Published in final edited form as:

Syst Biol Reprod Med. 2011 February ; 57(1-2): 54–62. doi:10.3109/19396368.2010.506000.

Interrogating the transgenic genome: Development of an interspecies tiling array

Graham D. Johnson¹, Adrian E. Platts^{1,2}, Claudia Lalancette^{1,2}, Robert Goodrich^{1,2}, Henry H. Heng¹, and Stephen A. Krawetz^{1,2}

¹The Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, MI, USA

²Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI, USA

Abstract

A single expressing copy of the human protamine domain was randomly inserted into an intron of Cyp2c38. The transgenic locus was shown to recapitulate the level of expression observed in normal human testis while not perturbing endogenous protamine expression. The development of an interspecies tiling array was pursued to enable direct comparison of the orthologous protamine domains in a single experiment. Probe design was adapted to generate species-specific high resolution probe sets that would tolerate repetitive elements. Results from competitive hybridizations demonstrate that interspecies tiling arrays are a valuable tool for parallel analysis of highly similar DNA sequences. This approach provides a rapid and reliable means of interrogating samples prior to deep sequencing analysis. These arrays should readily complement most DNA isolation techniques such as ChIP, nuclease sensitivity and nuclear matrix association assays.

Keywords

protamine; tiling array; transgenic

INTRODUCTION

Spermiogenesis, the differentiation of a haploid spermatid into a mature spermatozoon, is characterized by dramatic morphological changes including a marked reduction in nuclear volume [Hecht 1998]. In mammals, this remodeling of the sperm nucleus results from the compaction of the genome through the replacement of histones by sperm-specific proteins (reviewed in [Braun 2001; Miller et al. 2010]). The histone-protamine transition initiates with the global acetylation of histones facilitating deposition of the transition nuclear proteins (TNPs) and sperm-specific histone variants [Awe and Renkawitz-Pohl 2010; Churikov et al. 2004; Kurtz et al. 2007; Meistrich et al. 1992; Pivot-Pajot et al. 2003]. In elongating spermatids the transition proteins and the majority of histones that remain are replaced by protamines, resulting in the condensed paternal genome [Steger 1999]. Despite the persistence of nucleosome bound regions, sperm are transcriptionally quiescent [Kierszenbaum and Tres 1975].

Address correspondence to Stephen A. Krawetz, Ph.D., Wayne State University School of Medicine, 275 E. Hancock, C.S. Mott Center, Detroit, MI, USA 48201. steve@compbio.med.wayne.edu.

Protamines are the most abundant nuclear proteins in the sperm of many species [Oliva 2006; Oliva and Dixon 1991]. In mammals, the genes encoding these proteins are found in a multigenic cluster containing protamine 1 (*PRM1*), protamine 2 (*PRM2*), and transition nuclear protein 2 (*TNP2*) on chromosomes 16p13.2 and 16qA1 in human and mouse, respectively [Nelson and Krawetz 1993; 1995; Reeves et al. 1989]. The coding regions are well conserved between mice and men [Kramer et al. 1998a; Krawetz and Dixon 1988]. Following meiosis, the genes are transcribed in round spermatids and subsequently translated by the elongating spermatids [Hecht 1988; Kleene 1996]. Expression of these genes is essential as perturbation compromises fertility or results in sterility [Balhorn et al. 1988; Belokopytova et al. 1993a; 1993b; Cho et al. 2001; de Yebra et al. 1993; 1998; Steger et al. 1999; Yu et al. 2000; Zhao et al. 2001]. In several species, this gene cluster contains a fourth open reading frame, protamine 3 (*PRM3*) [Balhorn 2007; Nelson and Krawetz 1994]. In mouse, this gene is believed to encode a small cytoplasmic acidic protein transcribed in early round spermatids. Mice lacking *Prm3* are fertile despite deficiencies in sperm motility [Grzmil et al. 2008].

The human protamine locus, encompassing all *cis* elements necessary for expression, is approximately 40 kb in length. Introduction of this entire sequence into the mouse genome retains native temporal and tissue expression independent of the site of insertion [Choudhary et al. 1995; Martins et al. 2004]. The humanized mouse produces no phenotypic abnormalities. Within the extended locus lies a ~28 kb DNase I sensitive domain containing *PRM1*, *PRM2*, and *TNP2*. In both human and mouse this linear gene array is flanked by boundary elements, which have been shown to be essential for expression [Choudhary et al. 1995; Wykes and Krawetz 2003a]. Mutations in the 3' boundary element have been correlated with decreased protamine expression in two infertile males [Kramer et al. 1997]. A similar decrease in locus expression was observed in transgenic mice harboring a copy of the human protamine domain lacking the same 3' region [Martins et al. 2004].

Expression of the protamine domain is preceded by potentiation of the locus in pachytene spermatocytes from a closed repressed conformation to an open accessible state that is then permissible to the trans-factor binding necessary to initiate expression [Kramer et al. 1998; 2000]. Once potentiated the open chromatin conformation persists throughout spermiogenesis, which may reflect specific nucleosome retention [Choudhary et al. 1995; Kramer et al. 2000; Wykes and Krawetz 2003].

