Current Biology 22, 1524–1529, August 21, 2012 ©2012 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2012.06.028

Report

Genetic Consequences of Programmed Genome Rearrangement

Jeramiah J. Smith,[1,2,6,*] Carl Baker,[3] Evan E. Eichler,[3,5] and Chris T. Amemiya[1,4]
[1]Benaroya Research Institute at Virginia Mason, Seattle, WA 98101, USA
[2]Department of Biology, University of Kentucky, Lexington, KY 40506, USA
[3]Department of Genome Sciences
[4]Department of Biology
University of Washington, Seattle, WA 98195, USA
[5]Howard Hughes Medical Institute, Seattle, WA 98195, USA

Summary

The lamprey (*Petromyzon marinus*) undergoes developmentally programmed genome rearrangements that mediate deletion of ~20% of germline DNA from somatic cells during early embryogenesis. This genomic differentiation of germline and soma is intriguing, because the germline plays a unique biological role wherein it must possess the ability to undergo meiotic recombination and the capacity to differentiate into every cell type. These evolutionarily indispensable functions set the germline at odds with somatic tissues, because factors that promote recombination and pluripotency can potentially disrupt genome integrity or specification of cell fate when misexpressed in somatic cell lineages (e.g., in oncogenesis). Here, we describe the development of new genomic and transcriptomic resources for lamprey and use these to identify hundreds of genes that are targeted for programmed deletion from somatic cell lineages. Transcriptome sequencing and targeted validation studies further confirm that somatically deleted genes function both in adult (meiotic) germline and in the development of primordial germ cells during embryogenesis. Inferred functional information from deleted regions indicates that developmentally programmed rearrangement serves as a (perhaps ancient) biological strategy to ensure segregation of pluripotency functions to the germline, effectively eliminating the potential for somatic misexpression.

Results and Discussion

A Survey of Known Sequences

In lamprey, programmed genome rearrangement (PGR) events are known to occur during early stages of embryonic development (starting at approximately the midblastula transition: between day 2 and 3 of development), are inherited uniformly across all somatic tissues, and result in deletions that may individually encompass hundreds of kilobases of DNA (both single copy and repetitive) [1, 2]. To further resolve the nature of PGR, we surveyed all available lamprey germline sequence for evidence of somatic deletion using array comparative genomic hybridization (arrayCGH).

A signature of somatic rearrangement (programmed or otherwise) can be observed when a derived tissue lacks specific nucleotide sequences that were present in its progenitor cell population. To further survey for the changes that characterize lamprey PGR, we designed a customized oligonucleotide microarray to target all available germline sequence (BAC-end sequences) [2] and ~1% of the known somatic genome by arrayCGH. This microarray was used to measure the relative abundance of target sequences within an individual's germline (sperm) versus somatic (blood) DNA, using replicated, dye-swapped experiments. Analysis of relative hybridization intensities revealed that most target regions fell within the expected distribution for normalized data but also identified a substantial tail of the distribution, suggesting enrichment of several sequences within germline DNA (Figure 1A). Notably, the few sequences that had been previously classified as germline-specific [1] and that contained sufficient nonrepetitive sequence to be targeted, all fell within this tail of the distribution (n = 4). No significant differences were observed in arrayCGH comparisons of DNA from somatic tissues that were derived from these same animals (Supplemental Experimental Procedures).

In addition to these previously identified sequences, our analysis identified several other germline-enriched sequences. In total, ~13% of the surveyed germline sequence (259 of 2,100 fragments or 150 kb/1.08 megabases, including deleted repeats) showed evidence of somatic deletion. This percentage is consistent with previous flow cytometric estimates that compared nuclear DNA content in germline versus somatic tissues [1], confirming that programmed deletions result in extensive differentiation of germline versus somatic genomes. Candidate germline-enriched regions that were identified for the first time by arrayCGH included eight single/low-copy sequences and several tandemly repeated sequences that appeared to be uniquely enriched in the germline. Six of eight candidate sequences were clearly validated as germline-specific, five of which were observed to be expressed in adult and juvenile testes (Figure 1; Figure S3 and Supplemental Experimental Procedures available online) and one of which was expressed in cells that are identical to classical anatomical descriptions of migrating primordial germ cells in lamprey embryos [3–5] (Figure 2; Figure S2; Supplemental Experimental Procedures).

