Common variants in *DGKK* are strongly associated with risk of hypospadias

Loes F.M. van der Zanden[1,2], Iris A.L.M. van Rooij[1], Wout F.J. Feitz[3], Jo Knight[4,5], A. Rogier T. Donders[1], Kirsten Y. Renkema[2], Ernie M.H.F. Bongers[2], Sita H.H.M. Vermeulen[1,2], Lambertus A.L.M. Kiemeney[1,3], Joris A. Veltman[2], Alejandro Arias-Vásquez[2,6], Xufeng Zhang[7], Ellen Markljung[7], Liang Qiao[8], Laurence S. Baskin[8], Agneta Nordenskjöld[7,9], Nel Roeleveld[1,10], Barbara Franke[2,6,10], Nine V.A.M. Knoers[2,10]

[1] Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, [2] Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, [3] Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, [4] Division of Genetics and Molecular Medicine, King's College London School of Medicine, London, United Kingdom [5] National Institute for Health Research (NIHR), Biomedical Research Centre, Guy's and St. Thomas' NHS Foundation Trust and King's College London, London, United Kingdom, [6] Department of Psychiatry, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, [7] Department of Women's and Children's Health and Center of Molecular Medicine, Karolinska Institutet, Stockholm, Sweden, [8] Department of Urology, University of California, San Francisco, USA [9] Department of Pediatric Surgery, Astrid Lindgren Children Hospital, Karolinska University Hospital, Stockholm, Sweden, [10] These authors contributed equally to this work

Corresponding author

Barbara Franke, PhD
Department of Human Genetics, 855
Radboud University Nijmegen Medical Centre
P.O. Box 9101
6500 HB Nijmegen
The Netherlands
Tel: +31 24 3614017
Fax: +31 24 3668752
E-mail: B.Franke@antrg.umcn.nl

Abstract

Hypospadias is a common congenital malformation of the male external genitalia. We performed a genome-wide association study using pooled DNA from 436 individuals with hypospadias (cases) and 494 controls of European descent and selected the highest ranked SNPs for individual genotyping in the discovery sample, an additional Dutch sample of 133 cases and their parents, and a Swedish series of 266 cases and 402 controls. Individual genotyping of two SNPs (rs1934179 and rs7063116) in *DGKK*, encoding diacylglycerol kinase κ, produced compelling evidence for association with hypospadias in the discovery sample (allele-specific odds ratio (OR) = 2.5, $P = 2.5 \times 10^{-11}$ and OR = 2.3, $P = 2.9 \times 10^{-9}$, respectively) and in the Dutch (OR = 3.9, $P = 2.4 \times 10^{-5}$ and OR = 3.8, $P = 3.4 \times 10^{-5}$) and Swedish (OR = 2.5, $P = 2.6 \times 10^{-8}$ and OR = 2.2, $P = 2.7 \times 10^{-6}$) replication samples. Expression studies showed expression of *DGKK* in preputial tissue of cases and controls, which was lower in carriers of the risk allele of rs1934179 ($P = 0.047$). We propose *DGKK* as a major risk gene for hypospadias.

Article

Hypospadias is a common congenital hypoplasia of the penis, affecting approximately 1 in 750 births in Europe (see URLs). Due to developmental arrest of urethral fusion, the urethral opening is displaced along the ventral side of the penis. The opening can be located glandular, penile or even more posterior in the scrotum or perineum. Although most children with this condition undergo surgery in their second year of life, serious medical, social and sexual problems may still exist later in life. Hypospadias shows familial clustering[1], pointing toward genetic factors being important in its etiology[2]. Hypotheses about the multifactorial etiology of hypospadias mainly focus on hormonal disturbances. Polymorphisms in endocrine-related genes have been associated with hypospadias[3-11]. However, most of these associations have been found in small studies and were not replicated in the series used in the present study[12]. Therefore, our understanding of the molecular pathways leading to hypospadias is incomplete.

With the availability of SNP microarrays, genome-wide association studies (GWASs) have become feasible in elucidating the genetic basis of common complex disorders. Large sample sizes are needed in GWASs to detect genetic factors with modest effects on disease risk, having substantial implications in terms of costs. A useful solution is offered by DNA pooling, which has been proven to be feasible and accurate[13-15].

To identify genetic variants contributing to hypospadias susceptibility, we performed the first GWAS for this malformation using pooled DNA samples. We included 436 cases of European descent with isolated anterior or middle hypospadias from the AGORA (Aetiologic research into Genetic and Occupational/environmental Risk factors for Anomalies in children) project (Supplementary Fig. 1 and Supplementary Table 1) and 494 unaffected male controls of European descent from the Nijmegen Biomedical Study. In this discovery sample, we allelotyped 906,600 SNPs in duplicate using Affymetrix GeneChip 6.0 microarrays and calculated allele frequencies using *k*-corrected signal intensities (see Supplementary Note). The worst performing 5% of measurements, indicated by the biggest differences between the allele frequency estimates from the duplicate measurements, were excluded. Furthermore, we excluded SNPs based on several quality control criteria, such as high variance in case or control pools and minor allele frequencies (MAF) below 5%. A total of 574,400 SNPs passed quality control steps and were included in the analyses. We selected the 50 highest ranked SNPs based on the standard χ^2 statistic and a modified χ^2 statistic[16] applied to the raw and *k*-corrected allele frequency estimates. Of these 50 SNPs, we chose 20 based on several criteria, such as location near a gene and MAF (Supplementary Table 2). Seven of these 20 SNPs were located in the X-chromosomal gene *DGKK*, encoding diacylglycerol kinase κ. As most SNPs in this gene are in strong linkage disequilibrium (LD) with each other (Supplementary Fig. 2), we selected the intronic SNP that tagged the most other SNPs (rs1934179) for individual genotyping in the

discovery sample, as well as a potentially regulatory SNP in the 5' upstream region (rs7063116). Furthermore, nine SNPs in other genes were individually genotyped (Supplementary Table 3).

Table 1. Association of hypospadias with SNPs in *DGKK* in the discovery sample, the Dutch replication sample and the Swedish replication sample.

Study population	*rs1934179 in DGKK*			*rs7063116 near DGKK*		
	OR	**(95% CI)**	**χ^2 *P***	**OR**	**(95% CI)**	**χ^2 *P***
Dutch discovery sample	2.46	(1.88-3.21)	2.5×10^{-11}	2.25	(1.72-2.96)	2.9×10^{-9}
Dutch replication sample	3.92	(2.08-7.38)	2.4×10^{-5}	3.83	(2.03-7.24)	3.4×10^{-5}
Swedish replication sample	2.48	(1.80-3.42)	2.6×10^{-8}	2.16	(1.56-2.99)	2.7×10^{-6}

OR, odds ratio; CI, confidence interval.

Individual genotyping was completed with a success rate of ≥99%. All genotype frequencies in the controls were in Hardy-Weinberg equilibrium ($P > 0.05$). Both SNPs in *DGKK* showed genome-wide significant association in the discovery sample (OR for the A (risk) allele of these X-chromosomal SNPs in our male sample = 2.5, $P = 2.5 \times 10^{-11}$ for rs1934179 and OR = 2.3, $P = 2.9 \times 10^{-9}$ for rs7063116) (Table 1). These results were validated by genotyping the parents of the cases using the transmission disequilibrium test, a method robust to population stratification (Table 2). Eight of the other nine SNPs showed suggestive association with hypospadias ($P < 0.05$) (Supplementary Table 4).

Table 2. Association of hypospadias with SNPs in *DGKK* in cases in the discovery sample and their parents.

SNP	**Minor allele**[a]	**T**	**NT**	**OR**	**(95% CI)**	**χ^2 *P***
rs1934179 in DGKK	A	147	58	2.53	(1.87-3.43)	2.0×10^{-9}
rs7063116 near DGKK	A	134	65	2.06	(1.53-2.77)	1.7×10^{-6}

[a]*The least frequent allele in the controls of the discovery sample; NT, minor allele not transmitted; T, minor allele transmitted; OR, odds ratio; CI, confidence interval.*

For the ten associated SNPs, we subsequently attempted replication in an additional Dutch sample of 133 anterior or middle hypospadias cases of European descent and their parents. Seven SNPs showed similar ORs in this sample compared to the discovery sample, although most of these SNPs did not reach statistical significance, probably due to the small number of heterozygous parents. The only exceptions to this were the SNPs in *DGKK*, in which the same A alleles were again strongly associated with hypospadias (OR = 3.9, $P = 2.4 \times 10^{-5}$ for rs1934179 and OR = 3.8, $P = 3.4 \times 10^{-5}$ for rs7063116) (Table 1 and Supplementary Table 5). A second replication in a Swedish cohort of 266 anterior or middle hypospadias cases and 402 male controls convincingly confirmed the associations with the A alleles of the SNPs in *DGKK* (OR =

2.5, $P = 2.6 \times 10^{-8}$ for rs1934179 and OR = 2.2, $P = 2.7 \times 10^{-6}$ for rs7063116), whereas associations with the other eight SNPs did not reach statistical significance (Table 1 and Supplementary Table 6). We then performed a meta-analysis with both the discovery sample and the two replication samples (Supplementary Table 7). In addition to the SNPs in *DGKK*, SNPs in *PPARGC1B* (rs4705372) and *GRID1* (rs1880386) reached statistical significance in this analysis after correcting the critical *P* value for multiple testing (critical Bonferroni $P < 0.005$).