Though the total amount of histone remaining in mature spermatozoa varies between species, those regions that remain bound by nucleosomes are not random [Gatewood et al. 1987; Wykes and Krawetz 2003; Zalenskaya et al. 2000]. Promoter regions of genes essential to embryonic development are particularly enriched [Arpanahi et al. 2009; Gardiner-Garden et al. 1998; Hammoud et al. 2009]. Indeed, it has been suggested that nucleosome retention within specific genes may influence or direct the initial expression of the paternal genome during early embryonic development [Gatewood et al. 1987].

Three techniques have been instrumental in furthering our understanding of the relationship between primary sequence, epigenetic modification, and the resultant chromatin structure. PCR assays, now commonplace, are useful for determining the representation of a specific sequence within a pool of DNA. The specificity, sensitivity, and relative speed of this approach routinely permits the selective enrichment of singular DNA sequences up to 3 kb in length [Erlich et al. 1991]. However, this assay is an inefficient means of enriching all individual DNA sequences comprising a large genomic tract. DNA microarrays and next generation sequencing are currently the most reliable methods of efficiently achieving the coverage needed to capture entire genic domains, individual chromosomes, or even entire genomes. Though current sequencing technologies are capable of achieving greater

resolution and higher coverage than DNA microarrays, the associated cost is currently prohibitive for many laboratories. Microarray strategies are cost efficient, when restricting coverage to defined regions of interest, but this technology is not without limitations. Commercial array platforms commonly omit repetitive regions from analysis due to the inability to accurately distinguish between highly related sequences. Similarly, this logic would suggest that representation of orthologous sequences from more than one species on a single array should be avoided.

Prior to undertaking comparative studies of locus control in the transgenic system an efficient complimentary method of analysis was required. Accordingly, an orthologous mouse domain interspecies tiling array was developed to enable one to simultaneously query the many repetitive regions of the human transgenic protamine locus [Nelson and Krawetz 1994]. The maskless method of array design outlined in this study should be readily adaptable to the interrogation of most non-unique sequences. Importantly, interspecies tiling array technology fundamentally differs from cross-species hybridization (CSH) [Bar-Or et al. 2007]. An adaptation of traditional array technology, CSH utilizes a reference genome differing from that of the target DNA to determine the extent of sequence conservation between species [Dumas et al. 2007; Flynn and Carr 2007]. This technique has also been performed when a species of interest lacks a representative commercially available platform [Li et al. 2008]. Though useful in these instances CSH is unable to quantitatively ascertain the relative ratios of homologous targets within samples [Gilad et al. 2005]. The proof of principle demonstrated in this communication establishes the interspecies tiling array as a novel genomic tool capable of rapid parallel analysis of orthologous sequences within many transgenic models.

RESULTS AND DISCUSSION

Transgene Insertion and Localization

A 40 kb segment of human cosmid hp3.1 containing the protamine locus was released by digestion with SalI and EagI. This sequence containing the entire genic domain flanked up- and downstream by boundary elements was inserted into the mouse genome by DNA microinjection into fertilized oocytes [Martins et al. 2004]. As shown in Figure 1, *in situ* hybridization of a transgene-specific fluorescent probe localized the insertion site to region 3 of mouse chromosome 19. The presence of four foci within the stained metaphase chromosomes affirms a single insertion event. This was confirmed by establishing the ratio of the human and mouse gene copy number (Table 1; [Platts et al. 2008]) utilizing species-specific primers shown in Table 1. Gene equivalency coupled with the results from the FISH analysis confirmed that the transgenic construct inserted as a single copy.

Terminal transferase-dependent PCR (TTD-PCR) was subsequently utilized to establish the genomic coordinates of the site of insertion [Chen et al. 2000; 2001]. Sequencing of the TTD-PCR reaction products evidenced the faithful incorporation of the complete human protamine locus within the seventh intron of cytochrome P450 2c38 (Cyp2c38), as shown in Figure 2. Cyp2c38 is expressed throughout mouse, but not in testes [Su et al. 2002]. The absence of phenotypic abnormalities suggests that expression of the cytochrome gene was not perturbed and/or that the translated product from the inserted locus is functionally redundant in mice.

Testis Expression of a Domain within a Domain

The relative levels of mRNA for each member of the endogenous and transgenic protamine loci were assessed by quantitative real-time PCR. Endogenous protamine expression was not perturbed by the presence of the orthologous human locus (Fig. 3A). As illustrated in Figure

3B, the relative level of the transgenes mirrored that observed in the normal healthy male. Importantly, expression was confined to testis, recapitulating the tissue specific pattern of expression observed in men [Martins et al. 2004].

It remains to be ascertained whether the human orthologs are translated. Abnormally high or low protamine ratios are correlated with infertility in men [Aoki et al. 2005]. An increased pool of total protamine mRNA could affect the concentration of protamine 1 and 2. However, neither gross sperm chromatin packaging nor fertility were altered in the transgenic animals, suggesting that if the human proteins are translated they confer no deleterious effect.