Sequencing and Analysis of Lamprey Germline DNA

Our hybridization-based assays and earlier computational studies [1] hold the capacity to identify candidate deletion regions, yet both methods carry the same limitation in that they can only identify differences when sequences are known a priori. To address this limitation, we performed a single 454 Titanium shotgun sequencing run on lamprey germline (sperm) DNA. This sequence set consisted of 554,979 sequence reads with a minimum quality score of Q20 and a minimum trimmed length of 300 bp (median length: 484 bp, mean length: 428 bp, total length: 230 megabases), representing ~10% of the germline genome. The availability of a large whole-genome shotgun (WGS) data set from the lamprey genome project (liver DNA) [6] enabled us to develop a relatively simple analytical

[6]Present address: Department of Biology, University of Kentucky, Lexington, KY 40506, USA
*Correspondence: jjsmit3@uky.edu

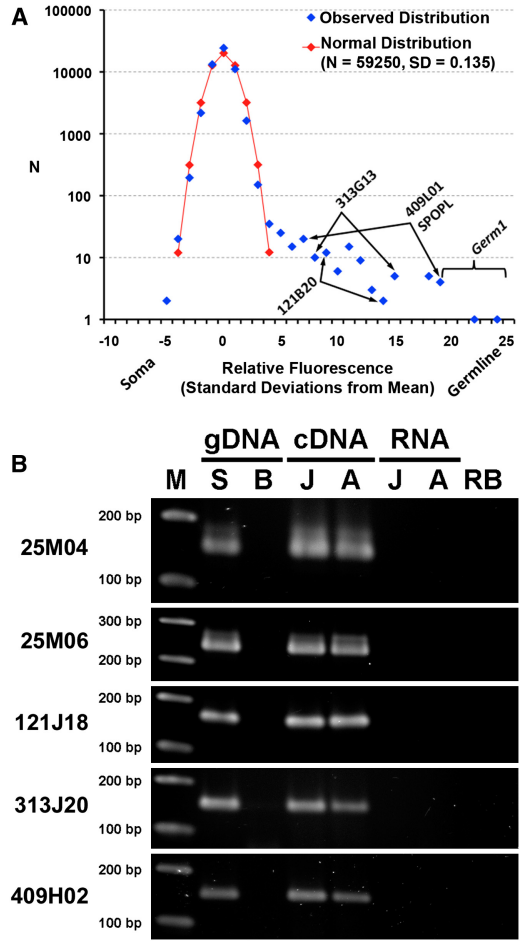

Figure 1. Summary of Germline-Specific Sequence and Gene Discovery Using ArrayCGH

(A) Germline-enriched sequences were identified by comparing observed relative hybridization intensities to a normal distribution with the same number of sampled regions (N) and SD. The y axis is plotted on $\log_{10}$ scale in order to magnify differences at the tails of the distribution. All previously discovered germline-specific sequences [1] (marked by arrows and brackets) and several additional germline-specific sequences were identified in this assay.

(B) Examples of PCR validation of single-copy sequences eliminated from soma and their expression in the germline. Sequences are present in testes gDNA (genomic DNA) but absent from blood gDNA. These same fragments can be amplified from testes cDNA, but not from the source RNA (a control for gDNA contamination) or reagent blank. S, sperm; B, blood; A, adult testes; J, juvenile testes, RB, reagent blank; M = 100 bp DNA ladder.

See also Figure S1 and Table S3.

pipeline in order to discover novel sequences and architectures present in the germline and absent from soma. This pipeline involved aligning all germline reads to all somatic (liver) reads and then computationally examining alignments to search for signatures of rearrangement (Figure 3). Different alignment patterns were considered indicative of: (1) "normal" single-copy or repetitive DNA, (2) WGS coverage gaps, (3) candidate deletion regions, and (4) candidate recombination sites. Numerous putative deletion and recombination regions were identified (41,996 "deletion" and 18,842 "recombination" reads: Table S1). Importantly, alignment of "deletion" reads to the human RefSeq data set identified 246 nonredundant gene hits for the deleted fraction ($E < 1e^{-20}$ with a total of 2,265 homology-informative reads, including several redundant alignments to zinc finger genes, which may represent multiple independent loci). This suggests that a substantial fraction of the somatically deleted DNA corresponds to single-copy and protein-coding DNA.

