Evaluation of the ISL1 gene in the pathogenesis of bladder exstrophy in a Swedish cohort

Samara Arkan1, Jia Cao1, Johanna Lundin1,2,3, Daniel Nilsson1,4, Thomas Källman5,6, Gillian Barker7, Gundela Holmdahl8, Christina Clementsson Kockum9, Hans Matsson1 and Agneta Nordenskjöld1,9

Bladder exstrophy is a congenital closure defect of the urinary bladder with a profound effect on morbidity. Although the malformation is usually sporadic, a genetic background is supported by an increased recurrence risk in relatives, higher concordance rates in monozygotic twins and several associated chromosomal aberrations. Recently, the ISL1 gene was presented as a candidate gene for bladder exstrophy and epispadias complex (BEEC) development in two different studies. In our study, we screened for genetic variants in the ISL1 gene in DNA from 125 Swedish patients using Sanger sequencing and array-CGH analysis. In addition, we evaluated ISL1 expression in RNA of human bladder during embryonic and fetal weeks 5–10 relative to that in lung tissue (week 9). In total, 21 single-nucleotide variants were identified, including a potentially novel missense variant, c.137C>G p. (Ala46Gly), substituting a conserved amino acid. This variant was inherited from an unaffected mother. No structural variants were identified. RNA sequencing revealed ISL1 mRNA expression during the critical time frame of human bladder development. In conclusion, we did not detect any known or likely pathogenic variants in the ISL1 gene in 125 Swedish BEEC patients, indicating that variation in the ISL1 gene is not a common genetic mechanism of BEEC development in the Swedish population.

A genome-wide association study of 110 classic bladder exstrophy (CBE) patients and 1,177 controls of European origin followed by a meta-analysis identified the rs9291768 variant in genome-wide significant association with isolated CBE.11 The ISL LIM homeobox 1 (ISL1) on chromosome 5q11.2 is the closest protein-coning gene to the genomic location of rs9291768. As the ISL1 gene was also shown to be expressed in the murine fetal bladder region, it was proposed as a susceptibility gene for CBE. The association to this region on chromosome 5q11.2 was subsequently replicated in another study of 268 CBE cases from different populations, including 116 DNA samples from Sweden, in which a significant association was identified between CBE and another variant, rs6874700 single-nucleotide polymorphism, located 16 kb from rs9291768 in the ISL1 gene locus.12 These two studies suggest ISL1 as a major candidate gene for the development of CBE. In this study, we screened the ISL1 gene for single-nucleotide variants (SNVs) in DNA from the same 116 Swedish samples and 9 additional cases with BEEC. Initially, we focused our analyses on the exons in the ISL1 gene. However, the ISL1 intron 1 contains marks for lysine 27 acetylation of the Histone 3 protein (H3K27Ac, ENCODE Jan 2011 Freeze), indicating a region of enhanced transcription activity. This finding motivated us to analyze the entire intron 1 for the presence of polymorphisms in DNA from all 125 bladder exstrophy cases.

Furthermore, we assessed the expression of the ISL1 gene based on our in-house RNA-sequencing data of embryonic and fetal
human urinary bladder tissues. Finally, to identify any copy number variation overlapping the ISL1 gene, the 5q11.2 region was analyzed with array-CGH using DNA from the 125 BEEC cases.

MATERIALS AND METHODS

Subjects

DNA was isolated from the blood or skin of 125 patients with BEEC who were recruited by the Pediatric Surgery Departments in Stockholm, Gothenburg, Uppsala and Lund. The Ethics Committee at Karolinska Institutet in Stockholm approved the study, as in all included centers. Only patients with a diagnosis of BEEC with no additional malformations or family history of BEEC were included in the study. The majority of patients are Caucasian. Control DNA samples were acquired from two different sources: placenta tissue, acquired after normal delivery of healthy newborns of European origin and without obvious malformations during the year 2006 at the Karolinska University Hospital; and peripheral blood from anonymous blood donors sampled at the Karolinska University Hospital.

Embryonic and fetal bladders and lung tissue were obtained from terminated pregnancies after informed consent and with ethics approval. Samples were taken from embryonic and fetal weeks 5, 6, 7, 7.5, 8, 8.5 and 10, and lung tissue from week 9.

