ORIGINAL ARTICLE

A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation

Marlene D Dalgaard,[1] Nils Weinhold,[2] Daniel Edsgärd,[2] Jeremy D Silver,[3] Tune H Pers,[2,4] John E Nielsen,[1] Niels Jørgensen,[1] Anders Juul,[1] Thomas A Gerds,[3] Aleksander Giwercman,[5] Yvonne L Giwercman,[5] Gabriella Cohn-Cedermark,[6] Helena E Virtanen,[7] Jorma Toppari,[7,8] Gedske Daugaard,[8] Thomas S Jensen,[2] Søren Brunak,[2] Ewa Rajpert-De Meyts,[1] Niels E Skakkebæk,[1] Henrik Leffers,[1,9] Ramneek Gupta[2]

ABSTRACT

Background Testicular dysgenesis syndrome (TDS) is a common disease that links testicular germ cell cancer, cryptorchidism and some cases of hypospadias and male infertility with impaired development of the testis. The incidence of these disorders has increased over the last few decades, and testicular cancer now affects 1% of the Danish and Norwegian male population.

Methods To identify genetic variants that span the four TDS phenotypes, the authors performed a genome-wide association study (GWAS) using Affymetrix Human SNP Array 6.0 to screen 488 patients with symptoms of TDS and 439 selected controls with excellent reproductive health. Furthermore, they developed a novel integrative method that combines GWAS data with other TDS-relevant data types and identified additional TDS markers. The most significant findings were replicated in an independent cohort of 671 Nordic men.

Results Markers located in the region of *TGFBR3* and *BMP7* showed association with all TDS phenotypes in both the discovery and replication cohorts. An immunohistochemistry investigation confirmed the presence of transforming growth factor β receptor type III (TGFBR3) in peritubular and Leydig cells, in both fetal and adult testis. Single-nucleotide polymorphisms in the *KITLG* gene showed significant associations, but only with testicular cancer.

Conclusions The association of single-nucleotide polymorphisms in the *TGFBR3* and *BMP7* genes, which belong to the transforming growth factor β signalling pathway, suggests a role for this pathway in the pathogenesis of TDS. Integrating data from multiple layers can highlight findings in GWAS that are biologically relevant despite having border significance at currently accepted statistical levels.

INTRODUCTION

Infertility, cryptorchidism, small testis size with poor semen quality, and hypospadias are relatively common among men in Western countries and are all risk factors for testicular germ cell cancer (TGCC),[1,2] which has been increasing in incidence and is the most common malignancy of young men in many countries.[3] Caucasian men are at particularly high risk of testicular cancer, which currently affects ∼1% of the Danish and Norwegian male populations.[4] It has been proposed that most cases of TGCC, and some cases of hypospadias, cryptorchidism and decreased spermatogenesis (excluding those unrelated to development of the fetal testis) may be symptoms of a testicular dysgenesis syndrome (TDS) that originates from perturbations during fetal development.[2,5]

Recent progress in basic and epidemiological research on testis cancer has provided evidence that TGCC, in spite of its occurrence in young adult men, has its origin in fetal germ cells with stem cell characteristics,[6,7] the so-called intratubular carcinoma in situ (CIS) germ cells.[8] According to a widely tested hypothesis, these cells are derived from gonocytes that failed to differentiate to spermatogonia because of dysfunction of Sertoli, peritubular and Leydig cells, the function of which is essential for differentiation of fetal germ cells.[6,9] However, the invasive potential of CIS cells is not attained until after the pituitary—gonadal axis is activated during puberty.

Little is known about the causes of TDS, although environmental and lifestyle factors appear to play an important role, as suggested by the rapid increase in incidence of TGCC over a couple of generations.[3] The prevalence of mild forms of cryptorchidism and poor semen quality also appear to be increasing.[10,11] There is evidence of a genetic contribution, as brothers and fathers of patients with TGCC have a significantly increased risk of TGCC.[12] Marked ethnic differences in risk of TGCC among men living in the same areas also indicate a clear genetic component in the aetiology of the disease.[3] Three genome-wide association studies (GWASs)[13–15] and a replication study[16] have recently identified several genetic loci that predispose to TGCC (*KITLG*, *SPRY4*, *TERT-CLPTM1L*, *ATF7IP* and *DMRT1*). Altogether, these loci can potentially explain up to 13% of the heritability. The variants with highest effect size were found at 12q21, implicating *KITLG/KIT* signalling as a pathway involved in TGCC susceptibility.

On the basis of our hypothesis that patients with TDS may be genetically predisposed to as yet unknown environmental factors affecting inter-related pathways during testis development, we conducted a novel GWAS in which we grouped cases with at least one of four TDS phenotypes: TGCC, cryptorchidism, hypospadias, or infertility with low sperm concentration. The GWAS was conducted on a discovery cohort of 926 Danish men, comprising 439 healthy controls and 488 patients affected by at least one TDS phenotype. The cohort was carefully selected with the help of detailed patient records, which provided background information, and additional phenotype characterisation including semen analysis. Candidate markers from the discovery stage were subsequently assessed in an independent replication cohort of 671 Nordic men. The markers were selected using three approaches: (1) conventional single-marker association; (2) an integrative systems biology approach based on augmenting the GWAS data with several complementary types of data; and (3) protein-complex-based pathway analysis, which examined whether groups of related genes in the same functional pathway were jointly associated with the trait of interest.

Systems biology approaches that exploit the complementarities of heterogeneous types of data have recently been successfully applied in GWASs to detect associations that may be missed by single-marker association.[17–19] Here we used two such approaches. First, we adapted the MetaRanker tool[20] to integrate gene–phenotype associations from several complementary data sources, including targeted knockout experiments in mice, transcriptome time-series expression studies of the developing testis, and protein–protein interactions. In a second approach, we tested protein complexes for association with TDS, in order to jointly examine multiple markers from genes inter-related by prior biological knowledge.[21–23]

With the above approaches we were able to: (1) identify pleiotropic risk variants for TDS—that is, variants associated with all of the phenotypic subtypes; (2) confirm the risk factors in the *KITLG* gene found by two recent GWASs on TGCC; and (3) identify genes and protein complexes through integrative systems biology approaches that would not be found by single-marker association.

MATERIALS AND METHODS
Study subjects
The discovery cohort comprised 927 individuals of Danish descent: 439 healthy young men with semen concentrations above 60 million sperm/ml (table 1) and 488 cases (212 men with germ cell tumours, 138 men with cryptorchidism (diagnosed by lack of testis descent at birth), 31 men with hypospadias, and 107 infertile men identified using the following criteria: absence of any known cause of infertility (no Y-chromosome microdeletions and no history of radio- or chemo-therapy); absence of varicocele; sperm count below 15 million sperm/ml; and testis volume below 15 ml). Cases with several symptoms of TDS were grouped according to severity (TGCC>hypospadias>cryptorchidism>infertility) (supplementary figure 1 online and supplementary table 1 online).