Interspecies Tiling Array

The development of an interspecies tiling array was pursued to simultaneously interrogate the orthologous protamine domains in the transgenic mouse model. Of particular interest were the boundary elements flanking the human protamine locus (Fig. 2). Both elements lie within repetitive regions, which are often excluded from commercial arrays. To address this constraint a maskless array synthesis strategy was adopted [Graf et al. 2007; source code is available under an open source license from <http://www.ebi.ac.uk/~graef/arraydesign/>]. As shown below, this strategy generated high resolution probe sets of sufficient specificity to discreetly capture sequences from functional domains independent of repetitive elements or sequence conservation. By utilizing this approach the largest gap within either of the protamine domains was 48 bp.

The efficiency of the tiling array is enhanced by the ability to achieve simultaneous high resolution coverage of multiple regions of interest (Table 2). Other loci were chosen for representation on the array based on their role in spermatogenesis or embryo development. These included the Hox A, B, C, and D clusters, acrosin, phosphoglycerate kinase 1 and 2 (*PGK1* and 2), and α -globin. The Hox genes, which encode a family of potent developmental regulators, have recently been shown in human and mouse spermatozoa to retain a significant level of histones [Arpanahi et al. 2009; Deschamps 2007; Hammoud et al. 2009; Wellik 2009]. Transcribed during the meiotic phase of spermatogenesis, acrosin is also nucleosome enriched in mature sperm [Kashiwabara et al. 1990]. *Pgk1* is somatically expressed. Its transcription is terminated by X-inactivation at the start meiosis at which point *Pgk2* is expressed from chromosome 17 [McCarrey et al. 1992]. In contrast to the closed chromatin conformations of acrosin, *PGK1/2* at the termination of spermatogenesis, the *PRM1*, *PRM2*, *TNP2*, and α -globin genes are found in a potentiated DNase I-sensitive open chromatin conformation [Kramer and Krawetz 1996; Kramer et al. 2000; Krawetz et al. 1999]. These established differences in histone retention, chromatin structure, and in the timing of potentiation provide a representative sampling of spermatogenesis.

As summarized in Figure 4, array design was a multistep process beginning with the initial indexing of the mouse genome and the extended human protamine domain. Using a 14 bp sliding window with a 1 bp step the number of iterations of all 14 bp subsequences was determined. Each genic domain was divided into potential probe sequences using a 55 bp sliding window with a 1 bp step. All potential probe sequences were then aligned to the genomes of interest. Probe sequences were rejected if they aligned outside of the region of interest from which they were derived. The remaining candidate probes were subsequently assigned a quality score reflecting overall sequence complexity and the influence of base distribution on hybridization thermodynamics.

The quality score of a probe was penalized for each multiple occurrence of a 14 bp subsequence within the murine genome or human protamine domain. Probes were also penalized for the presence of subregions containing sequences that are problematic to

synthesize or prone to GC clamping. In order to prevent the formation of intra-molecular bonds, probes exceeding a set palindromic sequence threshold were penalized. Lastly, probes with a salt-adjusted melting temperature deviating from a set isothermic range were penalized [Sambrook et al. 1989]. Probes specific to each domain were then selected on the basis of this score and their location with respect to neighboring probes. In some cases the lack of sequence complexity resulted in minimal representation of that segment by high quality probes. A lower quality probe was then used if its absence would otherwise result in a gap of coverage exceeding the maximum allowable length (99 bp). The assignment of a quality score to all probes allowed for *a priori* predictions of probe performance prior to sample hybridization. Such information was useful in instances of large signal discrepancies between neighboring probes. Utilizing this method of array design, 43,020 oligonucleotide probe sequences were generated and synthesized on the Agilent 4 × 44 K custom array platform. This approach can be readily adapted to any genome.

As shown in Figure 5, competitive hybridization of wild-type mouse and human genomic DNA samples to the interspecies tiling array demonstrated the efficacy of the species-specific probe design. Essentially the entire signal throughout the mouse protamine probe set corresponds to the hybridization of mouse DNA (Fig. 5 A). As expected, this is mirrored by the probe set targeting the extended human domain hybridized to human DNA (Fig. 5 B). Probe specificity due to hybridization of the transgenic DNA is clearly shown in Figure 5C. Upon competitive hybridization of the transgenic DNA containing the human locus to the 110,050 bp region of human DNA that encompasses the domain. The signals from the 2,282 probes representing the inserted domain are essentially exclusive to the human segment. As expected, probes corresponding to regions flanking the inserted sequence display levels of hybridization which are below background as these sequences are essentially equally matched between the samples. This clearly demonstrates the specificity of the probes and the utility of this probe design strategy.

Development of an interspecies tiling array provides an ideal tool to compliment the transgenic mouse model of the human protamine locus. The design strategy outlined in this study generated probe sets capable of simultaneously querying the orthologous domains in a single experiment independent of repetitive elements. With proof of principle now established this approach will move forward to begin dissecting the mechanisms that regulate the selective expression of the protamine locus.