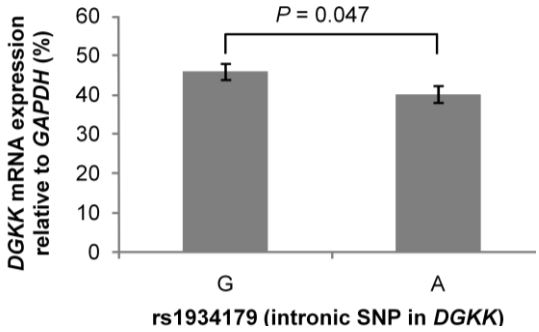
The pathogenesis of hypospadias probably includes many causal factors. We calculated the population attributable fraction (PAF) for hypospadias of rs1934179 in *DGKK* to be 32% in the Dutch population and 31% in the Swedish population, meaning that the variant underlying the association between rs1934179 in *DGKK* and hypospadias is one of the causal factors in nearly one third of hypospadias cases. As a comparison, the PAF for *APOE* in Alzheimer's disease is 26% (ref. 17), and that association is one of the strongest and best known genetic associations reported for a multifactorial disorder. However, the PAF calculated for hypospadias is based on data from our study only, and independent population-based studies should be performed to verify the validity of the estimate.

DGKK encodes a human type II diacylglycerol kinase[18]. Diacylglycerol kinases modulate the balance between diacylglycerol and phosphatidic acid, two signaling lipids, thereby playing an important role in signal transduction. *DGKK* mRNA is most abundant in testis and placenta[18]. Although *DGKK* has not previously been associated with hypospadias and there are other genes in close proximity to it, we suggest *DGKK* as the hypospadias susceptibility gene in the identified X-chromosomal locus as the LD block in which the gene is located encompasses only *DGKK* (Supplementary Fig. 2). Because the LD block also covers likely regulatory regions, variants regulating *DGKK* expression may underlie the association of *DGKK* with hypospadias. We performed real-time quantitative PCR analyses showing that *DGKK* is expressed in preputial skin of all investigated healthy boys (n = 10) and of hypospadias cases (n = 14) (Supplementary Fig. 3). Expression was lower in individuals with the A (risk) allele (n = 15) of rs1934179 ($P = 0.047$) (Fig. 1). These results suggest that variants regulating *DGKK* mRNA expression underlie the association of rs1934179 with hypospadias. A search for potential functional variants identified one SNP in a *FOXL1* transcription factor binding site, rs1934176, which is in high LD ($r^2 = 0.99$) with rs1934179; however, the significance of this SNP requires further study.

SNPs in *GRID1*, *PPARGC1B* and *KIAA2022* showed associations in the same direction in the discovery sample and in both replication samples, resulting in outcomes with a higher level of statistical significance in the meta-analysis. This suggests that nonsignificance of results may be due to lack of power. Indeed, the power of our study to detect associations with OR < 1.5 was limited and was even further reduced by the fact that we used DNA pooling. Using DNA pools instead of individual DNA samples results in less accurate allele frequency estimates, possibly producing more false-positive and false-negative findings. Validating our results by individual

genotyping enabled us to identify false-positive findings arising from such inaccurate estimates. False-negative results cannot be identified, however, and we may have missed additional associations that we would have detected with a GWAS based on individual genotyping. In addition, we may have missed associations with rare variants by excluding SNPs with MAF below 5%, that is, SNPs for which we had insufficient power. Therefore, individual GWAS (preferentially in larger samples) may identify additional hypospadias loci.

Figure 1. Quantification of *DGKK* mRNA expression (± s.e.m.) in preputial skin of hypospadias cases and controls relative to *GAPDH* mRNA expression.



Previously performed studies showed familial occurrence of hypospadias for anterior and middle forms of hypospadias but not for posterior types[1,19]. Because of this apparent etiologic heterogeneity, we included only anterior and middle cases in the current analyses. As expected, an additional analysis of the SNPs in *DGKK* in cases with posterior hypospadias showed weaker associations (Table 3), although the small number of cases used may have hampered a fair comparison. Nevertheless, these data are compatible with anterior and middle forms of hypospadias having an oligogenic or polygenic multifactorial etiology, including a crucial role for *DGKK*, and with posterior forms having a different etiology. These results warrant stratification by hypospadias phenotype based on location of the urethral opening in future genetic studies, which may reduce genetic heterogeneity and improve the reproducibility of the results.

In summary, we have identified a new X-chromosomal risk locus for hypospadias. We showed expression of *DGKK* in preputial skin, which was lower in boys with the risk allele. We propose *DGKK* as a major risk gene for anterior and middle forms of hypospadias. Because hypospadias is a fusion defect, *DGKK* might be important for other congenital closure defects as well.

Table 3. Association of different forms of hypospadias with SNPs in *DGKK*.

Study population ($n_{controls}/n_{cases}$)	**Minor allele (A) frequency (%)**				
	Controls	**Cases**	**OR**	**(95% CI)**	χ^2 ***P***
rs1934179 in DGKK					
Dutch sample					
Anterior and middle cases (494/436) (cases used in the discovery sample)	32.9	54.6	2.46	(1.88-3.21)	2.5×10^{-11}
Posterior cases (494/87)	32.9	38.1	1.26	(0.78-2.03)	0.348
Swedish sample					
Anterior and middle cases (402/266) (cases used in the replication sample)	31.0	52.7	2.48	(1.80-3.42)	2.6×10^{-8}
Posterior cases (402/62)	31.0	45.9	1.89	(1.09-3.26)	0.021
rs7063116 near DGKK					
Dutch sample					
Anterior and middle cases (494/436) (cases used in the discovery sample)	29.3	48.3	2.25	(1.72-2.96)	2.9×10^{-9}
Posterior cases (494/87)	29.3	26.2	0.86	(0.51-1.45)	0.565
Swedish sample					
Anterior and middle cases (402/266) (cases used in the replication sample)	29.4	47.3	2.16	(1.56-2.99)	2.7×10^{-6}
Posterior cases (402/62)	29.4	41.0	1.67	(0.96-2.91)	0.067

OR, odds ratio; CI, confidence interval.

Acknowledgements

We thank all cases and parents for their cooperation in this study. We are grateful to the Microarray Facility Nijmegen of the Department of Human Genetics (RUNMC), in particular to M. Steehouwer, for performing the microarray experiments, and to T. Vrijenhoek and J. Hehir-Kwa for their assistance in data analysis. We would also like to thank everyone involved in the data collection, S. van der Velde-Visser, C. Beumer, K. Kwak, J. Knoll, R. de Gier, B. Kortmann, A. Paauwen, H.G. Kho, J. Driessen and the anesthesiologists of OR 18. Finally, we are grateful to M. Coenen, J. Groothuismink, R. Makkinje and M. Schijvenaars for their practical guidance. This research is performed within a PhD project supported by the Radboud University Nijmegen Medical Centre and with extra budget from the Urology Foundation 1973. J. Knight is funded by the Department of Health through the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust.

Author contributions

L.F.M.v.d.Z. was the principal investigator who conducted the study. I.A.L.M.v.R., N.R., B.F. and N.V.A.M.K. designed the study and obtained financial support. L.F.M.v.d.Z., I.A.L.M.v.R., W.FJ.F., K.Y.R., E.M.H.F.B., S.H.H.M.V., L.A.L.M.K., N.R., B.F. and N.V.A.M.K. were involved in the collection of the discovery sample and the Dutch replication sample. J.A.V., A.A.-V. and B.F. collected the in-house controls. X.Z., E.M. and A.N. were responsible for the collection of the Swedish replication sample. L.Q. and L.S.B. collected the prepuce samples and performed the expression studies. L.F.M.v.d.Z. conducted all statistical analyses in collaboration with I.A.L.M.v.R., J.K. and A.R.T.D. L.F.M.v.d.Z. took primary responsibility for drafting the manuscript, with intellectual contributions, editing and approval from all other authors.

References

1. Fredell L, Kockum I, Hansson E, Holmner S, Lundquist L, Läckgren G, Pedersen J, Stenberg A, Westbacke G, Nordenskjöld A. Heredity of hypospadias and the significance of low birth weight. *J Urol* 2002:167;1423-1427.
2. Schnack TH, Zdravkovic S, Myrup C, Westergaard T, Christensen K, Wohlfahrt J, Melbye M. Familial aggregation of hypospadias: a cohort study. *Am J Epidemiol* 2007:167;251-256.
3. Watanabe M, Yoshida R, Ueoka K, Aoki K, Sasagawa I, Hasegawa T, Sueoka K, Kamatani N, Yoshimura Y, Ogata T. Haplotype analysis of the estrogen receptor 1 gene in male genital and reproductive abnormalities. *Hum Reprod* 2007:22;1279-1284.
4. Wang Y, Li Q, Xu J, Liu Q, Wang W, Lin Y, Ma F, Chen T, Li S, Shen Y. Mutation analysis of five candidate genes in Chinese patients with hypospadias. *Eur J Hum Genet* 2004:12;706-712.
5. Aschim EL, Nordenskjöld A, Giwercman A, Lundin KB, Ruhayel Y, Haugen TB, Grotmol T, Giwercman YL. Linkage between cryptorchidism, hypospadias, and GGN repeat length in the androgen receptor gene. *J Clin Endocrinol Metab* 2004:89;5105-5109.
6. Beleza-Meireles A, Omrani D, Kockum I, Frisén L, Lagerstedt K, Nordenskjöld A. Polymorphisms of estrogen receptor β gene are associated with hypospadias. *J Endocrinol Invest* 2006:28;5-10.
7. Beleza-Meireles A, Töhönen V, Söderhäll C, Schwentner C, Radmayr C, Kockum I, Nordenskjöld A. Activating transcription factor 3: a hormone responsive gene in the etiology of hypospadias. *Eur J Endocrinol* 2008:158;729-739.
8. Radpour R, Rezaee M, Tavasoly A, Solati S, Saleki A. Association of long polyglycine tracts (GGN repeats) in exon 1 of the androgen receptor gene with cryptorchidism and penile hypospadias in iranian patients. *J Androl* 2007:28;164-169.
9. Beleza-Meireles A, Kockum I, Lundberg F, Söderhäll C, Nordenskjöld A. Risk factors for hypospadias in the estrogen receptor 2 gene. *J Clin Endocrinol Metab* 2007:92;3712-3718.

