PLOS ONE

Genome-Wide Mapping of Loci Explaining Variance in Scrotal Circumference in Nellore Cattle

Yuri T. Utsunomiya[1], Adriana S. Carmo[1], Haroldo H. R. Neves[2], Roberto Carvalheiro[2,3], Márcia C. Matos[1], Ludmilla B. Zavarez[1], Pier K. R. K. Ito[4], Ana M. Pérez O'Brien[5], Johann Sölkner[5], Laercio R. Porto-Neto[6], Flávio S. Schenkel[7], John McEwan[8], John B. Cole[9], Marcos V. G. B. da Silva[10], Curtis P. Van Tassell[11], Tad S. Sonstegard[11], José Fernando Garcia[1,4]*

1 Departamento de Medicina Veterinária Preventiva e Reprodução Animal, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Jaboticabal, São Paulo, Brazil, 2 Departamento de Zootecnia, UNESP - Univ Estadual Paulista, Jaboticabal, São Paulo, Brazil, 3 GenSys Consultores Associados, Porto Alegre, Rio Grande do Sul, Brazil, 4 Departamento de Apoio, Saúde e Produção Animal, Faculdade de Medicina Veterinária de Araçatuba, UNESP - Univ Estadual Paulista, Araçatuba, São Paulo, Brazil, 5 Division of Livestock Sciences, Department of Sustainable Agricultural Systems, BOKU - University of Natural Resources and Life Sciences, Vienna, Austria, 6 CSIRO - Commonwealth Scientific and Industrial Research Organization, Food Futures Flagship, Brisbane, Queensland, Australia, 7 Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada, 8 Centre for Reproduction and Genomics, AgResearch, Invermay, Mosgiel, Otago, New Zealand, 9 Animal Improvement Programs Laboratory, ARS-USDA - Agricultural Research Service - United States Department of Agriculture, Beltsville, Maryland, United States of America, 10 Bioinformatics and Animal Genomics Laboratory, Embrapa Dairy Cattle, Juiz de Fora, Minas Gerais, Brazil, 11 Bovine Functional Genomics Laboratory, ARS-USDA - Agricultural Research Service - United States Department of Agriculture, Beltsville, Maryland, United States of America

Abstract

The reproductive performance of bulls has a high impact on the beef cattle industry. Scrotal circumference (SC) is the most recorded reproductive trait in beef herds, and is used as a major selection criterion to improve precocity and fertility. The characterization of genomic regions affecting SC can contribute to the identification of diagnostic markers for reproductive performance and uncover molecular mechanisms underlying complex aspects of bovine reproductive biology. In this paper, we report a genome-wide scan for chromosome segments explaining differences in SC, using data of 861 Nellore bulls (*Bos indicus*) genotyped for over 777,000 single nucleotide polymorphisms. Loci that excel from the genome background were identified on chromosomes 4, 6, 7, 10, 14, 18 and 21. The majority of these regions were previously found to be associated with reproductive and body size traits in cattle. The signal on chromosome 14 replicates the pleiotropic quantitative trait locus encompassing *PLAG1* that affects male fertility in cattle and stature in several species. Based on intensive literature mining, *SP4*, *MAGEL2*, *SH3RF2*, *PDE5A* and *SNAI2* are proposed as novel candidate genes for SC, as they affect growth and testicular size in other animal models. These findings contribute to linking reproductive phenotypes to gene functions, and may offer new insights on the molecular biology of male fertility.

Introduction

Reproductive performance has a high economic value in beef cattle, because fertility affects generation intervals, the intensity of selection pressure that can be applied to the population, and the amount of product that can be sent to the market [1]. Furthermore, reproductive wastage is a major reason for culling beef cows.

Domestic cattle are composed by two interfertile species: the humpless taurine cattle (*Bos taurus*) and the humped indicine or zebu cattle (*Bos indicus*). Indicine breeds, such as Nellore cattle, form the majority of the beef herds in tropical and subtropical countries. Zebus generally take longer to reach puberty than taurines [2], making the improvement of reproductive perfor-

mance an impending challenge in the production systems of these regions of the world.

Scrotal circumference evaluated at yearling (SC) is the most recorded reproductive trait in breeding programs for beef cattle, as the trait is inexpensive and easy to measure [3], is highly heritable [4], and is associated with testis development, quantitative and qualitative semen parameters [5], age at puberty in bulls and related heifers [6,7], heifer pregnancy [1], and body weight [8]. Consequently, SC is used in these programs as a major indicator of precocity and fertility.

Characterizing genomic regions that explain differences in SC in *B. indicus* can contribute to the identification of reproductive performance informative molecular markers to assist breeding, as well as to the mapping of loci implicated in reproductive biology. In this paper, we analyzed data of estimated breeding values (EBV)

from 861 Nellore bulls genotyped for over 777,000 single nucleotide polymorphism (SNP) markers. We aimed at identifying putative genomic regions explaining differences in SC in *B. indicus* cattle via genome-wide mapping.

Materials and Methods

Ethical statement

Local ethical committee approval was not necessary in the present study, because phenotypic data were obtained from a database [9], and DNA samples used for genotyping were obtained from commercialized semen straws.

Animals and phenotypes

Estimated breeding values for SC were obtained from routine genetic evaluations [9], comprising data from 542,918 animals born between 1985 and 2011, and raised in 243 grazing-based Brazilian herds. Scrotal circumference in yearlings (around 18 months of age) was measured as recommended by the *Society for Theriogenology* [3]. Genetic evaluation of SC was based on two single-trait animal models, both including a fixed effect of contemporary group (defined as animals from the same herd, born in the same year and season, and belonging to the same management group from birth until yearling), and random effects that included direct additive genetic, maternal additive genetic, maternal permanent environmental and residual error effects. In the first model (SC$_A$), the fixed effect of age at SC measurement was included as a covariate. In the second model (SC$_{AW}$), covariates accounting for differences due to age and weight at yearling were included. Estimated breeding values were deregressed by the method described by Garrick *et al.* (2009) [10] and treated as pseudo-phenotypes in the genome-wide mapping analysis.

