown in parentheses.

When the profiles of these marker transcripts were evaluated *in vivo*, transcripts marking the initially selected lam_B cells (Affymetrix microarrays) dominate the testis at early ages (Fig. 17) when gonocytes and type A spermatogonia are prevalent (Dettin et al., 2003; McLean et al., 2003). However, these same transcripts decrease in relative testicular abundance as the animal ages due to the fact that lam_B-like cells in the testis become out-numbered by more differentiated spermatogonia, spermatocytes and spermatids (see Fig. 17). Transcripts marking lam_{NB} cells coincided with mRNA that substantially

increased in relative abundance in the testis as a function of age (see Fig. 17), and thus, correlated strongly with the appearance of spermatocytes and spermatids. Transcripts exclusively present in testicular somatic cells decreased in relative abundance as a function of testis age (see Fig. 17). Thus, cultures of lam_B germ cells represent a relatively pure population of type A spermatogonia that is highly enriched in spermatogonial stem cell activity.

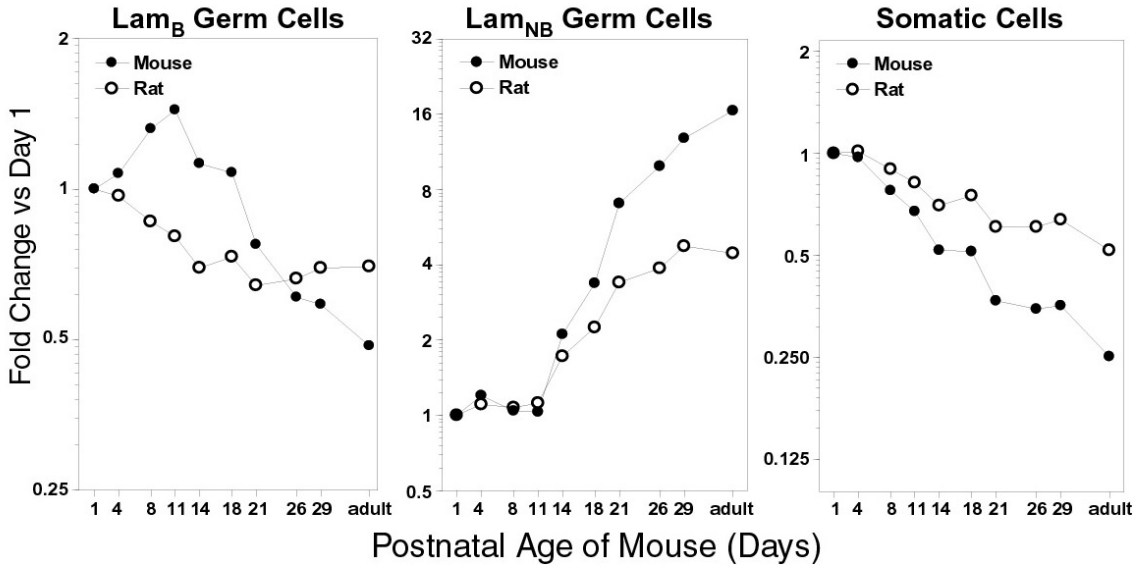

Figure 17. Age-dependent expression of marker transcripts for cultures of testicular lam_B, lam_{NB} and somatic cells in the postnatal mouse testis

Transcripts exclusively expressed in each testicular cell culture type (i.e. lam_B, lam_{NB} and somatic) were used as markers for identifying the expression profiles of the cell types *in vivo*. Transcripts exclusively marking each culture population were required to be at least 2-fold more abundant than in the other two culture populations of the same species, and to be present in at least one developmental time point of the mouse testis. Plotted are the average signal intensities for all transcripts per marker set at ten different developmental time points in the mouse testis (normalized to the day 1 time point). The numbers of exclusively expressed transcripts identified in cultures of mouse testis cells were: lam_B germ cells, 292 (280 genes); lam_{NB} germ cells, 1185 (1050 genes); somatic cells, 530 (508 genes). The total numbers of rat transcripts queried from each cell population were: lam_B=381 (375 genes), lam_{NB}=462 (453 genes) and somatic=206 (198 genes). Numbers of mouse homologues identified for the exclusively expressed rat transcripts per marker set were: lam_B germ cells, 308; lam_{NB} germ cells, 355; somatic cells, 174.

All culture microarray data can be queried in the NCBI GEO repository. Accession numbers are GSE829 (mouse testis cell cultures), GSE830 (rat testis cell cultures).

5.3.2 Testis colonization assay

To evaluate the effectiveness of various culture conditions to maintain spermatogonial stem cells within the lam_B population, we developed a quantitative assay for assessment of colonization of rat testes by GFP^+ cells from the GCS-GFP rats. After transfer of GFP^+ lam_B cells (maintained for 1 day on STO feeder layers) to recipient testes, followed by solubilization of testes 32 days later, a nearly linear relationship was evident between numbers of cells transplanted and subsequent GFP fluorescence (Fig. 18).