10. Thai HTT, Kalbasi M, Lagerstedt K, Frisén L, Kockum I, Nordenskjöld A. The valine allele of the V89L polymorphism in the 5-α-reductase gene confers a reduced risk for hypospadias. *J Clin Endocrinol Metab* 2005:90;6695-6698.

11. Ban S, Sata F, Kurahashi N, Kasai S, Moriya K, Kakizaki H, Nonomura K, Kishi R. Genetic polymorphisms of *ESR1* and *ESR2* that may influence estrogen activity and the risk of hypospadias. *Hum Reprod* 2008:23;1466-1471.

12. van der Zanden LFM, van Rooij IALM, Feitz WFJ, Vermeulen SHHM, Kiemeney LALM, Knoers NVAM, Roeleveld N, Franke B. Genetics of hypospadias: are single-nucleotide polymorphisms in *SRD5A2*, *ESR1*, *ESR2*, and *ATF3* really associated with the malformation? *J Clin Endocrinol Metab* 2010:95;2384-2390.

13. Butcher LM, Davis OSP, Craig IW, Plomin R. Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behav* 2008:7;435-446.

14. Macgregor S, Zhao ZZ, Henders A, Nocholas MG, Montgomery GW, Visscher PM. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. *Nucleic Acids Res* 2008:36;e35.

15. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL *et al.* Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. *Am J Hum Genet* 2007:80;126-139.

16. Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG, Hettema JM, An SS, Neale MC, van den Oord EJCG *et al.* A whole genome association study of neuroticism using DNA pooling. *Mol Psychiatry* 2008:13;302-312.

17. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B *et al.* Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat Genet* 2009:41;1094-1099.

18. Imai S, Kai M, Yasuda S, Kanoh H, Sakane F. Identification and characterization of a novel human type II diacylglycerol kinase, DGK kappa. *J Biol Chem* 2005:280;39870-39881.

19. Brouwers MM, van der Zanden LFM, de Gier RPE, Barten EJ, Zielhuis GA, Feitz WFJ, Roeleveld N. Hypospadias: risk factor patterns and different phenotypes. *BJU Int* 2010:105;254-262.

URLs

http://www.eurocat-network.eu/; R, http://www.R-project.org; Review Manager 5, http://www.cc-ims.net/revman.

Online methods

Pool construction

The AGORA (Aetiologic research into Genetic and Occupational/environmental Risk factors for Anomalies in children) project of the Radboud University Nijmegen Medical Centre (RUNMC), Nijmegen, The Netherlands, is building a data bank and bio bank with questionnaire data and DNA samples from individuals with a congenital malformation or childhood cancer, as well as their parents. DNA was available for 679 hypospadias cases from the AGORA project (Supplementary Fig. 1 and Supplementary Table 1) and 596 unaffected male controls of European descent from the Nijmegen Biomedical Study[20]. The Arnhem-Nijmegen Regional Committee on Research Involving Human Subjects approved both studies. All participants and/or their parents gave written informed consent for participation in the studies. A detailed description of the study populations and pooling process can be found in the Supplementary Note. Finally, DNA samples of 436 cases of European descent with isolated anterior or middle hypospadias and 494 unaffected male controls of European descent were pooled in eight case pools and eight control pools.

Genome-wide analysis

We allelotyped each pool in duplicate using Affymetrix GeneChip 6.0 microarrays containing 906,600 polymorphic SNPs. Array experiments were performed according to protocols provided by the manufacturer. We used the SNPMaP package[21] in the statistical software program R (see URLs) to calculate relative allele signal (*RAS*) scores and *k*-corrected raw allele frequency (RAF_k) estimates for each SNP, as is described in the Supplementary Note.

A total of 574,400 SNPs passed quality control steps (described in the Supplementary Note) and were included in the association analyses. We averaged data across case pools and control pools separately to obtain RAF_k and *RAS* estimates for cases and controls. We subsequently calculated the modified χ^2 statistic[16], which is expressed as:

$$Z = \frac{P_{cases} - P_{controls}}{\sqrt{P_p \times (1 - P_p) \times (1 / 2N_{cases} + 1 / 2N_{controls}) + \sigma^2_{pool} \times (1 / np_{cases} + 1 / np_{controls})}}$$

Where P_{cases} is the allele frequency estimate (RAF_k) in the case pools, $P_{controls}$ is the RAF_k in the control pools and P_p is an estimate of the allele frequency in the population, for which we used the allele frequency in 603 individually genotyped in-house controls of European descent[22]. N_{cases} is the number of individuals in the case pools and $N_{controls}$ is the number of individuals in the control pools, taking into account the loss of cases and controls after the quality control. A SNP-specific variance calculated across the two measurements for each DNA pool was averaged across the pools to obtain σ^2_{pool}. Finally, np_{cases} and $np_{controls}$ are the number of

measurements (pools) for the case and control pools. We selected the 5,000 SNPs with the highest *Z* scores, and calculated two additional statistics for these SNPs: the standard χ^2 statistic using the expected numbers of alleles in the cases and controls calculated from the RAF_k estimates and the adjusted χ^2 statistic using the *RAS* values. Combining the results from all three statistics enabled us to select the most promising SNPs.

We used several criteria to select SNPs from the pooled GWAS as eligible for individual genotyping in the discovery sample to validate the results as is described in the Supplementary Note. Ultimately, 20 SNPs were eligible. As seven of these 20 SNPs were SNPs in high LD within *DGKK* (Supplementary Fig. 2), we genotyped the SNP that tagged most of the other SNPs (rs1934179, an intronic SNP). In addition, we genotyped a potentially regulatory SNP in the 5' upstream region (rs7063116).

Validation of the results from the pooled GWAS

Taqman SNP genotyping assays could not be designed for four of the 15 SNPs that were selected for individual genotyping in the discovery sample. We excluded these SNPs from further analysis. The other 11 SNPs (Supplementary Table 3) were individually genotyped using 5' nuclease TaqMan SNP genotyping assays (Supplementary Table 8). In each 96-well plate, we loaded five wells with randomly selected duplicate DNA samples from the same and other plates for quality control purposes. In addition, we included four blanks in each plate. Genotyping was completed with a success rate of at least 99%.

All genotype frequencies in controls were in Hardy-Weinberg equilibrium, with *P* values ranging from 0.34 to 0.94. For the genotyped SNPs, we calculated ORs for hypospadias risk and the corresponding 95% confidence intervals (CIs) at a genotypic and an allelic level using the most frequent homozygous genotypes in controls as reference values. Furthermore, we performed χ^2 tests. When the expected cell numbers were below five, exact 95% CIs around the ORs were calculated using the Fisher exact method. Ten of the associations were statistically significant ($P < 0.05$), but only the association with the two SNPs in the X-chromosomal *DGKK* gene reached genome-wide significance ($P < 5.0 \times 10^{-8}$) (Supplementary Table 4). These results were validated by genotyping the parents of the cases. For these case-parent triads, we used the transmission disequilibrium test (TDT)[23] for statistical analysis of the data with the software program Haploview 4.1 (ref.24). Furthermore, we calculated the ORs for hypospadias risk and corresponding 95% CIs at the allelic level[25] (Table 2).

Replication studies

A detailed description of the study populations can be found in the Supplementary Note. The Arnhem-Nijmegen Regional Committee on Research Involving Human Subjects and the Ethics Committee at Karolinska Institutet approved the studies and all participants and/or their parents gave written informed consent for participation in the studies.

For the Dutch replication sample, *we* genotyped 133 cases of European descent with isolated anterior or middle hypospadias and their parents for the SNPs that were associated with hypospadias in the individually genotyped discovery sample ($P < 0.05$). Genotyping was completed with a success rate of at least 95%. All genotype frequencies in the parents were in Hardy-Weinberg equilibrium (with *P* values ranging from 0.16 to 0.99). For these case-parent triads, we again used the TDT[23] for statistical analysis of the data with the software program Haploview 4.1 (ref.24), and we calculated the ORs for hypospadias risk and the corresponding 95% CIs at the allelic level[25] (Supplementary Table 5).

For the Swedish replication sample, we genotyped 266 Swedish cases with anterior or middle hypospadias and 402 male Swedish controls for the SNPs that were associated with hypospadias in the individually genotyped discovery sample ($P < 0.05$). Genotyping was completed with a success rate of at least 98%. All genotype frequencies in the controls were in Hardy-Weinberg equilibrium (with *P* values ranging from 0.13 to 0.99), except for rs1022357 in *SLCO3A1* ($P = 0.01$). However, after correcting the critical *P* value for multiple testing, this result did not reach statistical significance (critical Bonferroni $P < 0.005$). We calculated ORs for hypospadias risk and the corresponding 95% CIs at genotypic and allelic level, performed χ^2 tests and calculated exact 95% CIs around the ORs using the Fisher exact method when expected cell numbers were below five (Supplementary Table 6).

Meta-analysis

We combined the results of both the discovery sample and the two replication samples in a meta-analysis in Review Manager 5 (see URLs) using the inverse-variance method and random effects models (Supplementary Table 7).

Expression of DGKK

We isolated RNA from preputial skin samples from 14 hypospadias cases and ten age-matched controls. The origin of the samples and the methods used are described in the Supplementary Note. We performed RT-PCR according to the standard protocol as described in the Supplementary Note. Expression levels of *DGKK* were reported relative to *GAPDH*. We also genotyped DNA samples for the two SNPs in *DGKK* and compared relative gene expression levels between boys with the G allele (45.9% of *GAPDH*, n = 9) and the A (risk) allele (40.1% of *GAPDH*, n = 15) of the X-chromosomal SNP rs1934179 in *DGKK* using the independent samples *t* test.