The replication cohort included 235 controls and 436 TDS cases from Denmark, Sweden and Finland, three Nordic populations that are genetically closely related (supplementary figure 2 online). Cases of cryptorchidism (n=103) were contributed from Finland (Turku area) (n=34) and from Rigshospitalet Denmark (n=69); TGCC samples (n=333) were collected in Sweden (from the Malmö region (n=112) and the Stockholm area (n=135)) and Denmark (n=86); controls were obtained

Table 1 Constitution of discovery and replication cohorts

Phenotype	Discovery cohort	Replication cohort
Controls	439	235
Infertile	107	0
TGCC	212	333
Hypospadias	31	0
Cryptorchidism	138	103

TGCC, testicular germ cell cancer.

from Rigshospitalet, Denmark (n=184) and the Turku area in Finland (n=51) (table 1).

For investigations of gene expression at the protein level, we used nine samples of adult orchiectomy specimens (from men operated on for testicular tumours), as well as three samples of normal human fetal testis tissue (originating from spontaneous miscarriages at 14–15 weeks of gestation). All tissue samples were paraffin-embedded and stored in the pathology archives at Rigshospitalet.

The local ethics committees approved the study, and all subjects provided informed consent. The tissue samples were investigated as coded (adult testes) or anonymous (fetal testes) sections.

Genotyping
The discovery cohort was genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0, following the manufacturer-supplied protocol. Genotype calling and subsequent analysis were performed using the Birdseed algorithm (of Affymetrix Power Tools), Plink v1.06 and R. Genotyping of 34 associated markers and their call rates (supplementary tables 2 and 3 online), selected from the discovery stage, was performed using the KBiosciences Competitive AlleleSpecific PCR SNP genotyping system (KASPar) as an in-house service at KBiosciences. KASPar assay primers were designed using the Primer Picker software available at http://www.kbioscience.co.uk (KBiosciences). p Values were obtained by the MAX test,[24] and OR with confidence limits were retrieved by logistic regression in R.

Quality control
From the 926 samples included in the initial discovery cohort, we removed 55 for which genotyping call rates were below 96% or their quality control contrasts were below 0.4. Individuals detected to be of non-Scandinavian ancestry (n=11) were excluded (supplementary figure 3 online). In addition, 21 samples with high potential for confounding the association analysis due to risk of consanguinity (n=17) or a large amount of missing data (n=4) were excluded.

Single-marker association statistics
We removed markers with low minor allele frequencies (<0.1), an excess of missing genotypes (>0.01) or strong deviation from Hardy–Weinberg equilibrium (<10e−5). The resulting 600 798 markers were tested for association with the disease phenotypes using MAX statistics.[24] Multiple testing correction was performed using the maxT permutation method.

Immunohistochemistry
The polyclonal antibody, transforming growth factor β receptor III (TGFBR3) (Santa Cruz Biotechnology, Santa Cruz, California, USA; sc-6199), was used according to a standard indirect peroxidase protocol based on a Zymed Histostain kit (Invitrogen, Carlsbad, California, USA), as previously described.[25]

All tissue samples were formalin-fixed and paraffin-embedded. For negative controls, the primary antibody was replaced with the dilution buffer. For fetal testis (14−15 weeks of gestation, three samples), the same procedure as above was used, except that the tertiary layer was streptavidin-coupled with alkaline phosphatase and development was performed with BCIP NBT as previously described.[26]

Gene-level meta-rank analysis

A three-layer MetaRanker analysis[20] was applied to integrate information from external phenotype-specific datasets. Ranked lists of all human genes were obtained in each layer, as described below, and then combined into a meta-score. Low ranked genes were considered candidate susceptibility genes.

Layer 1: GWAS evidence

For each gene, a p value was obtained as the minimal GWAS p value of all markers mapping to the gene according to Affymetrix Genome-Wide Human SNP Array 6.0 annotations. The p values were bias-corrected for gene size, linkage disequilibrium and marker density, on the basis of the (latent) number of independent markers.[20 27] All genes were ranked according to the corrected p values.

Layer 2: expression data from fetal testis

To identify differentially expressed genes in the developing fetal testis, time-series array expression data from two studies on mouse fetal testis and from one study on human fetal testis were retrieved from the Gene Expression Omnibus and normalised by q-spline normalisation.[28] The study on human fetal testis was the only one found in the Gene Expression Omnibus; it provided data on 17 testis samples taken between 9 and 20 weeks of gestation.[29] Genes were ranked according to the coefficient of variation (the SD of expression divided by the mean). The mouse studies included data from biological duplicates taken at five time points between the time of the indifferent gonad (11.5 days postcoitus (dpc)) and birth (18.5 dpc),[30] respectively at five time points between gestational day 11 and 18.[31] For both mouse studies, genes were ranked on the basis of one-way analysis of variance F-tests for differential expression at different time points. A single ranked list of genes that describes the transcriptional regulation in the developing testis was obtained by summing the gene ranks from the three studies (after mapping the mouse genes to their human orthologues).

Layer 3: mouse targeted genes with protein−protein complexes

All targeted alleles with their human orthologues and associated morphologies were obtained from the Mouse Genome Informatics (MGI) database via the BioMart interface. A set of 34 morphologies from the Mammalian Phenotype ontology were manually curated on the basis of their developmental defects of the testis and relation to TDS (supplementary table 4 online). To avoid a bias towards spermatogenesis-related genes, two morphologies, 'male infertility' and 'abnormal spermatogenesis',