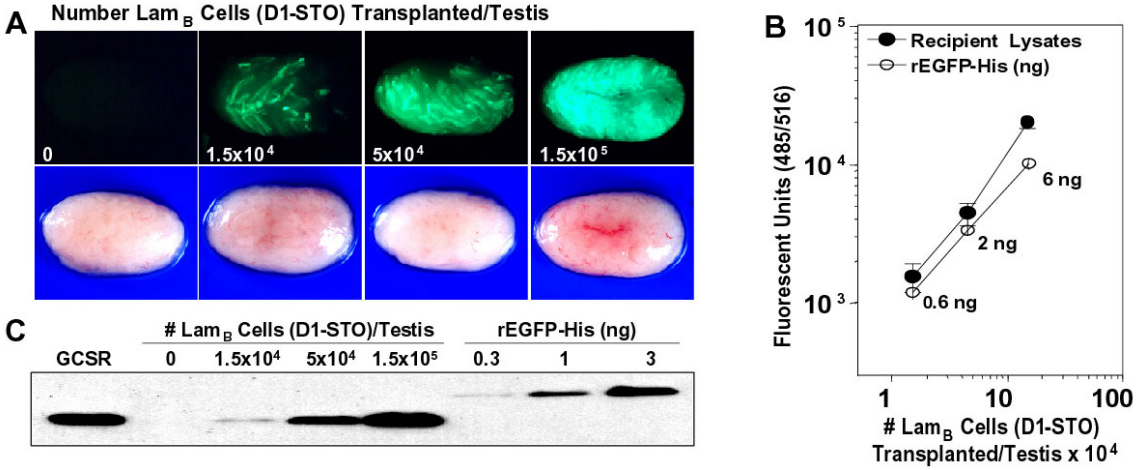

Figure 18. Fluorometric assay for GCS-GFP rat germ stem cell activity
(A) WT testes transplanted with increasing numbers (*0, 1.5x10^4, 5x10^4, 1.5x10^5* GFP^+ cells/testis, *Left to Right*) of GCS-GFP lam_B cells cultured for ~1 day on STO cells. *(Top)* Green fluorescence observed in the testis correlates with the number of donor cells transplanted. *(Bottom)* Bright-field microscopy images of testes in *Top*.
(B) Fluorescence units measured in extracts made from testes transplanted with increasing numbers of GFP^+, lam_B cells after ~1 day in culture on STO cells (*closed circles*). Values represent the mean number of fluorescence units in 10µl from 1.5 ml of testicular extracts (±SEM, n=3 rats/cell concentration transplanted). Also shown are fluorescence units plotted from increasing amounts of rEGFP-His (*open circles*).
(C) Immunoblot detection of GFP in extracts from a 23 day old GCS-GFP rat, WT recipient testes transplanted with increasing numbers of GCS-GFP^+ lam_B cells after 1 day of culture on STO cells (*0, 1.5x10^4, 5x10^4, 1.5x10^5* GFP^+ donor cells/testis) and of increasing amounts of rEGFP-His. Equal volumes from triplicate testis extracts were loaded per lane.

On average, extracts prepared from rats transplanted with $1.5x10^5$ GFP^+ cells/testis contained 1.8±0.3 µg equivalents of rEGFP/testis (n=3). Based on this assay, we could quantify spermatogonial stem cell activity, which was at least 60-fold higher in day 1 lam_B (221±59 ng rEGFP/testis/$3x10^4$ donor cells, n=5, ±SEM) than in day 1 lam_{NB} cultures (3.5±4.9 ng rEGFP/testis/$3x10^4$ donor cells, n=3, ±SEM). The stem cell activity measured in day 1 lam_{NB} cultures was not significantly above background levels for the

assay as measured in non-transplanted testes (1.7±3.1 ng rEGFP/testis, n=8, ±SEM). Cultures of donor lam_B cells (day 1 on STO) that colonized the seminiferous tubules of recipients in these studies also maintained the ability to differentiate into GFP^+ spermatocytes and spermatids, and the recipient males were able to transmit the GCS-GFP transgene (data not shown). Thus, the cultures of rat lam_B cells used to establish the colonization assay are highly enriched in functional spermatogonial stem cells.

5.3.3 Investigation of feeder cell lines

We investigated various cell feeder layers for ability to maintain spermatogonial stem cell activity. When the initially selected rat lam_B cells were incubated on STO fibroblasts, they remained GFP^+ (Fig. 19A), but lost ability to colonize a testis by day 10 in culture (Fig. 19B,C). In contrast, when lam_B GFP^+ cells were cultured on MSC-1 cells, colonization activity was maintained for at least 20 days (see Fig. 19B,C). A potentially important characteristic of male germ cells cultured on STO feeder layers was the common appearance of long chains of 8 to 64 interconnected cells (see Fig. 19A); this is similar to germ cell development characteristics *in vivo* (Huckins, 1971b; Huckins, 1971c; Oakberg, 1971; Lok and de Rooij, 1983b; Lok et al., 1983). In contrast, GFP^+ cells grown on MSC-1 cells continued to grow predominantly as either single or paired cells (see Fig. 19A); this characteristic is consistent with various stem cell models (Huckins, 1971b; Huckins, 1971c; Oakberg, 1971; Lok and de Rooij, 1983b; Lok et al., 1983).