MATERIALS AND METHODS

Transgenic Animals

All live animal protocols were approved by Wayne State University Animal Investigation Committee A 02-04-08. Transgenic animals were generated by pronuclear microinjection as described previously [Martins et al. 2004]. Briefly, restriction endonuclease digestion of cosmid hp3.1 with SalI and EagI released an approximately 40 kb fragment of DNA containing the complete human protamine domain. Following microinjection of purified DNA into fertilized C57BL/6 eggs, newborn pups harboring the transgene were detected by PCR [Brinster et al. 1985]. Technical replicates of triplicate reactions were repeated three times. Offspring from a transgenic founder and wild-type mates were bred to homozygosity.

Transgene copy number was established by real time PCR of serially diluted tail clip DNA. Primers sets specifically targeting a ~350 bp region within either the human or mouse protamine domain were utilized in separate reactions (Table 1). The relative template values from all reactions were determined using the KLab PCR algorithm [Platts et al. 2008]. A single copy insertion was considered to be represented by a 1:1 ratio of transgenic and endogenous template values.

FISH Analysis

Transgenic mouse lymphocytes were isolated from spleen and cultured in supplemented RPMI. Slides of colcemid treated cells were prepared by conventional means of hypotonic swelling and fixation. The inserted human locus was detected by labeled cosmid clone hp3.1 as described previously [Heng et al. 1992; Heng and Tsui 1993; Schmid et al. 2001]. One hundred fields were evaluated using a fluorescent microscope. Images were captured using a CCD camera and analyzed with RS Image software (Photometrics, Surrey, BC, Canada).

Site of Insertion

The site of insertion was fine mapped using terminal transferase dependent PCR (TTD-PCR) [Chen et al. 2000; 2001]. A distal end of the transgene and neighboring endogenous sequence was linearly amplified using a biotinylated primer and the Clontech Advantage HD Taq Polymerase system (*PRIMER SEQs*; Clontech, Mountain View, CA, USA). The resultant blunt end reaction products were enriched by streptavidin capture followed by the addition of a riboguanosine tail (Promega, Madison, WI, USA) by terminal deoxynucleotidyl transferase (Invitrogen, Carlsbad, CA, USA). Ligation of a known linker sequence by T4 DNA ligase (Roche; Madison, WI, USA) added additional priming sites to ribo-tailed DNA. Reaction products from nested primers were sequenced (Table 1).

Expression Analysis

Total RNA was isolated from whole transgenic and non-transgenic testes. Tissue was homogenized using a PRO Scientific 200 homogenizer (PROScientific Inc., Oxford, CT, USA). RNA was purified using the Qiagen RNeasy total RNA purification system (Qiagen, Valencia, CA, USA). Genomic DNA contamination was removed by DNase treatment with RNase-free DNase (Roche). Digestion was terminated by the addition of EDTA pH 8.0 to a final concentration of 5 mM. Integrity of the purified RNAs was determined using the Agilent Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). RNA was then quantified with Ribogreen (Invitrogen) [Goodrich et al. 2003]. A 5 µg aliquot of total RNA was reverse transcribed using oligo (dT)₁₂₋₁₈ and Superscript III™ reverse transcriptase (Invitrogen) as described by the manufacturer. Quantitative PCR reactions were performed for 50 cycles in a 20 µL volume utilizing 2 ng cDNA template, a final primer concentration of 1 µM and the Sigma SYBR® Green JumpStart™ Taq™ ReadyMix system (Sigma-Aldrich, St. Louis, MO, USA). Reactions were evaluated on the Chromo4 real time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA) and all results were analyzed using KLab-PCR to determine relative template values [Platts et al. 2008]. Values are presented as log transformed. Primer sequences are provided in Table 1.

Probe Design

A multistep design approach was implemented to generate high resolution species specific probe sets targeting functional domains, independent of repetitive elements [Graf et al. 2007; source code is available under an open source license from <http://www.ebi.ac.uk/~graef/arraydesign/>]. Targeted domains in mouse included: protamine (chromosome 16: 10,782,515 – 10,802,516), acrosin (chromosome 15: 89398500-89405500), Hox A (chromosome 6; 52080000 – 52220001), Hox B (chromosome 11; 96090000-96220001), Hox C (chromosome 15; 102720000-102920001), Hox D (chromosome 2; 74460000, 74600001), phosphoglycerate kinase 1 (chromosome 20; 102385000-102410001), and phosphoglycerate kinase 2 (chromosome 17; 39668500-39673501). Probes targeting sequences within an extended human protamine domain were also designed (chromosome 16: 11,312,500 – 11,452,413).

Genomes of interest were indexed utilizing a sliding 14 bp window with a 1 bp step. The number of iterations of each 14 bp subsequence throughout the entire genome was recorded. Regions of interest were divided into potential probes using a 55 bp sliding window with a 1 bp overlap. Each potential probe sequence was aligned to the genome(s) of interest and all nonspecific sequences were rejected. Remaining probes were ranked based on sequence complexity, adherence to an optimized T_m , and absence of GC rich subregions or sequences prone to intramolecular hybridization. A total of 43,020 oligonucleotide probe sequences were synthesized by Agilent utilizing their 4 × 44 K custom CGH array platform (Agilent Technologies, Inc.). The final suite of probe sequences is available in Supplemental Table 1 (see online edition).