Figure 1 Manhattan plots showing the association between all single-nucleotide polymorphisms and (A) testicular dysgenesis syndrome and (B) the subset of cases of testicular germ cell cancer.
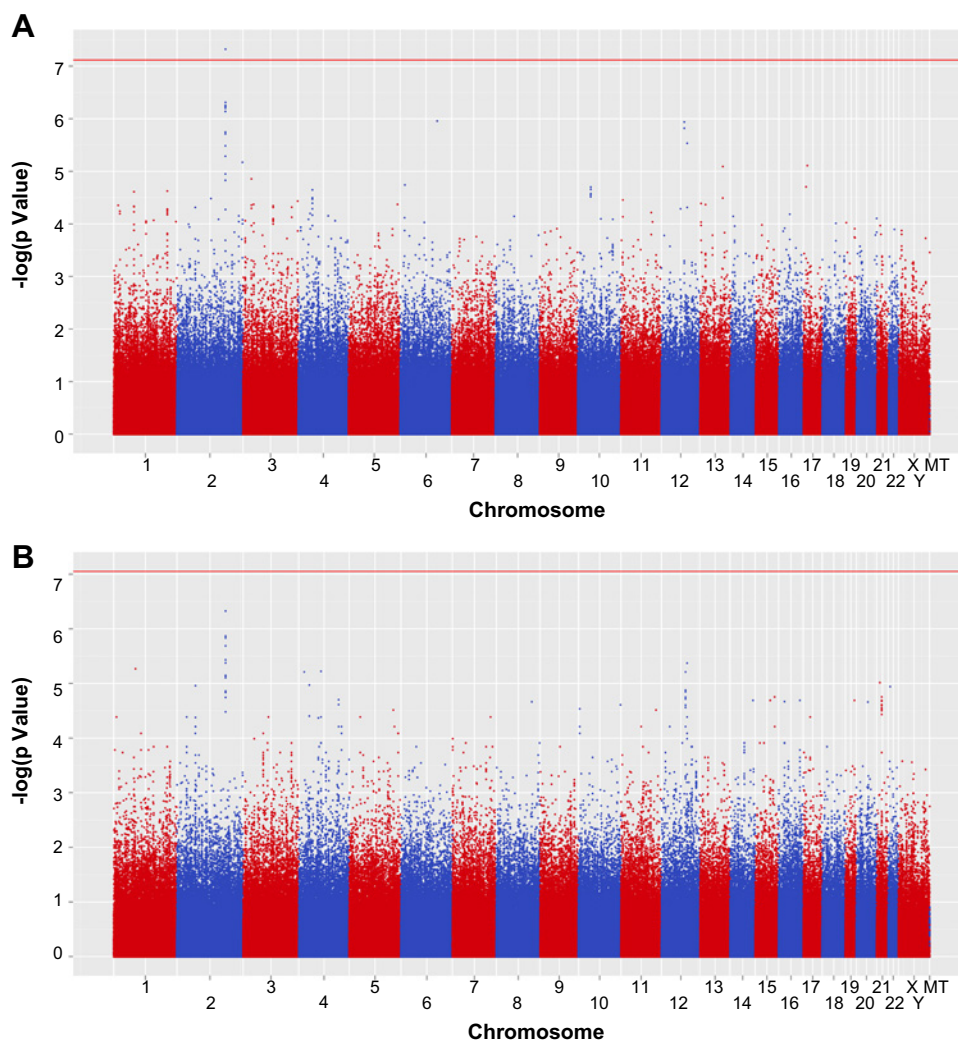
Table 2 Markers and genes with a possible association with the testicular dysgenesis syndrome (TDS) phenotype

Gene	Marker*	Selection method†	Phase	Risk allele	Control RAF‡	OR (95% CI) Per allele	OR (95% CI) Optimal genetic model	p_{unadj}	p_{adj}	Optimal genetic model§
KITLG	rs1352947	TGCC	Discovery	T	0.81	1.56 (1.19 to 2.05)	1.52 (1.18 to 1.97)	3.1×10^{-3}	1	Additive
	C/T		Replication		0.82	2.11 (1.48 to 3.03)	1.93 (1.39 to 2.69)	8.5×10^{-5}	2.0×10^{-3}	
TGFBR3	rs12082710	ISB	Discovery	T	0.58	1.35 (1.10 to 1.65)	1.77 (1.33 to 2.36)	2.4×10^{-4}	1	Recessive
	T/C		Replication		0.59	1.27 (0.99 to 1.63)	1.52 (1.08 to 2.15)	1.6×10^{-2}	3.8×10^{-1}	
BMP7	rs388286	Pathway	Discovery	C	0.47	1.34 (1.10 to 1.63)	1.36 (1.11 to 1.67)	2.3×10^{-3}	1	Additive
	C/T		Replication		0.47	1.29 (1.01 to 1.66)	1.28 (1.01 to 1.62)	4.1×10^{-2}	9.9×10^{-1}	
HOXDx	rs17198432	TDS	Discovery	A	0.07	2.31 (1.66 to 3.26)	2.58 (1.82 to 3.70)	4.7×10^{-8}	2.2×10^{-2}	Dominant
	C/A		Replication		0.11	0.96 (1.43 to 0.64)	0.97 (0.64 to 1.48)	8.7×10^{-1}	1	

*The marker with the lowest p value in the discovery cohort, among the markers tagging a gene, is presented.
†Four different approaches were used for the selection of markers: TDS, single-marker genome-wide association study (GWAS) on all TDS sub-phenotypes; TGCC, single-marker GWAS on the testicular germ cell cancer (TGCC) subset of cases; Pathway, aggregated effect in pathways; ISB, integrative systems biology by combing evidence of association from several data types.
‡Risk allele frequency.
§The genetic model with strongest association among the models tested by the MAX test (selected on discovery cohort).

were excluded. All descendants to the curated terms were then retrieved from the Mammalian Phenotype ontology, resulting in a total of 122 terms. This set of terms was mapped to 315 human orthologous genes that were targeted in MGI. Next, we ranked all human protein complexes according to their enrichment for these 315 targeted genes. A complex was defined as a central hub protein and all its first-order interaction partners in a network of protein–protein interactions.[32] Each complex was tested for enrichment of targeted genes using a hyper-geometric test. In this manner, all sets of genes interacting in a biologically active complex obtained a rank with respect to its enrichment of TDS targeted mouse genes.

Integration of three evidence layers into a meta-rank

For each gene, let L_j denote the rank from layer j. The meta-score was obtained as $S_{meta} = L_1 + L_2 + L_3$. Low S_{meta} rank indicates candidate disease genes.

### Pathway-based analysis of protein complex association

To complement the single-nucleotide polymorphism (SNP) GWAS and the gene-centric meta rank analysis, we applied a protein–protein interaction-based approach that leverages GWAS analysis by testing for the aggregate association across several gene products.[21] Protein complexes were defined and delineated by use of our previously published protein–protein interaction database, InWeb.[32] We ranked 8342 protein complexes on the basis of their enrichment in genes that harbour low TDS GWAS association p values. Of particular interest were complexes with low scores that also contained genes with low meta-rank from the three-layer MetaRanker analysis.