Search for potential functional variants

We searched different databases, such as dbSNP and HapMap, for nonsynonymous SNPs that might be causative variants, taking validation of the SNPs, conservation of amino acids, and Grantham scores of amino acid replacements into account. Concerning non-coding areas, we

used the UCSC genome browser to check for microRNA binding sites, enrichments of histone marks associated with enhancers or promoters, and for transcription factor binding sites.

Calculation of Population Attributable Fraction

We calculated the population attributable fraction (*PAF*) as[26]:

$$PAF = \frac{q\,(OR - 1)}{q\,(OR - 1) + 1}$$

where OR is the odds ratio and q is the proportion of exposed individuals (proportion of individuals with the risk allele) in the control group, which is the A allele of rs1934179 in *DGKK*.

References

20. Hoogendoorn EH, Hermus AR, de Vegt F, Ross HA, Verbeek ALM, Kiemeney LALM, Swinkels DW, Sweep FCGJ, den Heijer M. Thyroid function and prevalence of anti-thyroperoxidase antibodies in a population with borderline sufficient iodine intake: influences of age and sex. *Clin Chem* 2006:52;104-111.

21. Davis OSP, Plomin R, Schalkwyk LC. The SNPMaP package for R: a framework for genome-wide association using DNA pooling on microarrays. *Bioinformatics* 2009:25;281-283.

22. Franke B, Arias Vasquez A, Veltman JA, Brunner HG, Rijpkema M, Fernández G. Genetic variation in *CACNA1C*, a gene associated with bipolar disorder, influences brainstem rather than gray matter volume in healthy individuals. *Biol Psychiat* 2010:68;586-588.

23. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993:52;506-516.

24. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005:21;263-265.

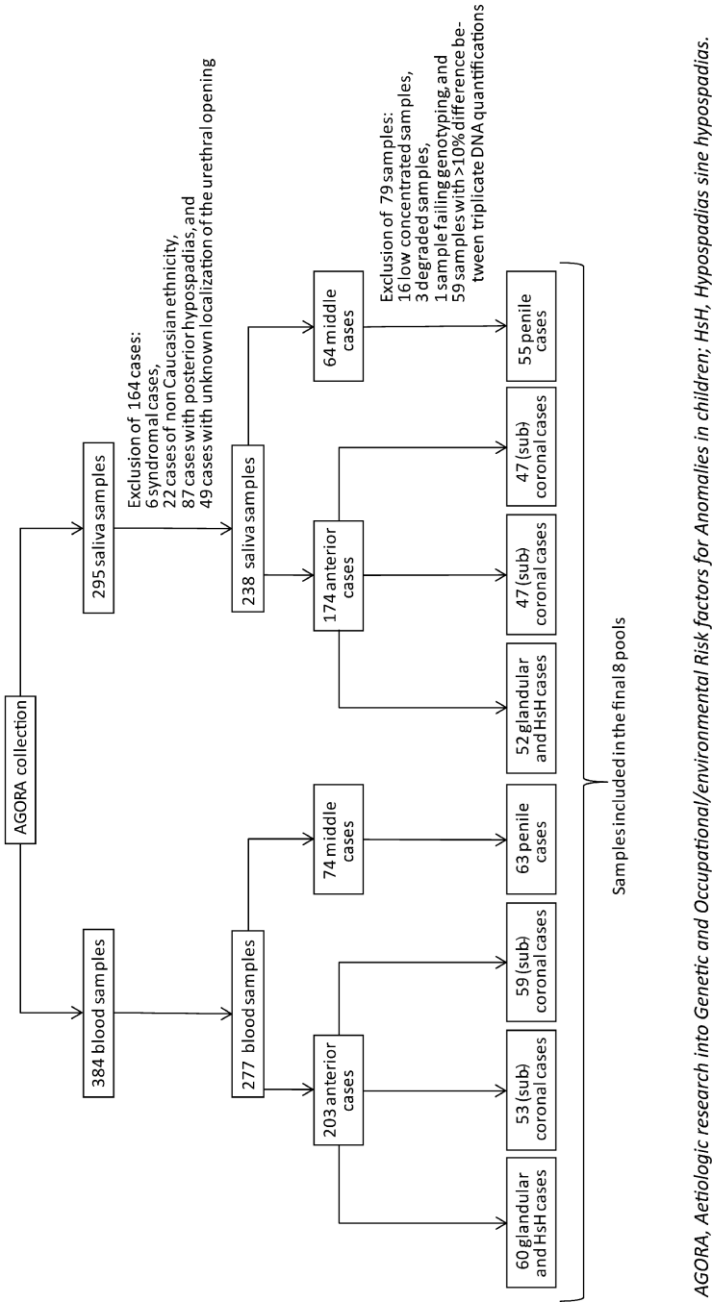
25. Kazeem GR, Farrall M. Integrating case-control and TDT studies. *Ann Hum Genet* 2005:69; 329-335.

26. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. (Lippincott Williams & Wilkins, Philadelphia, 2008).

Supplementary material

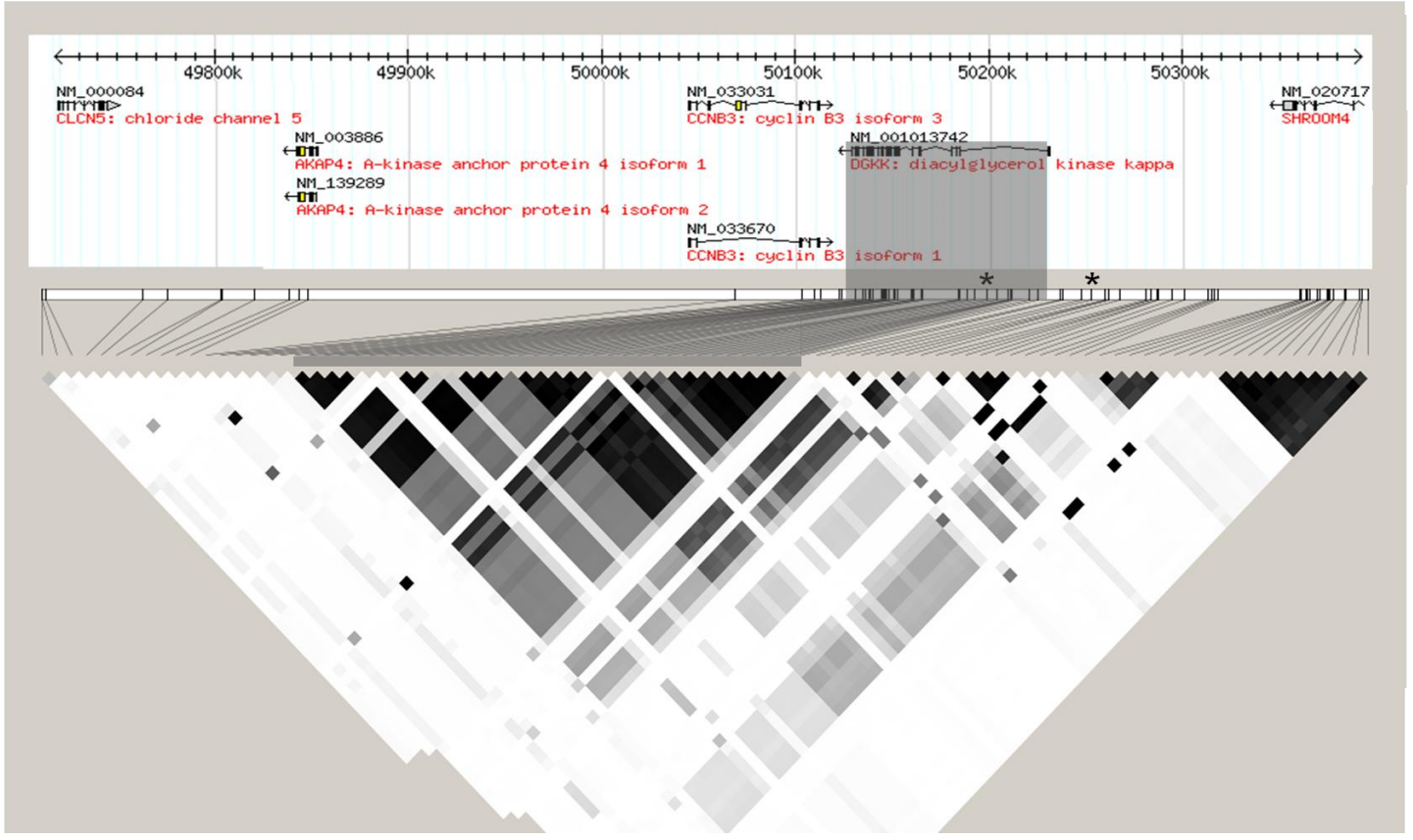
Supplementary figures

Suppl. figure 1. Flowchart showing the selection of case samples for inclusion in the discovery sample of the current study.

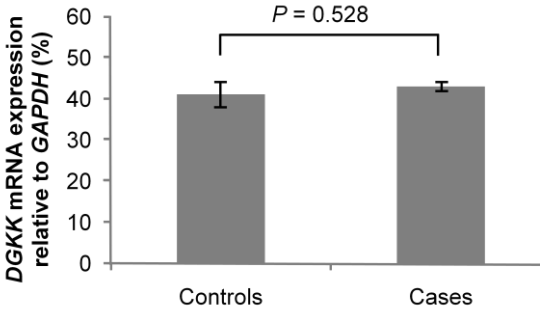


AGORA, Aetiologic research into Genetic and Occupational/environmental Risk factors for Anomalies in children; HsH, Hypospadias sine hypospadias.