CGH Array Profiling

Genomic DNA was isolated from HeLa cells, wild-type and transgenic tailclips by phenol-chloroform extraction and subsequently by ethanol precipitation [Sambrook et al. 1989]. Isolated samples were fragmented, labeled, and hybridized according to the manufacturer's instructions (User Manual version 3.1; Agilent Technologies, Inc.). Hybridizations and array scanning were performed by the Toxicology Core Facility at Wayne State University School of Medicine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

<i>Prm</i>	protamine
<i>Tnp</i>	transition nuclear protein
PCR	polymerase chain reaction
TTD-PCR	terminal transferase-dependent PCR
PGK	phosphoglycerate kinase
CSH	cross-species hybridization

Acknowledgments

This work is supported in part by the NIH grant HD36512. GDJ is supported by a Wayne State School of Medicine graduate fellowship. The authors would like to thank Frédéric Leduc for his review of this manuscript.

REFERENCES

- Aoki VW, Liu L, Carrell DT. Identification and evaluation of a novel sperm protamine abnormality in a population of infertile males. *Hum Reprod* 2005;20:1298–1306. [PubMed: 15705617]
- Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, et al. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* 2009;19:1338–1349. [PubMed: 19584098]
- Awe S, Renkawitz-Pohl R. Histone H4 Acetylation is Essential to Proceed from a Histone- to a Protamine-based Chromatin Structure in Spermatid Nuclei of *Drosophila melanogaster*. *Syst Biol Reprod Med* 2010;56:44–61. [PubMed: 20170286]
- Balhorn R. The protamine family of sperm nuclear proteins. *Genome Biol* 2007;8:227. [PubMed: 17903313]
- Balhorn R, Reed S, Tanphaichitr N. Aberrant protamine 1/protamine 2 ratios in sperm of infertile human males. *Experientia* 1988;44:52–55. [PubMed: 3350120]

- Bar-Or C, Czosnek H, Koltai H. Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends Genet* 2007;23:200–207. [PubMed: 17313995]
- Belokopytova IA, Kostyleva EI, Tomilin AN, Vorob'ev VI. Human male infertility may be due to a decrease of the protamine P2 content in sperm chromatin. *Mol Reprod Dev* 1993a;34:53–57. [PubMed: 8418817]
- Belokopytova IA, Kostyleva EI, Tomilin AN, Vorob'ev VI. Human male sterility can be caused by mutations in the protamine P2 gene. *Tsitologiya* 1993b;35:61–67. [PubMed: 8328025]
- Braun.
- Brinster RL, Chen HY, Trumbauer ME, Yagle MK, Palmiter RD. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *Proc Natl Acad Sci USA* 1985;82:4438–4442. [PubMed: 3892534]
- Chen HH, Castanotto D, LeBon JM, Rossi JJ, Riggs AD. In vivo, high-resolution analysis of yeast and mammalian RNA-protein interactions, RNA structure, RNA splicing and ribozyme cleavage by use of terminal transferase-dependent PCR. *Nucleic Acids Res* 2000;28:1656–1664. [PubMed: 10710433]
- Chen HH, Kontaraki J, Bonifer C, Riggs AD. Terminal transferase-dependent PCR (TDPCR) for in vivo UV photofootprinting of vertebrate cells. *Sci STKE* 2001;2001:11.
- Cho C, Willis WD, Goulding EH, Jung-Ha H, Choi YC, Hecht NB, et al. Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat Genet* 2001;28:82–86. [PubMed: 11326282]
- Choudhary SK, Wykes SM, Kramer JA, Mohamed AN, Koppitch F, Nelson JE, et al. A haploid expressed gene cluster exists as a single chromatin domain in human sperm. *J Biol Chem* 1995;270:8755–8762. [PubMed: 7721781]
- Churikov D, Zalenskaya IA, Zalensky AO. Male germline-specific histones in mouse and man. *Cytogenet Genome Res* 2004;105:203–214. [PubMed: 15237208]
- de Yebra L, Ballesca JL, Vanrell JA, Bassas L, Oliva R. Complete selective absence of protamine P2 in humans. *J Biol Chem* 1993;268:10553–10557. [PubMed: 8486708]
- de Yebra L, Ballesca JL, Vanrell JA, Corzett M, Balhorn R, Oliva R. Detection of P2 precursors in the sperm cells of infertile patients who have reduced protamine P2 levels. *Fertil Steril* 1998;69:755–759. [PubMed: 9548169]
- Deschamps J. Ancestral and recently recruited global control of the Hox genes in development. *Curr Opin Genet Dev* 2007;17:422–427. [PubMed: 17870464]
- Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, et al. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res* 2007;17:1266–1277. [PubMed: 17666543]
- Erich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction. *Science* 1991;252:1643–1651. [PubMed: 2047872]
- Flynn SM, Carr SM. Interspecies hybridization on DNA resequencing microarrays: efficiency of sequence recovery and accuracy of SNP detection in human, ape, and codfish mitochondrial DNA genomes sequenced on a human-specific MitoChip. *BMC Genomics* 2007;8:339. [PubMed: 17894875]
- Gardiner-Garden M, Ballesteros M, Gordon M, Tam PP. Histone- and protamine-DNA association: conservation of different patterns within the beta-globin domain in human sperm. *Mol Cell Biol* 1998;18:3350–3356. [PubMed: 9584175]
- Gatewood JM, Cook GR, Balhorn R, Bradbury EM, Schmid CW. Sequence-specific packaging of DNA in human sperm chromatin. *Science* 1987;236:962–964. [PubMed: 3576213]
- Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP. Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res* 2005;15:674–680. [PubMed: 15867429]
- Goodrich RJ, Ostermeier GC, Krawetz SA. Multitasking with molecular dynamics Typhoon: quantifying nucleic acids and autoradiographs. *Biotechnol Lett* 2003;25:1061–1065. [PubMed: 12889815]
- Graf S, Nielsen FG, Kurtz S, Huynen MA, Birney E, Stunnenberg H, et al. Optimized design and assessment of whole genome tiling arrays. *Bioinformatics* 2007;23:95–204.