RESULTS

In this study we genotyped a discovery cohort comprising 439 controls and 488 cases affected by at least one of four TDS sub-phenotypes: testicular germ cell cancer, cryptorchidism, hypospadias or low semen concentration. Samples were analysed using the Affymetrix Genome-Wide Human SNP Array 6.0. After application of stringent quality control criteria, we compared the genotype frequency distributions between cases and controls for 600 798 markers using the MAX test,[24] which allows simultaneous investigation of additive, dominant and recessive inheritance. We observed only marginal confounding due to population stratification or other biases, as indicated by the genomic inflation factor (λ) of 1.019 and the quantile–quantile plot of observed versus expected χ^2 test statistics (supplementary figure 4 online).

Figure 2 Integration of genome-wide association study (GWAS) data with heterogeneous data types. The rationale behind the method is to prioritise genes if several data types show mildly pronounced associations with the phenotype, thereby identifying genes that would not have been found with GWAS analysis alone. TGFBR3, which was selected by this method and whose single-nucleotide polymorphism markers were further validated, is shown in red. The position of TGFBR3 is marked by red circles in the distributions of p values, and by red squares in the ranking of each data type layer. Three data types were used: (1) single-marker GWAS results of this study (grey); (2) targeted mutations from the Mouse Genome Informatics (MGI) database, filtered for testicular dysgenesis syndrome phenotypes, and ranked by their enrichment in protein–protein complexes (pink); (3) differential expression in the fetal testis of mouse and human (green). (A) A distribution of the p values of all human genes is shown for each of the data types. The p value ranges from 0 to 1, with 0 at the top. The p value associated with each gene is converted into a rank, such that each human gene is assigned a rank. The vertical line to the right of each distribution represents the ranking of all genes, where the top corresponds to rank 1. (B) The gene ranks of each individual data type are combined to a final meta-rank of each gene. As an example, TGFBR3 had the ranks 238, 42 and 408 among all human genes in the three data types: GWAS, targeted mutations with protein–protein interaction enrichment, and differential expression in the developing testis, respectively. After combination of these data type-specific ranks, TGFBR3 was ranked 3rd among all genes. PPI, protein-protein interaction.
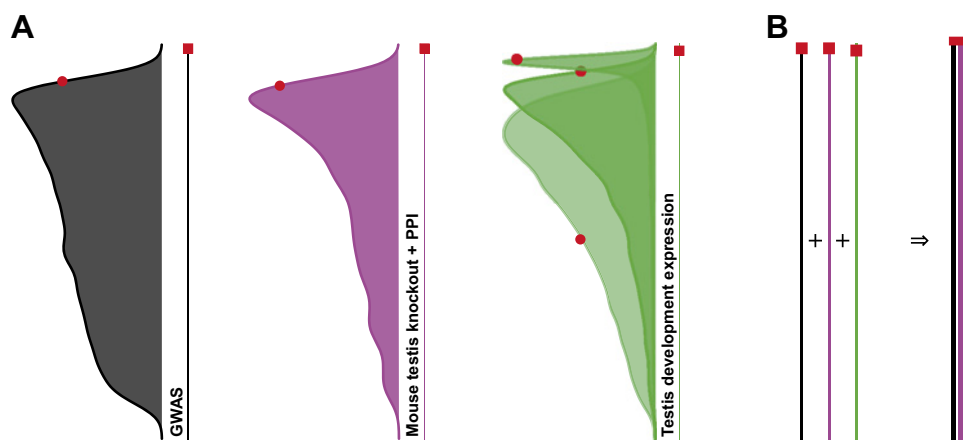
Single-marker association

We identified a strong association with all phenotypes of TDS ($p < 10^{-6}$) for nine markers, within the same linkage disequilibrium block, at 2q31.1, with the strongest association for rs17198432 (figure 1). The subset of 212 TGCC cases in our cohort was investigated in a separate single-marker analysis to compare the corresponding TGCC-specific results with those of two recent GWASs on TGCC.[13–15] The most significant marker at *KITLG* was ranked 12th among all markers ($p=1.17 \times 10^{-5}$). Furthermore, in the TGCC group, the observed genotype frequencies of the *KITLG* markers were comparable to those previously reported (table 2).

Integration of gene–phenotype associations

For each data type, all genes were ranked on the basis of their strength of association with the TDS phenotype. These data type-specific ranked lists of genes were then combined into a final ranked list of genes. In order to avoid high ranking of genes that were only supported by a single data type, we used a modification of the MetaRanker[20] tool by using the sum rather
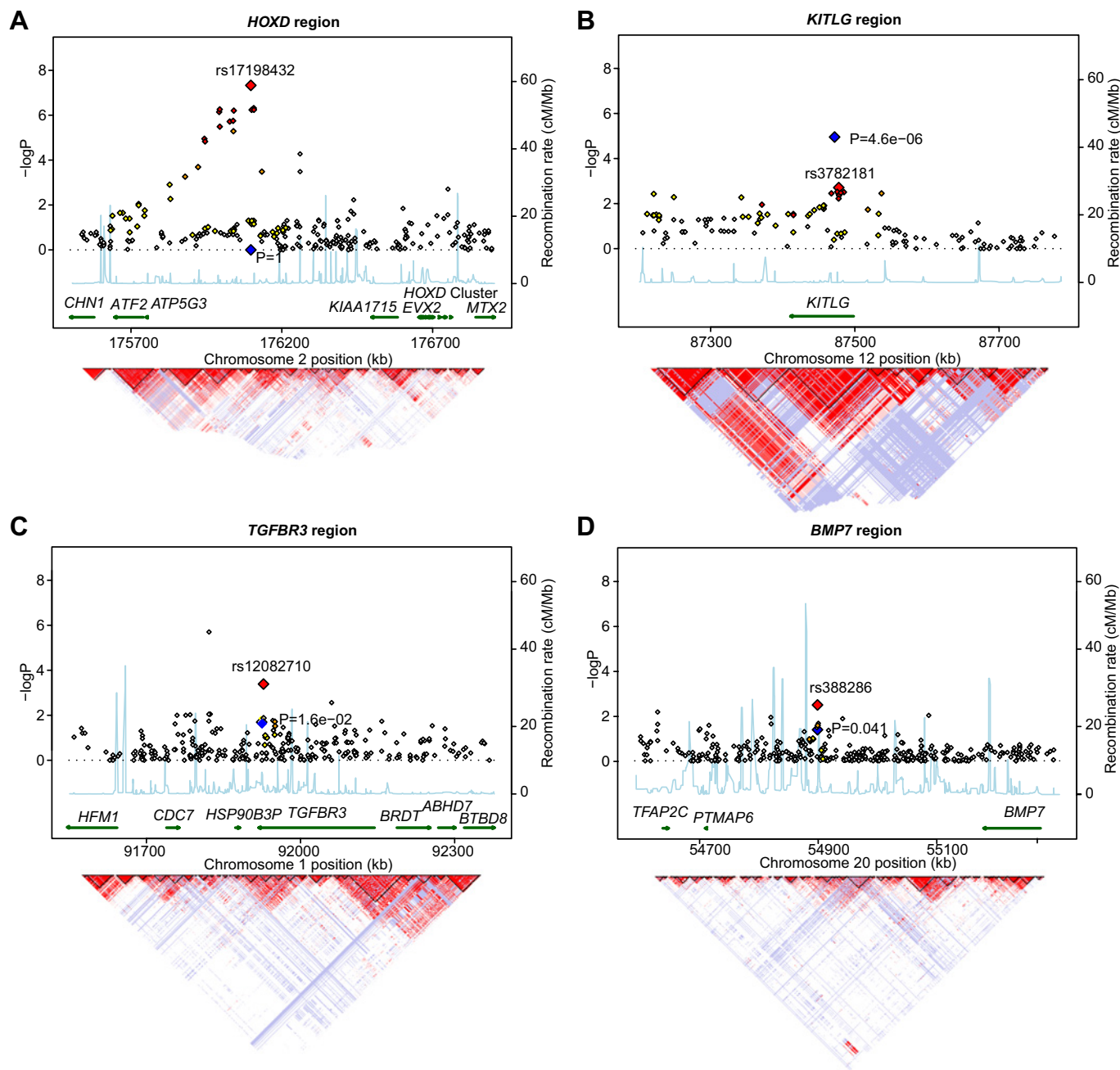

