Classic Bladder Exstrophy: identification of genetic markers and characterization of its associated *ISL1* gene

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Enrico Mingardo

aus

Padova

Bonn 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer: Prof. Dr. Rer. Nat. Benjamin Odermatt
Gutachter: Prof. Dr. Rer. Nat. Michael Pankratz

Tag der Promotion: 17.10.2023
Erscheinungsjahr: 2024

Dedicated to my family.

Summary

The bladder exstrophy-epispadias complex (BEEC) is a spectrum of congenital abnormalities which involves the urinary tract, abdominal wall, bony pelvis, the external genitalia, and in the worse cases also the gastro-intestinal tract. Classic bladder exstrophy (CBE) is the intermediate form with the highest occurrence of about 1:30.000. CBE is characterized by evaginated bladder plate template from the abdominal wall and can include kidney and other upper urinary tract anomalies. CBE management is surgical but has long-term complications such as decreased bladder continence, susceptibility of bladder infection and malignancies of the bladder comprising mainly urothelial cell carcinoma and adenocarcinoma.

CBE complex occurs during the first 8 weeks of fetal development when the cloaca tissues split to form urogenital sinus, bladder and rectum and the anterior abdominal wall matures to form muscles and connective tissue of the lower abdomen. The causes that imply a CBE developmental defect during this gestation time are still unknown. De facto, even if the CBE phenotype is well characterized, few is known about its genetic causes and no molecular mechanism are described in literature. Only recently, genetic studies identified one locus in chromosome 5q11.2 that suggested *ISL1* as major susceptibility gene for CBE, but these studies were limed to a restrict number of patient cohort.

The aim of this study is to shade better light in the genetic causes of CBE and in the contribution of the founded loci both for the development of CBE condition and its associated bladder malignancies, alongside with the molecular characterization on CBE development of the top associated locus on *ISL1.*

For that, in this study we show the genome-wide association study (GWAS) on CBE with the largest patient cohort to date, identifying 8 genome-wide significant loci, 7 of which are novel. Within these loci reside four non-coding and ten coding genes. To study the contribute of these genes in bladder development, we performed total RNA sequencing (RNA-seq) of mouse embryonic bladder tissues at stage E10.5, E12.5 and E15.5 analysed together with total RNA-seq of human embryo bladder from gestational week 7 to 9. The result showed that all genes are expressed and/or significantly regulated in both mouse and human bladder development. In addition, to study their contribution in bladder cancer susceptibility, the expression of these genes was queried in 5 different bladder cancers RNA-seq compared with normal healthy bladder tissue donor. As result, nine of the GWAS genes were differently expressed in bladder cancers. It is remarkable to observe a molecular switch of the genes which were downregulated in bladder development becoming upregulated in bladder cancers. This

study brings to light 7 new loci associated with CBE and evidences the connection of the related genes both with bladder development and CBE associated cancer.

The most significant GWAS locus for the CBE cohort reseeds in chromosome 5, a region also identified in previous genetic studies. Here is ISL LIM Homeobox 1 (*ISL1*), a transcription factor (TF) that belongs to the family of the Homebox genes. *ISL1* is known to play an important role for the development of heart, kidney, limb and neurons. Recently, mouse models have shown the relevance of *ISL1* for the correct development of the early urinary tract and genitalia. In addition, human embryo bladder RNA-seq revealed significant expression of *ISL1* that, together with our previous findings, suggests this gene to be highly expressed in the early stages of bladder development and strong downregulate in differentiated bladder tissue. Interestingly, an exome sequencing of *ISL1* in CBE cohort did not detect any pathogenic variant, suggesting that the contribute of *ISL1* to CBE reseeds in its regulation rather than variation. In fact, our GWAS shows association with the regulatory variant rs2303751 in *ISL1*. This region has been suggested by regulome database (regulomebd) to be targeted by Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit (EZH2) and recently though, EZH2 has been shown to act as a TF. A sliding window luciferase assay in the rs2303751 locus identified a 1.2 kilobases (kb) human intragenic promoter (called in this study Fragment 2) residing between 6.2 and 7.4 kb downstream of the ISL1 transcription starting site (TSS) and active in the reverse strand. The A to G variant rs2303751 does not influence the promoter regulation neither its orientation but overexpression of EZH2 enhance its activity. This study shows that in HEK 293 cells EZH2 is occupying the Fragment 2 region and its presence is associated specifically with *ISL1* regulation; in detail, the absence of EZH2 in Fragment 2 diminishes *ISL1* expression.

To check whether the *ISL1* EZH2-mediated dysregulation could affect the development of the urinary tract, further *in vivo* experiments are performed in *ezh2*$^{-/-}$ knock out (KO) zebrafish (zf) larvae (*ezh2*^{ul2}). Since previous *in situ* study proved *isl1* expression in the pronephron of developing zf wild type (WT) larvae, this study replicates the experiment in the *ezh2*^{ul2} line observing a strong tissues specificity regulation of *isl1* with a reduced signal in the pronephric tissue. This finding is later confirmed with retro transcription quantitative PCR (RT-qPCR). To access the location of this tissue specificity regulation, the *ezh2*^{ul2} is crossed with the nephron reporter line *Tg(wt1ß:eGFP)* obtaining the double transgenic line *Tg(wt1ß:eGFP)-ezh2*^{ul2}. Here, immunohistochemistry (Ihc) of both Isl1 and GFP showed a decreased level of Isl1 protein that locates specifically in the glomeruli and in the nephric duct structures. Finally, confocal images of the nephric structures in the *Tg(wt1ß:eGFP)-ezh2*^{ul2} revealed developmental defect of the nephric duct, suggesting that the decreased level of Isl1 mediated by Ezh2 is tissue specific and contribute to developmental defect of the urinary tract.

Since in human we find the Fragment 2 promoter targeted by EZH2 active in the reverse strand, we tested for expression of *ISL1-DT* via qPCR under EZH2 knock down. *ISL1-DT* is a human specific long non coding RNA gene that spaces 115 bp from *ISL1,* and divergently orientated from *ISL1.* Our meta-analysis of RNA-seq in different human tissues suggests *ISL1* and *ISL1-DT* to share the same promoter and, given these conditions, we showed that the EZH2 knock down reduces the expression of *ISL1* but does not alter the expression of *ISL1-DT*, confirming the Fragment 2 as a specific promoter for the regulation of *ISL1*.

All these findings suggest that our GWAS identifies a CBE-associated promoter that reseeds in *ISL1*, this promoter is active in the reverse strand and can be targeted by EZH2 enhancing specifically *ISL1* transcription. We observed that this mechanism appears to be conserved in zebrafish, further supporting its importance. The zebrafish model provided evidence for a tissue-specific mode of action showing that the knocking out Ezh2 resulted in specific downregulation of *isl1* in the nephric region. In this study we propose a model where a miss-binding of EZH2 in the *ISL1* internal promoter decreases the expression of *ISL1* during early bladder development, altering the expression of the its downstream target such as *SHH*. This could lead to a developmental defect in the cloaca and primitive bladder for example enhancing an early differentiation of smooth muscles tissue. On the other side, our GWAS suggests that this mechanism is not the main cause for bladder congenital defect but it is rather a contribution together with the other different identified CBE-associated loci.

Abstract

Abstract

Classic bladder exstrophy (CBE) is a congenital anomaly with an occurrence of 1 in 30.000 new born. It is characterized by the protrusion of the bladder plate through an open abdominal wall, often accompanied by kidney and upper urinary tract abnormalities. Its long-term complications include different bladder malignancies.

While the phenotype of CBE is well understood, the genetic and molecular causes remain largely unknown. Previous genetic studies have identified one locus associated with CBE, but these studies had limited sample sizes. The aim of this study is to identify new risk loci and gain a better understanding of their contributions to CBE and CBE-associated cancer. Furthermore, the study aims to elucidate the molecular mechanisms through which the associated *ISL1* gene contributes to CBE.

For that, this study shows here the largest genome-wide association study (GWAS) conducted on CBE to date, identifying eight genome-wide significant loci, seven of which are novel. Within these loci, ten coding and four non-coding genes are found. RNA-seq analysis of mouse and human embryonic bladder tissues at different developmental stages reveal those genes to be expressed and differentially regulated in bladder development. Furthermore, those genes are differentially expressed in various bladder cancers with down-regulated genes in bladder development being up-regulated in bladder cancer and vice versa. These findings suggest genetic drivers for classic bladder exstrophy and their potential role in CBE-associated bladder cancer susceptibility.

The most significant CBE-associated locus is found on chromosome 5, housing the homeobox transcription factor *ISL1* gene, known to be involved in the development of different tissues. Mouse models have been recently shown its role in urinary tract and genitalia development, and our RNA-seq also indicates high expression of ISL1 during early bladder developmental stages, followed by a strong decreased expression in differentiated bladder tissue. Exome sequencing of *ISL1* in CBE cohort did not find any pathogenic variants in *ISL1*, suggesting that its contribution to CBE lies in gene regulation rather than genetic variation.

Our GWAS identified a regulatory variant rs2303751 in *ISL1* that is suggested to be target of EZH2, recently shown to act as a transcription factor. Here, we describe a novel 1.2 kb intragenic promoter (called Fragment 2) residing between 6.2 and 7.4 kb downstream of the *ISL1* transcription starting site.

This promoter is active in the reverse DNA strand of *ISL1* and in HEK 293 cells harbours a binding side for EZH2 where the rs2303751 marker reseeds. Here we show, that EZH2 enhance Fragment 2 activity and that its silencing reduces *ISL1* expression. Further in vivo experiments in zebrafish *ezh2*$^{-/-}$ ko larvae display tissues specificity of *isl1* regulation with reduced expression that locates specifically in the pronephric region of zebrafish larvae. In addition, a shorter and malformed nephric duct is observed in *ezh2*$^{-/-}$ ko zebrafish *Tg(wt1ß:eGFP)* reporter lines.

Since in human Fragment 2 is active in the reverse orientation of *ISL1*, we showed that in HEK 293 its EZH2 regulation is not altering the expression of *ISL1-DT*, the immediate divergent transcript from ISL1, but rather disrupts the ratio of the *ISL1*/*ISL1-DT* cassette. Our study proposes that EZH2 is a key regulator of *ISL1* during early urinary tract formation and suggests tissue specific *ISL1* dysregulation as an underlying mechanism for CBE formation.

Table of content

Summary I

Abstract IV

1. Introduction 1

1.1 Embryology of the urinary bladder development 1

1.2 Molecular pathways of the urinary bladder development 3

1.3 Classic bladder exstrophy (CBE) 5

1.3.1 CBE etiology 6

1.3.2 CBE genetic 7

1.4 A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy 9

1.4.1 Summary 9

1.4.2 Statement of the candidate contribution 10

1.5 ISL1 in urinary tract development 12

1.6 ISL1 as candidate region for CBE 13

1.6.1 rs2303751 as candidate variant for CBE in ISL1 13

1.6.2 rs2303751 is target of EZH2 13

1.7 The zebrafish 14

1.7.1 Zebrafish as model organism to study urinary tract development. 16

1.6 Aim of the study 17

2. Materials 17

2.1 Consumables 17

2.2 Chemicals 18

2.3 Enzyme 21

2.4 Kit and commercial assays 21

2.5 Plasmids 22

2.6 Oligonucleotide 23

2.6.1 Oligonucleotide for zebrafish genotyping 23

2.6.2 Oligonucleotide for cloning 23

2.6.3 Oligonucleotide for plasmid sequencing 24

2.6.4 Oligonucleotides for RT-qPCR 24

2.6.5 Oligonucleotide for gene knock down 24

2.6.6 Oligonucleotide for ChIP-qPCR 25

2.7 Primary antibody 25
2.8 Secondary antibody 26
2.9 Zebrafish lines 26
3. Methods 26
3.1 Cell culture 26
3.2 Cloning and library preparation 27
3.2.1 PCR amplification of genomic fragments 27
3.2.2 Enzymatic digestion 29
3.2.3 Ligation 29
3.2.4 Transformation and plasmid extraction 30
3.3 Dual Luciferase assay 30
3.3.1 Seeding specific number of cells 31
3.3.2 Transfection 31
3.3.3 Luciferase assay and data normalization 32
3.4 Chromatin immunoprecipitation quantitative PCR (ChIP-qPCR) 32
3.5 EZH2 knock down 34
3.6 Western blot 35
3.7 Zebrafish maintenance 36
3.8 Zebrafish genotyping 36
3.9 quantitative PCR (qPCR) 38
3.9.1 Retrotranscription quantitative PCR (RT-qPCR). 39
3.9.2 Quantitative PCR on shared chromatin (qPCR) 40
3.10 In situ hybridization 40
3.11 Immunohistochemistry 41
3.12 Zebrafish larvae imaging 42
3.12.1 Live zebrafish imaging 42
3.13 RNA-seq analysis. 42
4 Results 42
4.1 Luciferase assays identifies presence of a promoter in the reverse strand of the ISL1 locus 42
4.2 EZH2 protein binds to the GWAS associated CBE region and enhance ISL1 expression 44
4.2.1 ChIP qPCR reveals binding of EZH2 in the rs2303751 locus 44
4.2.2 EZH2 enhance ISL1 expression thrugh binding on fragment 2 45
4.3 Whole mount in situ hybridization on *ezh2*$^{ul2-/-}$ larvae displays tissue specific regulation of *isl1.* 46
4.4 Immune histochemistry on *ezh2*$^{ul2-/-}$ larvae reveals tissue specificity of ISL1 regulation in the *Tg(wt1b:eGFP)* line. 48
4.5 Fragment 2 promoter is specific for ISL1 expression 50

4.5.1 Fragment 2 interacts within ISL1 and ISL1-DT 50
4.5.2 Fragment 2 regulates ISL1 via EZH2 50
5. Discussion 52
5.1 Main findings 52
5.2 HEK 293 and Zebrafish as model organism for this study 53
5.3 Luciferase assay identify promoter activity on Fragment 2 54
5.4 EZH2 regulates ISL1 expression 55
5.4.1 EZH2 binding on Fragment 2 regulates ISL1 in HEK 293 55
5.4.2 ISL1 regulation via EZH2 is tissue specific 56
5.4.3 Nephric malformations in the *ezh2* knock out line 57
5.5 Fragment 2 is specific for ISL1 regulation via EZH2 57
6. Conclusion and outlooks 58
7. Acknowledgment 60
8. Bibliography 61
9. List of publications 68
10. Appendix: A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy 70

1. Introduction

1.1 Embryology of the urinary bladder development

The urinary bladder is a crucial organ in the urinary system that plays a vital role in storing and eliminating urine from the body. To this porpoise, the urinary bladder consists of urothelium layer surrounded by smooth muscles layers. Its development starts early in embryonic life and continues through to adulthood, and it is a complex process that involves various stages of growth and differentiation. Understanding the development of the urinary bladder is important for comprehending the normal anatomy and function of the urinary system, as well as for identifying and addressing any congenital anomalies that may occur.

The formation of the urinary bladder begins during fourth week of embryonic development. Here the caudal end of the endoderm derived hindgut and allantoids are fused in a structure called cloaca, which, in lower vertebrates, serves as a common termination of the digestive and urogenital systems. At this stage, allantoids and hindgut are slightly separated by a shelf of mesodermal derived tissue called urorectal septum (Figure 1 A). The cloaca ends in a structure called proctodeum that acts as a barrer between cloaca and ectodermal depression and consists of apposed endoderm-mesoderm layers (Figure 1 A).

At week six, the urorectal septum evaginates and situates between the hindgut and the base of allantois. During week seven and eight the urorectal septum elongates towards the proctodeum, dividing the cloaca into urogenital sinus and anorectal canal (or rectum) (Figure 1 B). According to classic embryology, the urorectal septum fuses with the cloaca membrane defining the anal and urogenital membranes. Other studies propose that cloaca undergo apoptosis and breaks down without its fusion urorectal septum. At around week eight and nine the urogenital sinus and allantoids are fused in a primitive bladder (Figure 1 B), this structure continues to expands forming the urinary bladder (Liaw et al. 2018).

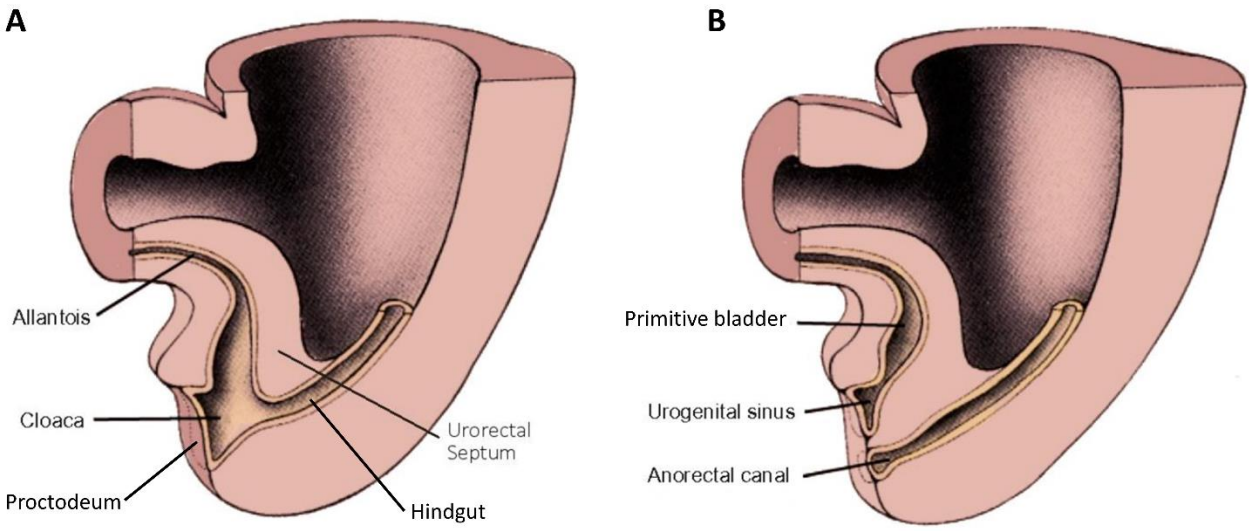


Figure 1. Early embryo development of urinary tract. Figures shows the saggital plane of the tail end of human embryo. A) Gestational week 4 and 5, allantois and hindguts are fused in a structure called cloaca that ends in the ectodermal proctodeum. The urorectal septum slightly invaginate inside the cloaca. B) Gestational week 6 and 7, the urogenital sinus migrates and fuses with the proctodeum splitting the cloaca in the anorectal canal (rectum) and primitive bladder. Images modified from Liaw et al., 2018.

While the bladder grows, it incorporates the mesonephric (Wolffian) duct and its epithelial outgrowth called ureteric bud (Figure 2 A). In this timeframe the ureteric bud grows and get fused into the bladder by opening in its posterior wall forming the ureters (Figure 2 B to D). The region of the bladder composed mostly by smooth muscles where the ureters find its opening is called trigone of the bladder. At the entrance of the mesonephric duct, the bladder gets sharpened and forms the urethra, which serves as outlet of the bladder (Figure 2 E). While the urethra elongates, in females the mesonephric duct regresses while in males it forms the ejaculatory ducts.

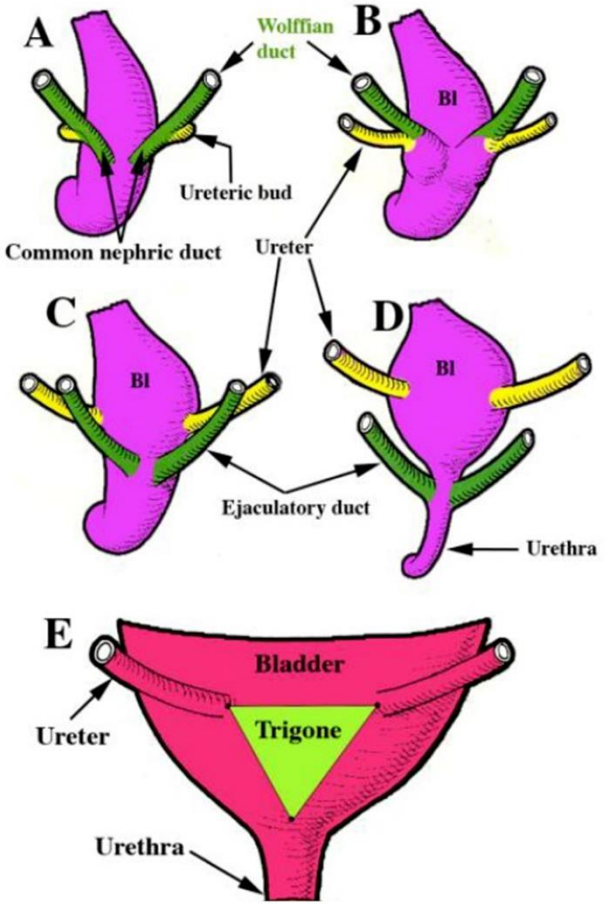


Figure 2. Late embryonic bladder development. A to E, posterior view of the urinary bladder in the timeframe between gestational week 7 to 9. A) The mesonephric (Wolffian) duct with its epithelial outgrowth ureteric bud is incorporated in the posterior part of the primitive bladder. B) The ureteric bud separates from the mesonephric duct and grows getting fused by opening into the posterior wall of the bladder forming ureters. C) and D) while the bladder grows, the mesonephric duct migrates together with uretere forming in males the ejaculatory duct. In female the mesonephric duct regresses. E) At the entrance of the mesonephric duct, the bladder gets sharpened and forms the urethra, which serves as outlet of the bladder. Image modified from Liaw et al., 2018.