Genotyping and data filtering

Only sires widely used via artificial insemination whose accuracies (i.e., square root of reliability, calculated based on prediction error variance estimates) for SC$_A$ and SC$_{AW}$ were greater than 0.5 were considered for analysis. A total of 861 progeny-tested Nellore bulls were genotyped for 777,962 SNPs with the Illumina® BovineHD Genotyping BeadChip assay, according to the manufacturer's protocol. This dataset builds on the data reported by Utsunomiya et al. (2013) [11]. As a first filtering criterion, only samples with call rate greater than 0.9 and SNPs with GenTrain score greater than or equal to 0.7 were considered for analysis. Mitochondrial DNA and unmapped markers were also excluded.

As males are hemizygous for both sex chromosomes, observation of heterozygous X and Y genotypes are only possible for SNP probes that hybridize against the XY pseudo-autosomal (PAR) region. As the *UMD v3.1* bovine genome assembly [12] does not allow for clear distinction of PAR markers, all heterozygous X- and Y-linked genotypes were considered as genotyping errors and set to missing. Next, SNPs were removed from the dataset if they did not exhibit minor allele frequency greater than or equal to 0.02 or call rate of at least 0.98. These procedures and many others described later were performed using customized functions and the *base* and the *GenABEL* v1.7-6 packages in *R v2.15.0* [13,14].

Genome-wide mapping

We adapted the two-steps *Fast Association Score Test-based Analysis* (FASTA) [15] to compute allele substitution effects accounting for relatedness, population structure and heterogeneity of variance in deregressed EBVs (dEBVs). In the first step, the variance-covariance matrix for the pseudo-phenotypes was estimated using an animal model that included random additive genetic and residual effects. In the second step, the estimated variance-covariance matrix was used to compute allele substitution effects for each SNP via generalized least squares. A detailed description of this analysis can be found in **File S1**.

Next, aiming at mapping loci explaining differences in SC, we investigated chromosome windows where the average phenotypic variance explained by SNPs deviated substantially from the genome background. First, the percentage of phenotypic variance explained by each SNP was calculated as:

$$\%\hat{\pi} = \frac{2p_i q_i \hat{\beta}_i^2}{\sigma_T^2} \times 100$$

where, relative to SNP i, $\hat{\beta}_i^2$ is the estimated allele substitution effect, p_i and q_i are the allele frequencies, and σ_T^2 is the total trait variance.

Second, in order to reduce noise and improve mapping, the phenotypic variance explained by SNPs was smoothed across the genome by averaging $\%\hat{\pi}$ in sliding windows of 1 Mb, sliding 50 kb at a time. Only windows containing at least 10 SNPs were averaged, and we considered as outliers the windows where $\%\hat{\pi} > 5 \times IQR + Q_3$, where IQR is the interquartile range and Q_3 is the third quartile of the distribution. Third, we used *BEDTools v2.12.0* [16] to merge the intervals of overlapping outlier windows. These merged windows were considered as candidate loci for SC.

We assessed the relative mapping resolution gain when using sliding windows instead of single SNPs by calculating the signal-to-noise ratio between the two strategies. For each approach, we let the signal-to-noise coefficient be represented by the reciprocal of the coefficient of variation $SN = \mu/\sigma$, where μ and σ are the mean percentage of phenotypic variance explained and its standard deviation, respectively, by either single SNPs or windows. Then, we calculated the ratio of the signal-to-noise coefficients between strategies as SN_{window}/SN_{SNP}.

Assessment of functional relevance

The *cattle QTLdb* database [17] was examined to find out if any genomic region identified here overlapped with a previously described bovine quantitative trait locus (QTL), in particular those related to body size and reproductive traits. Gene coordinates in the *UMD v3.1* assembly [12] were obtained from the *Ensembl genes 73* database using the *BioMart tool* [18], and overlaps between the boundaries of candidate loci and gene coordinates were determined using *BEDTools v2.12.0* [15]. Finally, we conducted intensive literature review to propose functionally sound candidate genes associated with SC.

Results

A total of 525,961 SNPs (67.6%) and 861 individuals (100.0%) passed the filtering criteria and were retained in the dataset. After filtering, the average and median gap size between consecutive markers were 5.1 kb and 3.3 kb, respectively, indicating high density coverage of the genome. In spite of the effort and interest to analyze both sex chromosomes, Y SNPs did not present sufficient variability to allow for estimation of allele substitution effects.