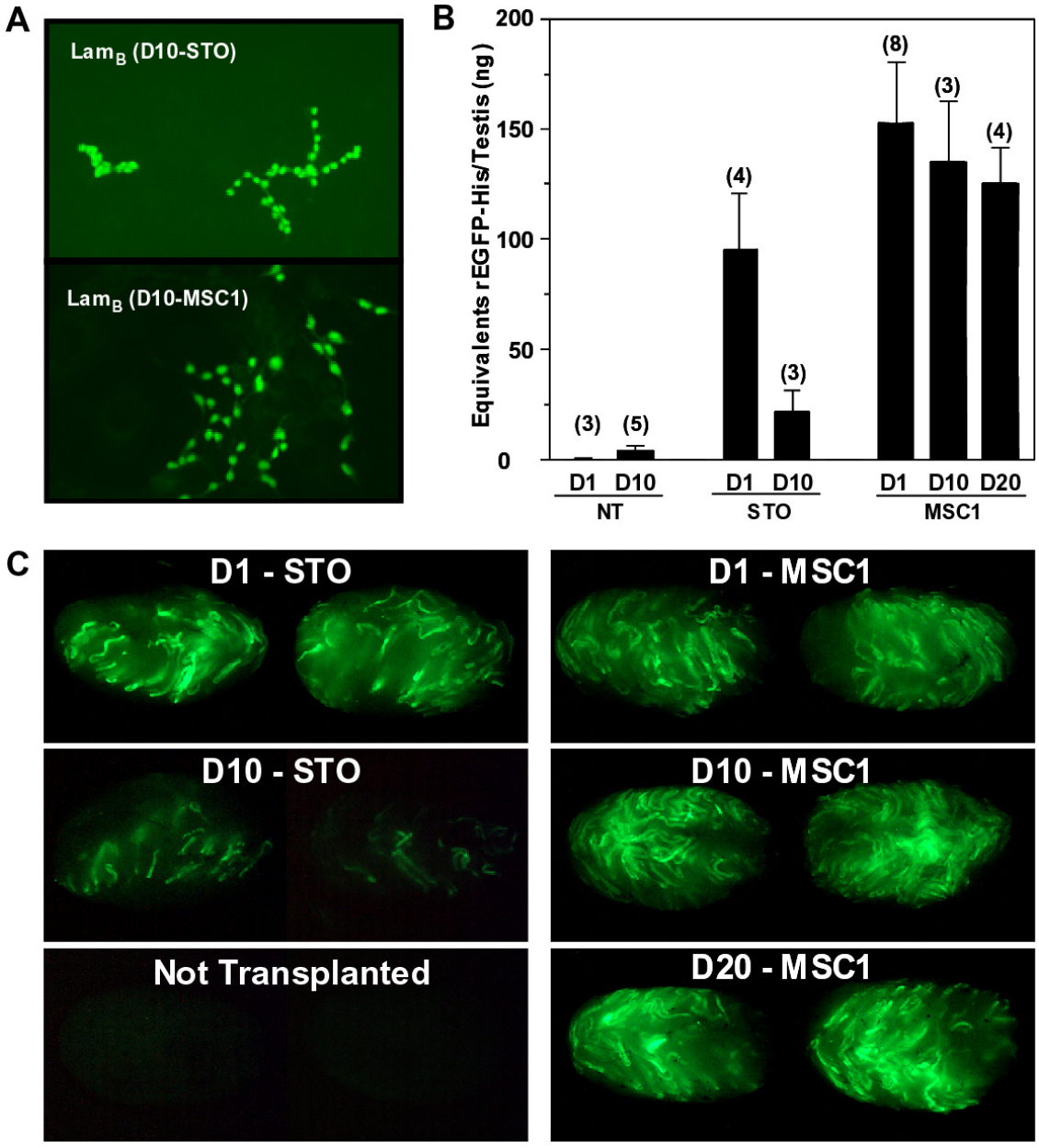

Figure 19. Maintenance of spermatogonial stem cell activity of rat lam$_B$ cells on different feeder layers.

Spermatogonial stem cells maintain stem cell activity when cultured on MSC-1 cells and lose it when cultured on STO cells. (A) Images of GCS-GFP lam$_B$ cells (green fluorescence) after maintenance for 10 days in culture on STO fibroblasts (*Top*) or on MSC-1 Sertoli cells (*Bottom*).