It should be noted that different individuals were used for the somatic WGS and germline 454 projects. This is because a female was selected for the lamprey WGS project, whereas pure germline DNA is much more readily accessible from sperm. Therefore, apparent deletion and recombination signatures could also reflect polymorphic insertion/deletion events that segregate in the lamprey population and were differentially inherited by the sequenced individuals. In order to address this potential issue, we performed further analyses on a subset of predicted gene deletions (n = 20) and recombination events (n = 28). We used PCR to test several candidate regions, focusing on predicted genes and recombination sites (Figure 3). These validation experiments revealed seven sites of programmed deletion, three recombination breakpoints, three segregating insertion/deletion polymorphisms and five WGS sequence coverage gaps. Comparison of PCR-validated breakpoint sequences reveals the presence of short 5′/3′ palindromes near the predicted breakpoint position, but no defined consensus sequence (Figure 4). The potential functionality of these is as yet unclear, but the presence of such sequences is considered strong evidence that site-specific recombination events facilitate the elimination of DNA from the lamprey genome (though chromosome loss cannot be ruled out as a contributing factor). Genes present within validated deletions included: *APOBEC-1 Complementation Factor*, *RNA Binding Motif 46* (*cancer/testis antigen 68*) and *47*, *KRAB Zinc Fingers 79* and *180*, *Lysophosphatidic Acid Receptor 1*, and *WNT7A/B*. Summaries of functional information for homologs of these genes (NCBI gene: http://www.ncbi.nlm.nih.gov/gene/) indicate their functional roles in maintenance of cell fate, cell proliferation, and oncogenesis/tumorigenesis.

To gain a better perspective on gene functions within the larger predicted deletion data set, we compared homology-derived ontology information [7] for all candidate deletions to ontology information for the remainder of the 454 shotgun data set (Table S2). Several ontologies were statistically overrepresented among predicted deletions, including categories related to regulation of gene expression, chromatin organization, and development of germ/stem cells (Figure 3; Table S2). A subset of these regions, with validated expression in meiotic testes, was also similarly enriched in transcriptional regulatory and germline developmental functions (Table S2). Coupled with the above studies, ontology analyses imply that the genomic differentiation of germline versus somatic lineages leads to differentiation in their capacities to deploy specific transcriptional programs, thereby regulating germline versus somatic cell fate.

Analyses of our germline 454 data corroborate previous findings that lamprey deletes ~20% of its genome through PGR [1, 2], though the method does not identify deletions of repetitive sequences when one or more members are retained in the soma. It is known that repetitive elements constitute a substantial fraction of lamprey germline-specific (and somatically retained) DNA [1, 2], although these and other nonfunctional single-copy regions do not necessarily contribute to the development or maintenance of germline. More importantly, our studies indicate that the deleted fraction contains a substantial complement of functional or potentially