1.2 Molecular pathways of the urinary bladder development

The cellular and molecular mechanisms of which cloaca get patterned into rectum and urogenital sinus by urorectal septum (week four and five) and the docking of ureters into bladder (week eight and nine) are poorly understood. Classic embryology describes it as tissue fusion and/or possible mechanical tensions between those. Studies shows that Homeobox genes Hoxa-13 and Hoxd-13 plays a role, in

fact mutant mice lack of the proper cloaca pattering showing a fused urogenital sinus and anus (Warot et al., 1997). Moreover, rather than tissues fusion, it seems that apoptosis plays a role in the separation of cloaca, where highly presence of apoptotic bodies located in the urorectal septum (Qi et al., 2000). Recently, in addition to mechanical tension of kidney and metanephric duct for the docking of the mesonephric duct in cloaca, a new study has shown that the precise spatiotemporal apoptosis gradient from both tissues is essential for the trigone formation (Hoshi et al., 2018).

The general pattering of bladder differentiation is driven by epithelium and mesenchyme interactions. *SHH* produced by inner epithelial cells signals to the peripherical mesenchyme to differentiate into smooth muscles. Vice versa, smooth muscles differentiation is inhibited in the inner epithelium where the SHH signal is strongest (Baskin et al., 1996) (Figure 3). The absence of SHH induces the expression of BMP4 that initiates the cascade of SMADs protein for the smooth muscles differentiation (Tasian et al., 2010) (Liaw et al., 2018a). It is thought that this SHH and SMADs gradient orchestrate the formation of an inner epithelium surrounded by smooth muscles generating the main structure of the urinary bladder.

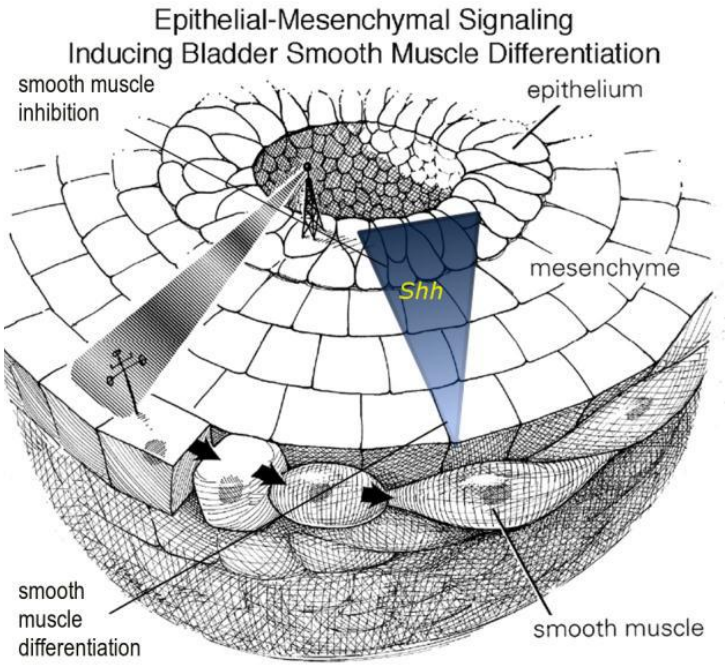


Figure 3. Pattering of bladder smooth muscle differentiation. SHH gradient orchestrate the differentiation of mesenchyme into smooth muscles: absence of SHH indices the differentiation of mesenchyme into smooth muscle, while its high expression induces the manteinante of the inner bladder epithelium.

1.3 Classic bladder exstrophy (CBE)

The bladder exstrophy-epispadias complex (BEEC) is a spectrum of congenital abnormalities which involves the abdominal wall, bony pelvis, the urinary tract, the external genitalia, and in the worse cases also the gastro-intestinal tract. Depending on severity, BEEC shows three different spectra with associated prevalence. Cloaca exstrophy (CE) shows two exstrophied bladder, as well as omphalocele, an imperforate anus and spinal defects (Figure 4 A). CE is the most severe spectrum of BEEC and has the rarest occurrence in new born with a prevalence of 1:300.000. Epispadias (E) is the mildest form and is shown with an open urethra in male and a cleft in female (Figure 4 B). It has an occurrence of 1:100.000. Classic bladder exstrophy (CBE) is the intermediate form with the highest occurrence of 1:30.000. New estimations give a prevalence of CBE in Germany of about 1:30.700 individuals, and counting the European population of about 450.000.000 citizens (https://ec.europa.eu/), it is assumed that around 15.000 CBE patients live in Europe. CBE is associated with an evaginated bladder plate from a non-properly closed abdominal wall, many cases present an epispadic urethra and some affected individuals shows kidney and other upper urinary tract anomalies (Figure 4 C). The management is entirely surgical, with bladder reconstruction and closure of the abdominal wall. Later surgery might be needed to adapt reconstructed bladder with the growth of the individual. Despite this, bladder continency is reduced with 80% of capacity compared to healthy individuals (A. K. Ebert et al., 2009). Associated long-term complications include malignancies of the bladder comprising mainly urothelial cell carcinoma and adenocarcinoma (Dahm & Gschwend, 2003). The management is entirely surgical, with bladder reconstruction and closure of the abdominal wall. Later surgery might be needed to adapt reconstructed bladder with the growth of the individual. Despite this, bladder continence is reduced with 80% of usual bladder capacity (A.-K. Ebert et al., 2009) (Williamson et al., 2011). On the side of the pathological implication of BEEC, it is of an important consideration the psychological and psychosexual outcome of affected individuals during the puberty and adulthood.

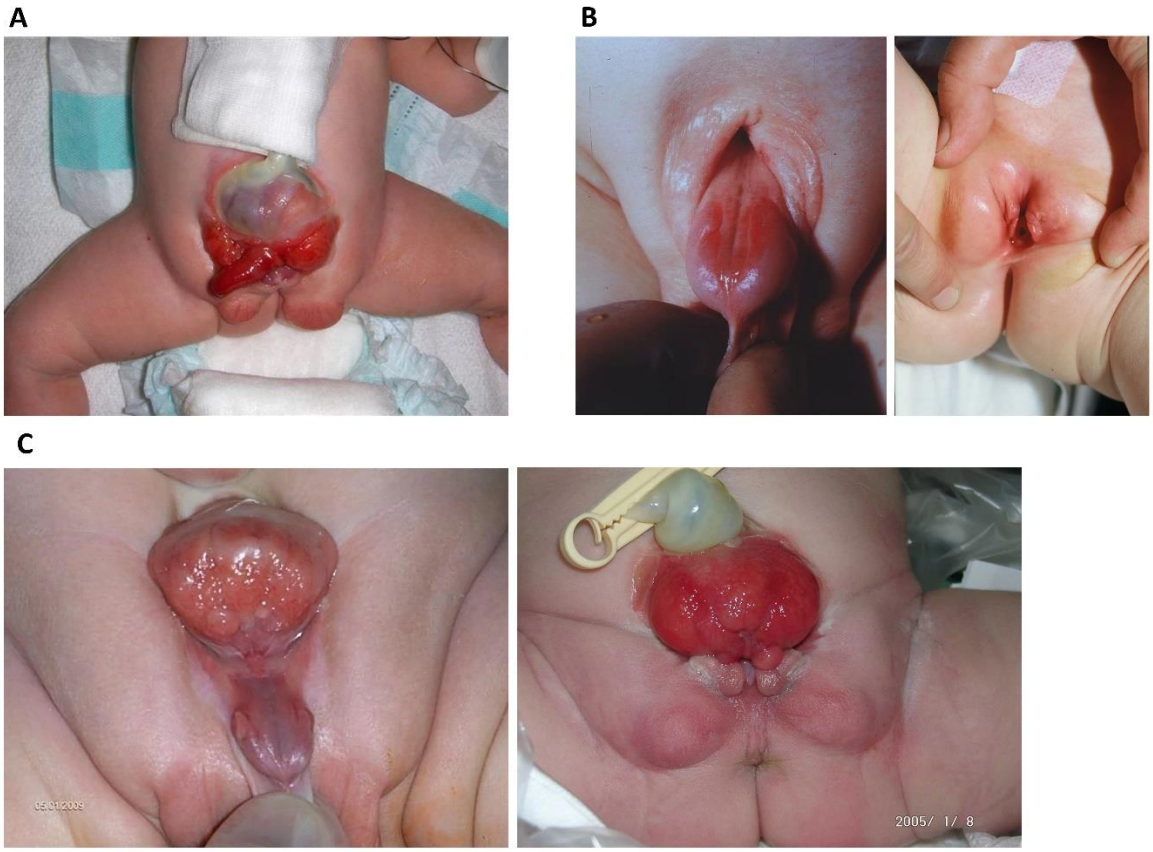


Figure 4. Phenotypes of the Bladder exstrophy epispadias complex spectrum. Representative pictures of BEEC phenotype complex in newborns: A) Cloaca exstrophy in new born (male). B) Epispadias in new born, male on the left and female on the right panel. C) Classic bladder exstrophy in new born, male on the left and female on the right panel. Images re-arranged from Ebert et al., 2009.

1.3.1 CBE etiology

Even though CBE phenotype is well characterized, its etiology is advanced only with hypothesis. In fact, CBE complex occurs in the early fetal development in between the first 6 weeks of gestation and the earlies parental CBE diagnosis can be made around the 23rd gestational week (Gearhart et al., 1995; Weiss et al., 2020). Nonetheless, most of the CBE is diagnosed at birth or in the last gestational weeks. Due to this, the access to the CBE development lacks of information, and functional studies are prevented by lack of early human fetal samples and the impossibility to access affected tissues. So far, the most quoted theory affirms that an overdevelopment of the cloaca would prevent the correct migration of the mesenchyme tissue resulting in a partially not-formed lower abdominal wall from which the bladder will protrude (Marshall & Muecke, 1962). Other mechanisms have been proposed involving pubic diastasis that would prevent abdominal wall closure, ending with a bladder located between the abdominal wall (Satish Kumar et al., 2015). Kulkarni and Chaudhari advanced an

abnormal genital tubercle migration that would drive cloaca out of its locus with the consequences of a bladder that protrudes out from the abdominal wall (Kulkarni & Chaudhari, 2008).

1.3.2 CBE genetic

Although most individuals affected by BEEC do not have a positive family history of the condition, there have been 30 multiplex families identified with BEEC, despite its rarity in familial occurrence. Some of these families appear to follow a Mendelian mode of inheritance, while for the majority of affected individuals, the genetic basis of BEEC is consistent with a multifactorial etiology. In most multiplex families, only two members are affected, while two families have reported three affected members, including males and females with varying degrees of BEEC severity (Beaman et al., 2021). In addition, an unique Moroccan family of three males (two cousins and a maternal uncle) being affected with CBE was recently described (Reutter et al., 2003). In these rare multiplex families, the inheritance of BEEC may be consistent with autosomal dominant with reduced penetrance, autosomal recessive, or X-linked patterns. The lack of recurrence could be partially attributed to reduced reproductive fitness.

Recently, few studies have shade light on the genetic alteration that are associated with CBE. Draaken et al. reported a chromosome 22q11 micro duplication (Draaken et al., 2010) and a recent copy number variant (CNV) analysis displayed many chromosomal duplication or deletions in a Swedish CBE cohort (Nordenskjöld et al., 2023). In a whole genomic overview, 2 genome wide association studies (GWAS) were performed in CBE cohort. The first, with a total of 98 patients and 526 controls, suggested association with WNT3 and WNT9b genes (Reutter et al., 2014) and the second, with a total of 110 CBE patients and 1177 controls with subsequent meta-analysis from the previous one (Draaken et al., 2015a), identified one significant association locus in chromosome 5q11.1, suggesting ISL1 as candidate gene for this region (Figure 5). An additional association study that gained a total of 268 CBE patients and 1354 controls, confirmed, after meta-analysis, the involvement of ISL1 locus (Zhang et al., 2017). These studies were limited to a restricted patient's cohort, for that the study here proposed shows the largest CBE GWAS to date.

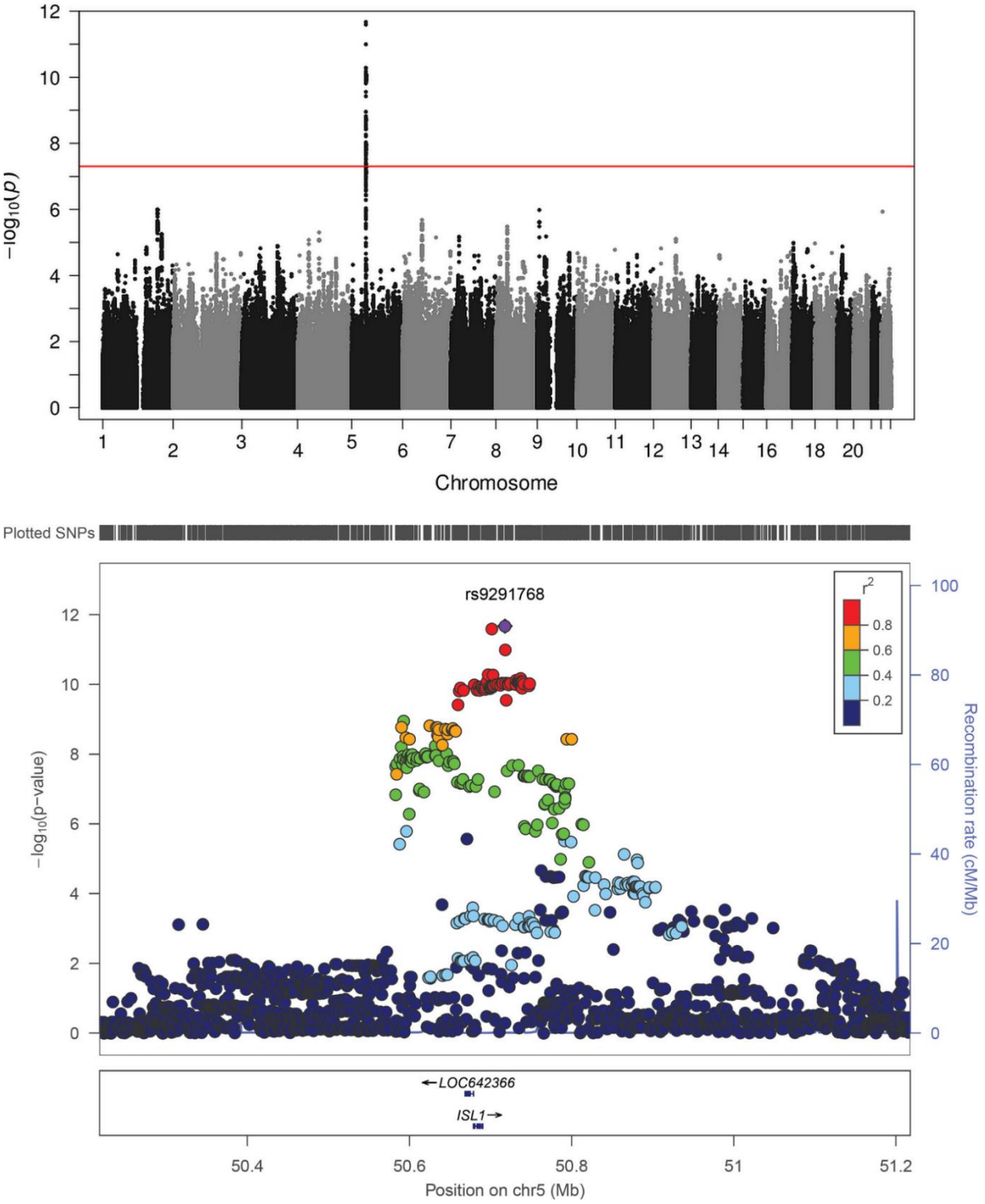


Figure 5. GWAS in CBE cohort of 110 CBE patients and 1177 controls identifies association in chromorome 5. Top panel shows the Manhattan plot of the GWAS in CBE, y-axes indicate the -log10(p-value) association, x-axes shows chromosome number. Every dot is a single nucleotide polymorphism (SNP). Significance is above the red line and indicates the locus in chromosome 5 where *ISL1* reseeds. Bottom panel shows the detail of the associated variants in chromosome 5 in a range of 1 mega base (Mb). Position of the variants in the chromosome (x-axes) are plotted in function of its p-value (y-axes). Top significant variant (rs9291768) is hglited in purple and their colour indicates the linkage disequilibrium (LD) r^2 value (Draaken et al., 2015b). The lowest p-value variant reseeds in *ISL1* gene.

1.4 A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy

1.4.1 Summary

Classic bladder exstrophy (CBE) is a congenital condition in which the bladder is abnormally herniated outside of the abdominal wall and the bladder plate is exposed. This rare condition affects 1 in 40.000 live births and requires surgical intervention to repair the bladder and surrounding structures. Associated long term complications are malignancies of the bladder comprising mainly urothelial cell carcinoma and adenocarcinoma. CBE occurs during the early fetal development in a timeframe between the first 6 gestational weeks. Here different tissues develop and participate to form the urinary tract. Although CBE phenotype and complications are well characterized, nothing is known about its causes and only hypothesis are advanced on its etiology. Only recently, two genome wide association studies (GWAS) were made to shade light on the genetic contribution for CBE and successfully identified a locus on chromosome 5q1.11, highlighting ISL1 as associated genes. On the other hand, those studies were limited to a restricted patient's cohort with a total of 110 CBE patients and 1177 ethnically matched controls.

The study presented here aims to increase the patient's cohort to access stronger loci association and understood their contribution in bladder development and CBE cancer susceptibility comparing those with RNA-seq data. Hence, we preformed the largest GWAS on CBE to date with 628 patients from central Europe, Italy, Spain, Sweden, UK and 7352 ethnically matched controls. This study identified 8 significant risk loci, 7 of which are novel. To prioritize candidate genes, we determined the one that reseeds in the linkage disequilibrium block for European population from the top significant variant of each locus. This identified 10 coding (*LPHN2, EFNA1, SLC50A1, DPM3, KRTCAP2, ISL1, TRIM29, SYT1, PAWR, GOSR2*) and 4 non coding genes (one pseudogene and three long non-coding RNA, respectively, *HMGB1P47, ISL1-DT, LINC01974,* and *LINC01716*). Since *EFNA1* is known to be strongly expressed in mouse embryonic genital tubercle, urethra and primitive bladder, we re-sequenced it in 580 CBE patients used in this study and we identified 14 rare variants in 14 independent patients. Four of these variants were novel: 2 were heterozygous missense 1 homozygous missense and 1 heterozygous loss of function variant. Parental transmission was observed in the one for which parents DNA was available for sequencing, confirming the inheritance of the 2 heterozygous missense variants.

To access the involvement of all the founded genes during bladder development we performed RNA-seq of cloaca, primitive bladder and bladder in mouse during embryonic stages E10.5, E12.5 and E15.5. Since non coding RNA are not conserved between mouse and human, we access their expression using

human fetal bladder RNA-seq from gestational week 7 to 9. In addition, to study a possible contribute in CBE bladder cancer development, we analysed the expression of these genes in RNA-seq of 5 different bladder cancer (bladder cancer, bladder squamous cell carcinoma, bladder transitional cell carcinoma, ureter urothelial carcinoma, and muscle invasive urothelial cancer) compared to healthy bladder tissue. Since the healthy bladder tissue samples is sequenced with polyA-RNA-seq, it was not possible to compare the expression of the non-coding genes.

Our investigation revealed that in mouse embryonic urogenital tissues, four of the candidate genes (*Isl1, Trim29, Syt1, Pawr*) displayed differential expression. Similarly, in human embryonic urogenital tissues, five of the candidate genes (*DPM3, ISL1, TRIM29, SYT1,* and *PAWR*) along with two non-coding genes (*HMGB1P47* and *ISL1-DT*) exhibited differential expression. As previously reported, *Isl1* and *Syt1* knock out mice, display CBE-like phenotypes.

Interesting we show that genes that were up regulated in bladder development (*TRIM29, DPM3, KRTCAP2*) are downregulated in different bladder cancer; vice versa, the ones that were downregulated in bladder development, are strongly upregulate in bladder cancers (*ISL1, SLC50A1, EFNA1, SYT1*), indicating a molecular switch of genes typical of cancer stem cell progression.

In synthesis, this study presents the largest GWAS on CBE to date and it identifies 8 genome wide significant loci 7 of which are novel. The genes that reseed in these loci shows differential expression both in bladder development and in different bladder cancers. Genes differentially expressed in bladder development have been previously shown to be involved in bladder malformations. In addition, genes up regulated in bladder development are down regulate in bladder cancer and vice versa.

1.4.2 Statement of the candidate contribution

The Ph.D. candidate in Molecular Biomedicine Enrico Mingardo (E.M.), author of this thesis and first shared author of the here presented peer-reviewed publication in Communications Biology (Nature portfolio) titled “A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy”, (DOI: https://doi.org/10.1038/s42003-022-04092-3) contributed at the manuscript as following.

E.M. worked on all the wet lab experiments and preparation. This includes: DNA extraction from saliva samples of the newly recruited patient; DNA measuring and dilution in 96 well plates; labelling, bar code assignment and storage of the newly prepared samples in the UKB genomic facility. E.M. prepared

the dilution for the EFNA1 sanger sequencing of 580 CBE patients, managed the shipment and managed the data quality control. After sequencing, E.M. was responsible for data repository in GeneBank (BankIt).