DNA isolation, sequencing and analysis

DNA samples from cases were prepared either by standard chloroform extraction or by using a Centra Puregene kit (Qiagen, Hilden, Germany) from either blood or skin. All six coding exons and intron 1 of the annotated ISL1 gene (RefSeq, NM_002202.2) genome assembly hg19) were individually amplified by PCR. Primer pairs were designed with the Primer Quest Tool (http://eu.idtdna.com/PrimerQuest/Home/Index). The primers used to amplify and sequence exon 2 containing the novel variant were 5'-GGCCTATAAGAAGCAGCACTAAA-3' (forward PCR and Sanger primer) and 5'-GGTTGGTATGACTAAGTGGG-3' (reverse PCR and Sanger primer). Additional primer sequences and details of the PCR conditions are available on request. The amplified products were purified using illustra ExproStar 1-STEP kit (GE Healthcare, Chicago, IL, USA) according to the standard protocol, followed by capillary DNA sequencing using a BigDye Terminator v3.1 Cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) on an ABI 3730 DNA Sequencer (Applied Biosystems). SNVs were queried against the ISL1 reference (NM_002202.2) and analyzed using CodonCode Aligner V3.7.1 (CodonCode, Centerville, MA, USA).

Allele frequencies were queried against all available populations (ALL) and the European (EUR) population sample included in the 1000 Genomes Phase 3 (http://browser.1000genomes.org/index.html) and the Exome Aggregation Consortium (http://exac.broadinstitute.org/), containing variants from 60,706 individuals. In addition, we used allele frequencies from the beta version of the Genome Aggregation Database (gnomAD), including 167 million sequences and 10 million whole-genome sequences from unrelated individuals (http://gnomad.broadinstitute.org/), and the SweGene Variant Frequency browser (https://swegen-exac.nbis.se/), containing variants from a cross-population sample of 1,000 Swedish whole genomes. For gnomAD frequencies, we chose the non-Finnish European population sample as the most representative control population. For the c.137C>G exon 2 variant, we also assessed the allele frequency by genotyping 714 control DNA samples (358 placenta and 356 blood donor samples) collected in Sweden.

Potential damaging effects of variants were assessed (access date 7 October 2016) using the online tools MutationTaster (http://www.mutationtaster.org/) and Combined Annotation Dependent Depletion, assessed via SeattleSeq Annotation 138 (http://snp.gs.washington.edu/SeattleSeqAnnotation138/index.jsp) as well as the Mendelian Clinically Applicable Pathogenicity score (http://bejerano.stanford.edu/mcap/), for rare missense variants and PANTHER (http://pantherdb.org/) for the ISL1 chr5: c.137C>G variant. Genetic information regarding this variant is available in ClinVar as accession number SCV000320699. Information regarding the novel c.29-123A>G variant in ISL1 (NM_002202.2) intron 1 was submitted to dbSNP and was given the NCBI ss number 213754377. The Regvsdb version 1.1 (http://www.regulomedb.org/) database was accessed on 6 November 2017 to annotate SNVs with known or predicted regulatory elements from ENCODE data sets (ChIP-seq peaks, DNase I hypersensitivity peaks and DNase I footprints) and additional data sources (ChIP-seq information from the Reference Genome Read Archive, conserved motifs, expression quantitative trait loci, chromatin states from the Roadmap Epigenome Consortium and experimentally validated functional variants). Regulome scores are based on the confidence of the functionality of variants with low scores corresponding to high confidence. Subcategories are used do denote additional functional annotations; http://www.regulomedb.org/help.

Array-CGH analysis

For array-CGH analysis, a 180K custom oligonucleotide microarray with whole-genome coverage and a median probe spacing of ~18 kb was used (Oxford Gene Technology, Yarrton, Oxfordshire, UK). This array design is used in both clinical investigations at the Department of Clinical Genetics, Karolinska University Hospital, Sweden. Genomic DNA isolated from the patients’ blood and sex-matched pooled reference DNA isolated from healthy controls (Promega, Madison, WI, USA) were analyzed, and sample labeling (CGH labeling kit for oligo arrays, Enzo Life Sciences, Farmingdale, NY, USA), hybridization and slide washing (Oligo aCGH/Chip-on-Chip Wash Buffer Kit, Agilent Technologies, Wilmington, DE, USA) were performed according to the manufacturers’ recommendations. The array slide was scanned on a microarray scanner with 3 mm resolution, and initial data analysis was performed with the Feature Extraction software v 11.5.1.1 (Agilent Technologies), followed by analysis with the CytoSure Interpret Software (Oxford Gene Technology).