Suppl. figure 2. Haplotype block structure of *DGKK* and surrounding area, determined using data from 603 in-house controls of European descent assayed on the Affymetrix GeneChip 6.0 microarrays individually[1]. SNPs in *DGKK* are indicated in grey shading. Regions with high r^2 values are dark grey, and regions with low r^2 values are lighter shades of grey. *SNPs genotyped individually in the discovery and replication samples.



Suppl. figure 3. Quantification of *DGKK* mRNA expression (± s.e.m.) in preputial skin of hypospadias cases and controls relative to *GAPDH* mRNA expression. Relative gene expression levels were compared between cases (43.1% of *GAPDH*, n = 14) and controls (41.2% of *GAPDH*, n = 10) using the independent samples t-test.



Supplementary tables

Suppl. table 1. Clinical characteristics of the hypospadias patients in the discovery sample.

	Anatomical location of urethral opening					
	Anterior n=318 (73%)			***Middle n=118 (27%)***		
	HsH	*Glandular*	*(Sub)-coronal*	*Distal penile*	*Midshaft*	*Proximal penile*
Total	33	88	197	63	48	7
Chordee	21 (64%)	30 (34%)	115 (58%)	41 (65%)	38 (79%)	6 (86%)
Cryptorchidism[a]	4 (12%)	2 (2%)	11 (6%)	3 (5%)	1 (2%)	0
Other deformities[b]	4 (12%)	10 (11%)	8 (4%)	4 (6%)	0	0

HsH, Hypospadias sine hypospadias; [a]*including only cases that underwent orchidopexy or orchidectomy;* [b]*including cases with heart-, renal- and gastro-intestinal deformities, neural tube defects, cleft lip and/or palate and deformities of the lower extremities.*

Suppl. table 2. SNPs appearing in the top 200 ranks for all three statistical tests used. The grey-shaded SNPs are the SNPs that were selected based on the criteria used. As seven of these 20 SNPs were SNPs in high LD within the *DGKK* gene, we selected the SNP that tagged most of the other SNPs (rs1934179) for individual genotyping in the discovery sample. In addition, we genotyped a potentially regulatory SNP in the 5'-upstream region (rs7063116). Taqman SNP genotyping assays could not be designed for four of the fifteen SNPs that were thus selected for individual genotyping in the discovery sample.

Suppl. table 2.

SNP	Chr	Rank adj χ^2-test using RAS	Rank adj χ^2-test using RAF_k	Rank χ^2-test using RAF_k	Nearest gene	Location	MAF in in-house control samples	Sup-porting evidence from SNPs in LD[a]	Reason exclusion from individual geno-typing
rs1934179	X	46	21	4	*DGKK*	intron	0.337	yes	
rs16985967	22	26	4	12	*TTC28*	intron	0.080	yes	
rs7085066	10	35	19	31	*SLC16A12*	intron	0.105	yes	
rs1022357	15	199	176	79	*SLCO3A1*	intron	0.166	yes	
rs10184015	2	32	17	83	*SOS1*	intron	0.066	yes	
rs6607499	X	38	33	133	*KIAA2022*	12 kb up	0.065	yes	
rs4705372	5	165	119	168	*PPARGC1B*	intron	0.123	yes	
rs3755071	2	31	81	16	*SRBD1*	intron	0.226	-	
rs1880386	10	180	42	35	*GRID1*	intron	0.152	-	
rs12660161	6	30	11	56	*MTHFD1L*	39 kb down	0.085	-	
rs7896487	10	19	144	55	*CDH23*	intron	0.123	-	assay
rs9735061	11	73	186	139	*FAM86A*	43 kb up	0.122	-	assay
rs7321040	13	21	13	75	*DIAPH3*	31 kb down	0.055	-	assay
rs10997978	10	28	2	26	*MYPN*	intron	0.056	yes	assay
rs1934171	X	3	1	1	*DGKK*	intron	0.336	yes	other SNP
rs1934188	X	52	5	2	*DGKK*	intron	0.314	yes	other SNP
rs1934176	X	8	3	3	*DGKK*	intron	0.337	yes	other SNP
rs4826629	X	99	79	5	*DGKK*	intron	0.316	yes	other SNP
rs1934175	X	77	57	8	*DGKK*	intron	0.336	yes	other SNP
rs17281440	X	182	169	34	*DGKK*	intron	0.240	yes	other SNP
rs17051831	13	85	95	118	*NBEA*	intron	0.096	-	LD/MAF
rs7229296	18	126	103	137	*RAB31*	9 kb down	0.092	-	LD/MAF
rs4869674	5	13	27	32	*SPEF2*	intron	0.091	-	LD/MAF
rs17126762	1	34	14	80	*CACHD1*	intron	0.072	-	LD/MAF
rs17021660	4	16	9	22	*POU4F2*	17 kb up	0.062	-	LD/MAF
rs527834	15	193	18	13	*CHRM /AVEN*	intron	0.053	-	LD/MAF
rs16841107	1	196	88	69	*RGS7*	intron	0.050	-	LD/MAF
rs12465925	2	191	151	72	*TMEFF2*	intron	0.132	no	LD
rs17236770	15	135	106	21	*TMEM85*	intron	0.123	no	LD
rs3771698	2	9	32	17	*BAZ2B*	intron	0.088	no	LD
rs16842909	2	83	50	97	*MAP2*	intron	0.072	no	LD
rs764125	3	78	40	146	*CMTM8*	intron	0.057	no	LD
rs9308056	4	6	93	172	*TKTL2*	54 kb down	0.083	yes	distance
rs2114929	5	48	102	62	*TARS*	63 kb up	0.141	yes	distance
rs1560398	18	10	8	40	*MC4R*	123 kb up	0.063	-	distance
rs1397151	16	79	39	76	*CES7*	130 kb up	0.057	yes	distance
rs16888346	4	63	38	18	*HSP90AB2P*	135 kb up	0.113	-	distance
rs6847625	4	5	10	52	*GPRIN3*	197 kb up	0.069	-	distance
rs10497759	2	81	72	6	*SLC39A10*	244 kb up	0.297	-	distance
rs2174232	10	53	37	67	*ZWINT*	816 kb up	0.070	-	distance

Suppl. table 2. (continued)

SNP	Chr	Rank adj χ^2-test using RAS	Rank adj χ^2-test using RAF_k	Rank χ^2-test using RAF_k	Nearest gene	Location	MAF in in-house control samples	Sup-porting evidence from SNPs in LD[a]	Reason exclusion from individual geno-typing
rs6127262	20	101	185	100	*DOK5*	62 kb down	0.010	-	distance
rs4850833	2	98	69	64	*PLCL1*	73 kb down	0.114	yes	distance
rs12544855	8	86	45	70	*LOC100129963*	87 kb down	0.100	-	distance
rs16864680	2	4	23	164	*LOC400940*	91 kb down	0.080	-	distance
rs10733948	10	176	107	142	*ZWINT*	131 kb down	0.095	no	distance
rs2468265	12	177	110	7	*SUDS3*	199 kb down	0.477	-	distance
rs10024666	4	57	35	175	*TRAM1L1*	314 kb down	0.078	-	distance
rs12265647	10	17	16	182	*YWHAZ*	321 kb down	0.070	-	distance
rs11046765	12	1	6	33	*ETNK1*	356 kb down	0.061	yes	distance
rs5942497	X	120	104	157	*CPXCR1*	574 kb down	0.063	no	distance

Chr, chromosome; adj, adjusted; RAS, relative allele signal; RAF_k, k-corrected raw allele frequency estimate; MAF, minor allele frequency; LD, linkage disequilibrium; kb, kilobases; up, upstream; down, downstream; assay, assay unavailable; other SNP, other SNP in gene; LD/MAF, LD and MAF < 0.10; distance, gene distance; [a]yes = supporting evidence defined as at least half of the SNPs in high LD (r^2 > 0.8) having a low average P-value (< 0.22, the 10th percentile of the average P-values of all 574,400 SNPs), no = reduced evidence defined as at least half of the SNPs in high LD having a high average P-value, - = no SNPs in high LD.

Suppl. table 3. Single nucleotide polymorphisms individually genotyped in the discovery sample.

SNP	Chr	Nearest gene	Location
rs1934179	X	*DGKK*	intron
rs7063116	X	*DGKK*	21 kb upstream
rs1880386	10	*GRID1*	intron
rs6607499	X	*KIAA2022*	12 kb upstream
rs12660161	6	*MTHFD1L*	39 kb downstream
rs4705372	5	*PPARGC1B*	intron
rs7085066	10	*SLC16A12*	intron
rs1022357	15	*SLCO3A1*	intron
rs10184015	2	*SOS1*	intron
rs3755071	2	*SRBD1*	intron
rs16985967	22	*TTC28*	intron

Chr, chromosome.

Suppl. table 4. Association analyses with hypospadias for single nucleotide polymorphisms in selected genes in the discovery sample.