E.M. performed the quality control of the GWAS genotyped data by excluding all patients with a call rate below 0.8 and was responsible of orchestrating the Final Report between the UKB genomic facility and the bioinformatics for the association analysis.

E.M. worked on the CBE GWAS variants screening: he filtered all the variants with a p-value below e^-8 and wrote an UCSC script to have a graphic visualization on the genome for future studies. He calculated the LD block distances to prioritize the genes.

E.M. entirely worked, with initial supervision of Prof. Philip Grote, on the RNA-seq of mouse cloaca and bladder tissues. He calculated the average between replicates and the log2FoldChnagefor all the genes.

E.M. downloaded the raw data of the human fetal bladder tissues (EMBL-EBI #exptnum), of the Encicopedia of cancer cell lines, performed the TPM average and calculate the Log2FC between stages and control (for cancer). E.M. was responsible of the supervising of the correct data uploading by the partner company ImmunityBio^{TM} for the muscle urothelial carcinoma.

E.M. drafted the manuscript and organized the scheme of the article. For that he was the corresponding author through all the process of submission. E.M. orchestrated the tusks between collaborator for the manuscript revision to reply both at editors and reviewers and resubmitted the corrected draft to the journal. E.M. prepared independently all the final revision before the online publication.

E.M. designed all the graphic illustrations.

All the co-authors agreed on using this publication as part of this thesis.

1.5 ISL1 in urinary tract development

Insulin Gene Enhancer Protein ISL-1 (ISL1) is a transcription factor of 39 kilo Dalton (kDa) that belong to the LIM Hoemeobox family genes that are known to be involved in body formation and correct plan segmentation. *ISL1* is essential for the development of many tissues and organs such as limb, neurons and mesenchymal pancreas (Ahlgren et al., 1997)(Wilfinger et al., 2013) (Biemar et al., 2001). Many studies has shown that its expression plays a main role in the formation of heart and cardiomyocyte from embryonic endoderm tissue (Cai et al., 2003) (Gao et al., 2019). A switch-off mechanism of ISL1 was observed to orchestrate the differentiation of cardiomyocyte precursor cells in cardiomyocyte (ISL1 expressed) in smooth muscles cells (non-coding RNA-mediated ISL1 silecing), underlying the role of this gene in tissue specification (Plaisance et al., 2022).

The contribution of ISL1 in the development of the caudal embryo region and urinary tract is now a day poorly understood but few recent studies supported its role on the development of urinary tract. Draaken et al. reported *Isl1* expression in the developing cloaca, genital tubercle, ureteric bud and bladder of mouse embryo from stage E9.5 to E14.5 (Draaken et al., 2015a), its presence was observed in the pronephric region of developing zebrafish larvae at 56 hours post fertilization (hfp) (Zhang et al. 2017) (Figure 6). In human, fetal *ISL1* expression has been observed with RNA-seq of fetal bladder during gestation week 5 and 7 (Arkani et al., 2018). The fact that its presence plays a role in the urogenital tract formation was shown by Ching et al. using conditional *Isl1* knock out (KO) mouse in the genital mesenchyme. Those KO displayed various malformations of the urogenital system (Ching et al., 2018). Other inducible *Isl1* knock out in mouse experiments have reported abnormal development of the urethra and pointed out that ISL1 regulate directly the SHH expression in this region (Su et al., 2019).

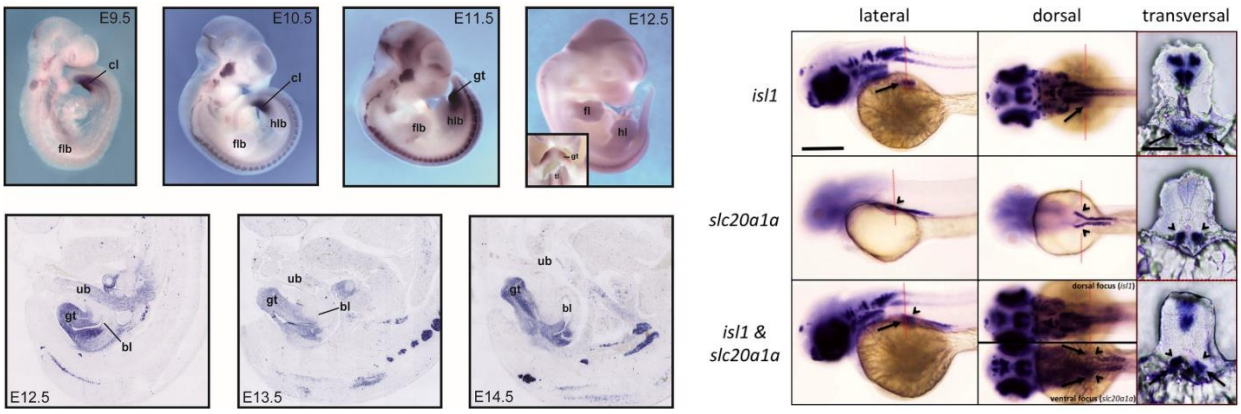


Figure 6. ISL1 is expressed in the urinary tract during embryogenesis. Left panel shows in situ hybridization of ISL1 in mouse from embryo stage E9.5 to E14.5, Isl1 shows expression in cloaca (cl), genital tubercle (gt), bladder (bl) and in the caudal tract of the ureteric bud (ub) (Draaken et al. 2015). In zebrafish embryo at 56 hours post

fertilization (hpf) (right panel), isl1 expression co-localize where the nephric duct marker slc20a1a reseeds (Zhang et al. 2017).

1.6 ISL1 as candidate region for CBE

Even though CBE is associated with multiple loci, the most significant reseeds in chromosome 5 where and highlight the *ISL1* gene. This gene seems to be involved in the early bladder development stages and its high expression on early tissue stage gets significantly downregulated during bladder development (Mingardo et al., 2022), in addition Draaken et al show higher *Isl1* presence in cloaca and urinary bladder from E9.5 to E12.5, after that its expression diminish and shift to genitalia tubercle (Draaken et al. 2015). Re-sequencing of ISL1 in a Swedish patient's cohort did not show any altering or disease-causing variant and for that suggest its pathogenic CBE role associated to gene regulation rather than variation (Arkani et al., 2018).

1.6.1 rs2303751 as candidate variant for CBE in ISL1

To identify candidate regulatory region on *ISL1*, CBE associated variants of chromosome 5 were screened in both in regulome database (regulome bd) and FORGEdb. Both gives score basing on results from deposited data on gene expression and chromatin sequencing. The regulatory potential score differs in output: for rs4865658 and rs2303751 displays the same high regulatory score in regulome bd (score of 2b) but only rs2303751 displays high score on FORGEdb (score of 8) as shown in table 1

	regulome db	FORGE db
rs2303751	2b	8
rs4865658	2b	7

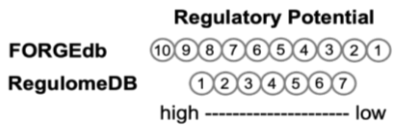


Table 1. Score of the two CBE associated regulatory variants in regulome bd and FORGEdb. Legend of the value on the right.

1.6.2 rs2303751 is target of EZH2

The regulatory variant rs2303751 reseeds in the coding region of the exon 4 of *ISL1*, but its presence does not alter the amino acid translation. Regulome db report from ENCODE a series of ChIP-seq that

identify EZH2 as the target protein of the locus comprehending rs2303751 in human embryonic stem cells (Figure 7).

Protein Binding Filter:

Method	Location	Bound Protein	Cell Type	Additional Info	Reference
ChIP-seq	chr5:50685364..50686401	EZH2	H1-hESC		ENCODE
ChIP-seq	chr5:50685211..50685995	EZH2	NH-A		ENCODE
ChIP-seq	chr5:50685364..50686375	EZH2	H1-hESC		ENCODE
ChIP-seq	chr5:50685364..50686375	EZH2	H1-hESC		ENCODE
ChIP-seq	chr5:50685436..50686106	RBBP5	H1-hESC		ENCODE

Figure 7. Regulome bd identifies EZH2 targeting rs2303751 region. Output of regulome db shows presence of EZH2 in human embryonic stem cell line H1 on the rs2303751 locus.

EZH2 is well characterize as the catalytic subunit of the Policomb Repressive Complex 2 (PRC2) responsible of the triple methylation of the Lysine 27 of Histone 3 (H3K27Me3) associated with chromatin and gene silencing (Margueron & Reinberg, 2011). Recently though, it has been demonstrate its PRC2-independent role as transcription factor directly contacting the DNA and activating gene expression (Kim et al., 2018).

1.7 The zebrafish

The zebrafish (Danio rerio) is a freshwater teleost of South-East Asian origin, belonging to the class Actinopterygii and the family Cyprinidae within the phylum Chordata. Zebrafish genes are highly conserved across other vertebrate species, with 71.4 % of human genes having at least one zebrafish orthologue and 69 % of zebrafish genes having at least one human orthologue (Howe et al., 2013) (Figure 8). This makes zebrafish an ideal vertebrate model system for studying human diseases, biology, and organs development (Driever et al., 1994).

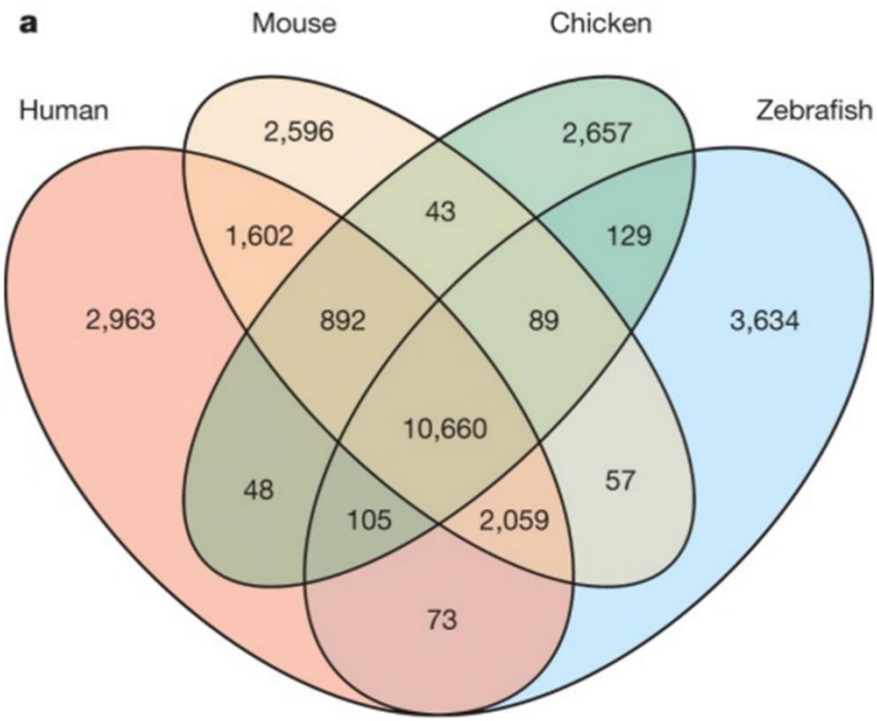


Figure 8. Orthologous genes shared between zebrafish, chicken mouse and human. Number of genes conservated between four different vertebrates, genes that are duplicated during evolution are grouped to be counted as single. (Howe et al. 2013).

Zebrafish mating generates a large number of offspring that develop rapidly ex utero, with around 100-200 eggs developing into freely swimming animals within three days from fertilization. Embryogenesis is completed after five days post-fertilization (dpf), and by this time, the larvae exhibit most of the mammalian organs (Kimmel et al., 1995). The small size and translucent nature of zebrafish during embryogenesis make them an excellent model for in vivo live imaging and whole-mount RNA and protein visualization. Therefore, various genetic modification techniques for generating KO lines have been adapted to zebrafish due to their external fertilization and development, including the Tol2 transposon system, modified Bacterial Artificial Chromosome system (BAC), GAL4/UAS binary transcription system, and knockout methodologies Zinc Fingers, TALENs, and CRISPR/Cas. Additionally, many transgenic zebrafish reporter lines were generated that express fluorescent exogenous proteins under the control of specific transcription factors to investigate tissue or cell-specific events (Choi et al., 2021). The combination of gene-specific KO with fluorescence transgenic lines makes possible to study the contribution of genes in the development of a single tissue or organ during the embryogenesis.

1.7.1 Zebrafish as model organism to study urinary tract development.

Zebrafish do not have a urinary bladder for the storage of urine but presents similar anatomical structures that work as secretion surrogates. The structure that defines the urinary tract of zebrafish is cranially defined by 2 glomeruli (Figure 9) and end caudally with pronephric ducts fused to the cloaca, the pronephric ducts are surrogates structure of what in human are the Wollfian ducts (Kolvenbach et al., 2019). To study the embryogenesis and the correlate malformation we used a zebrafish fluorescent transgenic line *Tg(wt1b:eGFP)* (Perner et al., 2007). This line was generated with transgenic insertion of the *wt1b* promoter region flanking the eGFP construct and resulted in a clear fluorescence of the gluomeruli, nephric ducts and exocrine pancreas (Figure 9) allowing studies for kidney and urinary tract development and diseases.

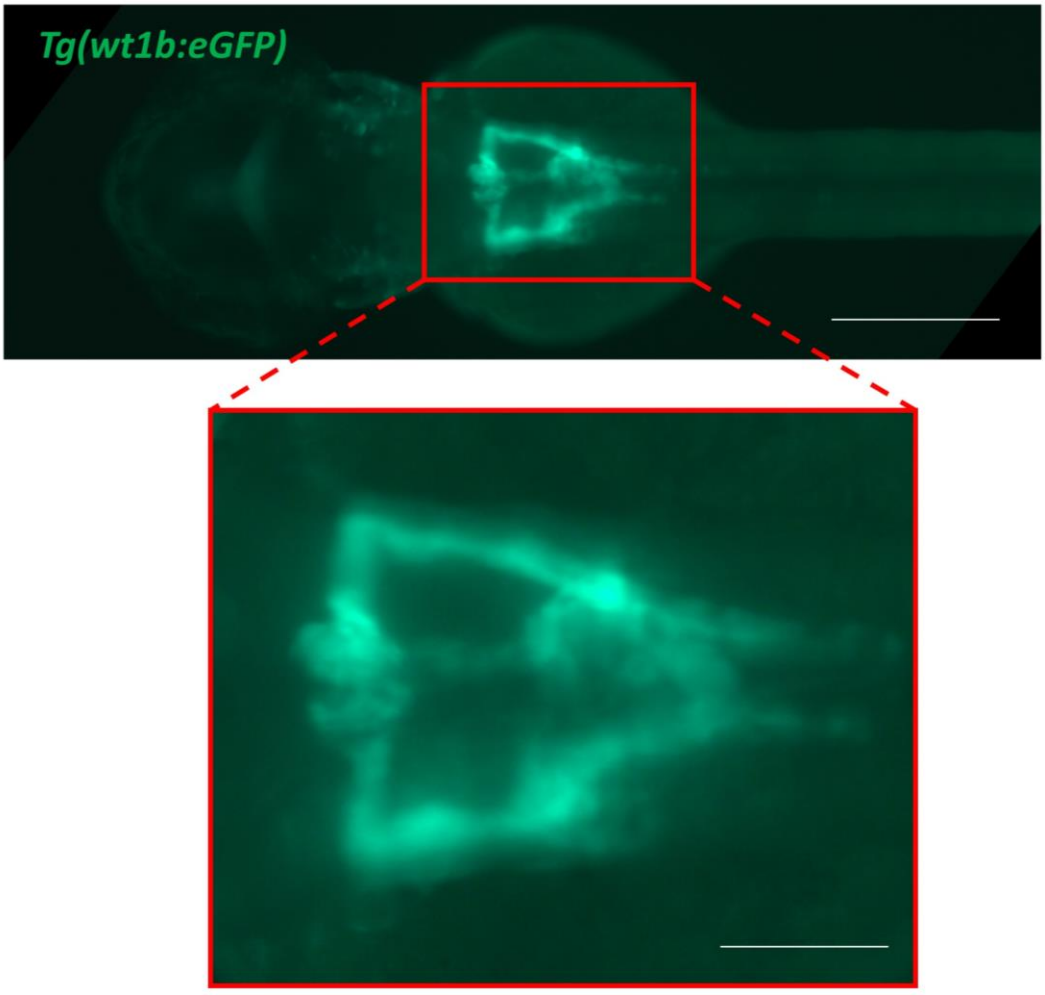


Figure 9. Fluorescent microscope image of Tg(wt1b:eGFP) zebrafish transgenic larva at 3 dpf. Red square shows the nephric structure from dorsal larva view, asterisks indicate the two glomeruli, white arrows indicate the nephric duct and red arrow indicates the exocrine pancreas. Scalebar is 300 µm for top image and 100 µm for the zoom image.

1.6 Aim of the study

Since the genetic and molecular mechanisms of CBE are brightly unknown, the work on this study aims to shade better light on the genomic loci involved in CBE and in the molecular characterization on CBE associated gene *ISL1*. For that, this study shows the largest GWAS on CBE to date, successfully identifying 8 loci, 7 of which are novel, and identifies genetic drivers combing GWAS with RNA-seq of human mouse fetal bladder. Since the molecular mechanism of CBE are unknown, this study shows for the first time the molecular mechanism of the associated *ISL1* gene in the pathology of the urinary tract development.

2. Materials.

2.1 Consumables

Consumables	Company	Product number
1.5 ml Tubes	Sarstedt	72690
10 cm petridish	Techno Plastic Product	93100
12 well plate	Techno Plastic Product	92012
15 cm petridish	Techno Plastic Product	93150
2 ml Tubes	Eppendorf	211-2120
6 well plate	Techno Plastic Product	92006
8 microTUBE-130 AFA Fiber H Slit Strip V2	Covaris	520239
96 well plate	Star Lab	19103
96 well plate	Cornig Costar	3917
Cell Scraper	Starstedt	833950
EppendorfR LoBind microcentrifuge tubes	Sigma Aldrich	Z666548
EppendorfR LoBind microcentrifuge tubes	Sigma Aldrich	Z666556
Falcon tubes 15 mL	Greiner Bio-One	188272

Falcon tubes 50 mL	Greiner Bio-One	227263
Immobilion-P Transfer Membrane	Merk	IPVH00010
Melek's whiskers	Öznur Yilmaz	-
Microloader 20 µl	Eppendorf	5242956003
Mini-PROTEAN TGX Gels	Biorad	4561084
Multiply®-µStrip Pro 8-Strip	Sarstedt	72991002
ParafilmTM	Labomedic	1447011
Pastette® Extended Fine Tip Mini	Alpha Laboratories Limited	Lw4231
Pasteur pipette, glass	Labomedic	447016
Pasteur plastic pipette	Ratio lab GmbH	2600111
PCR tubes	Starlabs	B1402-5500
Petri dish	Greiner Bio One	633180
Petri dish, small	Greiner Bio One	EL46.1
Pipette tips, 1000µl	Sartstedt	70762
Pipette tips, 10µl	Sartstedt	720031
Pipette tips, 2.5µl	Sartstedt	720025
Pipette tips, 200µl	Sartstedt	70760002
Precellys Bulk beads for 500pp Zirconium oxie beads	PEQLAB Biotechnologie	KT03961-1-103.BK
UVette	Eppendorf	30106300
Whatman paper	Biometra	GB005

2.2 Chemicals

Chemicals	Company/Sources	Product number
(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES)	VWR	441487M
10x Buffer for T4 DNA Ligase with 10 mM ATP	New Englad BioLabs	B0202S

10x Tris/Glycine/SDS	Biorad	1610772
6x DNA Loading buffer	Thermo Fisher Scientific	R0611
Agarose LE	Biozym	840004
Agarose Type IX-A, Ultra-low Gelling Temperature	Sigma-Aldrich	A2576
Ampicillin sodium	Sigma	A0166
Ampuwa water	Ampuwa	09016871/100
Buffer Tango	Thermo Fisher Scientific	BY5
Chloroform	PanReac Applichem	A1585
Cutsmart buffer	NEB	#B7204S
DEPC water	Roth	T143.3
Dimethyl sulfoxide (DMSO)	Sigma	D8418
Disodium phosphate	Merk	7558-79-4
dNTPs 100mM solutions	Thermo Fisher Scientific	R0182
DreamTaq™ buffer	Thermo Fisher Scientific	B65
Dulbecco's Phosphate-Buffered Saline (DPBS)	Gibco by Life Technolo-gies	14190-094
EDTA-disodium	Serva	39760.01
Ethanol absolute	PanReac Applichem	A1613
Ethidium bromide 1 %	Sigma Aldrich	46067
Ethyl 3-aminobenzoate methane sulfonate (MS222)	Fluka Analytical	A5040
Formamide	Sigma Life Science	47671
GeneRuler 1 kb DNA ladder 250 to 10000 bp	Fisher Scientific GmbH	SM0311
GeneRuler 100 bp plus DNA ladder	Fisher Scientific GmbH	SM0321
GeneRuler Ultra Low range DNA ladder	Fisher Scientific GmbH	SM1211
Gibco DMEM (1X) + GlutaMAX-I	Fisher Scientific GmbH	31966-021
Gibco Opti-MEM I	Fisher Scientific GmbH	11058-021