- Grzmil P, Boinska D, Kleene KC, Adham I, Schluter G, Kamper M, et al. Prm3, the fourth gene in the mouse protamine gene cluster, encodes a conserved acidic protein that affects sperm motility. *Biol Reprod* 2008;78:958–967. [PubMed: 18256328]
- Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 2009;460:473–478. [PubMed: 19525931]
- Hecht NB. Molecular mechanisms of male germ cell differentiation. *Bioessays* 1998;20:555–561. [PubMed: 9723004]
- Hecht NB. Post-meiotic gene expression during spermatogenesis. *Prog Clin Biol Res* 1988;267:291–313. [PubMed: 3070565]
- Heng HH, Squire J, Tsui LC. High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. *Proc Natl Acad Sci USA* 1992;89:9509–9513. [PubMed: 1384055]
- Heng HH, Tsui LC. Modes of DAPI banding and simultaneous in situ hybridization. *Chromosoma* 1993;102:325–332. [PubMed: 8325164]
- Kashiwabara S, Arai Y, Kodaira K, Baba T. Acrosin biosynthesis in meiotic and postmeiotic spermatogenic cells. *Biochem Biophys Res Commun* 1990;173:240–245. [PubMed: 1701633]
- Kierszenbaum AL, Tres LL. Structural and transcriptional features of the mouse spermatid genome. *J Cell Biol* 1975;65:258–270. [PubMed: 1127016]
- Kleene KC. Patterns of translational regulation in the mammalian testis. *Mol Reprod Dev* 1996;43:268–281. [PubMed: 8824926]
- Kramer JA, Adams MD, Singh GB, Doggett NA, Krawetz SA. Extended analysis of the region encompassing the PRM1-->PRM2-->TNP2 domain: genomic organization, evolution and gene identification. *J Exp Zool* 1998a;282:245–253. [PubMed: 9723181]
- Kramer JA, Krawetz SA. Nuclear matrix interactions within the sperm genome. *J Biol Chem* 1996;271:11619–11622. [PubMed: 8662749]
- Kramer JA, McCarrey JR, Djakiew D, Krawetz SA. Differentiation: the selective potentiation of chromatin domains. *Development* 1998b;125:4749–4755. [PubMed: 9806923]
- Kramer JA, McCarrey JR, Djakiew D, Krawetz SA. Human spermatogenesis as a model to examine gene potentiation. *Mol Reprod Dev* 2000;56:254–258. [PubMed: 10824979]
- Kramer JA, Zhang S, Yaron Y, Zhao Y, Krawetz SA. Genetic testing for male infertility: a postulated role for mutations in sperm nuclear matrix attachment regions. *Genet Test* 1997;1:125–129. [PubMed: 10464636]
- Krawetz SA, Dixon GH. Sequence similarities of the protamine genes: implications for regulation and evolution. *J Mol Evol* 1988;27:291–297. [PubMed: 3146639]
- Krawetz SA, Kramer JA, McCarrey JR. Reprogramming the male gamete genome: a window to successful gene therapy. *Gene* 1999;234:1–9. [PubMed: 10393233]
- Kurtz K, Martinez-Soler F, Ausio J, Chiva M. Acetylation of histone H4 in complex structural transitions of spermiogenic chromatin. *J Cell Biochem* 2007;102:1432–1441. [PubMed: 17471496]
- Li L, He H, Zhang J, Wang X, Bai S, Stolc V, et al. Transcriptional analysis of highly syntenic regions between *Medicago truncatula* and *Glycine max* using tiling microarrays. *Genome Biol* 2008;9:R57. [PubMed: 18348734]
- Martins RP, Ostermeier GC, Krawetz SA. Nuclear matrix interactions at the human protamine domain: a working model of potentiation. *J Biol Chem* 2004;279:51862–51868. [PubMed: 15452126]
- McCarrey JR, Berg WM, Paragioudakis SJ, Zhang PL, Dilworth DD, Arnold BL, et al. Differential transcription of P_{gk} genes during spermatogenesis in the mouse. *Dev Biol* 1992;154:160–168. [PubMed: 1426623]
- Meistrich ML, Trostle-Weige PK, Lin R, Bhatnagar YM, Allis CD. Highly acetylated H4 is associated with histone displacement in rat spermatids. *Mol Reprod Dev* 1992;31:170–181. [PubMed: 1372808]
- Miller.
- Nelson JE, Krawetz SA. Characterization of a human locus in transition. *J Biol Chem* 1994;269:31067–31073. [PubMed: 7983046]