For both SC$_A$ and SC$_{Aw}$, the distributions of pseudo-phenotypes were approximately normal, and the dEBVs for these two traits were fairly correlated, with R^2 = 0.74 (**Figure 1**). Accuracies were
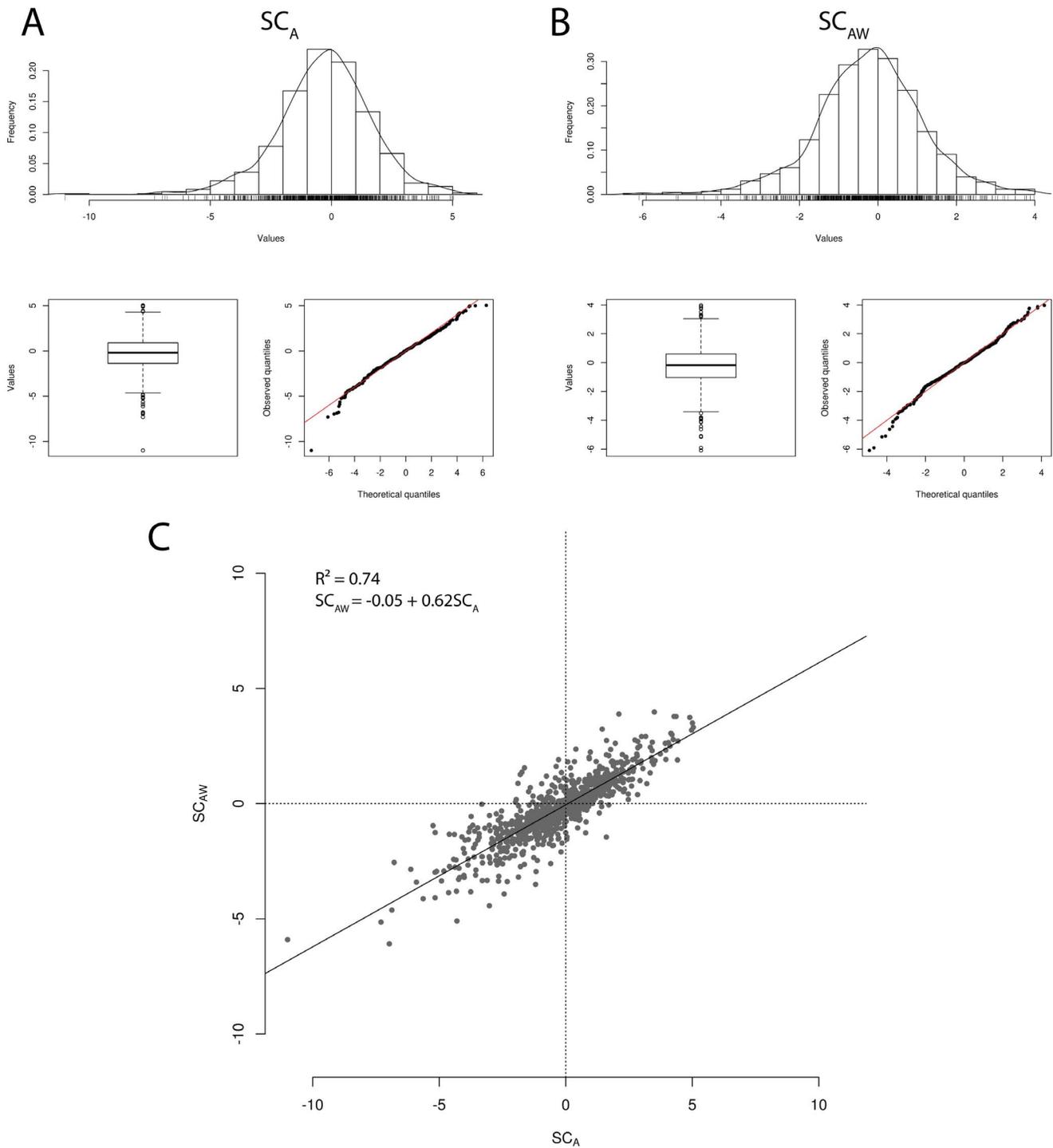
Figure 1. Descriptive statistics for scrotal circumference dEBVs of 861 Nellore bulls. Histograms (top), boxplot (bottom left) and normal quantile-quantile plots (bottom right) are provided for scrotal circumference A) corrected for age (SC_A) and B) corrected for age and weight at yearling (SC_{AW}). A scatter plot illustrating the linear relationship between the two dEBVs is also provided (C).

virtually equal for SC_A and SC_{Aw} (identical up to the second decimal place), with mean, minimum, median and maximum of 0.84 ± 0.11, 0.51, 0.86 and 0.98, respectively.

A total of 52,493 SNP windows of 1 Mb were built across the genome, with an average density of 198 ± 59 SNPs per window. The ratio between the signal-to-noise coefficients of sliding windows and single SNPs was 2.54, indicating that the smoothing strategy allowed for a 2.54 fold improvement in mapping resolution (**Figures 2, S1 and S2**). Considering the $5\times IQR+Q_3$ threshold for percentage of phenotypic variance explained ($SC_A = 0.40\%$ and $SC_{Aw} = 0.42\%$), 236 (0.45%) and 279 (0.53%) windows were declared outliers for SC_A and SC_{AW},