Gibco Trypsine-EDTA (0,5%)	Fisher Scientific GmbH	15400054
Glycerol	Merck	104093
Green GC Phusion Buffer	NEB	F539L
Heparin sodium salt from porcine Intestinal mucosa	Sigma-Aldrich	104093
Instant Ocean Sea Salt	Instant Ocean	SS15-10
Isopropyl alcohol	Calroth	6752.4
Lipofectamine 2000 Reagent	Thermo Fisher Scientific	11668-019
Lipofectamine RNAiMAX Reagent	Thermo Fisher Scientific	13778-075
Magnesium sulphate heptahydrate	Merck	1058861000
Methanol	PanReac Applichem	131091161
Methylene blue	Merck Darmstadt	6040
Mineral oil	Sigma	M5904
Normal Goat Serum	Sigma Aldrich	G9023
N-Phenylthiourea (PTU)	Sigma	P7629
Nuclease-free water	Qiagen	129144
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific	26616
Paraformaldehyde	Sigma-Aldrich	158127
Phenol red	Sigma-Aldrich	P0290
Phor Agarose	Biozym	850180
Phusion HF buffer	Thermo Fisher Scientific	F518
Pierce RIPA Buffer	Thermo Fisher Scientific	89900
Potassium chloride	Fluka Chemika	351861/1
Potassium dihydrogen phosphate	Merk	7778-77-0
Precision Plus Protein Dual Color Standards	Biorad	161-0374
Protease from Streoptomyces griseus XIV	Sigma-Aldrich	P5147

Proteinase K	PanReac AppliChem	A38300100
Roti-CELL 10x PBS	Roth	9150.1
Sodium chloride	AppliChem	A2942,1000
Sodium hydroxide	Merk	1310-73-2
SSC buffer 20x	Gibco by life technolo-gies™	15557-044
SuperSignal™ West Femto	Thermo Fisher Scientific	34094
TEMED	Biorad	1610800
Trans-Blot Turbo 5x Transfer Buffer	Biorad	10026938
Tris-HCl	Calroth	9090.3
TRIzol Reagent	Ambion by life technolo-gies™	T9424
Tween 20	Sigma Aldrich	P9416

2.3 Enzyme

Enzyme	Company	Product number
DreamTaq DNA polymerase	Thermo Fisher Scientific	EP0701
HOT FIREPol® Blend Master Mix Ready to Load	Solis Biodyne	04-25-00S20
MluI-HF	New Englad BioLabs	R3198L
Phusion	Thermo Fisher Scientific	F5305
T4 DNA-Ligase	New Englad BioLabs	M0202M
XhoI	New Englad BioLabs	R0146L

2.4 Kit and commercial assays

Kit/Assay	Company	Product number
BCA Preotein Assay Kit	Thermo Fisher	23225
DIG RNA Labeling Kit (SP6/T7)	Roche	11175025910

DNeasy Blood & Tissue Kit	Quiagen	69504
Dual-Luciferase® Reporter Assay System	Promega	E1910
iScript™ Reverse Transcription Supermix for qRT-PCR	Bio-Rad	1708841
iTaq Universal SYBR Green Supermix	Bio-Rad	1725121
Magna ChIP A/G	Merk	17-10085
mMessage mMachineTM T7 Ultra Kit	Thermo Fisher Scientific	AMB13455
NucleoSpin Gel and PCR Clean-UP	Macherey-Nagel	740609.250
NucleoSpin Plasmid	Macherey-Nagel	740588.250
Nucleospin RNA clean up kit	Macherey-Nagel	740948.50
QuikChange Lightning Site-Directed Mutagenesis Kit	Agilent	210518

2.5 Plasmids

Name	Size (bp)	Company/Deposition	Product number
(SK-)pBluescript_Isl1		Research group of Prof. Benjamin Odermatt, Institute of Anatomy and cell biology, University of Bonn, Germany	
pGL3-Basick	4818	Promega	E1751
pRL-SV40	3705	Promega	E2231
pCMVHA hEZH2	8805	Addgene	24230
pGL3-Isl1_Fr1_for	6116	Research group of Prof. Benjamin Odermatt, Institute of Anatomy and cell biology, University of Bonn, Germany	-
pGL3-Isl1_Fr1_FLP	6118	Research group of Prof. Benjamin Odermatt, Institute of Anatomy and cell biology, University of Bonn, Germany	-
pGL3-Isl1_Fr2_for	6076	Research group of Prof. Benjamin Odermatt, Institute of Anatomy and cell biology, University of Bonn, Germany	-
pGL3-Isl1_Fr2_FLP	6090	Research group of Prof. Benjamin Odermatt, Institute of Anatomy and cell biology, University of Bonn, Germany	-
pGL3-Isl1_Fr3_for	6079	Research group of Prof. Benjamin Odermatt, Institute of Anatomy and cell biology, University of Bonn, Germany	-

pGL3-Isl1_Fr3_FLP	6080	Research group of Prof. Benjamin Odermatt, Institute of Anatomy and cell biology, University of Bonn, Germany	-

2.6 Oligonucleotide

2.6.1 Oligonucleotide for zebrafish genotyping

Name	Sequence (5' to 3')
zf_EZH2_Talen_F	AAATCGGAGAAGGGTCCTG
zf_EZH2_Talen_R	ACACACATGCAACTGGACTC

2.6.2 Oligonucleotide for cloning

Name	Sequence (5' to 3')	Company
isl1_1.2kb_XhoI_FLP_F	catcatCTCGAGGACTTTGAGACCTGCTTCCCTTG	Sigma Aldrich
sl1+6.3_1.2kb_MluI_FLP_R	catcatACGCGTTAACTTCACCAGGAGGCCTGC	Sigma Aldrich
isl1+6.3_KpnI_F	catcatGGTACCGACTTTGAGACCTGCTTCCCTTG	Sigma Aldrich
isl1+6.3_XhoI_R	catcatCTCGAGTAACTTCACCAGGAGGCCTGC	Sigma Aldrich
Primer 8_XhoI_F	tcatACGCGTAGGTGTTGGCCTGACCCTAGGG	Sigma Aldrich
Primer 9_MluI_R	tcatCTCGAGGTTTGCGGCGTAGCAGGTCCG	Sigma Aldrich
Primer 8 FLP_MluI_F	tcatCTCGAGTAGGTGTTGGCCTGACCCTAGGG	Sigma Aldrich
Primer 9 FLP_XhoI_R	tcatACGCGTGGTTTGCGGCGTAGCAGGTCCGC	Sigma Aldrich
Primer 17_XhoI_F	tcatACGCGTTGCCCCTCATCCTTACCCCC	Sigma Aldrich
Primer 18_MluI_R	tcatCTCGAGTCGTGTCTCTCTGGACTGGCAG	Sigma Aldrich
Primer 17 FLP_MluI_F	tcatCTCGAGTGCCCCTCATCCTTACCCCC	Sigma Aldrich
Primer 18 FLP_XhoI_R	tcatACGCGTGTCGTGTCTCTCTGGACTGGCAG	Sigma Aldrich

2.6.3 Oligonucleotide for plasmid sequencing

Name	Sequence (5’ to 3’)	Company
pGL3_seq_F	CTAGCAAAATAGGCTGTCCC	Sigma Aldrich
pGL3_seq_R	CTTTATGTTTTTGGCGTCTTCCA	Sigma Aldrich

2.6.4 Oligonucleotides for RT-qPCR

Name	Sequence (5' to 3')	Company
ACTB_Hum_1F	CTTCCTTCCTGGGCATGGAG	Sigma Aldrich
ACTB_Hum_1R	AGCACTGTGTTGGCGTACAG	Sigma Aldrich
ISL1 qpcr F2	CAGCAACTGGTCAATTTTTCAG	Sigma Aldrich
ISL1 qpcr R2	CTCAATAGGACTGGCTACCATG	Sigma Aldrich
isl1_zf_qPCR_F	CGTGTTTGAAATGTGCAG	Sigma Aldrich
isl1_zf_qPCR_R	CCGTATAACCTGATGTAGTC	Sigma Aldrich
zf actb1 qpcr for.	GACACAGATCATGTTCGA	Sigma Aldrich
zf actb1 qpcr rev.	GCGTAACCCTCATAGATG	Sigma Aldrich
AC010478-201-2-3_F	GTGTGCTGACCCAAGTGGTG	Sigma Aldrich
AC010478-201-2-3_R	TGTTTCACTCTCCGGACTGC	Sigma Aldrich

2.6.5 Oligonucleotide for gene knock down

Name	Sequence (5' to 3')	Company
2146-EZH2	targeting NM_001203247, NM_001203248, NM_001203249, NM_004456, NM_152998, XM_005249962, XM_005249963, XM_005249964, XM_011515883, XM_011515884, XM_011515885, XM_011515886, XM_011515887, XM_011515888, XM_011515889, XM_011515890, XM_011515891, XM_011515892, XM_011515893, XM_011515894	siTOOLs BIOTECH

siPOOL control	siTOOLs BIOTECH

2.6.6 Oligonucleotide for ChIP-qPCR

Name	Sequence (5' to 3')	Company
ChIP -0,5 F	GAGCAGGGATTGGAGATATGGC	Sigma Aldrich
ChIP -0,5 R	CGAGAACTCTGCCAGAACGC	Sigma Aldrich
ChIP +0 F	TGCTGTGAACAGGGGGACAG	Sigma Aldrich
ChIP +0 R	CTTTGCTGAGTAATCCCGGCC	Sigma Aldrich
ChIP +0,3 F	GCCTCCAGCCCAGCGCTCAC	Sigma Aldrich
ChIP +0,3 R	TTCTCCGGCTGCTTGTGGACG	Sigma Aldrich
ChIP +0,6 F	AGCAGCAGCAGCCCAATGAC	Sigma Aldrich
ChIP +0,6 R	CCTGCGTACCAGGAACGCAC	Sigma Aldrich
ChIP +0,9 F	GAACCGGAGAAACGCCGTCC	Sigma Aldrich
ChIP +0,9 R	TCCCTCTTCTTGTGTACGTGAG	Sigma Aldrich
ChIP +1,2 F	AGATCACCCTCTGCTCCAGG	Sigma Aldrich
ChIP +1,2 R	CTTCACCAGGAGGCCTGCAG	Sigma Aldrich
ar F	CACAGGCTACCTGGTCCT	Sigma Aldrich
ar R	TCTGGGACGCAACCTCT	Sigma Aldrich
ACTB_ChIP_F	TCTGAACAGACTCCCCATCC	Sigma Aldrich
ACTB_ChIP_R	ACCATGTCACACTGGGGAAG	Sigma Aldrich

2.7 Primary antibody

Antibody	Spiecie	Company	Product number	Type
Beta Actine-peroxidase	mouse	Sigma Aldrich	A3854	primary
EzH2 (Human)	rabbit	invitrogen	49-1043	primary

GFP	rabbit	Invitrogen	A11122	primary
Islet1 (EP4182)	rabbit	Abcam	ab109517	primary
Mouse IgG	mouse	Diagedone	C15400001-15	primary

2.8 Secondary antibody

Antibody	Spiecie	Company	Product number	Type
Anti-rabbit Alexa Fluor 488	goat	Life Technologie	A11034	seconday
Anti-rabbit-HRP	goat	Kackson Immuni Research	111-035-144	seconday

2.9 Zebrafish lines

Name	Source
TU wildtype fishline	EZRC, KIT
TL wildtype fishline	EZRC, KIT
AB wildtype fishline	EZRC, KIT
Brass wildtype fishline	EZRC, KIT
Tg(wt1b:eGFP)	(Perner et al., 2007)
ezh2^{ul2}	(Dupret et al., 2017)

3. Methods

3.1 Cell culture

Human embryonic kidney 293 cells (HEK293) a cultured in Dulbecco's Modified Eagle Medium (DMEM) containing 10 % fetal bovine serum and without any antibiotic in incubator at 37 °C with 5 % CO_2. Cells

were split when reaching 90-95% of confluency of a 10 cm petri dish containing 10 mL medium as following:

- Medium is aspirated with a sterile pump and the cells are washed with 10 mL of 37 °C PBS 1X solution. The PBS get aspirated and 1 mL of tripsine/EDTA 1X solution is added to the cell layer and subsequently incubated for 5 min at 37 °C.
- To neutralize the tripsine reaction, 9 mL of medium are added to the cells reaching a total volume of 10 mL.
- Subcultures are set in 3 different splitting dilutions of 1:10; 1:5; 1:2 using respectively: 1 mL of trispsine/EDTA-separated cells in 9 mL medium; 2 mL of trispsine/EDTA-separated cells in 9 mL medium; 5 mL of trispsine/EDTA-separated cells in 5 mL medium. The remaining 2 mL of trispsine/EDTA-separated cells are discarded or liquid nitrogen frozen for back-up.

3.2 Cloning and library preparation

3.2.1 PCR amplification of genomic fragments

Genomic DNA (gDNA) was extracted from HEK293 using the DNeasy Blood & Tissue Kit according to manufactural procedures and 3 regions flanking or comprehending the region of interest on *ISL1* where rs2303751 reseeds were PCR amplified using primers with MluI and XhoI 5' sequences for cloning. All the 3 regions are PCR amplified in order to obtain forward and flipped orientation inserts, generating in total 6 different fragments (Fragment 1 Forward; Fragment 1 Flipped; Fragment 2 Forward; Fragment 2 Flipped; Fragment 3 Forward; Fragment 3 Flipped) (Figure 9).

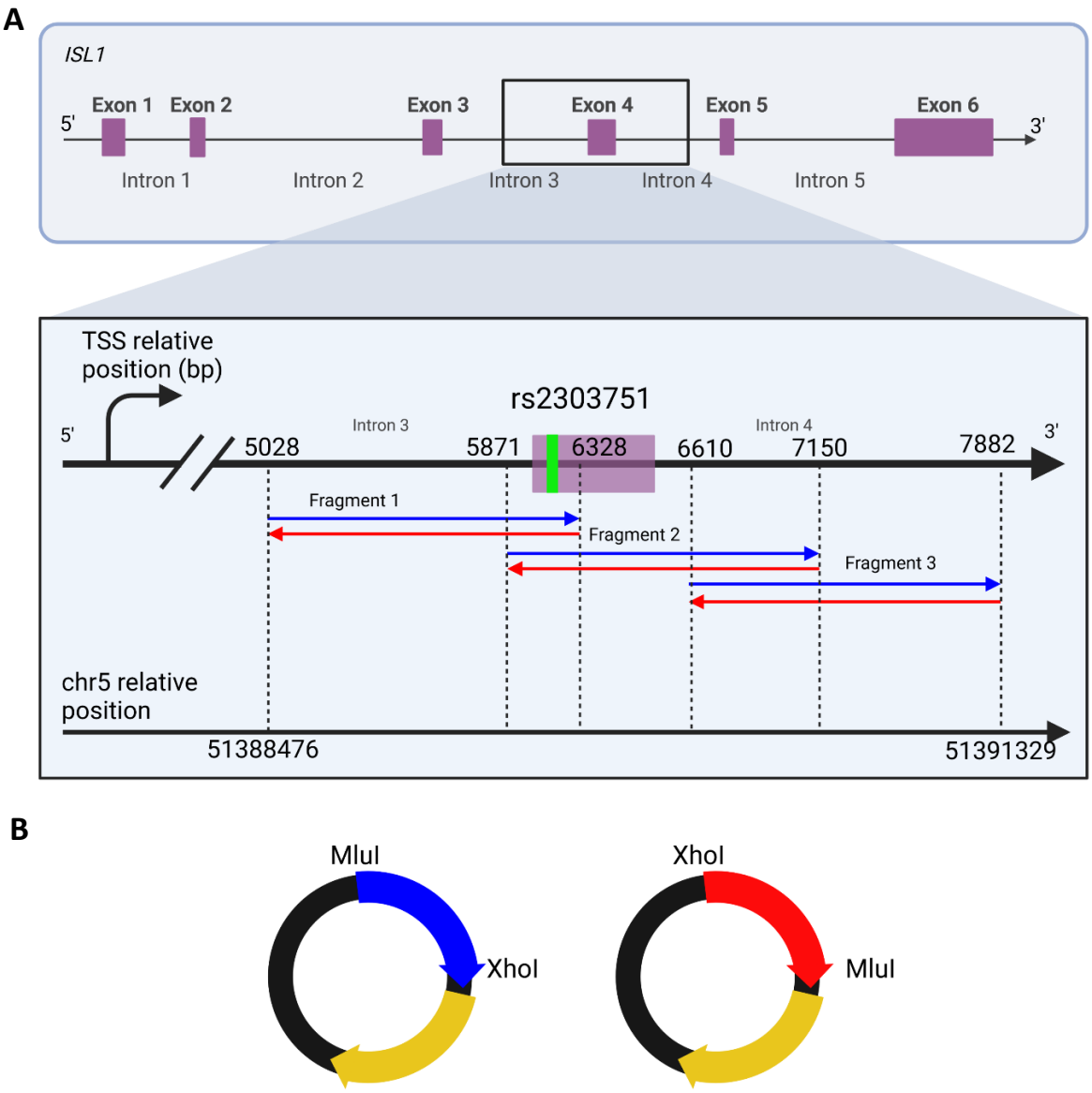


Figure 10. Overview of the genomic region referred to ISL1 and chromosome 5 cloned in the luciferase pGL3-basic plasmid. A) Top panel shows ISL1 gene and in the box the region where Fragment 1, 2 and 3 are located. In detail, coordinates (hg38) are shown relatively to the distance from the ISL1 TSS (top) and to chromosome 5 (lower). Blue and red arrows indicate the fragment orientation used for cloning. B) schematic representation of the fragments cloned in pGL3-basic luciferase plasmid for blue (forward-oriented) and red (flipped-oriented) Fragments. In yellow, luciferase gene.

Touch down PCR is performed with Phusion polymerase adding DMSO and gDNA according to manufactural procedure, touch down cycles are performed with annealing temperature from 69 to 55 °C as following:

Step	Temp. °C	Time (min:sec)
1 denaturation	98	00:30
2 denaturation	98	00:10
3 annealing	69-55	00:10
4 elongation	72	01:40
5 hold	4	∞

Step 3 is repeated 2 times for each annealing temperature with a decrease of 2 °C per cycle untill reaching 55 °C, this last cycle is repeated 20 times.

PCR products are purified with NucleoSpin Gel and PCR Clean-UP according to manufactural procedures.

3.2.2 Enzymatic digestion

The digestion of PCR amplicons and 1500 ng of pGL3 basic vector was performed as following:

COMPONENT	30 µL reaction
XhoI	1.5 µL
MluI	1.5 µL
DNA	x µL
CutSmart Buffer	3 µL
H2O	to 30 µL

Digestion is carried out at 37 °C for 1 hour.

Digested inserts and plasmid are cleaned up with a 0.7% gel electrophoresis running at 80 V for 1 hour with 1 mA current. Inserts and digested pGL3 vector were gel purified using NucleoSpin Gel and PCR Clean-Up according to manufactural procedure.

3.2.3 Ligation

Ligation was performed with a ratio backbone: insert of 1: 3 using 70 ng of pGL3 vector and 52.5 ng of insert. Amount of plasmid and insert are calculated in NEBioCalculator® online tool assuming the vector to be 4800 bp and the insert 1200 bp length.

Ligation is carried out at room temperature for 1 hour preparing the following mix:

COMPONENT	40 µl REACTION

T4 DNA Ligase Buffer (10X)	4 µl
Vector DNA	70 ng
Insert DNA	52.5 ng
Nuclease-free water	to 40 µl
T4 DNA Ligase	2.5 µl

3.2.4 Transformation and plasmid extraction

Transformation is carried out by incubating for 10 minutes 2 uL of ligation product with 50 uL of competent XL1-blue E. Coli strain at ice-temperature. Afterwards, bacteria and ligation product were heat shocked at 42 C for 40 seconds and immediately cooled down on ice for 10 minutes. After that, the heat shocked bacteria are plated in LB-agarose plates with 50 µg/mL Ampicillin and incubated overnight.

5 colonies per plate are picked up the following days and made growth in 5 mL LB liquid medium with 50 µg/mL Ampicilin for 24 hours. Plasmids are extracted with Nucleospin Plasmid kit according to manufactural procedure. The extracted plasmids were tested with PCR for the presence of insert using the same primers for cloning and sequenced both forward and reverse with primers pGL3_seq_F and pGL3_seq_R.

3.3 Dual Luciferase assay

Luciferase assay allows to detect functional area of the genome and quantify their activation testing specific transcription factors. The assay is based on the catalyzation of D-luciferin, ATP and O_2 in Oxyluciferin, pyrophosphate (PPi), AMP, CO_2 and a flash of light by the enzyme luciferase in presence of Mg^{2+}. The concept is to clone the synthetize a reporter luciferase vector where the enzyme is downstream to a putative promoter region and, when the promoter is activated by a transcription factor, luciferase get expressed and able to catalyze the flash of light reaction. The more the promoter is active, the more luciferase is expressed and the more flash lights will be emitted.

The luciferase assay requires (i) to seed a specific number of cells in a well plate (in this case a 24 well plate), (ii) an incubation time that allows cells to attach and grow, (iii) cell transformation with the luciferase reporter vector and luciferase normalization vector (iv) cell lysis and luciferase assay.