- Nelson JE, Krawetz SA. Linkage of human spermatid-specific basic nuclear genes. Definition and evolution of the P1-->P2-->TP2 locus. *J Biol Chem* 1993;268:2932–2936. [PubMed: 8428967]
- Nelson JE, Krawetz SA. Mapping the clonally unstable recombinogenic PRM1-->PRM2-->TNP2 region of human 16p13.2. *DNA Seq* 1995;5:163–168. [PubMed: 7612927]
- Oliva R. Protamines and male infertility. *Hum Reprod Update* 2006;12:417–435. [PubMed: 16581810]
- Oliva R, Dixon GH. Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Prog Nucleic Acid Res Mol Biol* 1991;40:25–94. [PubMed: 2031084]
- Pivot-Pajot C, Caron C, Govin J, Vion A, Rousseaux S, Khochbin S. Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol Cell Biol* 2003;23:5354–5365. [PubMed: 12861021]
- Platts AE, Johnson GD, Linnemann AK, Krawetz SA. Real-time PCR quantification using a variable reaction efficiency model. *Anal Biochem* 2008;380:315–322. [PubMed: 18570886]
- Reeves RH, Gearhart JD, Hecht NB, Yelick P, Johnson P, O'Brien SJ. Mapping of PRM1 to human chromosome 16 and tight linkage of Prm-1 and Prm-2 on mouse chromosome 16. *J Hered* 1989;80:442–446. [PubMed: 2614060]
- Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. 1989
- Schmid C, Heng HH, Rubin C, Ye CJ, Krawetz SA. Sperm nuclear matrix association of the PRM1-->PRM2-->TNP2 domain is independent of Alu methylation. *Mol Hum Reprod* 2001;7:903–911. [PubMed: 11574659]
- Steger K. Transcriptional and translational regulation of gene expression in haploid spermatids. *Anat Embryol (Berl)* 1999;199:471–487. [PubMed: 10350128]
- Steger K, Klonisch T, Gavenis K, Behr R, Schaller V, Drabent B, et al. Round spermatids show normal testis-specific H1t but reduced cAMP-responsive element modulator and transition protein 1 expression in men with round-spermatid maturation arrest. *J Androl* 1999;20:747–754. [PubMed: 10591614]
- Steger K, Wilhelm J, Konrad L, Stalf T, Greb R, Diemer T, et al. Both protamine-1 to protamine-2 mRNA ratio and Bcl2 mRNA content in testicular spermatids and ejaculated spermatozoa discriminate between fertile and infertile men. *Hum Reprod* 2008;23(1):11–16. [PubMed: 18003625]
- Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, et al. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA* 2002;99:4465–4470. [PubMed: 11904358]
- Wellik DM. Hox genes and vertebrate axial pattern. *Curr Top Dev Biol* 2009;88:257–278. [PubMed: 19651308]
- Wykes SM, Krawetz SA. Conservation of the PRM1 --> PRM2 --> TNP2 domain. *DNA Seq* 2003a; 14:359–367. [PubMed: 14756422]
- Wykes SM, Krawetz SA. The structural organization of sperm chromatin. *J Biol Chem* 2003b; 278:29471–29477. [PubMed: 12775710]
- Yu YE, Zhang Y, Unni E, Shirley CR, Deng JM, Russell LD, et al. Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proc Natl Acad Sci USA* 2000;97:4683–4688. [PubMed: 10781074]
- Zalenskaya IA, Bradbury EM, Zalensky AO. Chromatin structure of telomere domain in human sperm. *Biochem Biophys Res Commun* 2000;279:213–218. [PubMed: 11112441]
- Zhao M, Shirley CR, Yu YE, Mohapatra B, Zhang Y, Unni E, et al. Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice. *Mol Cell Biol* 2001;21:7243–7255. [PubMed: 11585907]

FIGURE 1. DETECTION OF THE SITE OF INSERTION BY FLUORESCENT *IN SITU* HYBRIDIZATION

Metaphase chromosomes were isolated from homozygous transgenic mouse lymphocytes and fixed onto slides. The site of insertion was detected using a fluorescently labeled cosmid clone hp3.1 (red and arrows). Individual chromosomes were identified by DAPI banding. FISH signals from 100 fields were observed under fluorescence. The hybridization signal was localized to mouse chromosome 19, region C3.

FIGURE 2. GENOMIC LOCATION OF INSERTION SITE

The cosmid DNA fragment encompassing the human protamine domain (chromosome 16; 11,349,856 - 11,390,141; red hatched box), including the *PRM1*, *PRM2*, and *TNP2* genes flanked by boundary elements, was determined by terminal transferase dependent PCR to have inserted within cytological band C3 of mouse chromosome 19. Integration occurred within an L1Md T repeat element of the seventh intron of the cytochrome P450c38 gene.