Figure 3 Regional association plots and linkage disequilibrium structure. The $-\log_{10}$ of the p values for the association of discovery markers are shown, and the markers are coloured in a white to red scale according to the strength of the pairwise linkage disequilibrium (r^2) to the most significant discovery marker at each locus. The blue marker represents the association in the replication stage. Light blue peaks indicate recombination rates from the CEU HapMap population. r^2-based linkage disequilibrium structures from the genome-wide association study data are displayed at the bottom. (A) The strongest association in the discovery stage was found at 2q31.1, close to the cluster of *HOXD* genes. (B) The most significant markers in the replication stage were found at the *KITLG* region. The genes *TGFBR3* (C) and *PBM7* (D) had single-nucleotide polymorphisms with mild association in the discovery stage, but were top-ranked by the integrative data analysis, and were further validated in the replication stage.

than the product of the individual gene ranks. Thus all human genes were ranked on the basis of a combination of microarray gene expression time-series studies of the developing fetal testis in mouse and human,[29–31] protein—protein interaction data,[32] targeted mice knockouts resulting in developmental defects of the testis (MGI), and gene scores based on the single-marker analysis of this GWAS (figure 2; see Materials and methods for details). From this analysis, we selected markers from 14 top-ranked genes, many of which are active during early development (supplementary tables 5 and 6 online). Genes implicated by previous GWASs of TGCC had relatively poor ranks (supplementary tables 7 and 8 online), partly because we used a TDS cohort rather than solely TGCC cases and partly because the genes did not exhibit any significant variable expression during testicular development.

Pathway analysis

Joint analysis of sets of markers may identify aggregated effects that are significant at a pathway level. Such analysis is motivated by the hypothesis that a set of genetic variations can perturb different components of a certain biologically functional entity, such as a signalling pathway or a protein complex, and each variation can influence the functional entity either individually or in combination. Thus we derived protein complexes from protein—protein interaction data, and tested each complex for significant enrichment of TDS-associated SNPs. A complex consisting of interacting members of the TGFβ superfamily (BMPER, BMP2, BMP7, BMP6 and BMP4) was found to be significantly associated (supplementary table 9 online).

Validation

On the basis of the findings from the marker selection approaches described above (TDS single-marker association, integration of gene—phenotype association and pathway analysis) and previously reported KITLG markers with association with TGCC, we selected a total of 39 SNPs (supplementary table 2 online) for subsequent validation in an independent replication cohort of 671 Nordic men. These included 16 SNPs from four independent loci based on single-marker association, and 23 SNPs based on integration of gene association ranks from different data types and protein complex ranks based on pathway analysis. Of the 671 samples, 23 did not meet the quality control standards for the genotyping system used by KBiosciences.

Markers at three genes showed a significant association in the replication cohort: KITLG, TGFBR3 and BMP7 with OR (95% CI) 1.93 (1.39 to 2.69), 1.52 (1.08 to 2.15), 1.28 (1.01 to 1.62) (table 2). Surprisingly, the nine SNPs located in the HOXD region that were identified by the single-marker association in the discovery cohort did not find support in the validation

cohort (table 2 and figure 3). Comparison of ORs for the different phenotypes of TDS showed that KITLG was mainly associated with TGCC, but not with the other TDS sub-phenotypes, whereas TGFBR3 and BMP7 showed similar association for all phenotypes (table 3).

For further investigation of TGFBR3 expression at the protein level, we performed an immunohistochemical analysis in fetal and adult testis (figure 4). Leydig cells and peritubular cells were TGFBR3-positive, predominantly at cell membranes, in both normal fetal testes and adult testes with dysgenetic features, such as the presence of CIS cells, hyperplastic Leydig cell micronodules and thickened peritubular membranes. Fetal gonocytes, CIS cells and other types of germ cells were negative, except for a very few scattered CIS cells in adluminal location and some staining in Golgi apparatus/pre-acrosome in round spermatids, which may be unspecific reactions. In contrast with a previous study,[33] TGFBR3 protein was not detected in Sertoli cells, probably because of the differences in tissue fixation methods or specificity of the antibodies used.

DISCUSSION

This study focused on understanding the genetics behind developmental TDS, which is typically associated with male infertility and TGCC, sometimes combined with mild forms of genital malformations. Some studies suggest that a common genetic background for TDS is doubtful,[1] and others have even implied that TDS does not exist.[34] We believe that TDS is a real syndrome with predominantly environment-dependent pathogenesis and a relatively weak component of genetic predisposition. This means that the prevalence of TDS among Caucasian populations probably depends on the strength of the adverse environmental or lifestyle factors. However, as this study suggests, it is likely that a subset of cases of TDS do share predisposing genetic variants. Here, we identified, using a combination of GWAS and systems biology approaches, genomic variants in TGFBR3 and BMP7 between several TDS phenotypes, most notably cryptorchidism and testicular cancer. TGFBR3 has not previously been reported to be directly associated with male infertility or TDS, but participates in pathways that are highly relevant in terms of involvement in regulation of embryogenesis and oncogenesis. In the single-marker approach, we also found evidence in the discovery cohort for a possible association for rs17198432 and other markers in the HOXD gene cluster.