3.3.1 Seeding specific number of cells

Technical triplicate of luciferase assay was done in 24 well plates where 1x10^5 cells in 500 µL DMEM per well were plated to reach confluence in 48 hours. For that, a seeding solution with an excess of 3 wells is prepared containing

$$(1 * 10^5) * (24 + 3) = 27 * 10^5 \ cells$$

In

$$500 * (24 + 3) = 13500 \ \mu L \ DMEM$$

The seeding solution comes from an HEK293 confluent 10 cm petridish culture. Here cells were tripsinated with 1mL of tripside/EDTA and 9 mL of DMEM are added to stop tripsination. The solution is then centrifuged at 300 rpm for 5 minutes, the supernatant discarded and the pellet of cells resuspended in a 1 mL of DMEM. To prepare the cell for counting, 10 µL of the resuspended pellet were stained in 90 uL of coomassie blue to distinguish livable cells from dead ones. To count cell 10 µL of the staining were placed in a Neubauer chamber and cells are counted under an inverted microscope in each of the four 1 mm^2 section. The total amount of cells $\boldsymbol{x}$ was counted as following:

$$\frac{n_1 + n_2 + n_3 + n_4}{4} = x * 10^5 \ cell/mL$$

Where n is the number of cells counted in a single square and 1, 2, 3 and 4 the number of square.

$27 * 10^5 \ cells$ are taken from the resuspended pellet in:

$$\frac{27 * 10^5 \ cells}{x * 10^5 \ cell/mL} = y \ mL$$

Seeding solution is prepared adding $y \ mL$ of cells in $(13500 \ \mu L \ DMEM - y \ mL)$.

Seeding is performed adding 500 µL of seeding solution in each of the 24 wells. Plates were incubated overnight.

3.3.2 Transfection

In each well a plasmid solution of 500 ng of reporter Firefly luciferase plasmid (pGL3 with Fragment 1, 2 and 3 forward and flipped) with 50 ng of Renilla luciferase normalization plasmid (pRL-SV40) is co-transfected with or without pCMVHA hEZH2. The amount of plasmid for the technical triplicate is calculated as following:

	Single well (ng)	duplicate (ng)	triplicate (ng)
Reporter Firefly (pGL3-Fragment #)	500	1000	1500

Renilla (pRL-SV40)	50	100	150
pCMVHA hEZH2	100	200	300

Plasmid solution is prepared for transfection by mixing the 1500 ng of reporter firefly to 150 ng of pRL-SV40 and with or without 100 ng of pCMVHA hEZH2.

Transfection master mix for 24 wells is prepared mixing 560 µL of OPTI-mem with 40 µL of Lipofectamine reagent 2000 (final volume 600 µL).

The triplicate solution is prepared mixing the plasmid solution with 75 µL of transfection master mix. After 10 minutes incubation at room temperature, 25 µL of the triplicate solution is added in each of the 3 wells. Cells are then incubated for 24 hours.

3.3.3 Luciferase assay and data normalization

The luciferase assay is carried out according to manufactural procedure, the light of renilla and firefly is automatically detected in a luminometer according to manufactural procedure.

For each well the luminescence of the firefly is normalized with the one of renilla and the final value consists on the average of the technical triplicate as described here below:

$$\frac{\left(\frac{firefly\ intensity_1}{renilla\ intensity_1} + \frac{firefly\ intensity_2}{renilla\ intensity_2} + \frac{firefly\ intensity_3}{renilla\ intensity_3}\right)}{3}$$

Where 1, 2 and 3 are the wells of the technical triplicate.

This procedure is repeated for at least 3 times to gain a biological replicate of at least N = 3.

3.4 Chromatin immunoprecipitation quantitative PCR (ChIP-qPCR)

Chromatin immunoprecipitation followed by qPCR is a technique that allow to quantify weather a protein of interest binds to a specific region of the genome without resorting to DNA sequencing. This method is based on cross-linking the cell with formaldehyde, sharing the chromatin to a specific fragment size, immunoprecipitating the protein-DNA fragments with antibody specific for the protein of interest, isolation and purification of the DNA fragments and qPCR to analyze the presence and the quantity of the fragments that were bandied to the protein of interest (Figure 10)

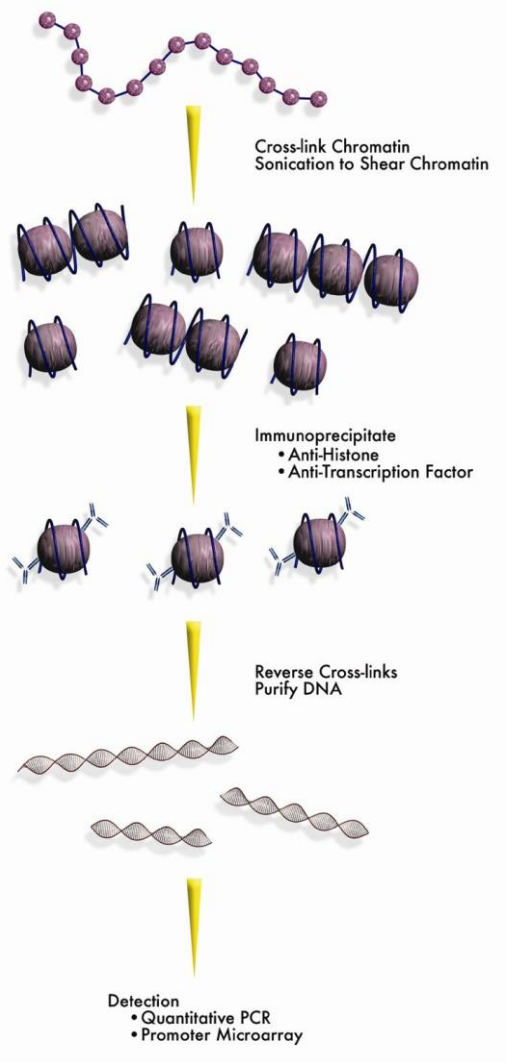


Figure 10. Schematic overview of the chromatin immunoprecipitation qPCR protocol. From top to bottom: chromatin is cross linked and fixed with formaldehyde; DNA-protein complex is shared with sonication in small fragments; immunoprecipitation is performed with specific antibodies, chromatin is purified, detection is performed via qPCR. Image from MERK.

Chromatin immunoprecipitation is performed with the Magna ChIP A/G kit with small adjustment from the manufactural procedure. For this study, 1.5x10^7 cells were fixed in 1% formaldehyde for 10 min at room temperature and chromatin was shared with sonication using a Covaris LE220. Sonication condition were adjusted to reach a chromatin fragment average size of around 200-300 bp. Shared chromatin is immunoprecipitated with two antibodies:

- Anti-EZH2 antibody: to test the presence of this protein to the region of interet
- Anti-mouse IgG: control for non-binding or unspecific binding of the anti EZH2 antibody

After shared chromatin purification, the immunoprecipitated fragments with the anti-EZH2 and anti-mouse IgG antibody are tested via qPCR. Multiple primers are designed on ISL1 to investigate a region

of 3 kb comprehending the rs2303751 locus. In addition, a positive and negative controls are set up for the binding of EZH2:

- Primers specific for the androgen receptor promoter as positive control. EZH2 has been shown to bind this region in HEK293, positively activating its transcription (Kim et al., 2018)
- Primers specific for actin beta exon 2 as negative control for the presence of EZH2.

For each specific pare of primers, qPCR is performed as technical triplicate on both the anti-EZH2 and anti-mouse IgG precipitated fragments. The fold enrichment is calculated with the $\Delta\Delta C_t$ using as housekeeping control the anti-mouse IgG C_t. qPCRs are repeated in 3 different cross linked HEK293 to gain a biological triplicate.

3.5 EZH2 knock down

RNA interference is used to knock down EZH2 in HEK293 cells. For that, silencing RNA (siRNA) and control were purchased from siTOOLs BIOTECH (siPool). HEK293 cells are reverse transfected in 6 well plates with a final siPOOL concentration of 3 nano molar (nM) according to manufactural procedure with a reduced seeding density.

Reverse transfection consists on seeding the cells together with the siRNA, this technique yields to higher transfection efficiency and allows longer assays duration. Cell seeding number and cell seeding volume for 6 well plate is shown in tabel 1:

volume calculated for:	plate type	final volume per well (µL)	total cell seeding number in stock solution	cell seeding density (cell/mL)	Cell suspension volume to seed per well (µL)
singe well	6	2000	250 000	166 666	1500
triplicate + 1 excess	6	2000	1 000 000	166 666	1500

Table 1. HEK293 cell seeding number, volume and density seeded in a 6 well plate. Cells are counted as described in paragraph 3.3.1, a stock solution for seeding cells is prepared counting triplicate + 1 excess.

According to manufactural procedure a stock solution of 0,15 µM of EZH2 siPOOL and control are prepared to reach a final concentration of 3 nM for the transfection in the well. Volumes and concentration are resumed in tabel 2.

6 well plate. Volume calculated for:	final volume single well (µL)	Optimem for siPool dilution (µL)	0,15 µM siPool for 3 nM final concentration (µL)	Optimem for RNAiMax dilution (µL)	RNAiMax (µL)	Total transfection mix (µL)	transfection mix per well (µL)
singe well	2000	210	40	246	4	500	500
triplicate + 1 excess	2000	840	160	984	16	2000	500

Tbale 2. Transfection parameter in HEK293 cells. Values are volumes of Optimem, siPOOL 0,15 µM, RNAiMax and transfection for single well and triplicate + 1 excess.

Seeding and reverse transformation are performed with the following steps:

- siPOOL dilution: Optimem for siPOOL dilution (column 3 table 2) is mixed and vortexed with 0,15 µM siPool for 3 nM final concentration (column 4 table 2)
- RNAiMAX dilution: Opti-MEM for RNAiMax dilution (column 5 table 2) is mixed with RNAiMax. (column 6 table 2).
- Transfection mix: siPOOL dilution and RNAiMAX dilution are combined in a 1:1 ratio, mixed and incubated 15 minutes at room temperature.
- Transfection mix (500 µL) is added to the bottom of each well
- Cell suspension volume to seed per well (column 6 table 1) is added to the well on top of the transfection mix.

Well plate is gently mixed and incubated for 48 hours before functional assays.

3.6 Western blot

To test the efficiency of the siPOOL knock down and the overexpression of EZH2, HEK293 cells are collected for western blot after 48 hours of transfection. Treated HEK293 cells were collected with a cell scraper using 0.3 mL of RIPA buffer containing Protease Inhibitor Cocktail for each well of 6 wells according to manufactural procedure. To quantify the amount of extracted protein, the lysate was measured with Pierce BCA Protein Assay Kit according to manufactural procedure. Western blot is performed using 30 µg of protein diluted in Lemmli buffer and loaded in a 4 % - 15 % gradient mini-protean TGX gel. After electrophoresis run, proteins were transferred in a PVDF membrane with a Trans Turbo Blot (BioRad). Membrane was blocked in 1x TBS with 0.1 % Tween 20 and 5 % milk powder solution for 2 hours, cut to separate EZH2 and beta actin size and then incubated at 4 C with primary EZH2 anti body (1:1.000) and anti-beta actin antibody (1:50.000). Bands were visualized using enhanced chemiluminescence reaction according to manufactural procedure.

3.7 Zebrafish maintenance

The laboratory usage of zebrafish in the described study followed guidelines outlined by Westerfield (2007). Adult zebrafish were maintained at a temperature of 28 °C with a light/dark cycle of 14/10 hours in the zebrafish facility of the Institute of Anatomy at the University of Bonn. To mate zebrafish, a pair of male and female zebrafish were transferred and separated by a divider in a mating tank the evening before. The dividers were removed at 9 am the next morning to allow for fertilization of the eggs and obtain embryos with similar hour of fertilization.

After collecting the fertilized eggs, embryos were reared at 28°C in 0.3x Danieau's buffer supplemented with 0.00001 % methylene blue solution until 24 hours post-fertilization (hpf). From 24 hpf, embryos were reared in 0.3x Danieau's buffer, or when used for imaging, in 0.3x Danieau's buffer supplemented with 0.003 % phenylthiourea (PTU) to prevent pigmentation.

All experiments were conducted in accordance with Belgium and European laws, and ethical commission protocols ULg1076 and ULg624 were followed.

3.8 Zebrafish genotyping

Ezh2 knock out line is obtained from the research group of Prof. Pierre-Oliver Angrand at the University of Lille, CNRS, Inserm, CHU Lille, UMR9020-U1277 - CANTHER - Cancer Heterogeneity Plasticity and Resistance to Therapies, F-59000 Lille, France. The line is obtained using the using the transcription activator-like effector (TALE) nuclease (TALEN)-based technology (Boch, 2011). The construct was designed to target the ezh2 exon 2 and resulted in an insertion of 22 base pare causing a frameshift and an appearance of a premature stop codon (Dupret et al., 2017). The *ezh2* knock-out (KO) zebrafish genotype (*ezh2*$^{ul2-/-}$) is not viable after 12-13 days post fertilization (dpf), for that it is not possible to obtain a stable adult knock out from which the genotype is transmitted by mating (Dupret et al., 2017). To study the KO effects, adult heterozygous *ezh2*$^{ul2+/-}$ are crossed and the offspring is genotyped to screen for the 22 bp insertion that characterize the *ul2* line. The genotyping described by Dupret et al. 2017 involves a PCR followed by enzymatic digestion of the amplicon to distinguish WT from heterozygous and homozygous. Since the caring out of this procedure for a large number of offspring would result in a high time-consuming protocol, we set up a screening based on the fragments size length of the PCR product from TALEN-targeted EZH2 exon 2. This protocol, based on DNA extraction, PCR and gel electrophoresis allows to distinguish the genotypes of 96 larvae in around 9 hours (Figure). Primers are designed to amplify a 199 bp region encompassing the 22 bp TALEN insertion on EZH2. Genotypes are distinguished as following:

- WT: single band at 199 bp
- Heterozygous (ezh2 +/-): double band, one at 199 bp and one at 221

- Knock out (ezh2-/-): single band at 221 bp

This procedure starts by enzymatic dechorionation of 40 hpf zebrafish larvae with Tripsine, when embryos are hatched, they are anesthetized with Tricaine solution. For each embryo, a fragment of caudal tissue was surgically severed with a scalpel under an optical microscope. After this docking, the larva is collected and placed with 300 µL of Daniau water without Tricaine in a well of a 96-well plate starting from position A1. The caudal fragment of that larva is collected with 15 µL of 5 mM NaOH solution using a P20 pipette and placed in the same well coordinates of a 96-well plate for PCR. This procedure is repeated for all the 96 larvae, after that, the plate with the docked larvae is placed in an incubator at 28 °C during all the time for the downstream genotyping processes to allow the proper development of embryos. Genomic DNA is extracted by heating the 15 µL of NaOH solution containing the tissue for 20 min at 90 °C in a thermocycle and cooled down at room temperature. The solution is neutralized by adding 1.5 µL of 10 mM Trish-HCL pH 8.5 (Meeker et al., 2007).

PCR mix per well is prepared as following:

COMPONENT	Concentration	µL
HOTFIRE Pol master mix	5X	4
zf_EZH2_Talen_F	10 mM	0,4
zf_EZH2_Talen_R	10 mM	0,4
H2O	-	13,2
Total mix per reaction		18

PCR is performed adding 2 µL of genomic DNA solution per well reaching a final volume of 20 µL. The cycles are set as:

Step	Temp. °C	Time (min:sec)
1 Taq activation	95	12:00
2 denaturation	95	00:30
3 annealing	62	00:30
4 elongation	72	00:30
5 go to 2 x32 times		
6	72	5:00
7	4	∞

PCR products are run in a 2.7% Phor agarose gel for 2 hours at 60 V. The genotype screening is based on the fragments amplified: one lower band corresponds to WT, double bands to heterozygous line and one upper band to knock out (Figure 11)

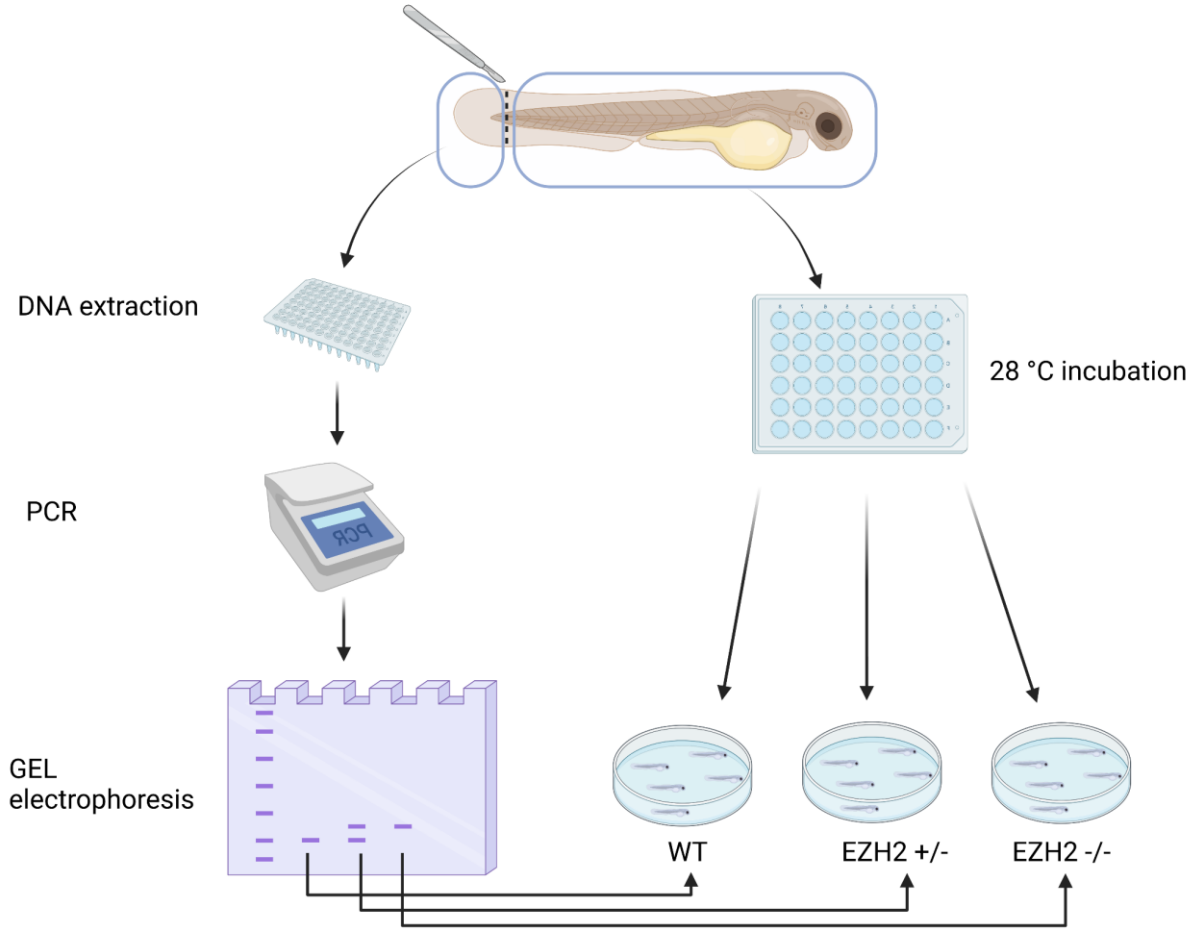


Figure 11. Genotyping protocol for embryo at 40 hpf of the offspring derived from ezh2$^{ul2+/-}$ line. Caudal tissue is dissected from a zebrafish embryo and placed in a 96 well plate for DNA extraction, while the docked larva is placed in a plate and raised at 28 C. PCR is performed on gDNA and a gel electrophoresis allow to distinguish the genotyping of the larvae. Embryos are then separate according to their genotype for future analysis.

Segregation of alleles displays usually between 18 to 25 individuals resulting WT or knock out and around 50 heterozygous.

3.9 quantitative PCR (qPCR)

In this study qPCR are performed both to check gene expression on RNA level and in DNA from shared chromatin. The first is done in HEK293 and zebrafish embryos RNA and requires retro transcription of RNA in cDNA, while the latter is directly performed on shared DNA.

3.9.1 Retrotranscription quantitative PCR (RT-qPCR).

For HEK293 cells, 1 mL of Trizol reagent was added in confluent culture ($1.5*10^6$ cell/well) trated with EZH2 siPOOL and control and RNA was extracted according to manufactural procedures. For zebrafish, 20 larvae were harvested after genotyping and RNA was extracted adding 1mL of Trizol. Whole or trunked larvae tissue was mechanically disrupted in the Trizol solution with a Percellys 24 tissue homogenizer (peqlab) using 3 cycles of 10 seconds at 6000 rpm. After that, RNA was extracted as manufactural procedures as for HEK293.

Both for HEK293 and zebrafish larvae, cDNA was synthetized with 1µg of total RNA using iScript™ Reverse Transcription Supermix for qRT-PCR according to manufactural procedures with the following incubation cycles:

For qPCR, 100 ng of cDNA was used for expression analysis with iTaq Universal SYBR Green Supermix (Bio-Rad, Cat. No. 1725121). qPCR was performed in a CFX96 Bio Rad reader with following cycles:

Step	Temp. °C	Time (min:sec)
1 denaturation	95	00:30
2 denaturation	95	00:05
3 annealing and elongation	65	00:30
4 reading		
5 go to 2 x39 times	+0.5/cycle	
6	65	0:45
7	4	∞

10 min of initial denaturation at 95°C followed by 40 amplification cycles consisting of 10s at 95 °C and 1 min at 60 °C.