Protein alignment

The human ISL1 protein was aligned against homologs in different species using ClustalW v.2.1 (http://www.genome.jp/tools/clustalw/) with the default settings. The species and protein accession numbers included were Homo sapiens, NP_00193.2; Canis lupus familiaris, XP_837321.2; Mus musculus, NP_067434.3; Dario reio, NP_571037; and Drosophila melanogaster, AAB49892.1.

RNA preparation, sequencing and analysis

Fresh tissues were dissected and immediately submerged in five volumes of RNA later Stabilization Solution (Catalog No. AM7020, Thermo Fisher Scientific, Waltham, MA, USA), incubated at 4 °C overnight and then transferred to a temperature of ~20 °C. RNA was extracted from one fetal lung and seven urinary bladder samples by using TissueLyser LT (Catalog No. 85600, QIAGEN, Hilden, Germany) and an RNeasy Fibrous Tissue Mini Kit (Catalog No. 74704, QIAGEN). The RNA was measured both by spectrometry (NanoDrop) and Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) instruments. High-quality RNA from eight samples (one fetal lung and four embryonic and three fetal bladder tissue samples) was used to generate TruSeq PCR-free (Illumina, San Diego, CA, USA) libraries for pair-end sequencing on an Illumina HiSeq2500 machine with a read length of 126 bp. Reads were mapped with TopHat 2.0.4 (http://ccb.jhu.edu/software/tophat/index.shtml) to the human genome assembly, build GRCh37. Between 27 and 32 million fragments were sequenced in each library. A quality control using the FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) software revealed no major issues with any of the obtained data and suggested that the data can be mapped without any filtering or trimming of reads. Between 73 and 85% of reads from individual samples mapped to known exons (median 83.15%). The Ensemble annotation and genome sequence version GRCh37 used are available on the Illumina iGenomes web page (http://support.illumina.com/sequencing/sequencing_software/igenome.html). Star v. 2.5.1b with the default settings was used to map the short reads to the reference genome. Read pairs in which both reads map to the genome and overlapped with a gene annotation were counted using the function featureCounts from the R package Rsubread v. 1.2.23. Reported expression values are the observed read counts converted to counts per million mapped reads, as implemented in the cpm function from the R package edgeR. The RNA-sequencing raw data file was submitted to ArrayExpress and is available under accession number E-MTAB-5143.
A first Sanger sequencing approach was performed to screen for SNVs in the protein-coding exons, flanking intronic sequences and the 5’ untranslated region of the ISL1 gene (NM_002202.2). Altogether, 17 SNVs were found. Seven of the 17 variants were predicted to result in possibly damaging amino-acid substitutions; however, all but one were previously described (Table 1). One likely novel missense variant, chr5:50680483:c.137C>G, in ISL1 exon 2, was found in DNA from a male patient with 46, XY karyotype (Figures 1 and 2a). The alternative G allele was absent in the 1000 Genomes Phase 3 ALL populations, the Exome Aggregation Consortium, the beta version of the gnomAD and the SweGen Variant Frequency browser (Table 1). In addition, the variant was not detected in 714 in-house control samples; we therefore conclude that the variant is likely novel. Further analysis of DNA from parents revealed that this variant was transmitted from an unaffected father. This c.137C>G variant predicts an alanine to glycine substitution, p.(Ala46Gly) located in the first of two LIM domains (Figure 2a). The position of the alanine 46 residue is conserved in vertebrates (Figure 2b).

We used prediction tools to assess the potential damaging effect and pathogenicity of the variants identified (Table 1). The c.137C>G variant was scored as “disease causing” by MutationTaster, “possible pathogenic” by Mendelian Clinically Applicable Pathogenicity (score = 0.52) and “probably damaging” by PANTHER. The evolutionary analysis in PANTHER assumes that the longer the position has been preserved, the more likely it will exert a deleterious effect. The calculated position-specific evolutionary preservation time of the c.137C allele was 1036 million years (the threshold for probably damaging is > 450 million years).