It should be mentioned that SNP markers found to be significant in GWASs are more often than not in intergenic regions and not in coding or known regulatory regions. Picking the closest gene as the affected one is an assumption; however, cis-eQTL studies[35] have shown that it is a reasonable

Table 3 Associations with TDS and its sub-phenotypes, TGCC, cryptorchidism, hypospadias and infertility for the top markers in the discovery stage and for replicated markers

Gene	Marker	Phase	Risk allele	OR (95% CI)				
				TDS	TGCC	Cryptorchidism	Hypospadias	Infertility
KITLG	rs1352947	Discovery	T	1.56 (1.18 to 1.97)	2.74 (1.82 to 4.26)	1.08 (0.77 to 1.55)	1.00 (0.52 to 2.11)	1.26 (0.86 to 1.90)
		Replication		1.93 (1.39 to 2.69)	2.23 (1.55 to 3.28)	1.38 (0.89 to 2.24)	na	na
TGFBR3	rs12082710	Discovery	T	1.77 (1.33 to 2.36)	1.55 (1.07 to 2.23)	1.62 (1.07 to 2.44)	2.04 (0.90 to 4.56)	2.38 (1.54 to 3.70)
		Replication		1.52 (1.08 to 2.15)	1.49 (1.04 to 2.14)	1.64 (0.99 to 2.71)	na	na
BMP7	rs388286	Discovery	C	1.36 (1.11 to 1.67)	1.38 (1.06 to 1.79)	1.49 (1.11 to 2.01)	1.24 (0.70 to 2.22)	1.21 (0.89 to 1.66)
		Replication		1.28 (1.01 to 1.62)	1.37 (1.06 to 1.78)	1.09 (0.78 to 1.52)	na	na
HOXDx	rs17198432	Discovery	A	2.58 (1.82 to 3.70)	2.85 (1.86 to 4.38)	2.9 (1.81 to 4.63)	2.83 (1.11 to 6.64)	1.76 (1.01 to 3.01)
		Replication		0.97 (0.64 to 1.48)	0.96 (0.61 to 1.50)	1.00 (0.52 to 1.83)	na	na

na, these phenotypes were not assessed in the replication stage.
TGCC, testicular germ cell cancer; TDS, testicular dysgenesis syndrome.

Figure 4 Localisation of transforming growth factor β receptor type III (TGFBR3) protein in normal fetal testis (A), adult testis with complete spermatogenesis (B), and dysgenetic adult testis with carcinoma in situ and Leydig cell hyperplasia (C). Leydig cells are marked by arrowheads (A, B, C), gonocytes (A) and CIS cells (C) by large arrows and peritubular cells by small arrows (B & C). Inserts (lower right) are without primary antibody. Bars indicate 100 μm. TGFBR3 is seen to be expressed in Leydig and peritubular cells.
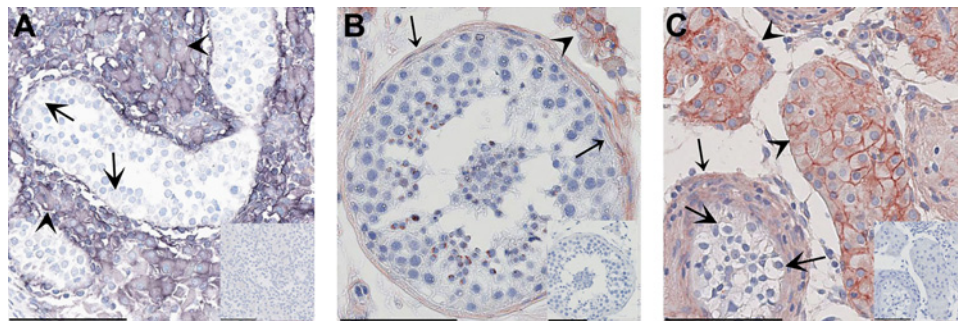
assumption. One of the strengths of our integrative approach is that it takes a gene focus that increases the probability of the gene being involved, as it requires the gene to have relevant expression, as well as protein—protein interactions of relevance to the phenotype.

The integrative systems biology analysis, which combined data from the GWAS with gene—phenotype associations found in complementary data types, including testis developmental expression data and protein—protein complexes associated with testicular developmental defects, identified several disparate loci that are functionally linked to the TGFβ superfamily signalling pathway, which were supported by the replication study. Using GWAS alone, these SNPs would fall into a 'grey zone',[18] and would not have been selected for replication. Many GWASs suffer from this phenomenon because the number of samples required to attain genome-wide significance can be unfeasibly high.

Among the loci with the highest significance across all four TDS phenotypes was the *TGFBR3* gene, which encodes the TGFβ receptor type III, a co-receptor for inhibins and TGFβ1—3, BMP2, BMP4 and BMP7. TGFBR3 and its co-receptors and ligands are expressed in most endocrine tissues, including the testis. TGFBR3 has been identified in human Sertoli cells and Leydig cells, both in the normal testis and in tubules with CIS.[36] In this study, we found that TGFBR3 was expressed in both Leydig and peritubular cells in fetal as well as adult testis, but was absent from Sertoli cells. This expression pattern supports the idea that TGFBR3 is essential for embryonic development of the reproductive system, as silencing of the murine *Tgfbr3* gene resulted in impaired function of fetal Leydig cells and testicular dysgenesis.[37] In addition, the signalling partners of TGFBR3, the family of activin receptors, are present in fetal human testis[38] and appear to be dysregulated in testes with TGCC.[33][39] Of note, among the loci that scored high in both the pathway analysis and the integration analysis were BMP7 and other bone morphogenetic proteins, which can bind TGFBR3. The presence of several components of the same pathway, which have all previously been implicated in the development of the testis and early reproductive system, reinforces the validity of the identified SNPs as possible genetic factors predisposing to TDS.

The analysis of the TGCC subset of cases confirmed the association of the *KITLG* locus (12q21) with increased risk of testicular cancer.[13][14] However, it has been suggested that SNPs in the *KIT* and *KITLG* genes may also be involved in infertility,[40] but we did not detect an association for other subtypes of TDS. Thus, the *KITLG* region should be further studied to pinpoint causative variants within *KITLG* or its regulatory regions. In the single-marker approach, we also found evidence for a possible association for rs17198432 and other markers in the *HOXD* gene cluster (supplementary table 10 online). *HOXD* genes encode a group of transcription factors well known for their importance in morphogenesis during early embryonic development of limbs and genitalia,[41] and a direct association with cryptorchidism has been shown in one study.[42] As *HOXD* was not supported in the replication cohort, we have not discussed this in detail, but we mention it as a variant locus that would be worth screening for in other cohorts.