(B) Equivalent rEGFP-His extracted from WT rat testes transplanted with GCS-GFP$^+$ lam$_B$ cells after their maintenance on feeder cell lines: *NT*, testes not transplanted with donor cells; *STO*, testes transplanted with 3x10^4 GFP$^+$ cells from cultures of lam$_B$ cells after their maintenance for 1 and 10 days on STO fibroblasts; *MSC-1*, testes transplanted with 3x10^4 GFP$^+$ cells from cultures of lam$_B$ cells after their maintenance for 1, 10 and 20 days on MSC-1 Sertoli cells. (n)=the number of testes analyzed.

(C) Images of testes 32 days after being transplanted using GFP$^+$ germ cell cultures as represented in “B”.

5.3.4 Germ cell differentiation on STO feeder layer

We determined the transcript profile of rat lam_B GFP^+ cells as a function of culture time on STO feeder layers. Initially isolated (day 0) GCS-GFP rat lam_B cells were plated onto monolayers of STO fibroblasts and subsequently maintained in culture for 1, 5, 10 or 20 days. After each time point, GFP^+ cells were sorted (FACS) from the GFP^- feeder cells. One-half of the GFP^+ cells harvested from each culture were also transplanted to recipient rat testes to measure stem cell activity (see legend to Fig. 19). The other half was used to isolate RNA, and microarray analysis was performed on these samples to identify transcripts expressed by each germ cell fraction. The relative expression level for two sets of germ line markers was then determined in each of the germ cell fractions. In terms of exclusively expressed marker transcripts initially identified in Fig. 16, the relative abundance of 381 transcripts initially expressed in the lam_B cells markedly decreased while the relative abundance of 462 transcripts initially expressed in the lam_{NB} cells markedly increased as a function of culture time on STO cells (Fig. 20A). Thus, the GFP^+ cells on STO feeders appeared to be moving towards the day 0 lam_{NB} transcript profile. This is supported by the time-dependent loss of stem cell activity within highly enriched cultures of type A spermatogonia (see Fig. 19), and as reported using suspensions of wild-type or cryptorchid mouse testis cells (Nagano et al., 1998; Nagano et al., 2000; Nagano et al., 2001; Nagano et al., 2003), or rat pup testis cells (Orwig et al., 2002a).

Our next goal was to determine whether we could identify genes whose expression would decrease prior to or coincident with a loss of spermatogonial stem cell activity. These gene products could be critical for the maintenance of stem cell activity and could also then serve as reliable molecular markers for the spermatogonial stem cell. Transcripts present in GFP^+ cells that yielded a time-dependent change in relative abundance by at least 2-fold (6639 of 12833) were grouped into 6 gene clusters (C1-C6; Fig. 20B). Clusters C1-C4 define 2167 transcripts that were detected at higher signal intensities in lam_B germ cells at day 1 relative to day 20 (STO feeder cells; see Fig. 9B). Clusters C5-C6 contain a total of 1145 transcripts that are detected at lower signal intensities in lam_B germ cells at day 1 relative to day 20 (STO feeder cells; see Fig. 9B). A number of transcripts (3327) did not fit into clusters C1-C6 and remained unclassified. The clusters containing genes of particular interest were those in C1 and C2, in that the relative abundance of these transcripts (C1, n=954; C2, n=231) decreased very rapidly and correlated closely with the time-dependent decrease in lam_B stem cell activity (see legend to Fig. 20B).