Reagent assembly for one reaction (10 µL) is set as following:

	Volume (µL)
SYBR green mix	5
Primer F 10 µM	0,3
Primer R 10 µM	0,3
cDNA	1
H2O	3,4

To study the expression of ISL1 data were normalized for beta-actin reporter gene; fold change is calculated using $\Delta\Delta C_t$ normalization.

3.9.2 Quantitative PCR on shared chromatin (qPCR)

qPCR is performed according to manufactural procedure using the following cycles:

Step	Temp. °C	Time (min:sec)
1 denaturation	94	03:00
2 denaturation	94	00:20
3 annealing	59	00:30
4 extension and reading	62	00:30
5 go to 2 x32 times		
6	72	00:30
7	4	∞

Reagent assembly for one reaction (25 µL) is set as following:

	Volume (µL)
SYBR green mix	12,5
Primer F 10 µM	0,8
Primer R 10 µM	0,8
ChIP DNA	2
H2O	8,9

Data and results are normalized to the mouse IgG imputed chromatin; fold change is calculated using $\Delta\Delta C_t$ normalization.

3.10 In situ hybridization

Embryos were gained by natural *ezh2*^{ul2} line spawning and genotyped according to paragraph 3.7., fixed in 4% PFA at 56 hpf and gradually dehydrated in 100% methanol followed by 24 hours incubation at 4 °C. The RNA *in-situ* probe was obtained from the SK(-)pBluescript_isl1_5UTR with the DIG RNA Labelink Kit (SP6/T7) according to manufactural procedure. Embryo were permeabilizated for the probe incubation by rehydration washes in PBT (PBS/tween 0.1%) and digestion with a 10 µg/mL PBS proteinase K solution for 30 min. The pre hybridization consists in incubating the embryos for 2 hours at 70 °C in 1 mL of hybridization mix (HM) that consists of:

	Volume 50 ml	Final concentration
Formamide	25,0 ml	50% formamide
20x SSC	12,5 ml	5x SSC

Heparin 5 mg/ml in DEPC water	0,5 ml	50 µg/ml
tRNA 10 mg/ml	2,5 ml	100 µg/ml
Tween 20 20%	0,25 ml 0,5	
Acide citrique 1M	0,46 ml	→ pH 6
H2O	to 50 ml	

To hybridize the DIG labelled probe, embryos are incubated overnight in 200 µL of HM buffer with 100 ng of probe at 70°C overnight.

To prepare embryos for the anti-DIG antibody labelling a series of washing steps with solutions containing HM in decreased concentration and PBS until reaching 100 % PBT increased is performed at room temperature. After that, embryos are incubated with anti-DIG antibody 1:20.000 in PBT containing 2mg/ml of sheep serum overnight at 4 °C. Staining occurs after several washes in PBT with incubation at dark using the Alkaline tris buffer prepared as following:

	Volume	final
Tris HCL 1M pH 9,5	10 ml	100 mM
$MgCl_2$ 1M	5 ml	50 mM
NaCl 5M	2 ml	100 mM
Tween 20 20 %	0,5 ml	0,1 %
H_2O	to 50 ml	

The reaction occurs in around between 2 and 5 hours where embryos are observed every 15 minutes to avoid overstaining or signal saturation. The reaction is stop with EDTA and embryos are fixed in 100 % glycerol for imaging.

3.11 Immunohistochemistry

Embryos are obtained by natural spawning from the double transgenic line Tg(wt1b:eGFP)-*ezh2*^{ul2} , selected for fluorescence and genotyped according to paragraph 3.7. Knock out and WT are fixed in 4 % PFA at 60 hpf and paraffin embedded according to standard procedure.

Slice are cut in the nephric region in 20 µm sections. Slices are dewaxed in Xylol and washed in gradient ethanol/PBS solution until 100 % PBS. Slices are incubated at 80 °C for 45 minutes in Tris-EDTA buffer (10 mM Tris base, 1 mM EDTA solution, 0.05 % Tween 20, pH 9.0) to heat-induce antigen retrieval and washed in PBS at room temperature. Permeabilization of sections are made with 10 minutes incubation in PBT (1 x PBS + 0,05% tween) followed by 30 minutes incubation in 1x PBS + 0,5 % TritonX-

100. Unspecific binding is prevented with 2-hour incubation incubation in blocking solution containing 10 % goat serum, 2 % BSA in PBS with 0,1 % TritonX-100.

Slices were incubated overnight at 4 C with

- Anti-ISL1 antibody diluted 1:250 in 2 % NGS + 2 % BSA + 0,1 % TritonX-100 in 1XPBS
- anti-EGFP antibody diluted 1:500 in 2 % NGS + 2 % BSA + 0,1 % TritonX-100 in 1XPBS

After this, red fluorescence for ISL1 is detected by incubation with Alexa Flour 546 goat anti-Rabbit; Invitrogen; A11035 and GFP.

3.12 Zebrafish larvae imaging

All larvae used for imaging were grown in 0.2 mM 1-phenyl 2-thiourea (PTU) Danieau solution to prevent pigmentation. *In-situ* images were taken mounting the larvae in 100 % glycerol using an Eclipse upright microscope (Nikon) equipped with a 10x and 40x objectives and a and a DS-Vi1 digital colour camera run by NIS-Element software (Nikon). Immunohistochemistry images were taken with a A1R HD25 ECLIPSE Ti2E laser scanning microscope using the NIS-Elements 5.21.02 software using a 40x oil immersion objective with scanning set every 2 µm.

3.12.1 Live zebrafish imaging

To check phenotype of nephric development, 3 dpf genotyped *Tg(wt1b:eGFP)-ezh2*^{ul2} larvae were anesthetized with 0.03 % tricaine (Merk) and mounted in 1.25 % low-melting temperature agarose with a help of Melek's hair for mounting the zebrafish. Images are taken with the A1R HD25 ECLIPSE Ti2E laser scanning microscope using the NIS-Elements 5.21.02 software.

3.13 RNA-seq analysis.

Normalized Reads per kilobase million (RPKM) were directly obtained for ISL1 and ISL1-DT from NCBI portal (gene ID respectively 3670 and 642366). Source of gene expression deposited data are from Fagerbert et al 2014 (Fagerberg et al., 2014).

4 Results

4.1 Luciferase assays identifies presence of a promoter in the reverse strand of the ISL1 locus

To investigate the activity of the CBE associated locus where rs2303751 reseeds, we performed a luciferase assay in HEK293 covering a region of 2854 bp. This region was divided in 3 partially overlapping fragments cloned both in forward and flipped orientation from ISL1 TSS (as shown in paragraph 3.2.1) (Figure 12 A). Significant luciferase signal was observed only in the flipped Fragment 2 harbouring variant rs2303751 (Figure 12 B). To study a possible regulatory effect of the variant, we introduced the variant in the Fragment 2 both in the forward and flipped orientation and tested the luciferase activity. Here neither the intensity nor the orientation of the promoter showed differences compared to the major allele (Figure 12 C).

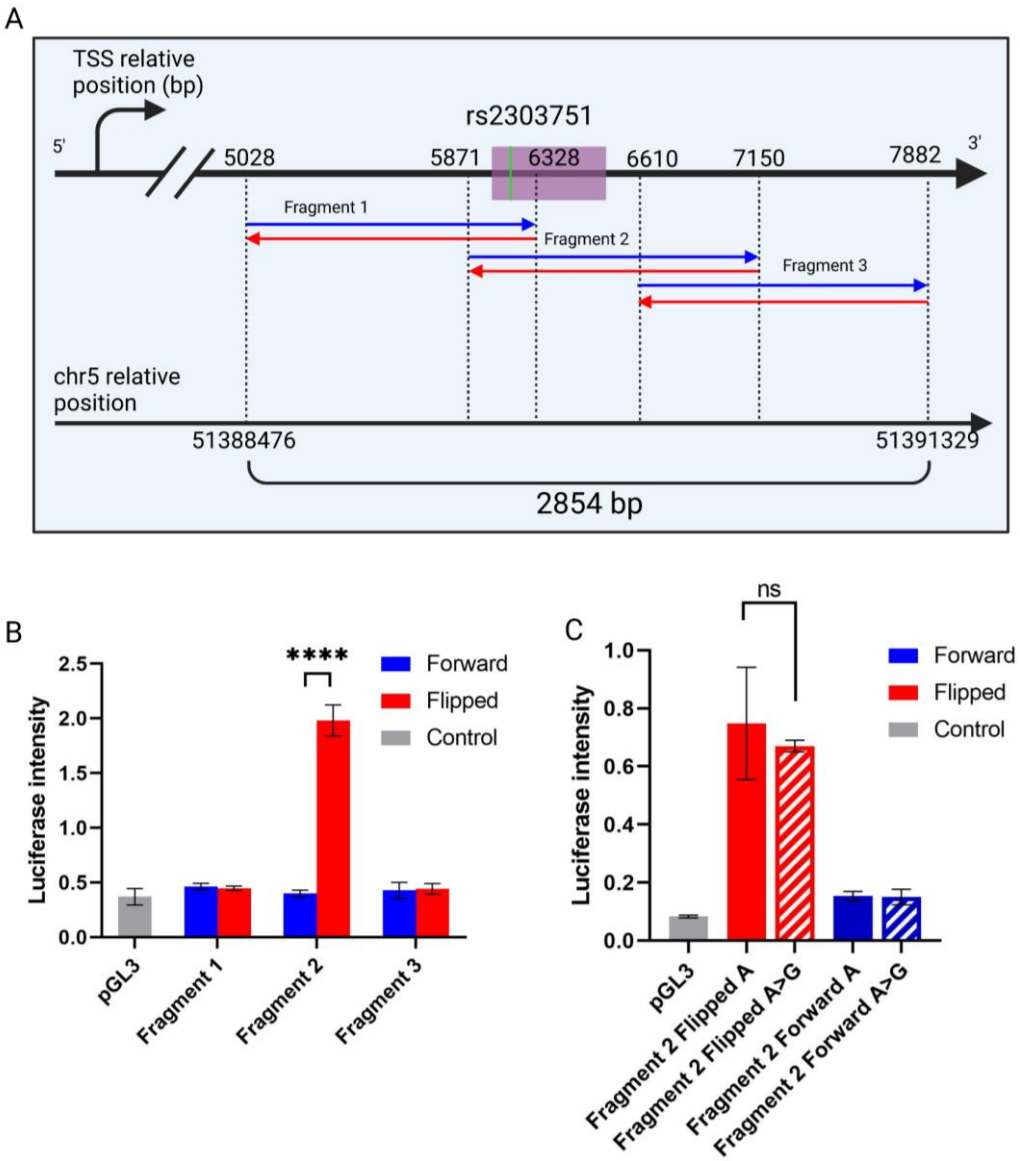


Figure 12. Luciferase sliding window approach identifies a promoter in the reverse strand of the rs2303751 CBE associated region. A) Genomic overview of the region where the luciferase is of the 3 partially overlapping fragments is perfomed. In detail, Fragment 1, Fragment 2 and Fragment 3

coordinates are shown relatively to the distance from the *ISL1* transcription starting site (TSS) (top) and also in chromosome 5 coordinates (hg38 /lower). B) Luciferase assay of the forward (blue) and flipped (red) fragments relative to the empty control vector pGL3 (in gray); significance over control is observed only for the flipped Fragment 2. C) Fragment 2 luciferase assay with the rs2303751 variant A>G in both forward and flipped orientations displays no significant difference between major (A) and minor (G) allele.

4.2 EZH2 protein binds to the GWAS associated CBE region and enhance ISL1 expression

4.2.1 ChIP qPCR reveals binding of EZH2 in the rs2303751 locus

To detect the presence of EZH2 in the CBE associated region we perfomed chromatin immunoprecipitation and, after chromatin sharing, we investigate 10 qpcr fragments equally distributed over a region that overlaps Fragment 2 from minus 1.6 kb upstream to 1.7 kb down-stream (Figure 13 A). The binding was observed to occur in each of the analysed fragments witch higher fold change in binding for the two tested fragments of minus 0.5 kb and plus 0 kb distance of Fragment 2 (Figure 13 B).

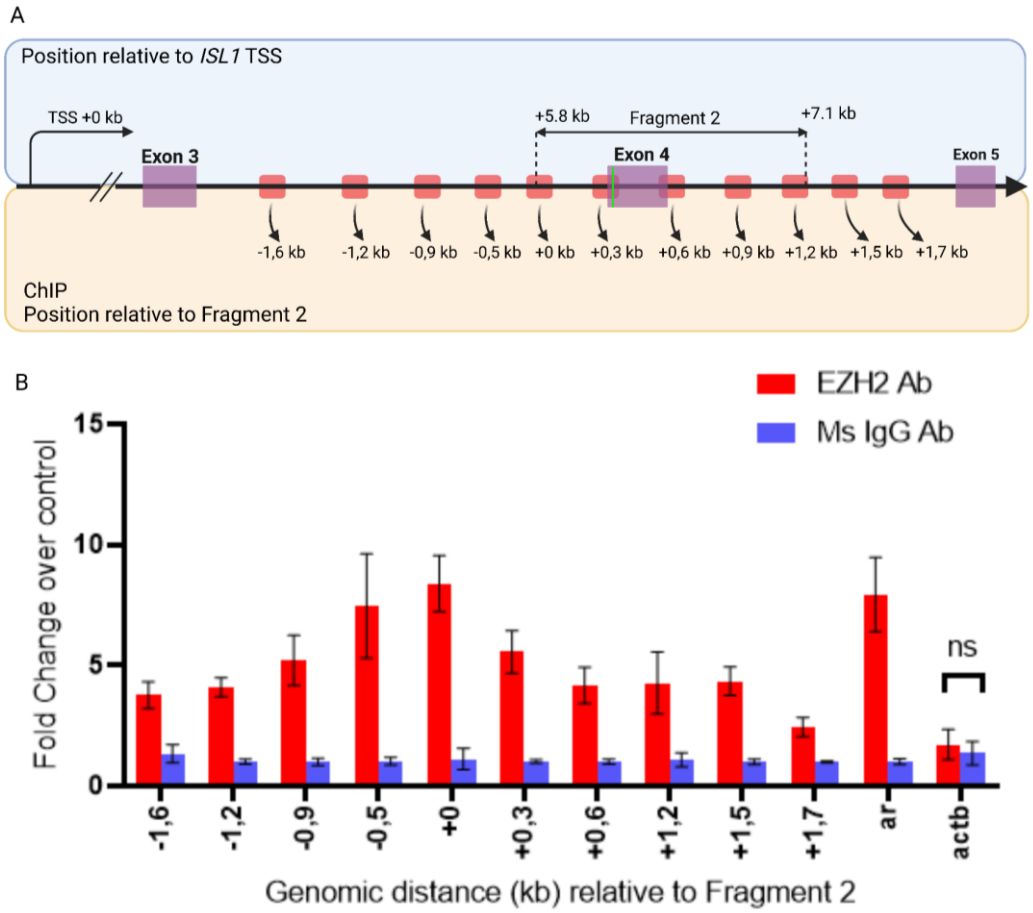


Figure 13. ChIP qPCR on the CBE associated locus reveal presence of EZH2 binding. A) schematic representation of the region tested for ChIP-qPCR in *ISL1* (red segments) with distance relative to the 5' of Fragment 2 in kb. B) ChIP-qPCR fold change representing EZH2 bound genomic DNA relative to mouse IgG bound DNA. Androgen receptor (ar) genomic DNA as positive control (Kim et al., 2018) and actin b as negative control. All fold change accept for actin b are significant over IgG control.

4.2.2 EZH2 enhance ISL1 expression thrugh binding on fragment 2

To investigate the regulatory role of EZH2 in the binding of *ISL1*, we knocked down EZH2 in HEK293 cells using siPOOLs and tested *ISL1* expression by qPCR. Successful knock down (Figure 14 A) showed a 1.5 fold change decrease of *ISL1* expression (Figure 14 B). To further check the regulation of the active Fragment 2 flipped, we overexpressed *EZH2* using the pCMVHA expression plasmid (Figure 14 C) together with the Fragment 2 forward or flipped luciferase plasmid. We observed a significant doubling of the of luciferase activity in the flipped fragment 2 when EZH2 is overexpressed but no effect on the

empty control and in the Fragment 2 forward version (Figure 14 D). Overexpression of *EZH2* does not influence *ISL1* expression in HEK 293 (Figure 14 E).

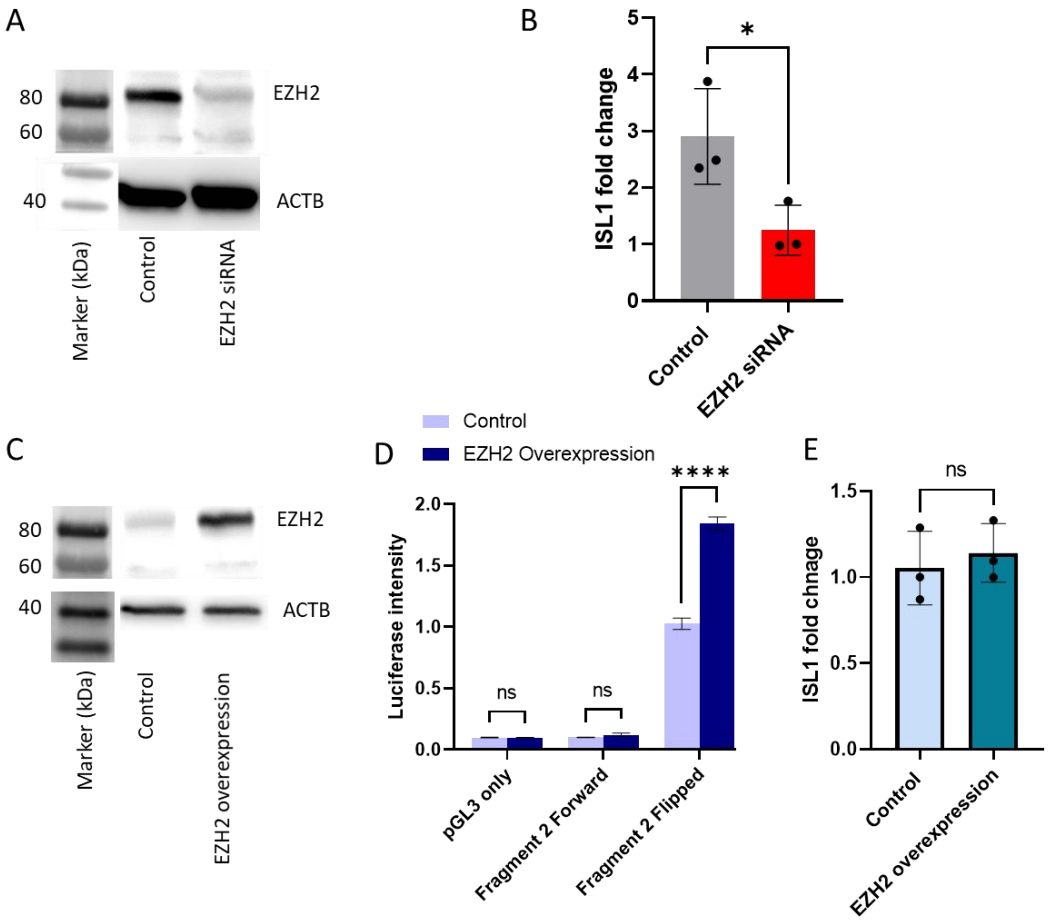


Figure 14. EZH2 regulates ISL1 expression through binding on Fragment 2. A) Western blot against EZH2 shows a succefull siRNA silencing of EZH2 compared to control. B) *ISL1* qPCR with EZH2 siRNA dispalys a reduced *ISL1* signal. C) Western blot on EZH2 with control vector and EZH2 overexpression plasmid confirms the efficence of the overexpression. D) Luciferase assay with EZH2 overexpression (red bars) and control (gray bars). As before vector (pGL3) and Fragment 2 forward shows only little expression in both cases but for the flipped orientated Fragment 2 EZH2 overexpression enhances reporter activity significantly.

4.3 Whole mount in situ hybridization on *ezh2*$^{ul2-/-}$ larvae displays tissue specific regulation of *isl1.*

Isl1 has been shown to be expressed in the nephric region of zebrafish embryo during development at 56 hpf (Zhang et al., 2017). To further investigate the *ezh2* mediated regulation of *isl1* in a vertebrate model, we performed mRNA *in situ* hybridization against *isl1* transcript in wt, *ezh2*$^{ul2+/-}$ and *ezh2*$^{ul2-/-}$ zebrafish larvae at 56 hpf. We observed a decreased *isl1* expression which locates specifically to the

nephron region of the *ezh2*$^{ul2-/-}$ larvae only. In fact, this strong decrease seems not to be such in the brain and spinal cord of the *ezh2*$^{ul2-/-}$ larvae (Figure 15).