Except for two synonymous variants, the variants in the ISL1 gene were either located in introns or in the 5’ untranslated region.

A second Sanger sequencing analysis of the entire ISL1 intron 1 identified an additional four SNVs; all but one were previously known (Table 1). A novel c.29-122G>A variant on chromosome 5:50680252 was detected in 2 out of 125 individuals, resulting in an alternative allele frequency of 0.8% (Table 1). This variant is absent in the 1,000 genomes of Swedish origin (SweGen), the gnomAD non-Finnish European population and the 1,000 Genomes populations.

In addition, we queried the RegulomeDB database to annotate the detected SNVs with known and predicted regulatory elements, including transcription factor-binding sites and promoter regions. Top-scoring variants included the intron 1 variant c.29-122G>A, the silent substitution c.504A>G; p.(Pro168 = ) and the intron 4 variants c.766-101A>G and c.766-21G>T. These variants presented a score of 2b, representing “likely to affect binding.” However, only the variant c.29-122G>A had a relatively high frequency in our bladder exstrophy cases and has not been reported previously in the available databases. Closer inspection revealed that the variant overlaps with a GATA2 protein-binding region and might affect interferon-regulatory factors binding sites. The absence of a score of “1” for any variant indicates that none of the variants are likely to affect the expression of gene targets as expression quantitative trait loci. We assessed in-house RNA-sequencing data from human embryonic and fetal bladder tissues and detected ISL1 gene expression in all samples from week 5 to 10, with a peak in expression level occurring at week 8 (206 counts per million reads (CPMR)). In contrast, one fetal lung tissue sampled as a control at week 9 showed very low ISL1 expression (3 CPMR).
DISCUSSION

On the basis of previous results suggesting that *ISL1* is a major candidate gene for bladder exstrophy, the purpose of this study was to investigate the genetic variation of the *ISL1* gene among 125 Swedish BEEC patients. We did not detect any known or likely pathogenic variants in the *ISL1* gene in our cohort. One potentially novel missense variant was detected in one patient; however, this variant was inherited from a healthy mother, reducing the pathogenicity score of this variant to that of a variant of unknown significance. Keeping in mind that BEEC is most likely a condition of multifactorial etiology and that the genetic influence likely involves some degree of variable penetrance and expressivity, we believe that although the c.137C→G was inherited from a healthy mother, that does not completely rule out this variant as disease-causing.

Analysis of *ISL1* intron 1 revealed a c.29-123G→A variant present in 2 out of 125 individuals. Although intriguing, further studies are needed using larger samples to assess the effect of intronic variations and the possible functional effect on *ISL1* gene regulation.

*ISL1*, a LIM-homeodomain transcription factor, regulates gene expression by binding via the HOX domain to enhancer regions of target genes, including insulin. LIM domains are dual zinc-finger...
structures that mediate protein–protein interactions, and LIM-homeodomain proteins exhibit distinct transcriptional activity during development. The Ala46Gly substitution replaces alanine to the less bulky glycine residue positioned in the linker between the first and second N-terminal zinc fingers in the first of two LIM domains in ISL1. Functional studies providing evidence of interference with normal ISL1 function are required to classify the variation as a pathogenic mutation.

Further support for the involvement of the ISL1 gene in human bladder development was gathered from RNA sequencing in four embryonic and three fetal human bladders collected during weeks 5–10. Our data indicated moderate-to-high ISL1 gene expression in human embryonic and fetal bladder during the period of closing of the abdominal and bladder wall, supporting the importance of the gene in human fetal bladder development. The availability of tissue samples restricted our analysis to one sample per time point. The interpretation of a detailed ISL1 gene expression profile during bladder development would require replicates from the same time point to allow for estimates of biological variance. However, the embryonic and fetal bladder samples are unique and present unique consecutive data regarding the RNA expression pattern of ISL1 during the most important embryonic period for bladder closure. The 5′ untranslated region of ISL1 partly overlaps with an uncharacterized non-coding RNA gene, LOC642366, transcribed in the opposite direction (minus strand) relative to ISL1. The expression level of LOC642366 in the fetal bladder is currently unknown, as the methodology used in this study for RNA sequencing does not allow for an assessment of the expression of non-coding RNA.