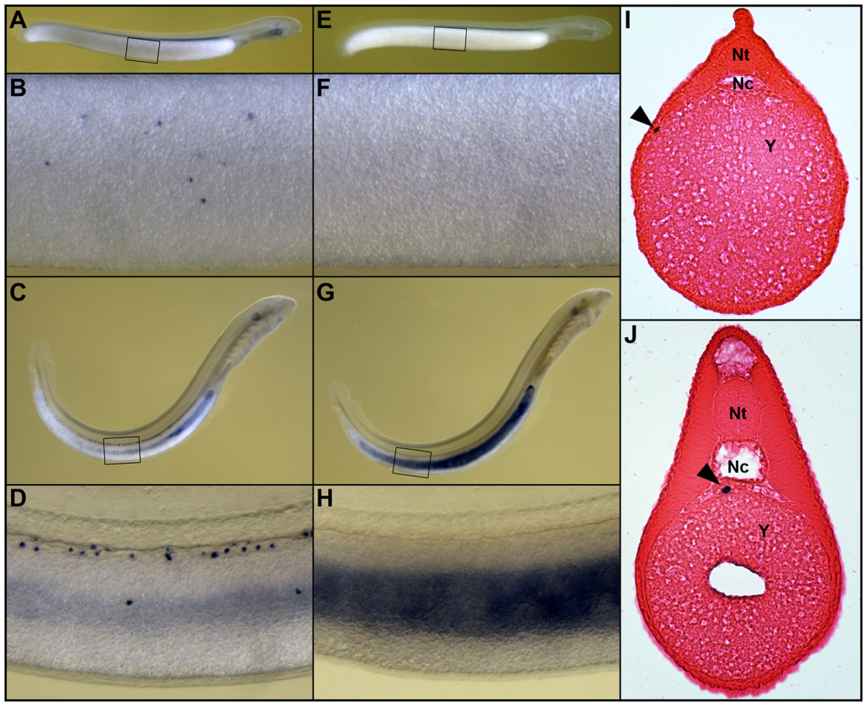

Figure 2. Expression of a Germline-Specific Marker in Embryonic Germline

In situ hybridization of an antisense probe of the germline-specific gene *25M04* (putative KRAB domain zinc finger protein) reveals expression in the developing germline cells at day 14 (A, B, I) and day 20 (C, D, J) postfertilization. Punctate staining reveals specific expression in the presumptive PGCs. Staining of PGCs is not observed in embryos that were hybridized with the sense strand probe (E and F), but some background staining is observed due to the presence of noncellular endogenous alkaline phosphatase activity in the developing gut, pharynx, notochord, and otic capsule. (B), (D), (F), and (H) correspond to the circumscribed regions in (A), (C), (E), and (G), respectively. (I) and (J) are transverse sections of the embryos shown in (A) and (C). Sections have been counterstained with eosin in order to enhance contrast; arrows mark the location of PGCs positive for the *25M04* marker. This expression pattern suggests that *25M04* is involved in some aspect of PGC differentiation and/or migration. Nc, notochord; Nt, neural tube; Y, yolk.
See also Figure S2.

functional genes: 7.6% of germline reads and 3.8% of germline gene homologies that are completely absent from the somatic WGS data set (Table S1). When interpreting these results, it is also important to note that homology information cannot identify all functional components within deleted regions. For example, the current analyses do not specifically identify recent gene duplicates, lamprey-specific genes, or functional noncoding sequences that are deleted via PGR.

Although the current data set does not identify the entire set of germline-specific genes, it seems clear that the developmentally regulated segregation of a few thousand protein-coding genes and associated regulatory elements should substantially limit the functional capacities of somatic cell lineages, relative to the germline. On the basis of gene homology, ontology, and gene expression data (Figures 1, 2, and 3), we hypothesize that DNA loss may be critical for segregating "totipotency" gene functions into the germline, thereby preventing the dysregulated deployment of germline-specific gene functions in somatic cell lineages. Notably though, several genes identified within the germline-specific fraction possess vertebrate homologs that are not currently known to function in either the development or maintenance of germline. We reason that their restriction to the germline-specific fraction of the lamprey genome, in itself, provides insight into their biological function. Specifically, the physical restriction of these genes to the lamprey germline genome implies that they (1) contribute to the development or maintenance of totipotent germline and (2) are dispensable (or deleterious) with respect to the maintenance and development of soma.

Conclusions

Genetic conflicts between germline and soma that are evident in our analyses of PGR are conceptually similar to the definition of cancer/testes genes (or cancer/testes antigens) [8, 9], although such conflicts are not necessarily limited to the development of cancer. Cancer/testes genes are diverse in evolutionary origin, but share a common feature in that they normally exhibit testes-restricted expression and are only observed in somatic tissues in the context of oncogenesis [8, 9]. From a biological standpoint, it seems plausible that misregulation of genes with germline-specific functions (recombination, unlimited proliferation, and a capacity for genomic reprogramming) could contribute to oncogenesis or other disease states [10]. Indeed, it has been shown empirically that ectopic expression of germline-specific genes can drive tumor growth in *Drosophila* [11] and *Hydractinia* [12]. In light of the differential and conflicting requirements of germline and soma, we hypothesize that PGR events might serve, in part, to segregate totipotency functions into the germline, thereby alleviating such untoward effects of these genes in the soma. Intriguingly, the conceptual similarities between cancer/testes and PGR models are seemingly further bolstered by our detection of *cancer/testis antigen 68* within the germline specific fraction of the lamprey genome. As such, the lamprey genome appears to present a large, readily identifiable, and evolutionarily informative collection of germline-limited genes that can be leveraged to understand the unique genetic requirements and pleiotropic liability of vertebrate germline.

Future studies aimed at dissecting the functionality of deleted lamprey genes and other molecular details of PGR should provide novel insights into molecular mechanisms of germline totipotency, somatic recombination, and biological tradeoffs between germline and soma. Notably, both extant lineages of jawless vertebrates (agnathans: lampreys and hagfish) are known to undergo PGR [1, 13], which would seem to indicate that the phenomenon is common to all extant agnathans and potentially represents an ancestral condition [14]. Thus, PGR may represent an ancient mechanism for moderating genetic conflict between germline and soma that evolved within an ancestral vertebrate lineage (alternately, repeated evolution of PGR in lamprey, hagfish, and