-

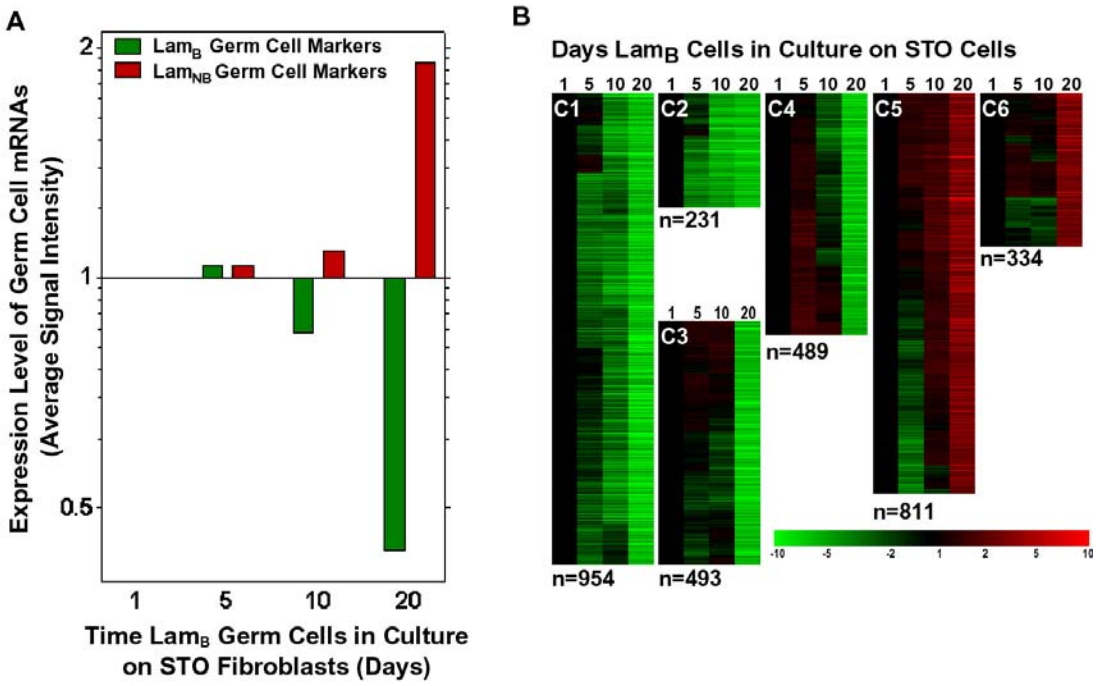

Figure 20. Transcriptional profile of lam$_B$ germ cells in culture on STO fibroblasts
(A) Expression pattern of germ line marker transcripts in rat lam$_B$ germ cells in culture on STO fibroblasts. The lam$_B$ germ cell marker set contained 381 transcripts (375 genes, *green bars*), and the lam$_{NB}$ germ cell marker set contained 462 transcripts (453 genes, *red bars*). Each set of germ line markers is fully described in the legend to Fig. 16. Plotted are average signal intensities for all transcripts of each marker set, which are detected in at least one time point, normalized to the day 1 time point.
(B) Heat maps showing the six clusters of transcripts that change during the culture of rat lam$_B$ germ cells on STO feeder cells. Each row of the heat map represents a gene, and each column represents a time point in development (as labeled at the top). The color saturation represents differences in gene expression compared with day 1. Red indicates an increase in gene expression, whereas green indicates a decrease.

If clusters C1 and C2 represent genes critical for maintenance of stem cell activity, these gene transcripts should be maintained in lam$_B$ germ cells cultured on MSC-1 cells. Two hundred and forty-eight of the putative genes (255 transcripts) in clusters C1 and C2 remained enriched (2-fold or greater relative to day 0 lam$_{NB}$ cells) in GFP$^+$ cells after maintenance for 20 days in culture on MSC-1 cells. Apparent mouse homologues or orthologs for 223 of the 248 rat genes were identified in the NCBI database. Based on mouse gene ontology these candidate 'stemness' genes were divided into 10 functional categories and a group of transcripts with unknown function or expressed sequence tags (ESTs) (Table 6). ESTs with unknown function represent the largest single category of genes in Table 6 (48%).

Table 6. Genes found to be closely associated with spermatogonial stem cell activity

Apoptosis (4)
Casp7, Card10[3], *Pmaip1, Sgk3*

Cell adhesion & cytoskeleton (25)
Efemp2, Enc1, Mbp, Nid2, Punc, Reln, Sdc2, Sdc4, Thbs4, Tubb5, Utrn, Waspip, Cldn4[1], *Cryga*[1], *Emb*[3], *Epb4.1l3*[3], *Epb4.1l4a*[3], *Lamb2*[3], *Myof1f*[1], *Plec1*[3], *Scarb1*[3], *Snap91*[3], *Spock1*[3], *Tjp2*[2], and 1 EST cluster

Cell cycle (1)
Cables1[1]

Channels & transporters (9)
Abcd4, Cacna1g, Slc25a13, Epb4.1l4b[3], *Nxf7*[2], *Rlbp1*[3], *Slc22a3*[3], *Slc6a4*[3], and 1 EST cluster

DNA modification & repair (3)
Dnmt3a, Hist1h4i, Polg2[3]

Metabolism & biosynthesis (16)
Elovl4, Elovl6, Enpep, Gus, Pccb, Prep, Rpl22, Ampd2[3], *Arts1-pending*[3], *C1qbp*[1], *Got2*[3], *Neu2*[1], *Psmb9*[1], *Psmb10*[3], *Smox-pending*[3], and 1 EST cluster

RNA binding & modification (7)
Rbpms, Rbms1[3], *Smfn*[3], and 4 EST clusters

Signal transduction (36)
Adcy6, Dusp6, F2r (Thrombin receptor), *Gfra1* (GDNF receptor α1), *H2-M3, Igbp1, Ltbp1, mSSH-1L, Pdk1, Ptprg, Ptprk, Rab12, Sema4d, Sh3d19, Chrna4*[1], *Cnp1*[1], *Dtnb*[3], *Ephb3*[1], *Fgfr3*[3], *H2-Q10*[3], *Hak-pending*[2], *Ick*[3], *Peli2*[3], *Ppp3ca*[3], *Ptgfrn*[3], *Ret*[1], *Rims4*[2], *Spry4*[3], *Stard13*[3], *Stat2*[3], *Tbc1d8*[2], and 5 EST clusters