Controls n=494 n (%)	Cases n=436 n (%)	OR (95% CI)	χ^2 *P*		Controls n=494 n (%)	Cases n=436 n (%)	OR (95% CI)	χ^2 *P*
Allele frequencies				***Genotype frequencies***				
rs1934179 in DGKK								
G 331 (67)	196 (45)	1.00 (ref)	ref	Not applicable, X-chromosomal SNP				
A 162 (33)	236 (55)	2.46 (1.88-3.21)	2.5×10^{-11}					
rs7063116 near DGKK								
G 348 (71)	224 (52)	1.00 (ref)	ref	Not applicable, X-chromosomal SNP				
A 144 (29)	209 (48)	2.25 (1.72-2.96)	2.9×10^{-9}					
rs1880386 in GRID1								
C 844 (86)	698 (81)	1.00 (ref)	ref	*CC*	360 (73)	285 (66)	1.00 (ref)	ref
T 140 (14)	162 (19)	1.40 (1.09-1.79)	0.008	*CT*	124 (25)	128 (30)	1.30 (0.97-1.75)	0.075
				TT	8 (2)	17 (4)	2.68 (1.14-6.31)	0.019
rs6607499 near KIAA2022								
G 474 (96)	402 (93)	1.00 (ref)	ref	Not applicable, X-chromosomal SNP				
A 18 (4)	31 (7)	2.03 (1.12-3.68)	0.018					
rs12660161 near MTHFD1L								
G 909 (92)	766 (88)	1.00 (ref)	ref	*GG*	419 (85)	343 (79)	1.00 (ref)	ref
A 79 (8)	100 (12)	1.50 (1.10-2.05)	0.010	*AG*	71 (14)	80 (18)	1.38 (0.97-1.95)	0.073
				AA	4 (1)	10 (2)	3.05 (0.95-9.82)	0.049
rs4705372 in PPARGC1B								
G 856 (87)	714 (82)	1.00 (ref)	ref	*GG*	371 (75)	298 (69)	1.00 (ref)	ref
A 132 (13)	154 (18)	1.40 (1.09-1.80)	0.009	*AG*	114 (23)	118 (27)	1.29 (0.96-1.74)	0.096
				AA	9 (2)	18 (4)	2.49 (1.10-5.62)	0.024
rs7085066 in SLC16A12								
G 898 (91)	769 (89)	1.00 (ref)	ref	*GG*	409 (83)	338 (78)	1.00 (ref)	ref
C 90 (9)	95 (11)	1.23 (0.91-1.67)	0.177	*CG*	80 (16)	93 (22)	1.41 (1.01-1.96)	0.043
				CC	5 (1)	1 (0)	0.24 (0.01-2.18)[a]	0.325
rs1022357 in SLCO3A1								
A 824 (84)	689 (79)	1.00 (ref)	ref	*AA*	344 (70)	277 (64)	1.00 (ref)	ref
G 162 (16)	179 (21)	1.32 (1.04-1.67)	0.020	*AG*	136 (28)	135 (31)	1.23 (0.93-1.64)	0.151
				GG	13 (3)	22 (5)	2.10 (1.04-4.25)	0.035
rs10184015 in SOS1								
G 923 (93)	781 (90)	1.00 (ref)	ref	*GG*	430 (87)	353 (81)	1.00 (ref)	ref
A 65 (7)	87 (10)	1.58 (1.13-2.21)	0.007	*AG*	63 (13)	75 (17)	1.45 (1.01-2.09)	0.044
				AA	1 (0)	6 (1)	7.31 (0.88-337)[a]	0.074
rs3755071 in SRBD1								
A 690 (70)	652 (75)	1.00 (ref)	ref	*AA*	238 (48)	242 (56)	1.00 (ref)	ref
G 292 (30)	214 (25)	0.78 (0.63-0.95)	0.016	*AG*	214 (44)	168 (39)	0.77 (0.59-1.01)	0.060
				GG	39 (8)	23 (5)	0.58 (0.34-1.00)	0.049
rs16985967 in TTC28								
C 924 (94)	771 (89)	1.00 (ref)	ref	*CC*	432 (88)	339 (78)	1.00 (ref)	ref
T 62 (6)	97 (11)	1.88 (1.34-2.62)	0.0002	*CT*	60 (12)	93 (21)	1.98 (1.39-2.82)	0.0001
				TT	1 (0)	2 (0)	2.55 (0.13-151)[a]	0.822

OR, odds ratio; CI, confidence interval; ref, used as reference; [a]exact 95% CI calculated using the Fisher exact method.

Suppl. table 5. Association analyses with hypospadias for single nucleotide polymorphisms in selected genes in the Dutch replication sample.

SNP	Minor allele[a]	NT	T	OR	(95%CI)	χ^2 *P*
rs1934179 in DGKK	A	12	47	3.92	(2.08-7.38)	2.4×10^{-5}
rs7063116 near DGKK	A	12	46	3.83	(2.03-7.24)	3.4×10^{-5}
rs1880386 in GRID1	T	24	33	1.38	(0.81-2.33)	0.233
rs6607499 near KIAA2022	A	6	7	1.17	(0.39-3.47)	0.782
rs12660161 near MTHFD1L	A	17	28	1.65	(0.90-3.01)	0.101
rs4705372 in PPARGC1B	A	21	31	1.48	(0.85-2.57)	0.166
rs1022357 in SLCO3A1	G	28	37	1.32	(0.81-2.16)	0.264
rs10184015 in SOS1	A	19	18	0.95	(0.50-1.81)	0.869
rs3755071 in SRBD1	G	35	40	1.14	(0.73-1.80)	0.564
rs16985967 in TTC28	T	7	14	2.00	(0.81-4.96)	0.127

OR, odds ratio; CI, confidence interval; NT, minor allele not transmitted; T, minor allele transmitted; [a]the least frequent allele in the controls of the discovery sample.

Suppl. table 6. Association analyses with hypospadias for single nucleotide polymorphisms in selected genes in the Swedish replication sample.

	Controls n=402 n (%)	Cases n=266 n (%)	OR (95% CI)	χ^2 *P*		Controls n=402 n (%)	Cases n=266 n (%)	OR (95% CI)	χ^2 *P*
Allele frequencies					***Genotype frequencies***				
rs1934179 in DGKK									
G	276 (69)	123 (47)	1.00 (ref)	ref	Not applicable, X-chromosomal SNP				
A	124 (31)	137 (53)	2.48 (1.80-3.42)	2.6×10^{-8}					
rs7063116 near DGKK									
G	284 (71)	137 (53)	1.00 (ref)	ref	Not applicable, X-chromosomal SNP				
A	118 (29)	123 (47)	2.16 (1.56-2.99)	2.7×10^{-6}					
rs1880386 in GRID1									
C	676 (85)	428 (82)	1.00 (ref)	ref	*CC*	288 (72)	179 (69)	1.00 (ref)	ref
T	124 (16)	92 (18)	1.17 (0.87-1.57)	0.293	*CT*	100 (25)	70 (27)	1.13 (0.79-1.61)	0.515
					TT	12 (3)	11 (4)	1.47 (0.64-3.41)	0.362
rs6607499 near KIAA2022									
G	372 (93)	235 (90)	1.00 (ref)	ref	Not applicable, X-chromosomal SNP				
A	29 (7)	27 (10)	1.47 (0.85-2.55)	0.165					
rs12660161 near MTHFD1L									
G	714 (89)	473 (91)	1.00 (ref)	ref	*GG*	320 (80)	213 (82)	1.00 (ref)	ref
A	84 (11)	49 (9)	0.88 (0.61-1.28)	0.502	*AG*	74 (19)	47 (18)	0.95 (0.64-1.43)	0.820
					AA	5 (1)	1 (0)	0.30 (0.01-2.72)[a]	0.473
rs4705372 in PPARGC1B									
G	709 (89)	453 (86)	1.00 (ref)	ref	*GG*	313 (78)	199 (76)	1.00 (ref)	ref
A	89 (11)	73 (14)	1.28 (0.92-1.79)	0.139	*AG*	83 (21)	55 (21)	1.04 (0.71-1.53)	0.833
					AA	3 (1)	9 (3)	4.72 (1.16-27.3)[a]	0.027
rs1022357 in SLCO3A1									
A	649 (81)	414 (81)	1.00 (ref)	ref	*AA*	271 (68)	163 (63)	1.00 (ref)	ref
G	151 (19)	100 (19)	1.04 (0.78-1.38)	0.794	*AG*	107 (27)	88 (34)	1.37 (0.97-1.93)	0.073
					GG	22 (6)	6 (2)	0.45 (0.18-1.14)	0.086
rs10184015 in SOS1									
G	738 (92)	477 (91)	1.00 (ref)	ref	*GG*	341 (85)	220 (84)	1.00 (ref)	ref
A	66 (8)	49 (9)	1.15 (0.78-1.69)	0.483	*AG*	56 (14)	37 (14)	1.02 (0.65-1.60)	0.917
					AA	5 (1)	6 (2)	1.86 (0.47-7.79)[a]	0.468
rs3755071 in SRBD1									
A	627 (78)	382 (75)	1.00 (ref)	ref	*AA*	241 (60)	143 (56)	1.00 (ref)	ref
G	173 (22)	130 (25)	1.23 (0.95-1.60)	0.115	*AG*	145 (36)	96 (38)	1.12 (0.80-1.55)	0.516
					GG	14 (4)	17 (7)	2.05 (0.98-4.28)	0.053
rs16985967 in TTC28									
C	720 (90)	467 (90)	1.00 (ref)	ref	*CC*	324 (81)	208 (80)	1.00 (ref)	ref
T	80 (10)	53 (10)	1.02 (0.71-1.47)	0.910	*CT*	72 (18)	51 (20)	1.10 (0.74-1.64)	0.629
					TT	4 (1)	1 (0)	0.39 (0.01-3.98)[a]	0.710

OR, odds ratio; CI, confidence interval; ref, used as reference; [a]exact 95% CI calculated using the Fisher exact method.

Suppl. table 7. Association analyses with hypospadias for single nucleotide polymorphisms in selected genes in the discovery sample, the Dutch replication sample and the Swedish replication sample, and results of the meta-analysis.