In conclusion, we did not detect any known or likely pathogenic variants in the ISL1 gene in 125 Swedish BEEC patients, indicating that variation in the ISL1 gene is not a common genetic mechanism of BEEC development in the Swedish population. Future ISL1 studies in additional populations would expand our knowledge of the contribution of the ISL1 gene to BEEC development.

ACKNOWLEDGEMENTS

We want to express our sincere gratitude to all participating families. We also acknowledge Christina Nyström for excellent technical support. This work was supported by the Swedish Research Council, Stiftelsen Frimurare Barnhuset i Stockholm, the Swedish Medical Association, Stockholm County Council and Karolinska Institutet. We acknowledge support from the Science for Life Laboratory, the National Genomics Infrastructure, NGI and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure.

COMPETING INTERESTS

The authors declare no conflict of interest.

PUBLISHER’S NOTE

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. DGJPJR. Exstrophy-Epispadias Complex and Bladder Anomalies. 7th edn. W. B. Saunders: Philadelphia, PA, USA, 1998.
2. Cervellione RM, Mantovani A, Gearhart J, Bogaert G, Gobert R, Caione P et al. Prospective study on the incidence of bladder/cloacal exstrophy and epispadias in Europe. J Pediatr Urol 2015; 11: e331–e336.
3. Shapiro E, Lepor H, Jeffs RD. The inheritance of the exstrophy-epispadias complex. J Urol 1984; 132: 308–310.
4. Boyadjiev SA, Dodson JL, Radford CL, Ashrafli GH, Beatty TH, Mathews RI et al. Clinical and molecular characterization of the bladder exstrophy-epispadias complex: analysis of 232 families. BJU Int 2004; 94: 1337–1343.
5. Ludwig M, Utsch B, Reutter H. [Genetic and molecular biological aspects of the bladder exstrophy-epispadias complex (BEEC)]. Urologe A 2005; 44: 1037–1038, 1040-1044.
6. Reutter H, Qi L, Gearhart JP, Boerms T, Ebert AK, Rösch W et al. Concordance analyses of twins with bladder exstrophy-epispadias complex suggest genetic etiology. Am J Med Genet 2007; 143A: 2751–2756.
7. D improve M, Reutter H, Schramm C, Bartels E, Boerms TM, Ebert AK et al. Microduplications at 22q11.21 are associated with non-syndromic classic bladder exstrophy. Eur J Med Genet 2010; 53: 55–60.
8. Lundin J et al. 22q11.2 microduplication in two patients with bladder exstrophy and hearing impairment. Eur J Med Genet 2010; 53: 61–65.
9. von Lowszto C, Hofmann A, Zhang R, Marsch F, Ebert AK, Rösch W et al. CNV analysis in 169 patients with bladder exstrophy-epispadias complex. BMC Med Genet 2016; 17: 35.
10. Baranowska Korberg I, Hofmeister W, Markljung E, Cao J, Nilsson D, Ludwig M et al. WNT3 involvement in human bladder exstrophy and cloaca development in zebrafish. Hum Mol Genet 2015; 24: 5069–5078.
11 Draaken M, Knapp M, Pennimpede T, Schmidt JM, Ebert AK, Rösch W et al. Genomewide association study and meta-analysis identify ISL1 as genome-wide significant susceptibility gene for bladder exstrophy. PLoS Genet 2015; 11: e1005024.

12 Zhang R, Knapp M, Suzuki K, Kajioka D, Schmidt JM, Winkler J et al. ISL1 is a major susceptibility gene for classic bladder exstrophy and a regulator of urinary tract development. Sci Rep 2017; 7: 42170.

13 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016; 536: 285–291.

14 Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 2016; 32: 292–294.

15 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013; 29: 15–21.

16 Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014; 30: 923–930.

17 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26: 139–140.

18 Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014; 11: 361–362.

19 Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics 2016; 32: 2230–2232.

20 Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 2016; 48: 1581–1586.

21 Boyle AP et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012; 22: 1790–1797.

22 Kadrmas JL, Beckerle MC. The LIM domain: from the cytoskeleton to the nucleus. Nat Rev Mol Cell Biol 2004; 5: 920–931.