Figure 2. Manhattan plots of scrotal circumference variance explained by SNP windows in Nellore cattle. Pseudo-phenotypes were based on dEBVs corrected for age (SC$_A$) and corrected for age and weight at yearling (SC$_{AW}$). Each dot represents a 1 Mb SNP window. Horizontal dashed lines represent adopted thresholds (SC$_A$ = 0.40% and SC$_{Aw}$ = 0.42%). Arrows indicate signals shared between the two 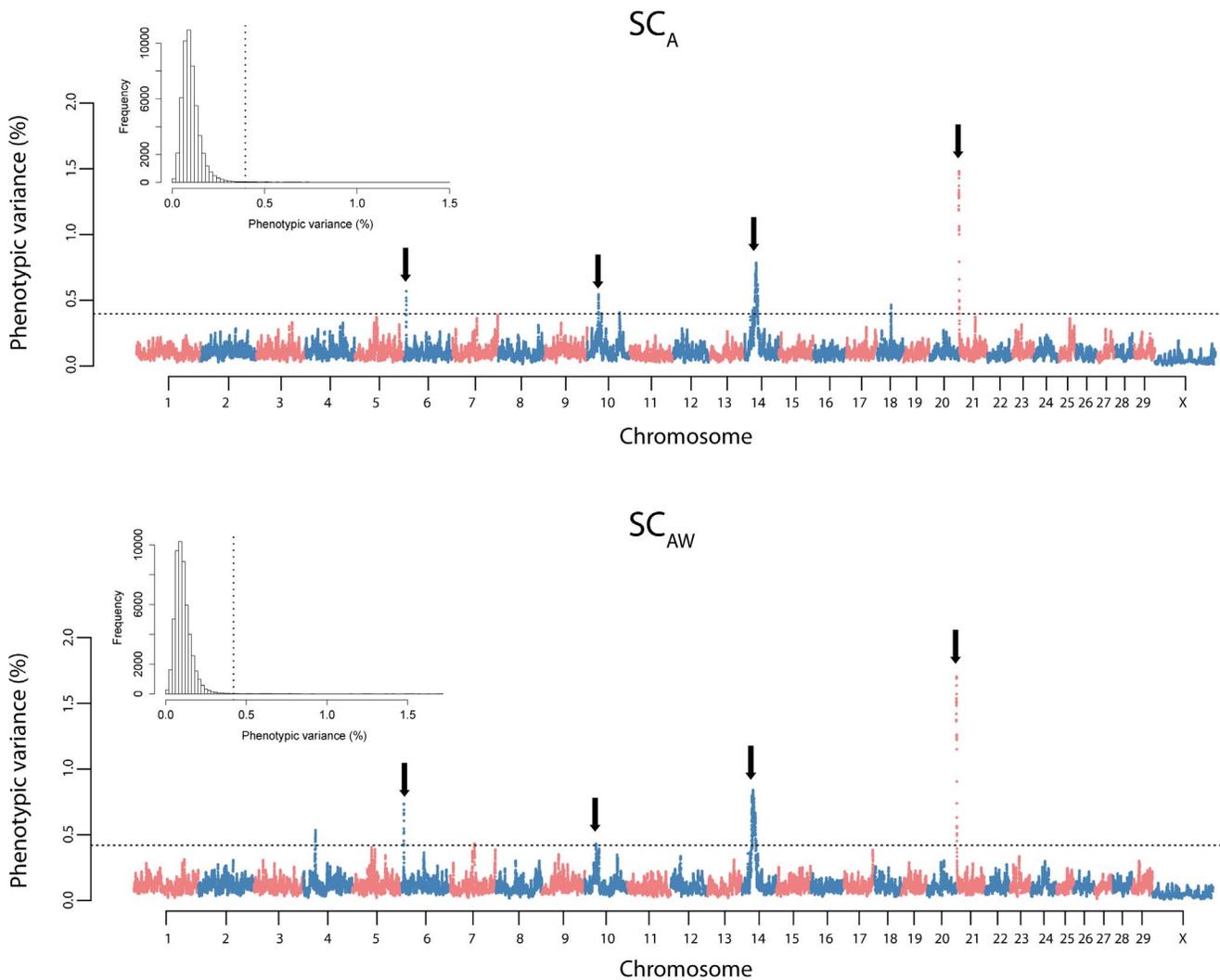models. Histograms represent the distribution of phenotypic variance explained by SNP windows, and the dotted vertical line marks the adopted thresholds.
doi:10.1371/journal.pone.0088561.g002

respectively. After merging overlapping outlier windows, we obtained a total of 8 and 6 candidate loci explaining approximately 4% of the variance in SC$_A$ and SC$_{AW}$, respectively (**Table 1**).

Overall, results from the genome-wide mapping analysis were strikingly similar between SC$_A$ and SC$_{AW}$, indicating that fitting the covariate for weight at yearling in the model did not cause substantial mapping differences. Indeed, four loci were shared between the two traits in chromosomes 6, 10, 14 and 21, respectively (**Figure 3**), which exhibited clear signals in the genome-wide plot of smoothed phenotypic variance explained by SNP windows (**Figure 2**). From the total of six differentially detected loci, one was located nearby the shared locus on chromosome 14 (**Table 1**), and the three SC$_{Aw}$ private loci on chromosomes 10 and 7 were close to be declared outliers for SC$_A$ (**Figure 2**).

A total of 285 and 190 genes were mapped against the major loci found for SC$_A$ and SC$_{AW}$, respectively, and a total of 309 unique genes were observed (**File S2**). From these, 246 protein coding, 25 snoRNA, 12 snRNA, 9 miRNA, 9 rRNA, 1 misc_RNA and 7 pseudo genes were observed (**File S2**). From this gene list, we filtered 9 functional candidates implicated in growth, testicular size and fertility (**Table 1**), which included: pleiomorphic adenoma gene 1 (*PLAG1*, ENSBTAG00000004022), carboxylesterase 4A (*CES4A*, ENSBTAG00000038325), Sp4 transcription factor (*SP4*, ENSBTAG00000014389), melanoma antigen family L 2 (*MAGEL2*, ENSBTAG00000045998), phosphodiesterase 5A (*PDE5A*, ENSBTAG00000024888), snail family zinc finger 2 (*SNAI2*, ENSBTAG00000013227), nuclear protein in testis (*C15ORF55*, ENSBTAG00000014948), fibrous sheath-interacting protein 1 (*FSIP1*, ENSBTAG00000012015), and SH3 domain containing ring finger 2 (*SH3RF2*, ENSBTAG00000006762).

A total of 76 production and reproduction QTLs, mined from 24 distinct publications, were mapped against the loci found here (**File S3**). The largest trait contingency observed was composed by traits related to body size (43 QTLs), followed by reproductive traits (23 QTLs). Furthermore, the locus detected on chromosome

Table 1. Detected major loci explaining variance in scrotal circumference in Nellore cattle.