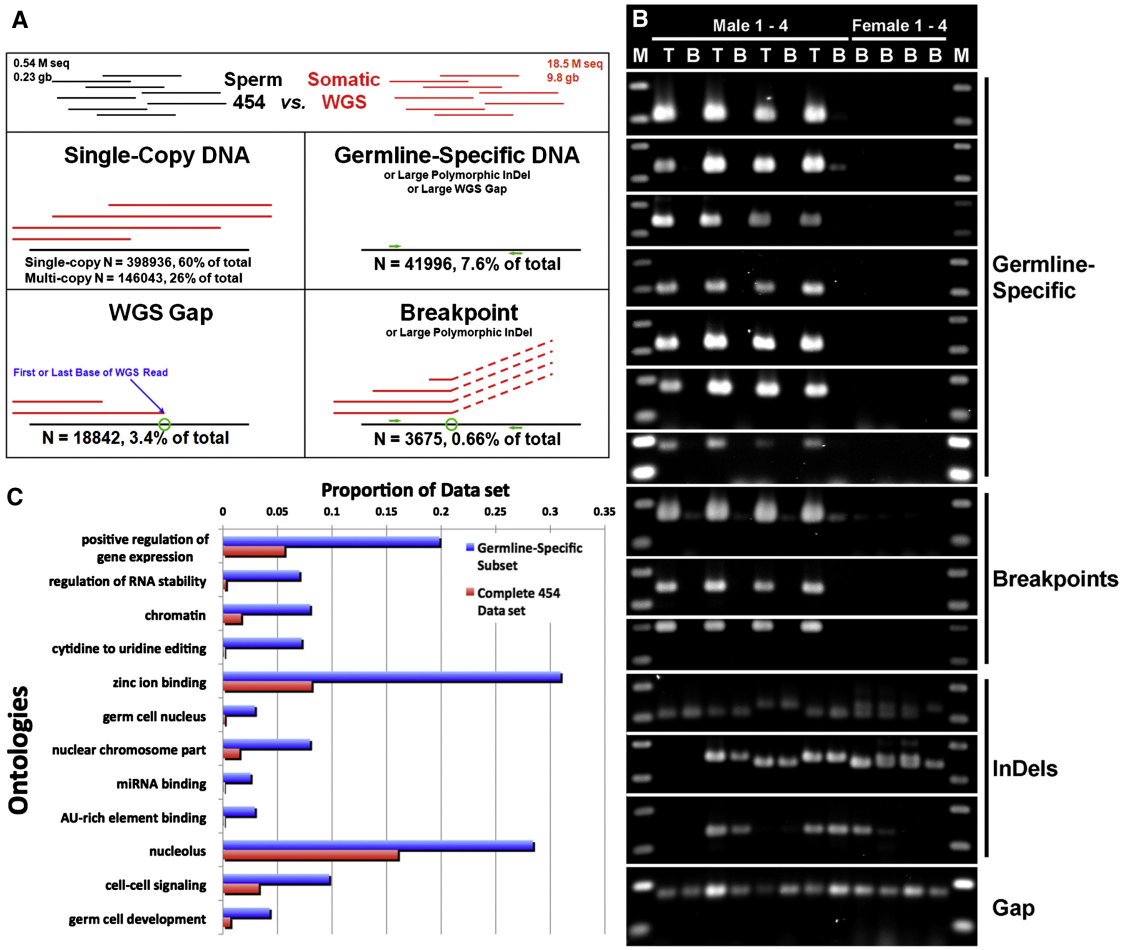

Figure 3. Analysis of Pilot 454 Sequencing Data

(A) All 454 reads were categorized on the basis of alignment patterns with the complete lamprey WGS data set (liver DNA). A majority (82%) of reads appeared as "normal" DNA (multicopy or single-copy). Other alignment patterns were consistent with coverage gaps in the WGS data set (3.4%), germline-specific DNA (7.6%), or recombination breakpoints (0.66%). Green circles depict the positions of alignment breaks and green arrows depict the generic locations of primer binding sites for validation PCRs.

(B) Results of PCR validations of germline-specific/gene-containing (BLAST hit) reads and breakpoint-flanking reads provided positive validation of members of both rearrangement classes and identified segregating (in the population) insertion/deletion (InDel) polymorphisms and WGS coverage gaps, which mimic programmed rearrangement outcomes. Note, the "Germline-Specific" and "Breakpoint" classes result in similar PCR validation patterns because one primer (breakpoint) or both primers (germline-specific) are designed to germline-specific regions. T, template is testes DNA; B, template is blood DNA; M, 100 bp DNA ladder.

(C) Overrepresented gene ontologies from 234 predicted germline-specific genes, relative to the entire 454 data set ($p > 1e-8$, corrected using false discovery rate control, as implemented by Blast2Go [9]).

