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

Dalgaard, Marlene D; Weinhold, Nils; Edsgard, Stefan Daniel; Silver, Jeremy D; Pers, Tune Hannes; Nielsen, John E.; Jørgensen, Niels; Juul, Ånders; Gerds, Thomas A.; Giwercman, Aleksander; Giwercman, Yvonne L.; Cohn-Cedermark, Gabriella; Virtanen, Helena E.; Toppari, Jorma; Daugaard, Gedske; Jensen, Thomas Skot; Brunak, Søren; Rajpert-De Meyts, Ewa; Skakkebæk, Niels E; Leffers, Henrik; Gupta, Ramneek

Published in:
Journal of Medical Genetics

Link to article, DOI:
10.1136/jmedgenet-2011-100174

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
ORIGINAL ARTICLE

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

Marlene D Dalgaard,1 Nil Weinhold,2 Daniel Edsgård,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 Gedsko Daugaard,8 Thomas S Jensen,2 Søren Brunak,2 Ewa Rajpert-De Meyts,1 Niels E Skakkebaek,1 Henrik Jeffers,1,9 Ramneek Gupta2

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 borderline significance at currently accepted statistical levels.

INTRODUCTION

Infertility, cryptorchidism, small testes 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.5 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.5 Three genome-wide association studies (GWASs)13–15 and a replication study16 have recently identified several genetic loci that predispose to TGCC (KITLG, SPRY4, TERT-CLPTM1L, ATP7P and DMRT4). 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 459 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 tool20 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=105) were contributed from Finland (Turku area) (n=34) and from Rigshospitalet, Denmark (n=69); TGCC samples (n=553) were collected in Sweden (from the Malmö region (n=112) and the Stockholm area (n=135)) and Denmark (n=86); controls were obtained 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 orchietomy 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 Ficker software available at [http://www.kbioscience.co.uk](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 (<0.01–5). 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 (Invi-rogen, Carlsbad, California, USA), as previously described.25

**Table 1** Constitution of discovery and replication cohorts

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Discovery cohort</th>
<th>Replication cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>439</td>
<td>235</td>
</tr>
<tr>
<td>Infertile</td>
<td>107</td>
<td>0</td>
</tr>
<tr>
<td>TGCC</td>
<td>212</td>
<td>323</td>
</tr>
<tr>
<td>Hypospadias</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>Cryptorchidism</td>
<td>138</td>
<td>103</td>
</tr>
</tbody>
</table>

TGCC, testicular germ cell cancer.
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 analysis20 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.26 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’, were excluded. Figure 1 Manhatten 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.

![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.](figure1.png)
Table 2 Markers and genes with a possible association with the testicular dysgenesis syndrome (TDS) phenotype

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

*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 515 human orthologous genes that were targeted in MGI. Next, we ranked all human protein complexes according to their enrichment for these 515 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.25 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 Lj 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.22 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 ($\lambda$) of 1.019 and the

---

**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.
quantile-quantile plot of observed versus expected $\chi^2$ test statistics (supplementary figure 4 online).

**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 ($p=1.17\times10^{-5}$). 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

---

**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 BMP7 (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.
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\(^2\) tool by using the sum rather 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 Kbioscience.

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,\(^3\) 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.

### 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

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

na, these phenotypes were not assessed in the replication stage.
TGCC, testicular germ cell cancer; TDS, testicular dysgenesis syndrome.
of several components of the same pathway, which have all morphogenetic proteins, which can bind TGFBR3. The presence analysis and the integration analysis were BMP7 and other bone note, among the loci that scored high in both the pathway zone in the samples required to attain genome-wide signi.

Leydig cells, both in the normal testis and in tubules with CIS.36 ligands are expressed in most endocrine tissues, including the testicular cancer.13 14 However, it has been suggested that SNPs study. Using GWAS alone, these SNPs would fall into a signalling pathway, which were supported by the replication result. Using GWAS alone, these SNPs would fall into a ‘grey zone’,10 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 uneasily 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 TGFβ1–5, 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.30 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.31 In addition, the signalling partners of TGFBR3, the family of activin receptors, are present in fetal human testis30 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.

Acknowledgements We would like to thank Betina F Nielsen, Ana Ricci Nielsen and Olga Regina for their excellent technical assistance, and Rasmus Nielsen and Anders AlbrechtSEN for statistical advice. We thank all patients for participating in this study, and the health professionals for facilitating our work. The population allele and genotype frequencies were obtained from the data source funded by the Nordic Center of Excellence in Disease Genetics based on samples regionally selected from Finland, Sweden and Denmark.

Funding This work was supported by the Villum Kann Rasmussen Foundation, a NABIT grant from the Danish Strategic Research Council, the Novo Nordisk Foundation, the Academy of Finland, Sigrid Juselius Foundation, Foundation for Paediatric Research, Turku University Hospital, the European Commission (FP7/2008-2012: DEER 212844), the Swedish Cancer Society (CAN 2008/817),
Gunnar Nissons Cancer Foundation, Malmö University Hospitals Cancer Foundation and King Gustaf V’s Jubilee Fund for Cancer Research.

Competing interests None.

Patient consent Obtained.

Ethics approval The local ethics committees in Denmark, Sweden and Finland.

REFERENCES

27. Kondo T, Kondo H, Shen J, Pfeiffer R, Greene MH. Variants in or near KITLG, BAK1, DMRT1, and TERT-CLPTM1L are associated with environmental aetiology. Nat Genet 2009;41:789–90.
A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation


doi: 10.1136/jmedgenet-2011-100174

Updated information and services can be found at:
http://jmg.bmj.com/content/49/1/58.full.html

These include:

**Data Supplement**

“Supplementary Data”
http://jmg.bmj.com/content/suppl/2011/12/07/jmedgenet-2011-100174.DC1.html
“Supplementary Data”
http://jmg.bmj.com/content/suppl/2011/12/02/jmedgenet-2011-100174.DC2.html

**References**

This article cites 42 articles, 17 of which can be accessed free at:
http://jmg.bmj.com/content/49/1/58.full.html#ref-list-1

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/