	Discovery sample				*Dutch replication sample*			
	MAF (%)				*TDT*			
SNP (minor allele[a])	Controls n= 494	Cases n= 436	OR (95%CI)	χ^2 *P*	NT	T	OR (95%CI)	χ^2 *P*
rs1934179 in DGKK (A)	32.9	54.6	2.5 (1.9-3.2)	2.5×10^{-11}	12	47	3.9 (2.1-7.4)	2.4×10^{-5}
rs7063116 near DGKK (A)	29.3	48.3	2.3 (1.7-3.0)	2.9×10^{-9}	12	46	3.8 (2.0-7.2)	3.4×10^{-5}
rs1880386 in GRID1 (T)	14.2	18.8	1.4 (1.1-1.8)	0.008	24	33	1.4 (0.8-2.3)	0.233
rs6607499 near KIAA2022 (A)	3.7	7.2	2.0 (1.1-3.7)	0.018	6	7	1.2 (0.4-3.5)	0.782
rs12660161 near MTHFD1L (A)	8.0	11.5	1.5 (1.1-2.1)	0.010	17	28	1.7 (0.9-3.0)	0.101
rs4705372 in PPARGC1B (A)	13.4	17.7	1.4 (1.1-1.8)	0.009	21	31	1.5 (0.9-2.6)	0.166
rs7085066 in SLC16A12 (C)	9.1	11.0	1.2 (0.9-1.7)	0.177				
rs1022357 in SLCO3A1 (G)	16.4	20.6	1.3 (1.0-1.7)	0.020	28	37	1.3 (0.8-2.2)	0.264
rs10184015 in SOS1 (A)	6.6	10.0	1.6 (1.1-2.2)	0.007	19	18	1.0 (0.5-1.8)	0.869
rs3755071 in SRBD1 (G)	29.7	24.7	0.8 (0.6-1.0)	0.016	35	40	1.1 (0.7-1.8)	0.564
rs16985967 in TTC28 (T)	6.3	11.2	1.9 (1.3-2.6)	0.0002	7	14	2.0 (0.8-5.0)	0.127

	Swedish replication sample				*Meta-analysis results*	
	MAF (%)					
SNP (minor allele[a])	Controls n= 402	Cases n= 266	OR (95%CI)	χ^2 *P*	OR (95%CI)	χ^2 *P*
rs1934179 in DGKK (A)	31.0	52.7	2.5 (1.8-3.4)	2.6×10^{-8}	2.6 (2.1-3.1)	2.8×10^{-21}
rs7063116 near DGKK (A)	29.4	47.3	2.2 (1.6-3.0)	2.7×10^{-6}	2.4 (1.9-3.0)	1.2×10^{-12}
rs1880386 in GRID1 (T)	15.5	17.7	1.2 (0.9-1.6)	0.293	1.3 (1.1-1.6)	0.003
rs6607499 near KIAA2022 (A)	7.2	10.3	1.5 (0.9-2.6)	0.165	1.6 (1.1-2.4)	0.011
rs12660161 near MTHFD1L (A)	10.5	9.4	0.9 (0.6-1.3)	0.502	1.3 (0.9-1.9)	0.242
rs4705372 in PPARGC1B (A)	11.2	13.9	1.3 (0.9-1.8)	0.139	1.4 (1.1-1.7)	0.001
rs7085066 in SLC16A12 (C)						
rs1022357 in SLCO3A1 (G)	18.9	19.5	1.0 (0.8-1.4)	0.794	1.2 (1.0-1.4)	0.027
rs10184015 in SOS1 (A)	8.2	9.3	1.2 (0.8-1.7)	0.483	1.3 (1.0-1.7)	0.075
rs3755071 in SRBD1 (G)	21.6	25.4	1.2 (1.0-1.6)	0.115	1.0 (0.7-1.4)	0.952
rs16985967 in TTC28 (T)	10.0	10.2	1.0 (0.7-1.5)	0.910	1.5 (0.9-2.4)	0.103

MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; TDT, transmission disequilibrium test; NT, minor allele not transmitted; T, minor allele transmitted; [a]the least frequent allele in the controls of the discovery sample.

Suppl. table 8. Primer and reporter sequences for the custom designed Taqman assays, ID numbers of the Taqman SNP genotyping assays, and oligonucleotide primers for the real-time PCR reactions.

SNP / Gene	Sequence / Taqman SNP genotyping assay ID
rs6607499 upstream of KIAA2022	
Forward primer	GAGTGGAATTTCTGGGTCATATGGA
Reverse primer	TGGTGCAGCTTCTATGGAAAACA
Reporter 1	CATTTGGAGTAACTATCAGAC
Reporter 2	ATTTGGAGTAACTGTCAGAC
rs7085066 in SLC16A12	
Forward primer	CCAGAGACCAAAATGGAACTGAAGT
Reverse primer	AAGAGTAAGGGCGTGAAATAACCT
Reporter 1	CACCAACAAGAGCACC
Reporter 2	CACCAACAACAGCACC
rs7063116 upstream of DGKK	
Assay ID	C_196656_10
rs1934179 in DGKK	
Assay ID	C_12116498_10
rs1880386 in GRID1	
Assay ID	C_12127943_10
rs12660161 downstream of MTHFD1L	
Assay ID	C_11414015_20
rs4705372 in PPARGC1B	
Assay ID	C_27948404_10
rs1022357 in SLCO3A1	
Assay ID	C_8719491_10
rs10184015 in SOS1	
Assay ID	C_30456791_10
rs3755071 in SRBD1	
Assay ID	C_2773842_10
rs16985967 in TTC28	
Assay ID	C_34310741_10
GAPDH (PCR product size: 160bp)	
Forward primer	CATGTTCGTCATGGGTGTGAACCA
Reverse primer	AGTGATGGCATGGACTGTGGTCAT
DGKK (PCR product size: 160bp)	
Forward primer	AGAAGAGATGAACACCCAGGGCAA
Reverse primer	GCAGGTTTGGCAAGGAGATGGTTT

bp, basepairs.

Supplementary note

Dutch discovery sample

AGORA project

The AGORA (Aetiologic research into Genetic and Occupational/environmental Risk factors for Anomalies in children) project of the Radboud University Nijmegen Medical Centre (RUNMC), Nijmegen, The Netherlands, is building a data- and biobank with questionnaire data and DNA samples from patients with a congenital malformation or childhood cancer and their parents. For the current study, DNA was available of 679 hypospadias cases born between 1980 and 2007, collected within the AGORA biobank until 2007. All cases were living in the catchment area of this hospital, which is Nijmegen and surrounding area. The majority of the cases were of European descent (83%), whereas descent was not European (3%) or unknown (14%) for the remaining cases. We excluded non-European cases from the analyses. Medical records of all cases were reviewed to identify information on syndromal hypospadias cases, to collect clinical characteristics of patients, and to determine anatomical location of the urethral opening. The location of the urethral opening was determined by experienced pediatric urologists before or during surgery. For the current study, anatomical location was subdivided into two categories: anterior and middle hypospadias (anterior cases with hypospadias sine hypospadias, glandular and (sub)coronal urethral openings and middle cases with penile urethral openings) and posterior hypospadias (cases with penoscrotal, scrotal, and perineal urethral openings). The Arnhem-Nijmegen Regional Committee on Research Involving Human Subjects approved the AGORA project. All participants and/or their parents gave written informed consent for participation in the study.

Nijmegen Biomedical Study

We obtained controls from the Nijmegen Biomedical Study, a population-based survey conducted by the Department of Epidemiology, Biostatistics and HTA and the Department of Clinical Chemistry of the RUNMC in 2003. Age and sex stratified randomly selected inhabitants of the municipality of Nijmegen (n = 22,500) received an invitation to fill out a postal questionnaire on lifestyle and medical history, and to donate blood samples. The response to the questionnaire was 42% (n = 9,373), whereas 72% (n = 6,747) of the respondents also donated blood samples[2]. We selected the 596 youngest males of Dutch descent for the current study. None of them reported hypospadias or penile surgery in the questionnaires. The Arnhem-Nijmegen Regional Committee on Research Involving Human Subjects approved the Nijmegen Biomedical Study. All participants gave written informed consent for participation in the study.

Dutch replication sample

Of the 515 cases that were assigned to one of the eight pools, only 436 could be pooled. The remaining 79 cases served as a replication sample, supplemented with 54 cases of European descent with isolated anterior or middle hypospadias collected in the AGORA project in 2008 and 2009.

Swedish replication sample

Swedish nonsyndromic hypospadias patients from the Karolinska University Hospital, Stockholm, Sweden, selected through medical records, served as a second replication sample. In total, DNA samples were collected from 356 patients for the replication study, of which we selected 266 samples from patients with anterior or middle hypospadias. DNA from 222 ethnically comparable anonymous healthy male voluntary blood donors at Karolinska University Hospital, Sweden, served as a control group, as well as 180 DNA samples from placentas taken after birth of a child without a visible malformation. The Ethics Committee at Karolinska Institutet approved the study protocol and all participants or parents gave written informed consent for participation in the study.

DNA Pooling

DNA was extracted from blood (n = 980) or saliva (n = 295, Oragene, DNA Genotek Inc.) using standard methods. We created eight case pools, after exclusion of six syndromal cases, 22 cases of non-European descent, 87 cases with posterior hypospadias, and 49 cases with unknown localization of the urethral opening. The remaining 515 cases were assigned to one of the eight pools based on the source of DNA (blood or saliva) and the localization of the urethral opening. Controls were randomly assigned to one of eight control pools.

For each pool, we diluted DNA samples to approximately 100 ng μl^{-1} in one 96-wells plate. Sixteen case samples were excluded due to low concentration (< 50 ng μl^{-1}). Before pooling, we verified DNA integrity of the samples by gel electrophoresis. Three case and nineteen control samples showed band smearing, which is evidence of DNA degradation, and were excluded. We tested DNA quality of the diluted samples using 5'-nuclease TaqMan SNP genotyping assays for two SNPs (rs6932902 in *ESR1*: C_2823640_10 and rs11119982 in *ATF3*: C_27262262_10, Applied Biosystems). One case sample failing both assays was excluded.