See also Figure S3 and Table S4.

numerous invertebrate and protist lineages [13, 15–21] may reflect recurrent selective advantages for PGR). Under either scenario, lamprey PGR holds the potential to fill an important gap in our understanding of the cause and consequence of dysregulated rearrangement of vertebrate genomes (e.g., in oncogenesis) [22–28] and the capacity for tight regulation of genome rearrangements in phylogenetically disparate lineages [13, 15–21, 29]. The availability of a draft genome [6] and established gene knockdown/replacement methods [30–32] for lamprey should promote future progress in resolving the causes, consequences, and evolutionary relevance of PGR.

```
                Retained in Soma                    Palindrome                  Germline-Specific

Breakpoint 1    TGACGTCACGCTTGTGCGTGTGTACATAGTAGTGATGCTACCATCACTACACATGTAACATTTTGTCTATGAGGGTCAAACATTGATTAAGTGTTCAAATCTAGAT
Breakpoint 2    ACACTTTGAAAATAGAATAACATTTTTTTTTCCTATTACATGAAATTACCACCGTTTATTGTAAGAACTGGTGAGTATTGTTTGAAATCTATTTTTTTACAGAGA
Breakpoint 3 GGGGCCAACGGCCCCGACAATTCGGTGATGGCGGGGGACGGTAGCCCCACAAATGCATTTTTCTGTTTCCACCACCATAACCACAACCTTATTAATGTGTGAACAACAGA
```

Figure 4. Sequence of PCR-Validated Breakpoint Regions

Breakpoints contain short 5′/3′ palindromes (green) at the junction between somatically retained (blue) and germline-specific (red) sequence. The breakpoint of junction 2 contains an imperfect 5′/3′ palindrome. It is as yet unclear if these are functionally related to programmed genome rearrangement.

Experimental Procedures

Microarray Design, Processing, Analysis, and Validation

We designed a customized NimbleGen (Roche) microarray consisting of 385,000 oligonucleotides targeted to lamprey germline and somatic sequences. DNA samples were prepared from agarose embedded nuclei (for sperm versus blood comparisons) or whole tissues by standard phenol/chloroform extraction [33]. Soma-germline array comparative genomic hybridization experiments were performed using DNA that originated from the same individual (blood versus sperm). Additional comparisons were performed between somatic tissues to assess the extent of somatic variation (Supplemental Experimental Procedures; Figure S1). Candidate regions identified by arrayCGH or from the analysis of 454 data were further evaluated by PCR and rtPCR. Detailed methods and validation procedures are provided as supplemental materials (Supplemental Experimental Procedures). Array data were deposited in the NCBI gene expression omnibus (Table S3; http://www.ncbi.nlm.nih.gov/geo/).

454 Sequencing and Analysis

We isolated sperm genomic DNA from a single individual and outsourced 454 sequencing to the Duke IGSP Sequencing Core Facility. Sequences were trimmed to Q20 using Phred [34, 35], and any sequences less than 300 bp in length were removed from the data set prior to analysis. The remaining 554,979 sequences were aligned (BLAST) [36] to a local database of 18,506,949 trimmed sequences from the lamprey genome project (liver DNA, average trimmed read length = 529 bp at Q20) [1, 6]. Alignments were parsed using custom scripts to identify various patterns that are indicative of rearrangement, insertion/deletion polymorphisms, or sampling artifacts (Figure 3A). Alignments to human RefSeq data sets were also performed using BLAST.

Germline Transcriptome

Germline transcripts were obtained from a single male lamprey collected early in the 2009 spawning run. RNA was extracted from testes using Trizol extraction, purified using QIAGEN RNeasy midi kit, and polyA selected via two rounds affinity purification (Dynabeads mRNA Direct purification kit, Invitrogen) of sequencing was performed by the UC Davis Genome Center using a single flow cell lane on a GAIIx machine. Reads were assembled using ABySS 1.2.0 [37] and aligned to local databases of germline and somatic reads using BLAST.

In Situ Hybridization

Whole embryo in situ hybridization was performed following the methods of Nikitina et al. [32]. A fragment of the 3′ untranslated region from *25M04* (KRAB domain zinc finger gene) was amplified by PCR then cloned into the plasmid vector pCRII-TOPO (Invitrogen). Following sequencing to verify orientation, amplicons were generated using the M13 primer and one of two gene-specific primers, 25M04_insitu.F or 25M04_insitu.R (Table S4), to generate templates for synthesis of sense and antisense probes, respectively. Digoxigenin-labeled riboprobes were synthesized from these templates using the SP6 polymerase and the MAXIscript (Ambion) in vitro transcription kit.