Scrotal circumference model	Chromosome	Position start (Mb)	Position end (Mb)	Peak position (Mb)	Segment length (Mb)	Number of SNPs	Average MAF[a]	Average %π̄[b]	Functional candidate gene
Corrected for age (SC_A)	6	5.20	6.65	5.80–6.00	1.45	42	0.28	0.52	PDE5A (ENSBTAG00000024888)
	10	26.90	28.80	28.00	1.90	387	0.23	0.49	C15ORF55 (ENSBTAG00000014948)
	10	34.80	35.80	35.30	1.00	140	0.19	0.40	FSIP1 (ENSBTAG00000012015)
	10	78.85	79.85	79.35	1.00	122	0.20	0.41	-
	14	20.25	21.45	20.90	1.20	310	0.24	0.41	SNAI2 (ENSBTAG00000013227)
	14	23.40	33.85	29.10	10.45	2236	0.23	0.55	PLAG1 (ENSBTAG00000004022)
	18	34.55	35.80	35.20	1.25	206	0.22	0.44	CES4A (ENSBTAG00000038325)
	21	0.00	2.50	1.20	2.50	127	0.17	1.07	MAGEL2 (ENSBTAG00000045998)
Corrected for age and weight at yearling (SC_AW)	4	28.95	30.80	30.15	1.85	301	0.20	0.48	SP4 (ENSBTAG00000014389)
	6	5.05	6.75	5.80–6.00	1.70	84	0.27	0.63	PDE5A (ENSBTAG00000024888)
	7	59.15	60.25	59.75	1.10	158	0.16	0.43	SH3RF2 (ENSBTAG00000006762)
	10	27.15	28.25	27.70	1.10	209	0.21	0.43	C15ORF55 (ENSBTAG00000014948)
	14	23.25	35.50	25.00 and 27.00	12.25	2687	0.23	0.63	PLAG1 (ENSBTAG00000004022)
	21	0.00	2.50	1.20	2.50	127	0.17	1.23	MAGEL2 (ENSBTAG00000045998)

[a]MAF = Minor allele frequency.
[b]%π̄ = Average percentage of phenotypic variance explained by overlapping 1 Mb SNP windows.
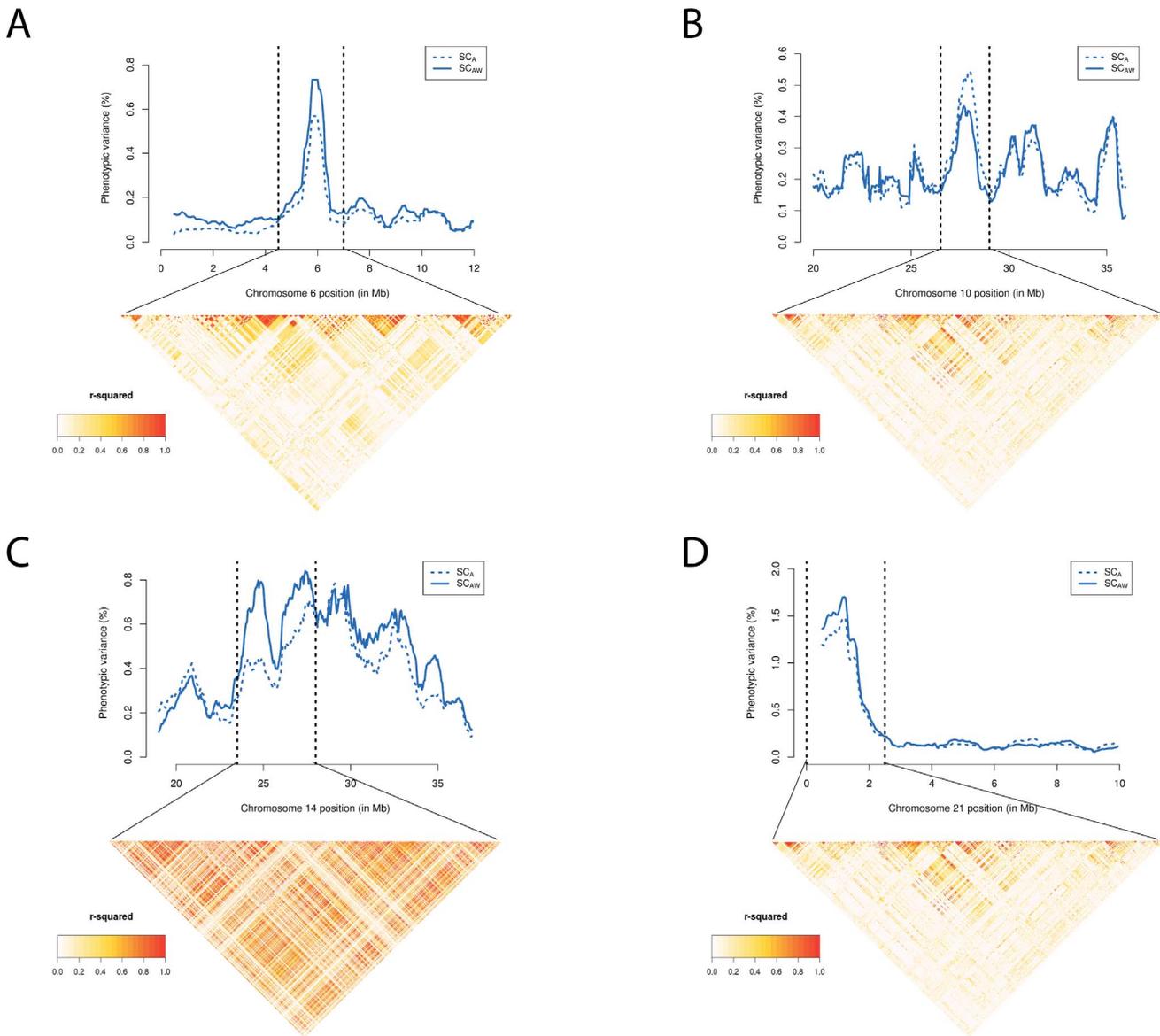doi:10.1371/journal.pone.0088561.t001

Figure 3. Regional plots of scrotal circumference variance explained by SNP windows in Nellore cattle. Pseudo-phenotypes were based on dEBVs corrected for age (SC$_A$) and corrected for age and weight at yearling (SC$_{AW}$). Clear common signals between SC$_A$ and SC$_{Aw}$ were found on chromosomes A) 6, B) 10, C) 14 and D) 21. Vertical black dashed lines delimit the regions where the highest variance explained were found. Linkage disequilibrium structure for these regions (bottom) is portrayed as a heatmap of r^2 values between SNPs.
doi:10.1371/journal.pone.0088561.g003

4 (**Table 1**) mapped against one previously described QTL for SC in Angus cattle (*B. taurus*) [19].