Genomic DNA of each sample was quantified in triplicate using PicoGreen (Molecular Probes). We only included DNA samples from which reliable triplicate readings could be obtained, indicated by less than 10% difference between the readings, in the pools. This led to the exclusion of 59 case and 83 control samples. Finally, 436 case and 494 control samples could be pooled in eight case and eight control pools. We added equal quantities of DNA from each sample to the pools, with a minimum volume of 5 µl, which we considered the minimum

volume that could be pipetted sufficiently accurately. Supplementary Figure 1 shows a flowchart of the collection, selection, and exclusion of cases for this study. Clinical characteristics of the pooled patients (discovery sample) are displayed in Supplementary Table 1.

We tested accuracy of the pools by comparing the allele frequencies of the individually genotyped SNPs (rs6932902 in *ESR1* and rs11119982 in *ATF3*) in the individuals contributing to a pool, with the allele frequencies of these SNPs determined in the respective pool. The pools and an individual heterozygote were allelotyped using the same allele-specific Taqman probes in a real-time PCR 20 times. PCR reactions were carried out in 96-wells plates in a 10 µl volume containing 10 ng genomic DNA, 5 µl Taqman Universal PCR master mix, 0.125 µl assay mix and MilliQ. The PCR reaction consisted of an initial denaturation step at 95°C for 10 min, followed by 40 cycles of denaturation at 92°C for 15 sec, and annealing and extension at 60°C for 60 sec. During PCR, allele-specific fluorescence was measured on an ABI 7500 FAST (Applied Biosystems). Ct values indicate the cycle time that an allele needs to produce a detectable amount of fluorescent signal. This is an indication of the initial quantity of the allele in the pool. The delta Ct value is the difference in cycle time that each of the alleles needs to produce a detectable amount of fluorescent signal, which is an indication of the proportion of both alleles in the pools[3]. We averaged delta Ct values in the pools across the 20 measurements and calculated allele frequencies after correcting for unequal detection of the two alleles based on the average delta Ct value in the heterozygote. The allele frequency determined in the individual samples differed on average only 1.2% from the allele frequency determined in the pools. Therefore, we concluded that the pools accurately reflected the allele distributions of the samples, meaning that the pools appear to be correctly constructed with equal amounts of DNA from all samples.

Derivation of allele frequency estimates from microarray analysis

We used the SNPMaP package[4] in the statistical software program R (http://www.R-project.org) to calculate Relative Allele Signal (*RAS*) scores for each SNP from the CEL files generated by the Affymetrix GeneChip Command Console. Using this package, *RAS* scores were calculated by dividing the signal intensity of allele A by the sum of the signal intensities of alleles A and B ($RAS = A / (A + B)$), the method most commonly used in the literature. The signal intensities of allele A and B should be the same in a heterozygous individual (equivalent to a pool with a 50% allele frequency), however, due to differential amplification and hybridization, this is rarely the case. Therefore, corrections need to be made to *RAS* scores to generate more reliable estimates of the allele frequency. The *k*-correction factor can be calculated for each SNP using the ratio of the average signal intensities across all known heterozygotes ($k = A / B$) to obtain a *k*-corrected raw allele frequency estimate for each SNP ($RAF_k = A / (A + k*B)$). We used the SNPMaP package to calculate the *RAS* values for each SNP from 603 healthy in-house controls of European descent assayed individually on the same genotyping platform[1]. Using the

average *RAS* value from all individuals heterozygous for a SNP, we calculated the *k*-correction factor for each SNP. Using these *k*-correction factors, RAF_k was calculated.

Quality control

We excluded the worst performing 5% of the measurements, indicated by the biggest difference between the RAF_k estimates from the duplicate measurements (> 0.07). Thereafter, we averaged RAF_k and *RAS* across the remaining duplicate measurements to obtain estimates for each SNP.

A SNP-specific variance was calculated across the RAF_k values of the case pools and of the control pools. Furthermore, we averaged RAF_k values across case pools with DNA extracted from blood and case pools with DNA extracted from saliva to obtain a RAF_k estimate for each of these two types of DNA origins. SNPs with a difference between these two estimates above the 95^{th} percentile (> 0.043) were excluded. In addition, we excluded SNPs with a variance in the case or control pools above the 95^{th} percentile (> 0.0018 or > 0.0017, respectively) and SNPs with a minor allele frequency below 5% in the individually genotyped in-house controls. A total of 574,400 SNPs passed quality control steps and were included in the association analyses.

Selection of SNPs eligible for individual genotyping in the discovery sample

We used several criteria to select SNPs from the pooled GWAS eligible for individual genotyping in the discovery sample to validate the results. First of all, SNPs were required to appear in the top 200 ranks for all three statistical tests used, which resulted in 50 eligible SNPs (Supplementary Table 2). Secondly, SNPs had to map in or near a known gene (< 50 kb), as determined by the annotation data provided by Affymetrix, leaving 32 eligible SNPs. This criterion increases the biological interpretability of findings, although we are aware of the risk of overlooking associations with intergenic SNPs that may lie in regulatory sequences or yet unknown genes. Finally, for the remaining eligible SNPs, we examined whether they were in high linkage disequilibrium (LD) (r^2 > 0.8) with other SNPs on the array using data from the in-house controls. If so, the average *P*-values from the three statistical tests were calculated for the SNPs in high LD with an eligible SNP.

We divided eligible SNPs into three groups; 1: A group with supporting evidence defined as at least half of the SNPs in high LD having a low average *P*-value (< 0.22, the 10^{th} percentile of the average *P*-values of all 574,400 SNPs), 2: A group with reduced evidence defined as at least half of the SNPs in high LD having a high average *P*-value, and 3: A group with no supporting evidence because of no SNPs in high LD. For individual genotyping in the discovery sample, we selected the fourteen eligible SNPs with supporting evidence and the four eligible SNPs with no supporting evidence, but with a minor allele frequency above 10% in the individually genotyped in-house controls. Furthermore, we selected two additional SNPs with no supporting evidence, the SNPs near *MTHFD1L* and *DIAPH3*, because we considered these to be

promising functional candidate genes. *DIAPH3* is an important factor in epithelial cell migration[5], and a previous study suggested an association of hypospadias with intake of certain nutrients related to one-carbon metabolism[6], in which the *MTHFD1L* gene plays a role. Thus, 20 SNPs were eligible for individual genotyping in the discovery sample.

Expression of DGKK

Preputial skin samples

Excess preputial skin was obtained from fourteen patients with hypospadias undergoing surgical repair and from ten age-matched controls undergoing elective circumcision between January 2004 and June 2004 at the Department of Urology of the University of California, San Francisco, CA, USA. Patients with undescended testis, intersex condition or known endocrine abnormalities were excluded from the study. No patients received preoperative testosterone treatment. The institutional Committee on Human Research at the University of California, San Francisco approved this study and all parents gave written informed consent for participation in the study.

RNA preparation, reverse transcription and real-time PCR

We isolated RNA from the preputial skin samples using the QIAGEN RNeasy Midi Kit (QIAGEN Inc.). The manufacturer's protocol was modified to eliminate any contaminating protein and DNA by adding Proteinase K (Roche Diagnostics GmbH) and incubation at 55°C for 20 min, and by adding RNase-free DNase I (QIAGEN Inc.), which was incubated at room temperature for 15 min. RNA quantity and purity were measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc.) and RNA integrity was visualized in agarose gels by the 28s and 18s ribosomal RNA bands. We performed the reverse transcription (RT) polymerase chain reaction (PCR) according to the standard protocol. Briefly, 2.5 µg RNA was reverse transcribed in a 20 µl reaction volume and RT products were diluted 4-fold with TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA).

We designed real-time PCR primers according to target gene sequences published in the NCBI nucleotide sequence database (Supplementary Table 8) and these were synthesized by Integrated DNA Technologies Inc. Real-time PCR was performed in a 20 µl reaction volume containing 20 ng template DNA, 1 × *Power* SYBR Green PCR master mix (Applied Biosystems) and 300 nM primers. Amplification was carried out in 96-wells plates in a 7300 Fast sequence detection system (Applied Biosystems) with the manufacturer's default thermal profile (50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 sec and 60°C for 1 min) followed by a dissociation stage (95°C for 15 sec, 60°C for 15 sec, followed by a slow ramp to 95°C). We performed primer titration and dissociation experiments to prevent primer dimers or false amplicons from interfering with the result. After the RT PCR experiment, Ct numbers were extracted for both the reference gene and the target gene with auto baseline and manual

threshold using standard 7300 Fast software. Expression levels of *DGKK* were reported relative to *GAPDH*. We repeated the PCR three times for each sample and averaged results across the three measurements when no exceptional outliers were seen. Consequently, two samples were averaged across two measurements.

References

1. Franke B, Arias Vasquez A, Veltman JA, Brunner HG, Rijpkema M, Fernández G. Genetic variation in *CACNA1C*, a gene associated with bipolar disorder, influences brainstem rather than gray matter volume in healthy individuals. *Biol Psychiat* 2010:68;586-588.
2. Hoogendoorn EH, Hermus AR, de Vegt F, Ross HA, Verbeek ALM, Kiemeney LALM, Swinkels DW, Sweep FCGJ, den Heijer M. Thyroid function and prevalence of anti-thyroperoxidase antibodies in a population with borderline sufficient iodine intake: influences of age and sex. *Clin Chem* 2006:52;104-111.
3. Germer S, Holland MJ, Higuchi R. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res* 2000:10;258-266.
4. Davis OSP, Plomin R, Schalkwyk LC. The SNPMaP package for R: a framework for genome-wide association using DNA pooling on microarrays. *Bioinformatics* 2009:25;281-283.
5. Gupton SL, Eisenmann K, Alberts AS, Waterman-Storer CM. mDia2 regulates actin and focal adhesion dynamics and organization in the lamella for efficient epithelial cell migration. *J Cell Sci* 2007:120;3475-3487.
6. Carmichael SL, Yang W, Correa A, Olney RS, Shaw GM. Hypospadias and intake of nutrients related to one-carbon metabolism. *J Urol* 2009:181;315-321.