Here, we have successfully applied systems biology methodologies to prioritise important findings in a GWAS of a carefully collected discovery cohort of individuals of Danish ethnic background with detailed clinical records, and a subsequent replication cohort with similar phenotypes. The GWAS data were complemented by integration of multiple data types. This approach enabled us to identify potentially significant variations that would normally not be prioritised on the basis of genome-wide multiple testing corrected p values. Two of the three replicated loci were selected on the basis of systems biology, indicating that integrative analyses can be useful for candidate selection of markers that do not reach genome-wide significance in a single-marker association analysis. The candidates identified in this study, notably those involved in TGFβ superfamily signalling pathways, provide evidence that at least a subset of TDS cases may have a common genetic predisposition. The paucity of informative markers indicates the predominant role of environmental and lifestyle factors in the pathogenesis of this syndrome. However, association with the TGFβ signalling pathway has not been shown previously, and this should be examined further. The findings should be corroborated in further independent GWASs, fine-mapping and sequencing studies, or by mechanistic investigations of the implicated pathways. In any case, integration of prevailing information via systems biology approaches holds promise for enhancing future GWASs.

Author affiliations
[1]Department of Growth and Reproduction, Rigshospitalet, Copenhagen, Denmark
[2]Centre for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark
[3]Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark
[4]Institute of Preventive Medicine, Copenhagen University Hospital, Centre for Health and Society, Copenhagen, Denmark
[5]Department of Clinical Sciences, Molecular Genetic Reproductive Medicine, University of Lund, Malmö, Sweden
[6]Department of Oncology-Pathology, Karolinska Institute and University Hospital, Stockholm, Sweden
[7]Departments of Physiology and Pediatrics, University of Turku, Turku, Finland
[8]Department of Oncology, Rigshospitalet, Copenhagen, Denmark
[9]Department of Biology, The August Krogh Building, Universitetsparken 13, Copenhagen, Denmark