Transcriptional regulation (37)
Aebp1, Bcl6b, Btbd3, Egr3, Foxo1, Irf1, Lmo4, Midn, Nr1d1, Pbx2, Satb1, Ski, Sox13, Zfp278, Ank2[3], *Ankrd6*[3], *Bhlhb3*[1], *D3Jfr1*[3], *Dri2*[3], *Egr4*[1], *Foxa2*[1], *Hdac5*[3], *Hey1*[3], *Klf3*[1], *Pou3f1*[1], *Sall4*[2], *Tcfe3*[3], and 10 EST clusters

Translation and posttranslational modification (3)
Galnt2 and 2 EST clusters

Other (107)
Asb13, Bklhd1, Gig1-pending, H19, Lphn1, Tafa5-pending, Tera-pending, Tpbg, Rtn1[3], *Upk1b*[1], and 98 unknown EST clusters

Transcripts present in cultures of testis somatic cells were not excluded from this table. Mouse gene symbols are presented where mouse homologues or orthologs to rat transcripts could be identified (See Supporting Tables 2 and 3).

[1]not detected on the mouse arrays
[2]no oligonucleotide probe set for transcript on mouse arrays
[3]no change of expression or up-regulated during postnatal mouse testis development

Surprisingly, some of the surface markers that have been successfully used in cell sorting and subsequent transplantation assays to enrich spermatogonial stem cell activity show contradicting expression patterns. Transcripts for Itga6 (integrin-α6), Itgb1 (integrin-β1) (Shinohara et al., 1999; Shinohara et al., 2000b) and Thy1 (thymus cell surface antigen) (Kubota et al., 2003) were detected as present in rat lam_B cells before (day 0) and after 20 days in culture on MSC-1 cells, but we find the mRNA for these genes to be relatively abundant in cultures of rat lam_{NB} germ cells and testicular somatic cells (Fig. 21), and in lam_B cells after 20 days of culture on STO feeder cells.

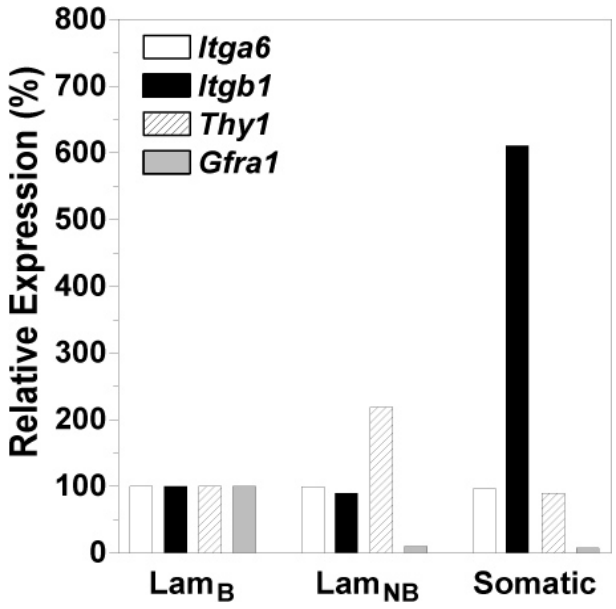

Figure 21. Expression of transcripts for known spermatogonial stem cell antigens in cultures of rat testis cells

The relative abundance of *Itga6, Itgb1, Thy1* and *Gfra1* in cultures of rat lam_B germ cells, lam_{NB} germ cells and somatic cells isolated from the seminiferous tubules are graphed as the percentage of the signal intensity determined for each transcript in the lam_B population. Signal intensities for *Itga6* (76), *Itgb1* (632), *Thy1* (106) and *Gfra1* (206) in the lam_B population were normalized to 100%. Each transcript was detected as present in each culture population. *Gfra1*, one of the genes listed in Table 6, is clearly enriched in lam_B cells compared to lam_{NB} cells.

5.3.5 Spermatogonial stem cell index

We developed a unit of measure for spermatogonial stem cell activity, which we named spermatogonial stem cell index (SSCI). SSCI values for cultures of rat germ cells (rSSCI) were determined using the mean signal intensity identified using the Affymetrix microarrays for the 248 rat genes enriched in spermatogonial stem cells. The rSSCI for day 0 lam_B GFP^+ cells dropped dramatically with time in culture on STO cells (810 at day

1 to 187 at day 20; Fig. 22A). The rSSCI at day 20 is similar to that seen with day 0 lam_{NB} germ cells (rSSCI=144; see Fig. 22A). In contrast, a rSSCI of 650 for cultures of day 0 lam_B GFP^+ cells was maintained at 786 after 20 days of culture on MSC-1 feeder cells (see Fig. 22A).