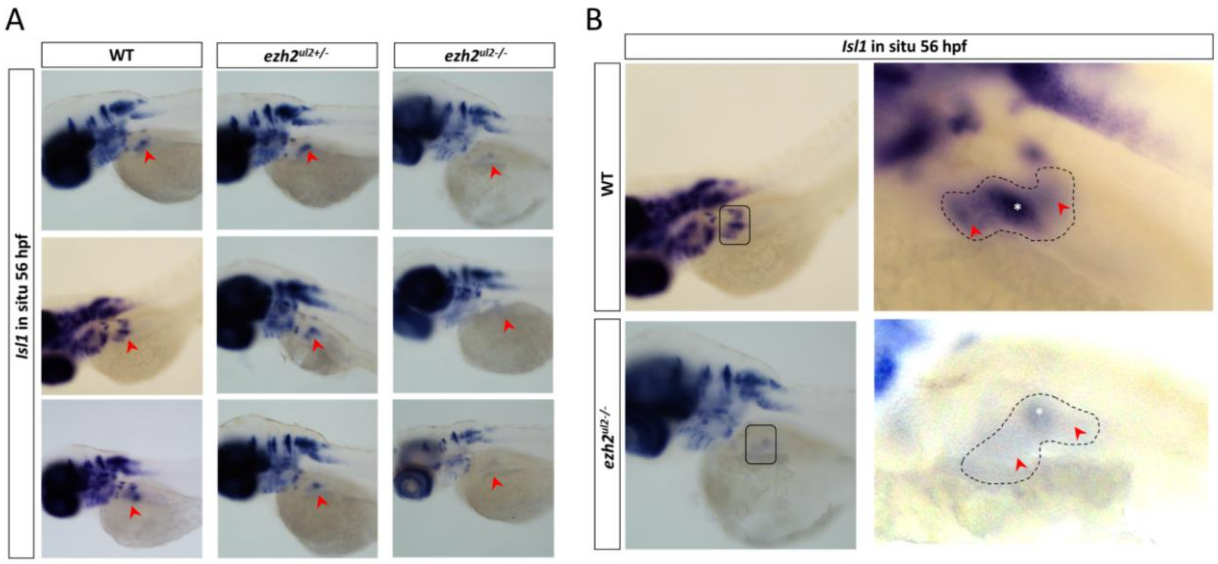


Figure 15. mRNA *in situ* hybridization of *isl1* in WT, *ezh2*$^{ul2+/-}$ and *ezh2*$^{ul2-/-}$ at 56 hpf reveals tissue specific regulation that locates in the pronephron. A) *isl1 in situ* in WT (left), *ezh2*$^{ul2+/-}$ (center) and *ezh2*$^{ul2-/-}$ (right) larvae at 56 hpf. Red arrows indicate the nephric region and shows a clear reduction of *isl1* signal in the *ezh2* KO line. Such decrease signal do not shows in the brain and spinal cord. B) Details of the nephric region (black box) in WT and *ezh2*$^{ul2-/-}$ at 56 hpf. White asterisks indicate the glomeruli and red arrows the nephric ducts.

To detect tissue specificity, we performed qRT-PCR on trunk from head-chopped embryos and on whole larva. The first, which still included the pronephric region, confirmed a significant decreased in *isl1* transcript signal in the knock out larvae (Fig 16 A). Instead, the same qRT-PCR protocol on whole larvae did not display any significant differences in *isl1* expression between WT and *ezh2*$^{ul2-/-}$ larvae (Fig 16 B). This confirms our WISH finding of mainly pronephric regulation of *isl1*, possibly mediated via Ezh2.

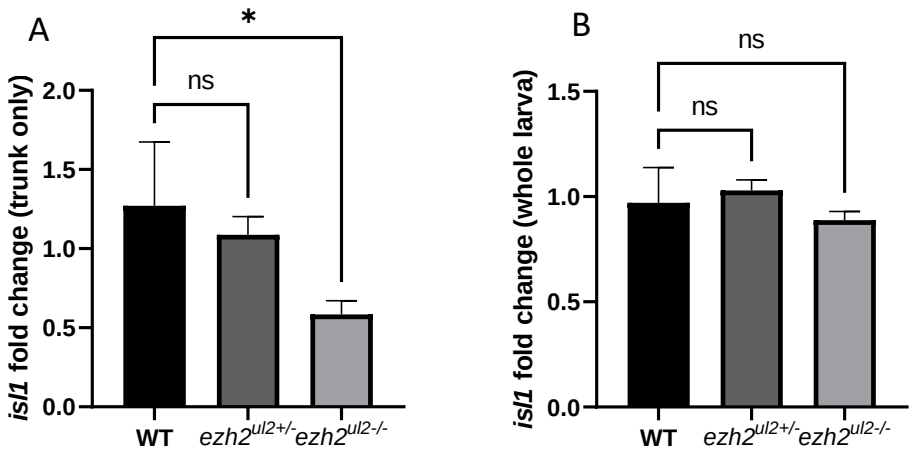


Figure 16. RT-qPCR on trunk and whole larva at 56 hpf shows a decreased *isl1* expression. A) RT-qPCR on trunked larvae comprehending the nephric region shows a decreased *isl1* expression. B) RT-qPCR on whole larva does not show a decreased *isl1* expression.

4.4 Immune histochemistry on *ezh2$^{ul2-/-}$* larvae reveals tissue specificity of ISL1 regulation in the *Tg(wt1b:eGFP)* line.

To access tissue specificity of *ezh2-isl1* mediated regulation, we performed Isl1 immune histochemistry on sagittal paraffin sections of the double transgenic lines *Tg(wt1b:eGFP)- ezh2$^{ul2-/-}$* and *ezh2$^{ul2+/+}$* lines at 56 hpf. Here we confirmed the presence of Isl1 protein within the pronephric region and observed a strong reduction of Isl1 protein specifically in the cells of the glomeruli and nephric duct of the *ezh2$^{ul2-/-}$* line (Figure 17 A). Therefore, we further investigate the *ezh2* knock out effect on the development of the pronephric structure and we observed an irregular and malformed nephric duct in the KO larvae from 3 dpf onwards (Figure 17 B)

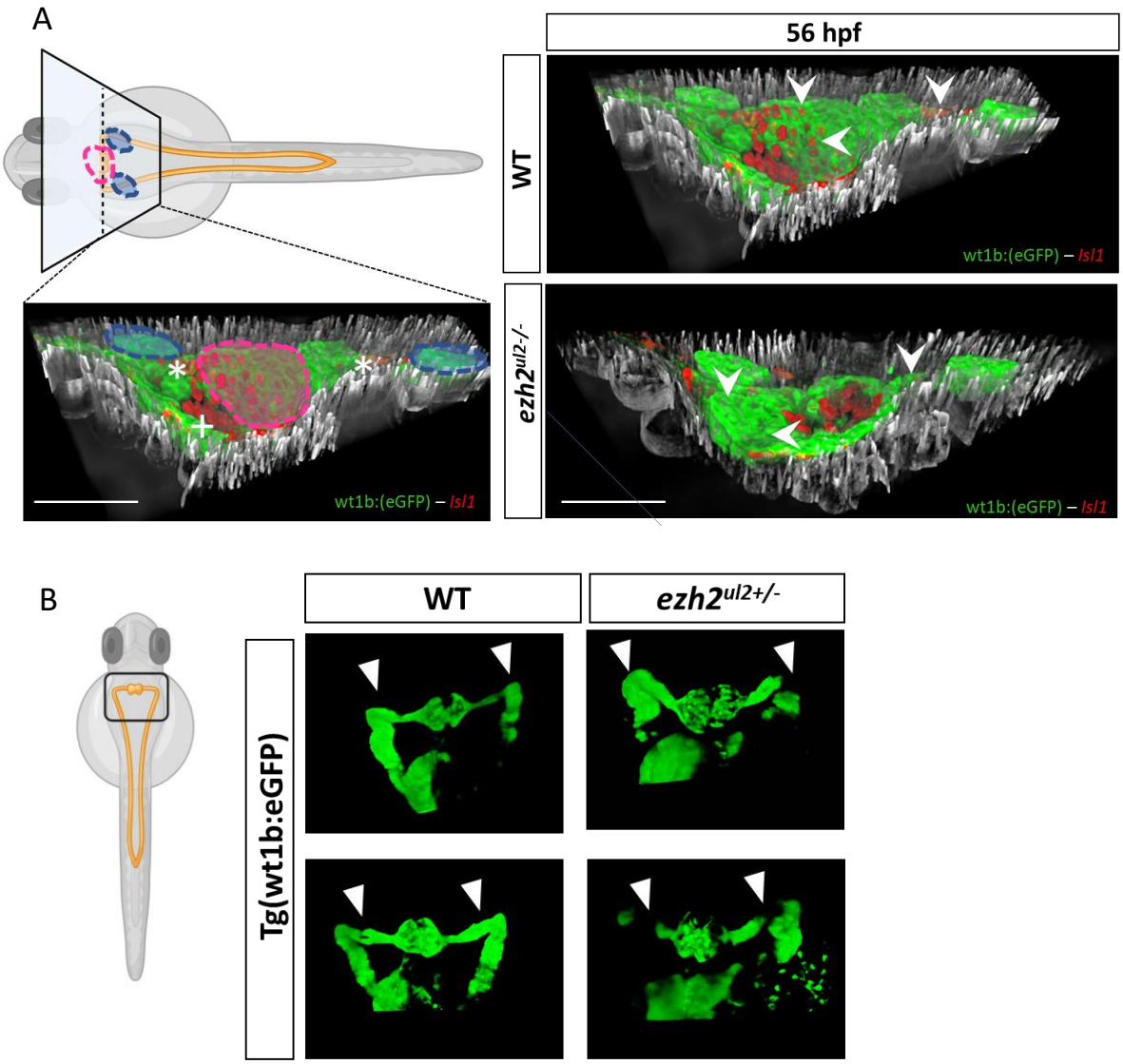


Figure 17. *Ezh2* mediates *isl1* regulation with tissue specificity on the nephric region and causes nephric duct developmental defects. A) *Isl1* immune histochemistry (red cells) in *Tg(wt1b:eGFP)* line (in green) and double transgenic *Tg(wt1b:eGFP)* – Ezh2 KO line. Left panel (top) indicates the location of the transversal paraffin section in reference of the whole larva and nephric region. Blue circles indicate the nephric ducts; purple circle indicates the glomeruli region; white asterisks (bottom) the nephric ducts and the white plus the pancreas. Right panels show the 3D co-localization of *Isl1* protein (red) on the glomeruli and nephric ducts (green) in the WT (top) and *Ezh2* KO (lower) 56 hpf larvae. A clear absence of *Isl1* signal locates to the glomeruli and nephric ducts of the *Ezh2* KO line compared to WT. B) Nephric ducts of the *Ezh2* KO larvae display developmental defects and malformation at 3 dpf. White arrows indicate the correct protrusion of the nephric duct in the WT and absence of growth or malformations in the Ezh2 KO.

4.5 Fragment 2 promoter is specific for ISL1 expression

4.5.1 Fragment 2 interacts within ISL1 and ISL1-DT

To further investigate this tissue specific effect and cause of *Isl1* regulation we came back to the fact that the Fragment 2 promoter only active in its flipped orientation. The genomic region where Fragment 2 reseeds has also been shown to interact on the 115 bp region that spaces *ISL1* from *ISL1-DT* (Domcke et al., 2020) (Figure 18).

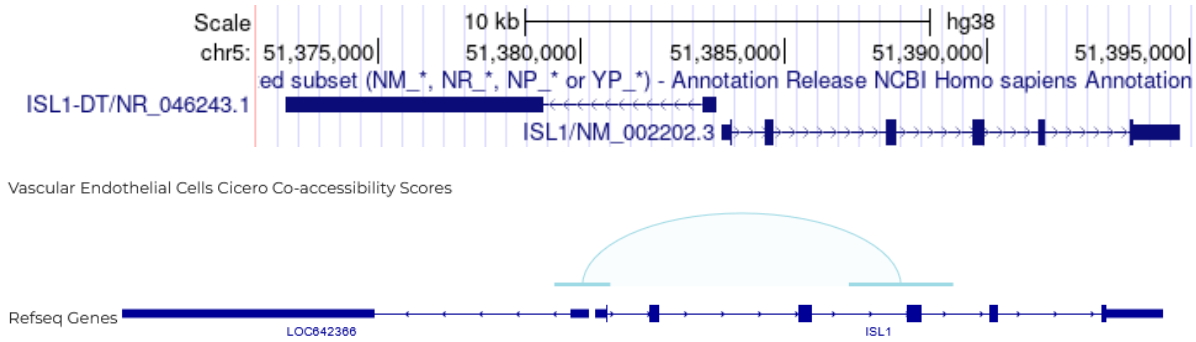


Figure 18. Fragment 2 interacts with the 115 bp that space ISL1 from ISL1-DT. Upper panel shows genomic location of ISL1 and ISL1-DT, arrows indicate the orientation of the genes. Lower panel shows the detected interaction in the genomic locus of ISL1 (light blue bridge) that reseeds in the Fragment 2 locus.

Our GWAS shows *ISL1-DT* to be associated with CBE and the relative RNA seq analysis has proved this gene, together with ISL1, to be significantly downregulated over time during bladder development (Mingardo et al., 2022).

4.5.2 Fragment 2 regulates ISL1 via EZH2

We checked for Fragments 2 EZH2 mediated regulatory influence on the opposite oriented gene *ISL1-DT* 115 bp upstream of *isl1*. To investigate the expression pattern of *ISL1* and *ISL1-DT* we analyzed RNA-seq deposited data provided by the NCBI project "HPA RNA-seq normal tissues" (BioProject: PRJEB4337) (Fagerberg et al., 2014). Here we found that *ISL1* and *ISL1-DT* are expressed with the same basal expression pattern in the same analysed tissues but *ISL1* always shows a higher RPKM than ISL1-DT (Figure 19 A). This mechanism identifies those two genes as divergent lncRNA/mRNA transcripts (Sigova et al., 2013). To investigate whether the EZH2 mediated Fragment 2 promoter regulates both *ISL1-DT* and *ISL1* in a shared fashion, we performed qPCR in HEK293 with and without EZH2 siRNA mediated knockdown. In the scrambled control we observed that *ISL1-DT* and *ISL1* shows the previous described basal expression of the analysed "HPA RNA-seq normal tissue" with *ISL1* showing approximately double the expression strength than *ISL1-DT*. In the EZH2 knock down though *ISL1-DT*

does not change its expression level, while *ISL1* confirms its decrease in expression (Figure 19 B). For this reason, we are suggesting here, that the EZH2 mediated Fragment 2 promoter does not act on *ISL1-DT* expression but exclusively on *Isl-1* expression at least in its closest neighbourhood (Figure 19 C).

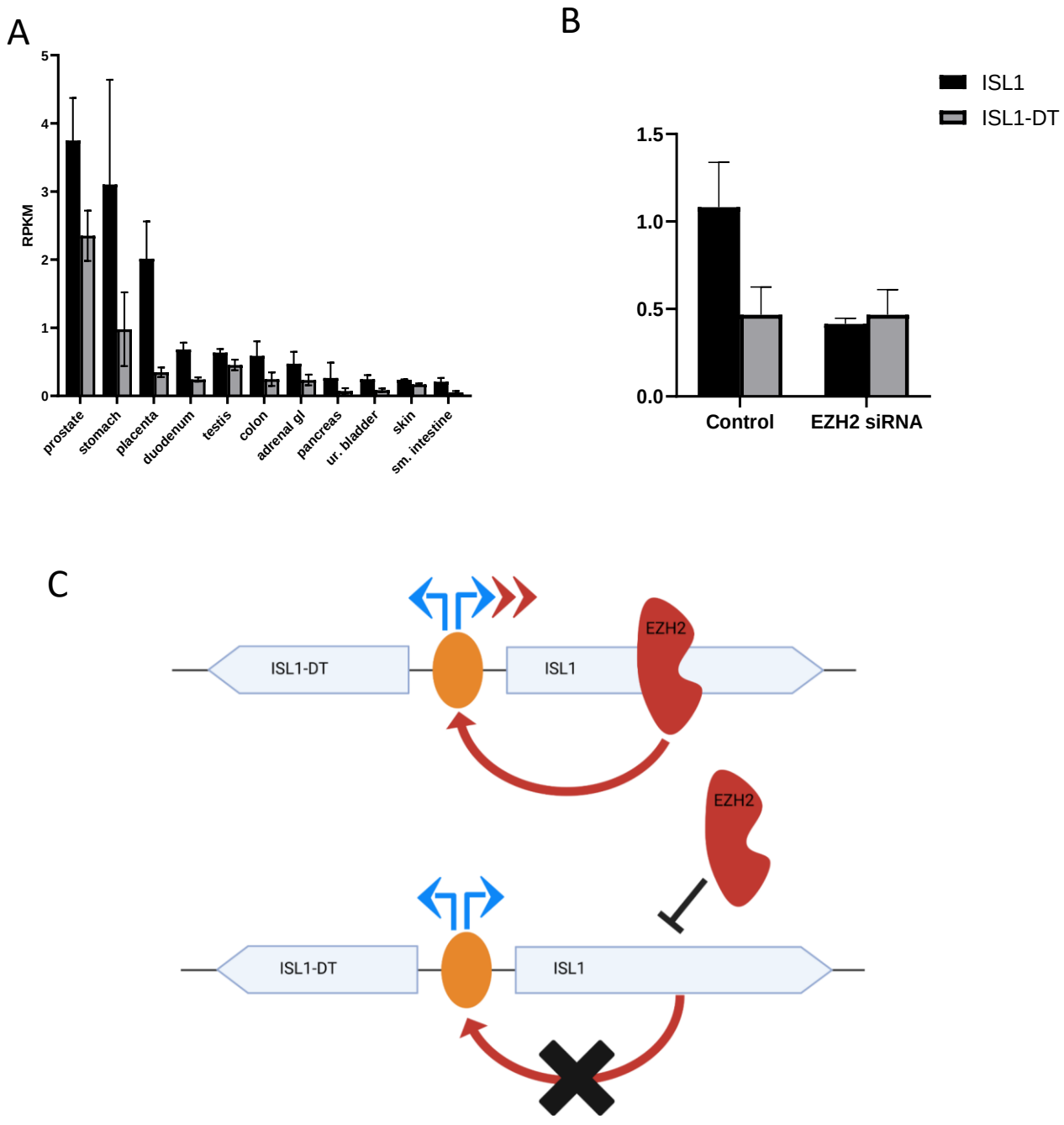


Figure 19. Fragment 2 specifically regulates ISL1 via EZH2. A) RNA-seq shows the expression pattern of ISL1 and ISL1-DT in different tissues. These genes are generally expressed with higher counts for ISL1 and lower for ISL1-DT. B) qPCR of ISL1 and ISL1-DT with control and EZH2 siRNA shows a reduced signal of ISL1 but non-changed ISL1-DT expression. C) Proposed molecular mechanism for the

expression pattern of ISL1 and ISL1-DT mediated by EZH2. Top panel shows physiological condition, EZH2 sits in the Fragment 2 promoter and specifically enhance the ISL1 expression (red arrows). In pathological condition, EZH2 does not bind on Fragment 2 and the expression of ISL1 is reduced, but the amount of ISL1-DT remains unalerted.

5. Discussion

5.1 Main findings

In this study we performed the largest GWAS on CBE to date and we identified 8 CBE significant loci, seven of which are novel. Among those loci reseeds different genes, of which RNA-seq of human and mouse fetal bladder at different developmental stages shows that those genes are differentially expressed during bladder development. Among all the genes, we decided to study ISL1, in fact this gene (i) resides where the most significant variant is (rs6874700, *p* value of 1.48×10^{-24}) (ii) has been shown to be involved in early bladder development (Draaken et al. 2015b; Zhang et al. 2017; Ching et al. 2018; Liaw et al. 2018; Su et al. 2019a) (iii) shows differential expression during bladder development with high expression in early stages and strong downregulation in mature bladder. ISl1 has been re-sequenced in a Swedish CBE cohort but no diseases causing variation has been found. For this reason and the fact that this gene is strongly downregulated during bladder development, we assumed and searched for its pathogenicity role in its regulation. To do that, we screened for regulatory variant in LD with rs6874700 and found rs2303751 to have high regulatory score in two different databases (regulobeDB and ForgeDB). In addition, regulomeBD shows ChIP seq experiments that display this region to be target by EZH2 in different cell lines and samples from ENCODE. EZH2 is the catalytic subunit of the policomb repressive complex 2 (PRC2) that is responsible to tri-methylate the H3K27 that follow chromaitne silencing (Margueron & Reinberg, 2011). Recently EZH2 has been shown to act as a transcription factor by directly contacting the DNA without the PRC2 complex and activating gene expression. Our luciferase assays experiments identified an active promoter in the reverse DNA strand from ISL1 TSS where rs2303751 reseeds (here called Fragment 2) and that the variant rs2303751 does not influence the promoter activity. Indeed, in HEK cells ChIP qPCR reveals presence of EZH2 in Fragment 2 and EZH2 knock down diminish ISL1 expression. On the other hand, the overexpression of EZH2 does not increase ISL1 expression but it enhances the luciferase activity of Fragment 2. We then investigated the EZH2 regulatory role on ISL1during development in vivo using an *ezh2* knock out zebrafish larvae, here whole mount in situ hybridization, qPCR and immunohistochemistry on the KO line (*ezh2*$^{ul2-/-}$) showed that in the *isl1* expression and protein is diminished only in the nephric region, suggesting a tissue specific regulator mechanism of ISL1 made by EZH2. We therefore checked for phenotype on the double transgenic line *Tg(wt1b:eGFP)-ezh2*^{ul2} and confirmed different malformation

and development defect in the glomeruli and nephric duct of the knock out. To access the role of the Fragment 2. To further investigate this tissue specific effect and cause of *Isl1* regulation we came back to the fact that the Fragment 2 promoter is only active in its flipped orientation. In fact, previous ATAC-seq (Domcke et al., 2020) has shown the genomic region where Fragment 2 reseeds to interact on a 115 bp region thatspaces *ISL1* from *ISL1-DT*. *ISL1-DT* is the non-coding divergent transcript of *ISL1* and, for its characteristic, is shown to share the same promoter and expression pattern in different tissues. This is indeed confirmed by the fact that both genes share expression in specific tissues with an expression pattern where *ISL1* expression is circa the double of *ISL1-DT.* Therefore, we investigate if the EZH2 knock down could affect also the expression of *ISL1-DT* and our qPCR confirmed a specific regulation on *ISL1*. Since *ISL1-DT* is not affected by EZH2 binding on fragment 2 and the EZH2 in vivo knock out shows *isl1* tissue-specific regulation, we propose that Fragment 2 is a promoter that enhance specifically *ISL1* expression through EZH2 which is the key for its regulation in tissue specificity of urinary tract

5.2 HEK 293 and Zebrafish as model organism for this study

In this study, we conducted a functional study of GWAS markers for CBE patient cohorts. CBE embryonic formation is believed to occur between the first four to six weeks of gestation, but due to the lack of human fetal biological samples, access to its study is prevented. At the time of this study, there were no human fetal bladder/cloaca cell lines available, but rather primary cultures from different bladder tissues of human adults. However, these were unsuitable models due to their heterogeneity, late differentiation stages and also were only viable for around 20 passages before senescence.