Discussion

The genome-wide mapping analysis detected positional candidate loci explaining approximately 4% of the dEBVs for SC (**Table 1**). Although this represents only a fraction of the trait variance, this percentage is substantial considering that 180 loci associated with human adult height, a highly heritable and classic polygenic trait, explain only 10% of the phenotypic variance together [20]. This is evidence that multiple loci across the genome are involved in the complex inheritance of SC, and the functional candidate genes filtered here may only scratch the surface of the molecular mechanisms underlying the trait. The dissection of the

pathways regulating precocity in *B. indicus* cattle will require multiple studies across breeds and trait models, with intensive multidisciplinary reasoning. Nevertheless, the loci reported here excel from the genome background, and represent important data in the context of bovine reproductive biology.

The region explaining the largest proportion of SC variance in the present study mapped to the beginning of chromosome 21, peaking around 1.5 Mb. The closest gene found in this region was *MAGEL2*. The orthologous human and murine genes regulate normal circadian output, and are highly expressed in the suprachiasmatic nucleus of the hypothalamus [21]. The human *MAGEL2* has been implicated in Prader-Willi Syndrome, a genetic disorder characterized by short stature, low muscle tone, cognitive disabilities, increased food intake, obesity, low levels of insulin and insulin-like growth factor 1 (IGF1), incomplete sexual develop-

ment, hypogonadism, and male infertility [21,22]. The disorder manifests when a segment on human chromosome 15, which encompasses seven maternally imprinted genes including *MA-GEL2*, presents a deletion or loss of expression of the paternal alleles. Inactivation of the mouse *MAGEL2* alone was shown to lead to abnormalities suggestive of hypothalamic dysfunction similar to the Prader-Willi Syndrome [22].

Of note, variation of copy number gain spanning the interval between 1.57 Mb and 2.99 Mb on bovine chromosome 21 has been found by the comparison of individual whole genome sequence data of Nellore with the *B. taurus* breeds Angus, Holstein and Hereford [23]. As the Prader-Willi Syndrome is caused by loss of the paternal copy of the orthologous sequence in humans, and *MAGEL2* is essential for proper hypothalamic control of growth and fertility [21], association of copy number variants with growth and reproductive traits seems to be a sensible hypothesis to be tested on this chromosome segment.

The locus detected on chromosome 7 encompasses *SH3RF2*. Rubin *et al.* (2010) [24] discovered a deletion removing all but the first exon of the orthologous chicken gene that is associated with body weight, and demonstrated that strong selection caused the deletion to reach fixation in a high growth lineage. Interestingly, using a mouse model of Prader-Willi syndrome, Stefan *et al.* (2005) [25] found that loss of expression of the *MAGEL2* region induces upregulation of *SH3RF2* and its flanking genes *TCERG1*, *LARS*, *RBM27* and *GPR151*. As both the *MAGEL2* and the *SH3RF2* regions were flagged in the present study, a trans-acting regulatory mechanism involving the loci on chromosomes 7 and 21 found here is likely to underlie SC variation. Hence, these signals are plausible candidates for weight and male fertility traits in Nellore cattle.

We identified a candidate locus on chromosome 14 with the highest percentage of phenotypic variance explained mapping to positions 25 Mb and 27 Mb. Fortes *et al.* (2013) [26] reported associations for IGF1 at 6 months and SC at 12 months in young Brahman bulls (*B. indicus*) in an overlapping region around 25 Mb, which was previously shown to correlate with age of Brahman bulls when they achieve 26 cm of SC [27]. This region has been well characterized in taurine cattle as harboring several human orthologues affecting stature and growth [28], especially *PLAG1* [29]. The locus has also been found to be associated with birth weight in Nellore cattle, and suggested to shelter polymorphisms with pleiotropic effects on traits that correlate with body size [11]. Furthermore, some first evidences for pleiotropism in body size and fertility traits in the *PLAG1* region have been recently found in Brahman cattle [30].

Although the human stature orthologues flanking 25 Mb are appealing candidates for SC, the chromosome 14 signal found here comprises a large segment spanning from 20.25 Mb to 35.85 Mb. This may be evidence that multiple genes and variants within this region are involved. For instance, *SNAI2* is located around 21.58 Mb and encodes for Slug (also known as Slugh), a Zinc-finger transcription factor that when mutated in mice produces males with testicular atrophy and marked decrease in seminiferous tubules sizes [31]. Although these mice are able to copulate, their offspring are small. Also, Fortes *et al.* (2013) [26] showed that ability to produce sperm at 18 months in Brahman bulls is not as significant around 25 Mb, and exhibits signals of association shifted towards the 35 Mb position instead.

Another possible justification for a signal coming from such a large chromosome segment is a long range linkage disequilibrium persistency within the region. In fact, we found a strong linkage disequilibrium structure underpinning the signal (**Figure 3**), which may be hampering the localization of the true locus

involved. In either case, these evidences together support the entire chromosome segment identified here as a key region affecting growth and fertility traits in cattle.

The locus detected on chromosome 6 is located 124 kb downstream of *PDE5A*. The phosphodiesterase encoded by *PDE5A* is substantially expressed in the testis, and mice overexposed to inhibitors of this protein present testicular tissue alterations, including decreased testis weight, degeneration, and atrophy of the seminiferous epithelium [32]. This genomic region also shelters genes that interact with other proteins previously linked to small testis size. For instance, the protein encoded by *MAD2* belongs to the mitotic checkpoint complex, and is recruited by the mitotic kinase Bub1. A residue change in the catalytic loop of Bub1 was shown to lead to male subfertility, with marked reduction in testicular size [33].