REFERENCES

1. **Schnack TH,** Poulsen G, Myrup C, Wohlfahrt J, Melbye M. Familial coaggregation of cryptorchidism, hypospadias, and testicular germ cell cancer: a nationwide cohort study. *J Natl Cancer Inst* 2010;**102**:187–92.
2. **Skakkebaek NE,** Rajpert-De Meyts E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum Reprod* 2001;**16**:972–8.
3. **Purdue MP,** Devesa SS, Sigurdson AJ, McGlynn KA. International patterns and trends in testis cancer incidence. *Int J Cancer* 2005;**115**:822–7.
4. **Engholm G,** Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, Køtlum JE, Olafsdóttir E, Pukkala E, Storm HH. NORDCAN—a Nordic tool for cancer information, planning, quality control and research. *Acta Oncol* 2010;**49**:725–36.
5. **Asklund C,** Jensen TK, Main KM, Sobotka T, Skakkebaek NE, Jørgensen N. Semen quality, reproductive hormones and fertility of men operated for hypospadias. *Int J Androl* 2010;**33**:80–7.
6. **Rajpert-De Meyts E.** Developmental model for the pathogenesis of testicular carcinoma in situ: genetic and environmental aspects. *Hum Reprod Update* 2006;**12**:303–23.
7. **Sonne SB,** Almstrup K, Dalgaard M, Juncker AS, Edsgard D, Ruban L, Harrison NJ, Schwager C, Abdollahi A, Huber PE, Brunak S, Gjerdrum LM, Moore HD, Andrews PW, Skakkebaek NE, Rajpert-De Meyts E, Leffers H. Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma in situ is an arrested gonocyte. *Cancer Res* 2009;**69**:5241–50.
8. **Skakkebaek NE,** Berthelsen JG, Giwercman A, Müller J. Carcinoma-in-situ of the testis: possible origin from gonocytes and precursor of all types of germ cell tumours except spermatocytoma. *Int J Androl* 1987;**10**:19–28.
9. **Sharpe RM,** Skakkebaek NE. Testicular dysgenesis syndrome: mechanistic insights and potential new downstream effects. *Fertil Steril* 2008;**89**:e33–8.
10. **Boisen KA,** Kaleva M, Main KM, Virtanen HE, Haavisto AM, Schmidt IM, Chellakooty M, Damgaard IN, Mau C, Reunanen M, Skakkebaek NE, Toppari J. Difference in prevalence of congenital cryptorchidism in infants between two Nordic countries. *Lancet* 2004;**363**:1264–9.
11. **Jørgensen N,** Carlsen E, Nermoen I, Punab M, Suominen J, Andersen AG, Andersson AM, Haugen TB, Horte A, Jensen TK, Magnus Ø, Petersen JH, Vierula M, Toppari J, Skakkebaek NE. East-West gradient in semen quality in the Nordic-Baltic area: a study of men from the general population in Denmark, Norway, Estonia and Finland. *Hum Reprod* 2002;**17**:2199–208.
12. **Hemminki K,** Li X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. *Br J Cancer* 2004;**90**:1765–70.
13. **Rapley EA,** Turnbull C, Al Olama AA, Dermitzakis ET, Linger R, Huddart RA, Renwick A, Hughes D, Hines S, Seal S, Morrison J, Nsengimana J, Deloukas P, Rahman N, Bishop DT, Easton DF, Stratton MR. A genome-wide association study of testicular germ cell tumor. *Nat Genet* 2009;**41**:807–10.
14. **Kanetsky PA,** Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, Letrero R, Ciosek SL, Doody DR, Smith LM, Weaver J, Albano A, Chen C, Starr JR, Rader DJ, Godwin AK, Reilly MP, Hakonarson H, Schwartz SM, Nathanson KL. Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. *Nat Genet* 2009;**41**:811–15.
15. **Turnbull C,** Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, Ricketts M, Linger R, Nsengimana J, Deloukas P, Huddart RA, Bishop DT, Easton DF, Stratton MR, Rahman N. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat Genet* 2010;**42**:604–7.
16. **Kratz CP,** Han SS, Rosenberg PS, Berndt SI, Burdett L, Yeager M, Korde LA, Mai PL, Pfeiffer R, Greene MH. Variants in or near KITLG, BAK1, DMRT1, and TERT-CLPTM1L predispose to familial testicular germ cell tumour. *J Med Genet* 2011;**48**:473–6.
17. **Zhong H,** Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet* 2010;**86**:581–91.
18. **Naukkarinen J,** Surakka I, Pietiläinen KH, Rissanen A, Salomaa V, Ripatti S, Yki-Järvinen H, van Duijn CM, Wichmann HE, Kaprio J, Taskinen MR, Peltonen L. Use of genome-wide expression data to mine the "Grey Zone" of GWA studies leads to novel candidate obesity genes. *PLoS Genet* 2010;**6**:e1000976.
19. **Hsu YH,** Zillikens MC, Wilson SG, Farber CR, Demissie S, Soranzo N, Bianchi EN, Grundberg E, Liang L, Richards JB, Estrada K, Zhou Y, van Nas A, Moffatt MF, Zhai G, Hofman A, van Meurs JB, Pols HAP, Price RI, Nilsson O, Pastinen T, Cupples LA, Lusis AJ, Schadt EE, Ferrari S, Uitterlinden AG, Rivadeneira F, Spector TD, Karasik D, Kiel DP. An integration of genome-wide association study and gene expression profiling to prioritise the discovery of novel susceptibility loci for osteoporosis-related traits. *PLoS Genet* 2010;**6**:e1000977.
20. **Pers TH,** Hansen NT, Lage K, Koefoed P, Dworzynski P, Miller ML, Flint TJ, Mellerup E, Dam H, Andreassen O, Djurovic S, Melle I, Børglum A, Werge T, Purcell S, Ferreira M, Kouskoumvekaki I, Workman C, Hansen T, Mors O, Brunak S. Meta-analysis of heterogeneous data sources for genome-scale identification of risk genes in complex phenotypes. *Genet Epidemiol* 2011;**35**:318–32.
21. **Jensen MK,** Pers TH, Dworzynski P, Girman CJ, Brunak S, Rimm EB. Protein interaction-based genome-wide analysis of incident coronary heart disease. *Circ Cardiovasc Genet* 2011;**4**:549–56.
22. **Segrè AV,** Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 2010;**6**:e1001058.
23. **Wang K,** Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010;**11**:843–54.
24. **Li Q,** Zheng G, Li Z, Yu K. Efficient approximation of P-value of the maximum of correlated tests, with applications to genome-wide association studies. *Ann Hum Genet* 2008;**72**:397–406.
25. **Novotny GW,** Nielsen JE, Sonne SB, Skakkebaek NE, Rajpert-De Meyts E, Leffers H. Analysis of gene expression in normal and neoplastic human testis: new roles of RNA. *Int J Androl* 2007;**30**:316–26; discussion 326–7.
26. **Kristensen DM,** Nielsen JE, Skakkebaek NE, Graem N, Jacobsen GK, Rajpert-De Meyts E, Leffers H. Presumed pluripotency markers UTF-1 and REX-1 are expressed in human adult testes and germ cell neoplasms. *Hum Reprod* 2008;**23**:775–82.
27. **Sidak Z.** On multivariate normal probabilities of rectangles: their dependence on correlations. *Ann Math Statist* 1968;**39**:1425–34.
28. **Workman C,** Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, Saxild H-H, Nielsen C, Brunak S, Knudsen S. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* 2002;**3**(9): research0048.1—research0048.16.
29. **Houmard B,** Small C, Yang L, Naluai-Cecchini T, Cheng E, Hassold T, Griswold M. Global gene expression in the human fetal testis and ovary. *Biol Reprod* 2009;**81**:438–43.
30. **Small CL,** Shima JE, Uzumcu M, Skinner MK, Griswold MD. Profiling gene expression during the differentiation and development of the murine embryonic gonad. *Biol Reprod* 2005;**72**:492–501.
31. **McMahon AP,** Aronow BJ, Davidson DR, Davies JA, Gaido KW, Grimmond S, Lessard JL, Little MH, Potter SS, Wilder EL, Zhang P. GUDMAP: the genitourinary developmental molecular anatomy project. *J Am Soc Nephrol* 2008;**19**:667–71.
32. **Lage K,** Karlberg EO, Størling ZM, Olason PI, Pedersen AG, Rigina O, Hinsby AM, Tümer Z, Pociot F, Tommerup N, Moreau Y, Brunak S. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* 2007;**25**:309–16.
33. **Dias V,** Meacham S, Rajpert-De Meyts E, McLachlan R, Manuelpillai U, Loveland KL. Activin receptor subunits in normal and dysfunctional adult human testis. *Hum Reprod* 2008;**23**:412–20.
34. **Thorup J,** McLachlan R, Cortes D, Nation TR, Balic A, Southwell BR, Hutson JM. What is new in cryptorchidism and hypospadias—a critical review on the testicular dysgenesis hypothesis. *J Pediatr Surg* 2010;**45**:2074–86.
35. **Veyrieras JB,** Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 2008;**4**(10):e1000214.
36. **Dias VL,** Rajpert-De Meyts E, McLachlan R, Loveland KL. Analysis of activin/TGFB-signaling modulators within the normal and dysfunctional adult human testis reveals evidence of altered signaling capacity in a subset of seminomas. *Reproduction* 2009;**138**:801–11.
37. **Sarraj MA,** Escalona RM, Umbers A, Chua HK, Small C, Griswold M, Loveland K, Findlay JK, Stenvers KL. Fetal testis dysgenesis and compromised Leydig cell function in Tgfbr3 (beta glycan) knockout mice. *Biol Reprod* 2010;**82**:153–62.
38. **Anderson RA,** Cambray N, Hartley PS, McNeilly AS. Expression and localization of inhibin alpha, inhibin/activin betaA and betaB and the activin type II and inhibin beta-glycan receptors in the developing human testis. *Reproduction* 2002;**123**:779–88.
39. **Cobellis L,** Cataldi P, Reis FM, De Palo G, Raspagliesi F, Pilotti S, Arcuri F, Petraglia F. Gonadal malignant germ cell tumors express immunoreactive inhibin/activin subunits. *Eur J Endocrinol* 2001;**145**:779–84.
40. **Galan JJ,** De Felici M, Buch B, Rivero MC, Segura A, Royo JL, Cruz N, Real LM, Ruiz A. Association of genetic markers within the KIT and KITLG genes with human male infertility. *Hum Reprod* 2006;**21**:3185–92.
41. **Kondo T,** Zákány J, Innis JW, Duboule D. Of fingers, toes and penises. *Nature* 1997;**390**:29.
42. **Wang Y,** Barthold J, Kanetsky PA, Casalunovo T, Pearson E, Manson J. Allelic variants in HOX genes in cryptorchidism. *Birth Defects Res Part A Clin Mol Teratol* 2007;**79**:269–75.