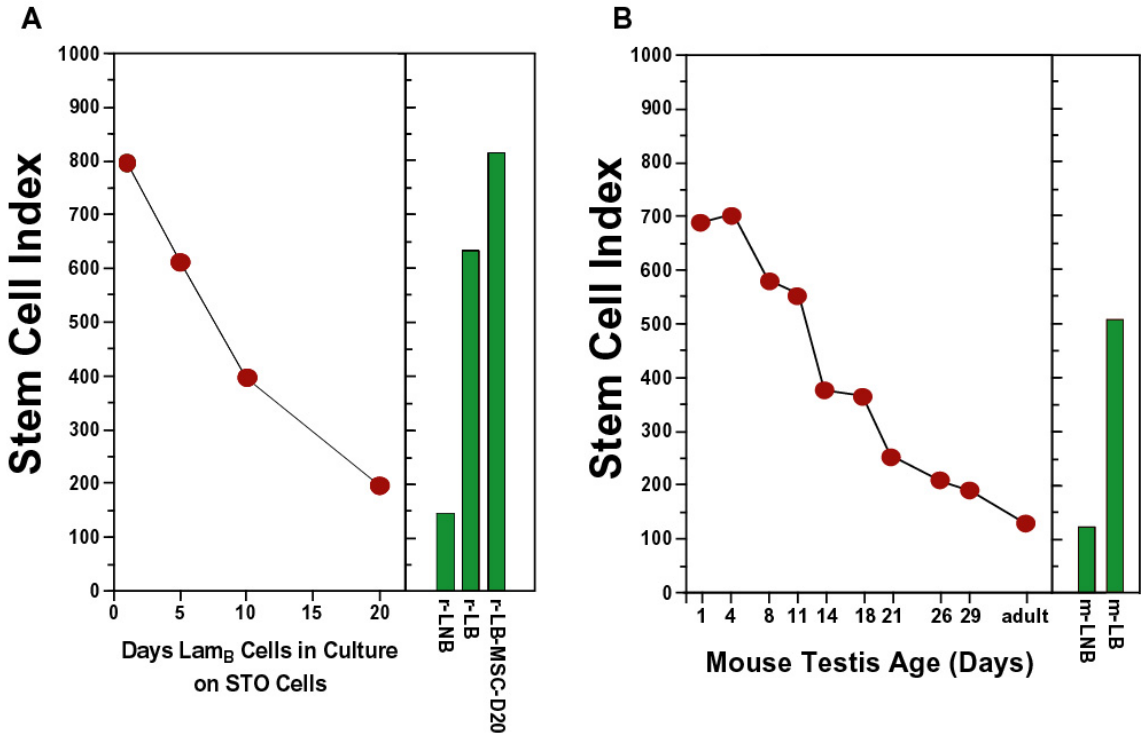

Figure 22. Mouse and rat spermatogonial stem cell indices (SSCI)
The mean signal intensity for transcripts enriched in cultures of spermatogonial stem cells (see Table 6) provides an index value for measuring spermatogonial stem cell activity. (A) The rat SSCI (rSSCI) rapidly drops over time in cultures of rat lam_B germ cells after maintenance on STO fibroblasts. Also shown are rSSCI values (green bars) for freshly isolated, day 0 cultures of GCS-GFP rat, lam_B germ cells (r-LB), lam_{NB} germ cells (r-LNB) and FACS-purified GFP^+ lam_B cultures after maintenance for 20 days on MSC-1 cells (r-LB-MSC-D20).
(B) The mouse SSCI (mSSCI) rapidly drops for testicular transcripts as a function of mouse testis age. Also shown are mSSCI values (green bars) for freshly isolated, day 0 cultures of mouse lam_B (m-LB) and lam_{NB} (m-LNB) germ cells.

To establish a SSCI in the mouse (mSSCI), we further analyzed the mouse homologues identified for the rat gene transcripts in Table 6. Homologues or orthologs for 202 of the genes in Table 6 (81% of genes) were identified on the Affymetrix Murine Genome U74v2 microarray set, of which 175 were detected as present in the mouse testis. Cluster analysis demonstrated that transcripts for 115 of these 175 genes (68%) showed a time-dependent decrease in relative abundance after postnatal days 1-8 in the mouse testis. The remaining 56 homologues showed a time-dependent increase or did not

change in relative abundance in the mouse testis. This could potentially be explained by species-specific differences in gene expression or by incorrect assignments of the authentic rat and mouse homologues, and therefore such transcripts were not used in mSSCI calculations. Based on mSSCI values, these 115 genes (145 transcripts) were most abundantly expressed in the testis at early ages, with a peak at day 4 (mSSCI=701); this was followed by a rapid decline as a function of testis age (mSSCI=122 in the adult testis; Fig. 22B). Thus, the mSSCI is high in the early testis at times when spermatogonial stem cells are prevalent in the mouse or rat. As the testis develops, differentiated germ cell populations dominate, and mSSCI values markedly decrease as expected (see Fig. 22B). In cultures of mouse testis cells, the mSSCI was 508 for lam_B cells and 120 for lam_{NB} cells (see Fig. 22B), which is in agreement with functional data reporting that mouse testis cells selected for on laminin are enriched in spermatogonial stem cell activity (Shinohara et al., 1999; Shinohara et al., 2000a).