Ontology Classification

All ontology analyses were performed using a custom ontology database, using Blast2GO [7]. The frequency of each ontology category within the predicted germline-specific subset was compared to the overall frequency of that ontology in the entire 454 data set, using best BLAST hits to SwissProt as a source of ontology information.

Accession Numbers

Microarray data for this study were deposited with the NCBI Gene Expression omnibus under accession numbers GSE23757, GPL10847 and GSM586211–GSM586216. 454 reads from sperm genomic DNA were deposited in the NCBI short read archive under accession number SRA023537.3. Transcriptomic reads were deposited in the NCBI short read archive under accession number SRA047838.2.

Supplemental Information

Supplemental Information includes three figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2012.06.028.

Acknowledgments

We thank M. Bronner and T. Sauka-Spengler for access to their lamprey husbandry facilities at the California Institute of Technology. We also thank M.C. Yao for his valuable discussions regarding programmed genome rearrangement. This work was supported by the National Institutes of Health (grant numbers GM079492, GM090049, and RR014085) and the National Science Foundation (grant number MCB-0719558) to C.T.A., the National Institutes of Health (grant numbers T32-HG00035, F32-GM087919) to J.J.S., and the National Institutes of Health (grant number GM58815) to E.E.E. E.E.E. is an Investigator of the Howard Hughes Medical Institute and on the scientific advisory boards for Pacific Biosciences, Inc., SynapDx Corp, and DNAnexus, Inc. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of NIH. Dissection of tissues was performed at Benaroya Research Institute under IACUC protocol 06AM01. Requests for reagents should be addressed to J.J.S. (jjsmit3@uky.edu) or C.T.A. (camemiya@benaroyaresearch.org).

Received: March 6, 2012
Revised: May 3, 2012
Accepted: June 7, 2012
Published online: July 19, 2012

References

1. Smith, J.J., Antonacci, F., Eichler, E.E., and Amemiya, C.T. (2009). Programmed loss of millions of base pairs from a vertebrate genome. Proc. Natl. Acad. Sci. USA *106*, 11212–11217.
2. Smith, J.J., Stuart, A.B., Sauka-Spengler, T., Clifton, S.W., and Amemiya, C.T. (2010). Development and analysis of a germline BAC resource for the sea lamprey, a vertebrate that undergoes substantial chromatin diminution. Chromosoma *119*, 381–389.
3. Okkelberg, P. (1921). The early history of the germ cells in the brook lamprey, *Entosphenus wilderi* (GAGE), up to and including the period of sex differentiation. J. Morphol. *35*, 1–151.
4. Butcher, E.O. (1929). The origin of the germ cells in the lake lamprey (*Petromyzon marinus unicolor*). Biol. Bull-Us *56*, 87–99.
5. Hardisty, M.W., and Cosh, J. (1966). Primordial germ cells and fecundity. Nature *210*, 1370–1371.
6. Washington University Genome Sequencing C. (2007). http://genome.wustl.edu/pub/organism/Other_Vertebrates/Petromyzon_marinus/assembly/Petromyzon_marinus-3.0. Washington University School of Medicine.
7. Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. *36*, 3420–3435.
8. Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T., and Old, L.J. (2005). Cancer/testis antigens, gametogenesis and cancer. Nat. Rev. Cancer *5*, 615–625.
9. Hofmann, O., Caballero, O.L., Stevenson, B.J., Chen, Y.T., Cohen, T., Chua, R., Maher, C.A., Panji, S., Schaefer, U., Kruger, A., et al. (2008). Genome-wide analysis of cancer/testis gene expression. Proc. Natl. Acad. Sci. USA *105*, 20422–20427.
10. Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57–70.
11. Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression of germline genes drives malignant brain tumor growth in *Drosophila*. Science *330*, 1824–1827.
12. Millane, R.C., Kanska, J., Duffy, D.J., Seoighe, C., Cunningham, S., Plickert, G., and Frank, U. (2011). Induced stem cell neoplasia in a cnidarian by ectopic expression of a POU domain transcription factor. Development *138*, 2429–2439.
13. Nakai, Y., Kubota, S., and Kohno, S. (1991). Chromatin diminution and chromosome elimination in four Japanese hagfish species. Cytogenet. Cell Genet. *56*, 196–198.
14. Smith, J.J., Saha, N.R., and Amemiya, C.T. (2010). Genome biology of the cyclostomes and insights into the evolutionary biology of vertebrate genomes. Integr. Comp. Biol. *50*, 130–137.
15. Yao, M.C., Yao, C.H., Halasz, L.M., Fuller, P., Rexer, C.H., Wang, S.H., Jain, R., Coyne, R.S., and Chalker, D.L. (2007). Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in *Tetrahymena*. J. Cell Sci. *120*, 1978–1989.