Several QTLs mapping to the loci detected here were related either to body size or reproductive traits that are associated with SC. In particular, the peak on chromosome 4 mapped against one previously reported QTL for SC [19], which encompasses *SP4*. This gene encodes for a zinc finger transcription factor that is predominantly expressed in the brain, but is also detectable in the testicular tissue [34,35]. Göllner *et al.* (2001) [36] showed that *SP4*-knockout mice develop until birth without obvious abnormalities, but two-thirds of them die within 4 weeks after birth and the remaining one-third present growth retardation. Surviving male mice exhibit reduced testis size, although complete spermatogenesis can be observed. Surviving female mice exhibit small-sized thymus, spleen and uterus, and all mice show pronounced delay in sexual maturation. As *SP4*-knockout mice present growth retardation mainly after birth, it is likely that variations in the bovine *SP4* affect body size and testicular growth from birth to yearling age, but they are unlikely to affect fetal development or spermatogenesis. Moreover, the evidence of delayed sexual maturation and reduced testicular size in surviving *SP4*-knockout mice is consistent with the known positive correlation between SC and precocity in cattle.

The functional candidate gene surrounding the peak on chromosome 18 was *CES4A*, a hydrolase member of the carboxylesterase large family (enzyme class EC 3.1.1.1.-), also known as *CES6*, *CES8* or Hydrolase A. Carboxylesterases act in the transesterification of a broad spectrum of substrates, and play an important role in the metabolism of endogenous lipid and foreign compounds such as drugs and pesticides [37]. Hydrolase A is known to be expressed in several tissues, including the testis [38]. Esterase activity (EC 3.1.1.-) has been found to be abundant in the testis and associated with androgen production [39]. Two intronic SNPs in the human *CES4A* were found to be correlated with high density lipoprotein levels (HDL) in an association analysis deposited in the *dbGaP* database (www.ncbi.nlm.nih.gov/gap, accession: pha002900.1, accessed on 21 Oct 2013), conducted in an expanded population sample from the original 1966 Northern Finland Birth Cohort (NFBC66) study [40]. Also, testosterone treatment of aging men with hypogonadism was demonstrated to lower HDL levels, and the mechanism underlying this relationship and its role in coronary disease risk have been targets of debate and controversy [41]. These evidences together point to *CES4A* as a functional candidate gene affecting differences in SC in Nellore cattle, and the underlying mechanism may involve the dynamics between HDL and androgen levels.

Some of the genes lying within the loci detected nearby positions 28 Mb and 35 Mb on chromosome 10 are related to the testicular tissue, but their association with scrotal circumference is unclear. The nuclear protein in testis gene (*C15ORF55* or *NUTM1*) is mainly known by its involvement with midline organs carcinoma

[42]. The fibrous sheath-interacting protein 1 (FSIP1) was shown to bind to Akap4 during spermatogenesis in order to form the fibrous sheath of the sperm flagellum [43]. The peak found around 79.35 Mb on chromosome 10 did not reveal an appealing functional candidate gene, but the region overlaps QTLs for dystocia, fetal death and birth weight (**File S3**). Further investigation of these loci is needed to elucidate if they contribute to phenotypic variation in SC, as well as to clarify the molecular mechanisms underlying this contribution.

Conclusions

In summary, this is believed to be the first study applying a high-density SNP panel in a genome-wide survey of loci affecting scrotal circumference in Nellore cattle, which contributes with important preliminary data to the dissection of molecular mechanisms regulating precocity in the bovine species. The loci identified here harbor known and novel functional candidate genes affecting scrotal circumference in *B. indicus* cattle. Fine mapping of these signals with whole genome sequence data and hypothesis-driven experiments may shed light on the genes and networks underlying phenotypic variation in fertility traits in cattle. In a broader perspective, as the majority of the genes found in eutherian mammals are orthologous, further investigation of these loci in cattle may offer new insights on complex aspects of mammalian reproductive biology.

Supporting Information

Figure S1   Manhattan plot of SC_A variance explained by single SNPs in Nellore cattle.
(TIF)

Figure S2   Manhattan plot of SC_{AW} variance explained by single SNPs in Nellore cattle.
(TIF)

File S1   Supporting methods: extended methods for weighted FASTA.
(PDF)

File S2   Genes mapping against the loci detected for scrotal circumference in Nellore cattle.
(XLS)

File S3   QTLdb hits for scrotal circumference corrected for age (SC_A) and age and weight (SC_{AW}).
(XLS)

Author Contributions

Conceived and designed the experiments: JFG TSS JS JM JBC CPVT FSS MVGBS LRPN. Performed the experiments: ASC YTU HHRN RC MCM LBZ PKRKI AMPO. Analyzed the data: YTU. Contributed reagents/materials/analysis tools: JFG TSS JS JM JBC CPVT FSS MVGBS LRPN RC HHRN ASC YTU. Wrote the paper: YTU. Coordinated the study: JFG. All authors read, approved and contributed to edit the final manuscript.

References

1. Van Melis MH, Eler JP, Rosa GJM, Ferraz JBS, Figueiredo LGG, et al. (2010) Additive genetic relationships between scrotal circumference, heifer pregnancy, and stayability in Nellore cattle. Journal of Animal Science 88:3809–3813.

2. Martin LC, Brinks JS, Bourdon RM, Cundiff LV (1992) Genetic effects on beef heifer puberty and subsequent reproduction. Journal of Animal Science 70: 4006–4017.

3. Ball L, Ott RS, Mortimer RG, Simons JC (1983) Manual for Breeding Soundness Examination of Bulls. J Soc Therio XII.

4. Dias LT, Faro LE, Albuquerque LG (2003) Estimativas de herdabilidade para perímetro escrotal de animais da raça nelore. Revista Brasileira de Zootecnia 32:1878–1882.