If the mSSCI is an excellent predictor of spermatogonial stem cell activity, then it should also define spermatogonial stem cell activity in another species. When the 115 genes of the mSSCI were now used for the rat germ cell cultures, a SSCI value of 709 for the GFP^+ day 0 lam_B cells was maintained at 898 after 20 days on MSC-1 feeder cells. Furthermore, SSCI values for the GFP^+ day 0 lam_B cells decreased dramatically as a function of time in culture on STO cells (day 1=850, day 5=668, day 10=415, day 20=196). The mSSCI, therefore, appears as an excellent predictor of spermatogonial stem cell activity across both rat and mouse.

5.3.6 Characterization of Egr3 protein expression

We selected a protein encoded by one of the robust signals in the SSCI to ask whether we could identify the spermatogonial stem cell in culture. *Egr3* represents a gene transcript that dramatically decreases in a manner similar to that for the average transcript expression pattern observed for all genes in Table 6 (see Fig. 22A; Fig. 23A). The Egr proteins, Egr1, Egr2, Egr3 and Egr4, are closely related members of a subclass of immediate early gene-encoded, inducible transcription factors that share a similar DNA-binding domain (Beckmann and Wilce, 1997). Lam_B cells grown on STO cells retain GFP fluorescence and Dazl immunoreactivity after 10 days in culture, but lose detectable Egr3 immunoreactivity (Fig. 23B). In contrast, when grown on MSC-1 cells for 10 days, the germ cells retain GFP fluorescence and Dazl and Egr3 immunoreactivity (Fig. 23C). Strongly positive Egr3 immunoreactive signals are restricted in most cases to either single or paired germ cells. Although it may be expected that one or both of the paired cells would normally commit to differentiation (Huckins, 1971b; Huckins, 1971c; Oakberg, 1971; Lok and de Rooij, 1983b; Lok et al., 1983), that the MSC-1 culture appears to lack

differentiation-inducing factors may explain the high Egr3 signals in both cells of a pair. On STO cells, in contrast, differentiation factors appear to exist, driving germ cells to form long chains, upon which Egr3 immunoreactivity is lost (see Fig. 23B).

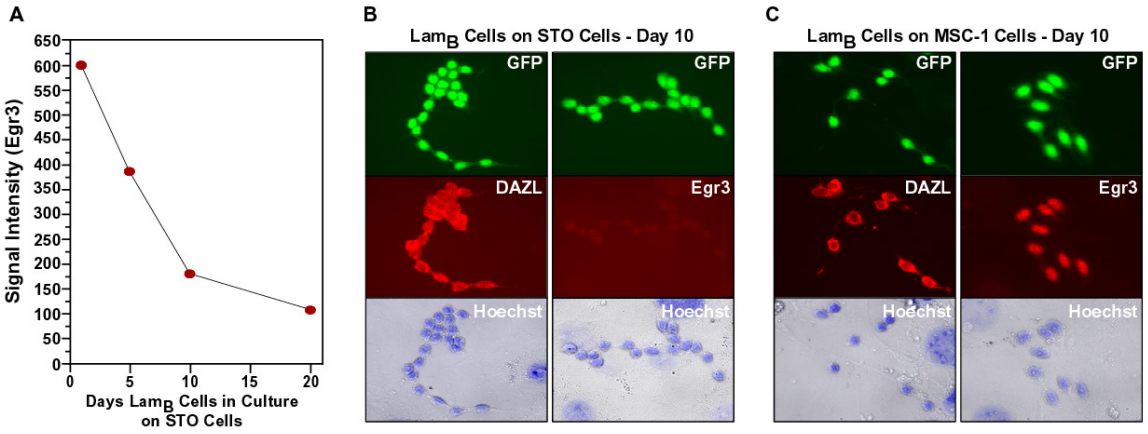

Figure 23. Egr3 expression in lam$_B$ germ cells after culture on STO or MSC-1 cells
(A) Time-dependent decreases in the relative abundance of the germ cell *Egr3* transcript after culture on STO cells. Plotted are signal intensities for *Egr3* on Affymetrix Rat Expression Set 230.
(B) (*Top Left and Right*) Expression of GFP (green) in germ cells derived form lam$_B$ cells after culture for 10 days on STO fibroblasts. (*Middle Left*) Anti-Dazl immunolabeling (red) of germ cells shown at top left. (*Middle Right*) Anti-Egr3 immunolabeling (red) of germ cells shown at top right. (*Lower Left* and *Lower Right*) Bright field microscopy image overlay of Hoechst 33342-labeled nuclei (blue fluorescence) in cells shown at top left and right.
(C) (*Top Left and Right*) Expression of GFP (green) in germ cells after 10 days in culture on MSC-1 Sertoli cells. (*Middle Left*) Anti-Dazl immunolabeling (red) of germ cells shown at top left. (*Middle-Right*) Anti-Egr3 immunolabeling (red) of germ cells shown at top right. (*Lower Left and Lower Right*) Bright field microscopy image overlay of Hoechst 33342-labeled nuclei (blue fluorescence) in cells shown at top left and right.