16. Nowacki, M., Higgins, B.P., Maquilan, G.M., Swart, E.C., Doak, T.G., and Landweber, L.F. (2009). A functional role for transposases in a large eukaryotic genome. Science *324*, 935–938.
17. Bachmann-Waldmann, C., Jentsch, S., Tobler, H., and Müller, F. (2004). Chromatin diminution leads to rapid evolutionary changes in the organization of the germ line genomes of the parasitic nematodes *A. suum* and *P. univalens*. Mol. Biochem. Parasitol. *134*, 53–64.
18. Rasch, E.M., Wyngaard, G.A., and Connelly, B.A. (2008). Heterochromatin endoreduplication prior to gametogenesis and chromatin diminution during early embryogenesis in *Mesocyclops edax* (Copepoda: Crustacea). J. Morphol. *269*, 387–397.
19. Boveri, T. (1887). Uber Differenzierung der Zellkerne wahrend der Furchung des Eies von *Ascaris megalocephala*. Anat. Anz. *2*, 688–693.
20. Yao, M.C., and Gorovsky, M.A. (1974). Comparison of the sequences of macro- and micronuclear DNA of *Tetrahymena pyriformis*. Chromosoma *48*, 1–18.
21. Lauth, M.R., Spear, B.B., Heumann, J., and Prescott, D.M. (1976). DNA of ciliated protozoa: DNA sequence diminution during macronuclear development of *Oxytricha*. Cell *7*, 67–74.
22. Kolomietz, E., Meyn, M.S., Pandita, A., and Squire, J.A. (2002). The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. Genes Chromosomes Cancer *35*, 97–112.
23. Flores, M., Morales, L., Gonzaga-Jauregui, C., Domínguez-Vidaña, R., Zepeda, C., Yañez, O., Gutiérrez, M., Lemus, T., Valle, D., Avila, M.C., et al. (2007). Recurrent DNA inversion rearrangements in the human genome. Proc. Natl. Acad. Sci. USA *104*, 6099–6106.
24. Strout, M.P., Marcucci, G., Bloomfield, C.D., and Caligiuri, M.A. (1998). The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia. Proc. Natl. Acad. Sci. USA *95*, 2390–2395.
25. Elliott, B., Richardson, C., and Jasin, M. (2005). Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. Mol. Cell *17*, 885–894.
26. Raphael, B.J., Volik, S., Yu, P., Wu, C., Huang, G., Linardopoulou, E.V., Trask, B.J., Waldman, F., Costello, J., Pienta, K.J., et al. (2008). A sequence-based survey of the complex structural organization of tumor genomes. Genome Biol. *9*, R59.
27. Stephens, P.J., McBride, D.J., Lin, M.L., Varela, I., Pleasance, E.D., Simpson, J.T., Stebbings, L.A., Leroy, C., Edkins, S., Mudie, L.J., et al. (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature *462*, 1005–1010.
28. Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., et al. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat. Genet. *40*, 722–729.
29. Nemetschke, L., Eberhardt, A.G., Hertzberg, H., and Streit, A. (2010). Genetics, chromatin diminution, and sex chromosome evolution in the parasitic nematode genus *Strongyloides*. Curr. Biol. *20*, 1687–1696.
30. Sauka-Spengler, T., Meulemans, D., Jones, M., and Bronner-Fraser, M. (2007). Ancient evolutionary origin of the neural crest gene regulatory network. Dev. Cell *13*, 405–420.
31. Nikitina, N., Sauka-Spengler, T., and Bronner-Fraser, M. (2008). Dissecting early regulatory relationships in the lamprey neural crest gene network. Proc. Natl. Acad. Sci. USA *105*, 20083–20088.
32. Nikitina, N., Bronner-Fraser, M., and Sauka-Spengler, T. (2009). The sea lamprey *Petromyzon marinus*: a model for evolutionary and developmental biology. In Emerging Model Organisms: A Laboratory Manual, *Volume 1* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 405–429.
33. Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual, *Volume 3* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
34. Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. *8*, 186–194.
35. Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. *8*, 175–185.
36. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
37. Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., et al. (2009). De novo transcriptome assembly with ABySS. Bioinformatics *25*, 2872–2877.