5. Bourdon RM, Brinks JS (1986) Scrotal circumference in yearling Hereford bulls: adjustment factors, heritabilities and genetic, environmental and phenotypic relationships with growth traits. Journal of Animal Science 62:958–967.

6. Toelle VD, Robison OW (1985) Estimates of genetic correlations between testicular measurements and female reproductive traits in cattle. Journal of Animal Science 60:89–100.

7. Evans JL, Golden BL, Bourdon RM, Long KL (1999) Additive genetic relationships between heifer pregnancy and scrotal circumference in Hereford cattle. Journal of Animal Science 77:2621–2628.

8. Bergmann JAG, Zamborlini LC, Procopio CSO, Andrade VJ, Vale Filho VR (1996) Estimativas de parâmetros genéticos do perímetro escrotal e do peso corporal em animais da raça Nelore. Arquivo Brasileiro de Medicina Veterinária e Zootecnia 48:69–78.

9. Aliança (2012) Sumário de touros 2012 Aliança Nelore. GenSys Consultores Associados.

10. Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution 41:55.

11. Utsunomiya YT, do Carmo AS, Carvalheiro R, Neves HH, Matos MC, et al. (2013) Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genetics 14:52.

12. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, et al. (2009) A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biology 10:R42.

13. R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: http://www.R-project.org. Accessed 2014 Jan 17.

14. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics 23:1294–1296.

15. Chen WM, Abecasis GR (2007) Family-based association tests for genomewide association scans. American Journal of Human Genetics 81:913–926.

16. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

17. Hu ZL, Park CA, Wu XL, Reecy JM (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acid Research 41:D871–9.

18. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, et al. (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011; Published online Jul 23.

19. McClure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, et al. (2010) A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. Animal Genetics 41:597–607.

20. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838.

21. Kozlov SV, Bogenpohl JW, Howell MP, Wevrick R, Panda S, et al. (2007) The imprinted gene *Magel2* regulates normal circadian output. Nature Genetics 39:1266–1272.

22. Bischof JM, Stewart CL, Wevrick R (2007) Inactivation of the mouse *Magel2* gene results in growth abnormalities similar to Prader-Willi syndrome. Human Molecular Genetics 16:2713–2719.

23. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, et al. (2012) Copy number variation of individual cattle genomes using next-generation sequencing. Genome Research 22:778–790.

24. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, et al. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–91.

25. Stefan M, Portis T, Longnecker R, Nicholls RD (2005) A nonimprinted Prader-Willi Syndrome (PWS)-region gene regulates a different chromosomal domain in trans but the imprinted pws loci do not alter genome-wide mRNA levels. Genomics 85:630–640.

26. Fortes MRS, Reverter A, Hawken RJ, Bolormaa S, Lehnert SA (2013) Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. Biology of Reproduction 87:58.

8

27. Fortes MRS, Lehnert SA, Bolormaa S, Reich C, Fordyce G, et al. (2012) Finding genes for economically important traits: Brahman cattle puberty. Animal Production Science 52:143–150.

28. Pryce JE, Hayes BJ, Bolormaa S, Goddard ME (2011) Polymorphic regions affecting human height also control stature in cattle. Genetics 187:981–984.

29. Karim L, Takeda H, Lin L, Druet T, Arias JA, et al. (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nature Genetics 43:405–413.

30. Fortes MRS, Kemper K, Sasazaki S, Reverter A, Pryce JE, et al. (2013) Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. Animal Genetics 44:636–647.

31. Pérez-Losada J, Sánchez-Martín M, Rodríguez-García A, Sánchez ML, Orfao A, et al. (2002) Zinc-finger transcription factor Slug contributes to the function of the stem cell factor c-kit signaling pathway. Blood 100:1274–1286.

32. Vezzosi D, Bertherat J (2011) Phosphodiesterases in endocrine physiology and disease. European Journal of Endocrinology 165:177–188.

33. Ricke RM, Jeganathan KB, Malureanu L, Harrison AM, van Deursen JM (2012) Bub1 kinase activity drives error correction and mitotic checkpoint control but not tumor suppression. Journal of Cell Biology 199:931–949.

34. Hagen G, Müller S, Beato M, Suske G (1992) Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. Nucleic Acids Research 20:5519–5525.

35. Supp DM, Witte DP, Branford WW, Smith EP, Potter SS (1996) Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. Developmental Biology 176:284–299.

36. Göllner H, Bouwman P, Mangold M, Karis A, Braun H, et al. (2001) Complex phenotype of mice homozygous for a null mutation in the Sp4 transcription factor gene. Genes Cells 6:689–697.

37. Satoh T, Hosokawa M (1998) The mammalian carboxylesterases: from molecules to functions. Annual Review of Pharmacology and Toxicology 38:257–288.

38. Yan B, Yang D, Brady M, Parkinson A (1995) Rat testicular carboxylesterase: cloning, cellular localization, and relationship to liver hydrolase A. Archives of Biochemistry and Biophysics 316:899–908.

39. Huggins C, Moulton SH (1948) Esterases of testis and other tissues. Journal of Experimental Medicine 88:169–179.

40. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, et al. (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature Genetics 41:35–46.

41. Langer C, Gansz B, Goepfert C, Engel T, Uehara Y, et al. (2002) Testosterone up-regulates scavenger receptor BI and stimulates cholesterol efflux from macrophages. Biochemical and Biophysical Research Communications 296:1051–1057.

42. French CA, Miyoshi I, Kubonishi I, Grier HE, Perez-Atayde AR, et al. (2003) BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma. Cancer Research 63:304–307.

43. Brown PR, Miki K, Harper DB, Eddy EM (2003) A-kinase anchoring protein 4 binding proteins in the fibrous sheath of the sperm flagellum. Biology of Reproduction 68:2241–2248.