FIGURE 3. RELATIVE LEVELS OF THE ENDOGENOUS AND HUMAN PROTAMINE TRANSCRIPTS IN TRANSGENIC MICE

Transcript levels were assessed by qRT-PCR of total testis RNA. A) The relative levels of the endogenous protamine gene transcripts *prm1*, *prm2*, and *tnp2* were not inhibited in transgenic animals harboring the human protamine construct compared to wild-type mouse. The level of α -actin provides a comparator of the variance among the animals. B) The relative levels of the human *PRM1*, *PRM2*, and *TNP2* transcripts. Transgene expression was similar to that observed in human males. Values are medians $\pm$ two standard deviations of triplicate reactions.

Probe Design Work Flow

FIGURE 4. PROCEDURAL FLOWCHART OF MASKLESS PROBE DESIGN

FIGURE 5. INTERSPECIES COMPARATIVE CGH ARRAY ANALYSIS OF HUMAN, WILD-TYPE, AND TRANSGENIC MOUSE GENOMIC DNA

Competitive hybridization of wild-type mouse and human genomic DNA to probes targeting A) the mouse protamine domain (chromosome 16: 10,782,515 – 10,802,516; GRCh37) and B) the extended human protamine domain (chromosome 16: 11,312,500 – 11,452,413; GRCh37). Both sets of probes exhibit species specificity. C) Competitive hybridization of wild-type and homozygous transgenic mouse genomic DNA to probes targeting the extended human protamine domain confirms insertion of the complete transgenic locus. Elevated transgenic signal is directly representative of the inserted human DNA sequence (chromosome 16: 11,349,856 - 11,390,141; GRCh37); signal ratio corresponding to the regions flanking the protamine domain is ~1:1. Y axis is Log₂.

TABLE 1

Primers used for determination of copy number and relative expression levels.

PRIMER NAME	FORWARD (5' – 3')	REVERSE (5' – 3')	AMPLICON SIZE (bp)	
<u>COPY NUMBER PRIMERS</u>				
Mm Ch16 10.8MB	CTCCAAGGCAACTCTTAAT	TTAAAGATAATGCGGGATGTAT	373	
PRM1	CAGAGCCGGAGCAGATATTAC	ATTTATTGACAGGCGGCATTGTT	359	
<u>TERMINAL TRANSFERASE DEPENDENT PCR PRIMERS</u>				
Biotinylated Extension Primer	TTCAATCTAGAATCCCACGTT			
Nested 1	AAATGCGAGGACGGTACAC	AGTAATAATAAGCAAACCTGG	96	
Nested 2	TACACGGCGACCCAAG	ACCTGGAAGACCCACACG	67	
<u>EXPRESSION ANALYSIS PRIMERS</u>			<u>cDNA</u>	<u>Genomic</u>
PRM1*	CGGAGCTGCCAGACACGGA	CTACATCTCGGTCTGTACCTGGG	65	156
PRM2*	AAGACGCTCCTGCAGGCAC	GCCTTCTGCATGTTCTCTT	71	233
TNP2	GAGTCCCAACACTAGTCCAC	TTGCGTAGAAATCACCATAGT	314	1164
Prm1	ACCATGGCCAGATACCGATG	TCAAGATGTGGCGAGATGCTC	237	329
Prm2	GCACCATGGTTCGCTACCG	GTGGCCTCACATGATGTTGCT	420	528
Tnp2	CCGTGCACTCTCGACACTCAC	ACATCCTGGAGTGCCTCACTTGT	176	347
ActB	CGTTGACATCCGTAAAGACCTCTAT	GTGTAAAACGCAGCTCAGTAACAGT	176	347

* Primers obtained from [Steger, et al. 2008].

TABLE 2

Genomic coordinates of the regions of interest represented on the tiling array.

REGION OF INTEREST	GENOMIC COORDINATES USED; BUILD 36.1	GENOMIC COORDINATES; BUILD 37.1
Acrosin	Chr 15: 89,395,836 – 89,402,836	Chr15: 89,398,500 – 89,405,500
α - Globin	Chr11: 32,158,235 – 32,218,234	Chr11: 32,158,235 – 32,218,234
Cytochrom P450 2c38	Chr19: 39,442,867 – 39,516,386	Chr19: 39,464,046 – 39,537,565
HoxA Cluster	Chr6: 5,2080,000 – 52,220,001	Chr6: 52,100,416 – 52,240,417
HoxB Cluster	Chr11: 9,606,000 – 96,220,001	Chr11: 96,105,224 – 96,265,225
HoxC Cluster	Chr15: 102,720,000 – 102,920,001	Chr15: 102,722,397 – 102,922,398
HoxD Cluster	Chr2: 74,460,000 – 7,460,001	Chr2: 74,497,218 – 74,637,219
Phosphoglycerate Kinase 1	ChrX: 102,385,000 – 102,410,001	ChrX: 103,382,463 – 103,399,038
Phosphoglycerate Kinase 2	Chr17: 39,668,500 – 39,673,501	Chr17: 40,341,974 – 40,346,975
Protamine Domain (Human)	Chr16: 11,159,950 – 11,270,000	Chr16: 11,312,500 – 11,452,413
Protamine Domain (Mouse)	Chr16: 10,696,000 – 10,716,001	Chr16: 10,782,515 – 10,802,516