For this reason, we opted for the HEK 293 model, which is kidney fetus derived and immortalized. The drawback of using these cells is that they do not represent the bladder tissue of human fetuses and do not derive from endoderm but rather mesoderm (Stepanenko & Dmitrenko, 2015). Additionally, since their first establishment, their genome becomes enriched in mutation on their karyotype such as duplications, deletions, and variations (Stepanenko and Dmitrenko 2015) (Y. C. Lin et al., 2014), making them unsuitable for developmental biology studies. Furthermore, HEK 293 cells may vary from batch to batch and their phenotype may even assume epithelial or fibroblast characteristics (Stepanenko and Dmitrenko 2015). However, we used HEK 293 to focus on a specific genomic regulation between EZH2 and the ISL1/ISL1-DT cassette. We did not use HEK 293 to detect any genomic structure, cell fate, or cell migration but rather to check the regulatory mechanism of EZH2 and its target gene ISL1.

To study the contribution of EZH2 in ISL1 regulation during development, we used zebrafish larvae. Although zebrafish do not have a urinary bladder, they have structures similar to those that contribute to urinary tract and cloaca formation (Kolvenbach et al., 2019). These structures are made by nephron

and nephric ducts, which fuse with the cloaca during development. The advantage of using zebrafish is that experiments can be performed in a whole vertebrate organism during development and allow for the study of tissue-specific gene regulation. In fact, in this study, we demonstrated that the mechanism of EZH2 and ISL1 affects urinary tract formation and is not present in other tissues where ISL1 is expressed. However, it was not possible to demonstrate or study the regulation of ISL1 by EZH2 from the homologous of Fragment 2 that would be present in zebrafish. In fact, ChIP-qPCR would not have been possible due to the lack of zebrafish Tg(wt1b:eGFP) developing larvae. In detail, we would have needed to perform FACS sorting of fluorescent cells from the nephric structure in the transgenic line, but the small number of developing embryo fluorescent positive would not have led to a sufficient amount of cells required for the chromatin immune precipitation protocol (around 1 x 10^7 cells).

Another noteworthy point is the impossibility of studying the long non-coding RNA ISL1-DT in zebrafish, which might play a regulatory role on ISL1 as well. In detail, long non-coding RNAs are not conserved between species but rather show species-specific conservation (Johnsson et al., 2014). Therefore, the usage of zebrafish did not allow the study of the role of ISL1-DT during development and its contribution to ISL1 regulation, but this gene is not present in any other species rather than human.

5.3 Luciferase assay identify promoter activity on Fragment 2

The luciferase assay is an efficient method for detecting promoter activity on DNA fragments, it also helps to identify the transcription factor targeting a specific fragment by its overexpression and luciferase quantification. We therefore applied this technique to investigate the rs2303751 locus by analysing 3 partially overlapping DNA fragments located where the rs2303751 reseeds to investigate a total of 2854 bp region. Since many human promoters shows different orientations (Duttke et al., 2015), we cloned these 3 fragments both in forward and flipped orientation from the ISL1 TSS direction and we found luciferase activity only in a flipped fragment that comprehends the rs2303751 variant (Fragment 2). We concluded that rs2303751 identified a promoter that was not described in ISL1 before. Since variants and mutations are known to affect promoter activity (Garieri et al., 2017), we introduced the rs2303751 A to G variant in the Fragment 2 both in forward and reverse but we found no difference in luciferase activity.

To gain a more comprehensive understanding of the variations that may impact the Fragment 2 promoter activity, it would be beneficial to perform sequencing of the intron-flanking regions of ISL1, specifically introns 3 and 4. These regions have not been extensively sequenced in previous studies

that focused primarily on the exons of ISL1. Sequencing the intron-flanking regions may uncover additional variations that could potentially affect the activity of the Fragment 2 promoter.

On the other hand, luciferase assay reveals only DNA activity and does not take in consideration neither the DNA methylation nor the histone marker modifications. Both modifications are important for dynamic regulation of gene expression (Cedar & Bergman, 2009) during development and over life course (Maunakea et al., 2010). DNA methylation, particularly in CpG-rich regions like Fragment 2, has been shown to regulate gene expression (Holliday & Pugh, 1975) (Bestor et al., 2015). Variation on these region has been shown to influence the methylome of individuals affecting gene regulation (Villicaña & Bell, 2021) (Hawe et al., 2022).

Additionally, histone modifications play a crucial role in gene regulation by activating or repressing gene expression and modulating chromatin accessibility for transcription factors and the transcription machinery (Cedar and Bergman 2009; Dong and Weng 2013). GWAS markers have been shown to highlight loci with associated chromatin changes during cell differentiation (Tak and Farnham 2015; Soskic et al. 2019). Since the regulatory effect of Fragment 2 in CBE patients does not appear to be directly linked to the GWAS rs2303751 variant, it is possible that an altered promoter activity possibly observed in CBE patients is influenced by differential DNA methylation or histone modifications in this region.

To further investigate the regulatory effects of Fragment 2 in CBE, future studies should consider examining DNA methylation patterns and histone marker modifications in patient samples. This integrative approach would provide a more comprehensive understanding of the regulatory mechanisms underlying ISL1 dysregulation in CBE.

5.4 EZH2 regulates ISL1 expression

5.4.1 EZH2 binding on Fragment 2 regulates ISL1 in HEK 293

After identifying the presence of EZH2 binding in the rs2303751 locus and its surroundings on HEK 293 cells via ChIP-qPCR, we further investigated the role of ISL1 as a transcription factor independent of PRC2. To do so, we conducted two experiments: (i) overexpression of EZH2 followed by luciferase activity measurement on Fragment 2 and assessment of ISL1 expression, and (ii) EZH2 knockdown followed by analysis of ISL1 expression.

In the first experiment, we observed that overexpression of EZH2 successfully activated the luciferase activity of the flipped Fragment 2, confirming the direct binding of EZH2 to this region and its role as a

transcription factor. However, the overexpression of EZH2 did not alter ISL1 expression. We speculate that this is because the endogenous levels of EZH2 in HEK 293 cells are already sufficient to saturate the binding on the Fragment 2 promoter, and the addition of exogenous EZH2 does not significantly impact its regulation. This hypothesis is supported by the fact that EZH2 knockdown resulted in downregulation of ISL1 expression, indicating that the endogenous EZH2 is indeed involved in regulating ISL1.

Overall, these findings suggest that EZH2 functions as a transcription factor in activating the flipped Fragment 2 promoter, but its overexpression does not further enhance ISL1 expression due to the already saturating levels of endogenous EZH2. However, EZH2 knockdown confirms the role of endogenous EZH2 in the regulation of ISL1 expression.

5.4.2 ISL1 regulation via EZH2 is tissue specific

We conducted further investigations into the role of EZH2 in *ISL1* regulation using zebrafish ezh2 Knock Out larvae as a vertebrate model. Through in situ hybridization, immunohistochemistry, and qPCR analyses (Figure 15; 16; 17 A), we observed a specific downregulation of ISL1 in the pronephron of zebrafish Ezh2 knock- out larvae at 56 hours post-fertilization (hpf). Moreover, in the *Tg(wt1b:GFP)* line, which highlights the nephric region, we observed malformations from 3 days post-fertilization (dpf) onwards. Based on these findings, we concluded that dysregulation of *ISL1* mediated by EZH2 specifically affects the development of the nephric region. This mechanism provides further evidence that the contribution of ISL1 to CBE lies in its regulation rather than genetic variation.

Indeed, ISL1 is known to play a role in the formation and differentiation of various organs and tissues, including the limb, heart, neurons, and pancreas (Ahlgren et al., 1997; Cai et al., 2003; Liang et al., 2011; Narkis et al., 2012; Wilfinger et al., 2013). Therefore, disease-causing variants in *ISL1* may affect these different tissues. However, a tissue-specific dysregulation of the gene could lead to defects only in the affected tissue, as observed in our study. This concept is supported by a study conducted by Ching et al., where tissue-specific knockdown of ISL1 in the cloaca tissue of mice resulted in defects in the entire urinary tract formation (Ching et al., 2018).

Considering that CBE occurs during the first 4 weeks of gestation in humans, the pathogenetic dysregulation of *ISL1* should be investigated within the early timeframe of cloaca development, as evidenced by its strong expression in the early cloaca and primitive bladder of mice (Draaken et al., 2015; Mingardo et al., 2022).

In our HEK 293 model, we demonstrated that EZH2 binds to an internal promoter of ISL1 and regulates its expression. The notion of tissue-specific regulation of *ISL1* through EZH2 arises from our observation

of isl1 expression in the ezh2 knock-out line of zebrafish, as mentioned earlier. Although we have not determined whether zebrafish possess a homolog promotor to the human Fragment 2 in isl1 for Ezh2-mediated regulation, we observed downregulation of ISL1 in both human and zebrafish models in the absence of EZH2. This suggests that this mechanism may be conserved for nephric formation, as observed in the nephric duct of zebrafish and HEK 293 cells. To further support this hypothesis, it would be beneficial to study another human cell line, unrelated to the urinary tract (e.g., neurons), and investigate the effect of EZH2 knockdown on ISL1 regulation. In such a cell line, ISL1 expression would be expected to remain unaffected by the absence of EZH2, as observed in the brain and spinal cord of zebrafish embryos.

5.4.3 Nephric malformations in the *ezh2* knock out line

Our study reveals that the absence of Isl1 protein in the nephric duct of zebrafish ezh2 knock-out larvae leads to nephric malformation starting from 3 days post-fertilization (dpf). It is known that one of the targets of ISL1 is the widely studied SHH (L. Lin et al., 2006), and that dysregulation of SHH mediated by ISL1 is crucial for urethral development (Su et al., 2019b). Therefore, the possibility that the contribution to the phenotype lies in the downstream dysregulation of Shh caused by reduced levels of Isl1, which should be further investigated.

On the other hand, considering that Ezh2 is the catalytic subunit of the PRC2 complex, it is possible that this phenotype could be influenced by other genes targeted by this chromatin modelling complex. To address this, performing an RNA-seq analysis of fluorescent cells from the nephric region of *Tg(wt1b:eGFP)-ezh2*$^{ul2-/-}$ compared to *+/+* larvae obtained through FACS sorting would be extremely helpful. This approach would allow us to study all the differentially regulated genes specific to the nephric duct at 3 dpf and understand the contributions of Shh or other genes. Unfortunately, the number of knock- out and wild-type larvae obtained from genotyping (approximately 18 to 22 out of 96) is insufficient to obtain the necessary amount of fluorescent nephric cells for the RNA-seq experiment so far.

5.5 Fragment 2 is specific for ISL1 regulation via EZH2

Considering that Fragment 2 is active only in the reverse orientation of the ISL1 transcription start site (TSS) direction, we investigated whether there might be a regulatory effect on ISL1-DT, a long non-coding RNA located 115 base pairs upstream of ISL1 and expressed in the opposite direction. Our findings revealed that these two genes share the same promoter and are co-expressed in the same

tissues, with ISL1 showing double the expression level compared to ISL1-DT. This pattern is characteristic of divergent transcripts, where a long non-coding RNA shares the promoter with a divergent coding transcript (Seila et al., 2008). Such gene cassettes of lincRNA/mRNA are known to be important for endoderm tissue differentiation, where the long non-coding divergent RNA typically regulates the expression of its associated mRNA (Fernandes et al., 2019; Sigova et al., 2013; Statello et al., 2020). Therefore, we investigated whether Fragment 2 could also regulate ISL1-DT by knocking down EZH2. Surprisingly, only ISL1 expression was affected, showing a reduction that brought the amount of ISL1 transcripts to the same level as ISL1-DT expression. The role of ISL1-DT has not been reported in the literature thus far. However, since its expression pattern follows the lincRNA/mRNA cassette, we speculate that there may be a cross-regulatory mechanism between the two genes that could play a role in tissue differentiation. Plaisance et al. reported a similar phenomenon, where long non-coding RNA-mediated ISL1 silencing in cardiac progenitor cells led to the differentiation of smooth muscle cells (Plaisance et al., 2022). We believe that dysregulation of the ISL1 and ISL1-DT cassette mediated by the Fragment 2 promoter during the early stages of cloaca and primitive bladder development could result in developmental defects in urinary tract formation, contributing to the CBE phenotype. This dysregulation is likely tissue-specific, and EZH2 appears to be a key factor in the correct expression of ISL1 in the ISL1-DT/ISL1 cassette. However, the exact reasons why EZH2 might not interact properly with Fragment 2 remain unknown.

6. Conclusion and outlooks

This study demonstrates that CBE exhibits multiple genetic variables. Specifically, our GWAS identified different loci associated with both coding and non-coding genes. These genes are expressed and/or differentially regulated during bladder development and in bladder cancers associated with CBE. We focused our investigation on the most highly associated locus, which highlights the gene *ISL1.* Our aim was to explore its role in CBE contribution through the founded internal promoter, Fragment 2. Our findings reveal that EZH2 binds to this genomic sequence and activates *ISL1* expression. The presence of EZH2 in Fragment 2 is crucial for proper ISL1 expression in both HEK 293 cells and this mechanism is conserved in zebrafish. In the zebrafish model, the expression of *isl1* via Ezh2 is tissue-specific and is localized in the nephric region, here reduced expression of *isl1* results in malformed developing nephric ducts. We hypothesize that a similar dysregulation of *ISL1* through EZH2 may occur during the early stages of cloaca/primitive bladder formation in humans, leading to alterations in the expression of the ISL1/ISL1-DT gene cassette. The ISL1/ISL1-DT cassette consists of non-coding/coding divergent transcripts that are known to be essential for endoderm specification (Sigova et al., 2013). Considering

the significance of ISL1 in the formation of various organs and tissues, we believe that its contribution to CBE involves tissue-specific dysregulation mediated by EZH2. In fact, when EZH2 fails to bind to Fragment 2, *ISL1* expression decreases, resulting in malformation of the nephric duct in zebrafish larvae.

In humans, we observed that the absence of EZH2 in Fragment 2 only affects the *ISL1* gene expression and not the ISL1-DT gene. We speculate that disruption of this regulatory cassette could impact the proper formation of the cloaca and primitive bladder. Manipulating or editing this promoter using CRISPR/Cas9 technology would likely disrupt the *ISL1* gene, as Fragment 2 encompasses intron 3, exon 4, and intron 4 of ISL1. Since previous re-sequencing efforts focused solely on the exons of *ISL1* (Arkani et al., 2018), a more comprehensive re-sequencing of Fragment 2 in the CBE cohort would be beneficial in identifying variations that may affect this promoter and potentially disrupt EZH2 binding in this region.

As *ISL1-DT* is not well-described in the literature, we cannot exclude the possibility of its involvement in regulating *ISL1* expression. Future studies should prioritize investigating this gene to identify its genomic or transcriptomic targets and potential gene-mediated regulations. To accomplish this, studying *ISL1/ISL1-DT* expression in primary cell cultures or in stem cell differentiation, where gene expression drives differentiation and defines specific tissues, would be advantageous, as this aspect is lost in HEK 293 cells (Y. C. Lin et al., 2014; Stepanenko & Dmitrenko, 2015).

ISL1 has been shown to activate the SHH pathway(L. Lin et al., 2006; Su et al., 2019) that is known to orchestrate the mesenchyme-epithelial interaction where high SHH defines inner bladder epithelial cells and low SHH smooth muscles (Baskin et al., 1996; Liaw et al., 2018b; Tasian et al., 2010). CBE more accredited theory supports an abnormal growth of the cloaca in the early stages, that prevents the abdominal wall closure and for that the bladder protruding out of the abdomen(Marshall & Muecke, 1962). We think that, since the lack of ISL1 leads to smooth muscles differentiation (Plaisance et al., 2022), an alteration of *ISL1* expression leaded by the tissue specificity lack of EZH2 binding could start an early differentiation of the smooth muscles bladder-out layer that disturbs the primitive bladder “in locus” formation and for that forming the CBE phenotype.

7. Acknowledgment

In primis, I would like to thanks Prof. Benjamin Odermatt and Prof. Dr. Med. Heiko Reutter for giving me this great opportunity of studying, for their guidance, supervision and in addition for having been great persons to talk with. In my experience, they showed me how academia life would work at the best.

I would like to thank also Prof. Dr. Karl Schilling for his nice way of being with me and with everyone by taking care of all the persons that have been working during the time of his direction of the Anatomical Institute of Bonn.

Truly I thank my family: my mother Antonella Lion, father Marino Mingardo and brother Alessandro for always having supported me, making me grow even if living abroad for many years. I thank my grandparents Antonia and Bruno that raised me since I was a kid and that I will always keep them in my heart. I thank my girlfriend Maria for obvious reason. I would like to thank my friends Nicole Cesarato and Luca Ricci, great person, thank you for your support and for always being there in time of need. And also for the carbanora evenings. I thank Jesh that has always been not only a colleague but a real friend for many years, for having listened to me in the hardest time, for taking care of me and for the insane night in Berlin. Of course, I thank also Giordano Cotti for the discussions and for the many fun evenings in Bonn and Lisbon during night. I thank Tobias Lindenberg and Nathalie for being super cool neighbors and Tobi for also his help in my experiments. Öznur Yilmaz for her great help in the lab and her super tasty bread. I thank Luca, Ricarda and Sofia that have always included me making me feeling at home even as very far away from home. I thank Florian Wallau that even if I meet for one year has left me a mark on my soul. Ugur, thanks for all the drinks, for the political talks and the PhD support. I thank my friend and flatmate Anooshay Abid, for the talk the drinks the gossip the time. I thank Erica Noventa that is a great listener and person, I wish you all the best. I thank Fabi and Baris for all the great talk and evenings. Pizzeria Spaccanapoli, for having a piece of Italian environment in Beuel. I thank Beuel, I thank Thomas for the gaming and the drinking. Marco Sensi (MS) that wow tuns out a great friend for hunting, menare le mani on hunt, catching birds sharing so many things lately. I thank my roman friends that saved me in COVID times and the great time spent together in Rome (Eugenio, Leo, Yuri). Specially Leo for hosting me and Eugenio for the evenings out in Rome. I thank Nina 2 and Julia for the awesome time out. I thank Bitburger, Hunt: Showdown, my PC, my cats Mefisto and Mimmo. I thank every person that works in the Anatomical Institute of Bonn that made my stay a great time.

8. Bibliography

Ahlgren, U., Pfaff, S. L., Jessell, T. M., Edlund, T., & Edlund, H. (1997). Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature 1997 385:6613*, *385*(6613), 257–260. https://doi.org/10.1038/385257a0

Arkani, S., Cao, J., Lundin, J., Nilsson, D., Källman, T., Barker, G., Holmdahl, G., Clementsson Kockum, C., Matsson, H., & Nordenskjöld, A. (2018). Evaluation of the ISL1 gene in the pathogenesis of bladder exstrophy in a Swedish cohort. *Human Genome Variation*, *5*. https://doi.org/10.1038/hgv.2018.9

Baskin, L. S., Hayward, S. W., Sutherland, R. A., DiSandro, M. J., Thomson, A. A., Goodman, J., & Cunha, G. R. (1996). Mesenchymal-epithelial interactions in the bladder. *World Journal of Urology*, *14*(5), 301–309. https://doi.org/10.1007/BF00184602/METRICS

Beaman, G. M., Cervellione, R. M., Keene, D., Reutter, H., & Newman, W. G. (2021). The Genomic Architecture of Bladder Exstrophy Epispadias Complex. *Genes 2021, Vol. 12, Page 1149*, *12*(8), 1149. https://doi.org/10.3390/GENES12081149

Bestor, T. H., Edwards, J. R., & Boulard, M. (2015). Notes on the role of dynamic DNA methylation in mammalian development. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22). https://doi.org/10.1073/pnas.1415301111

Biemar, F., Argenton, F., Schmidtke, R., Epperlein, S., Peers, B., & Driever, W. (2001). Pancreas development in zebrafish: early dispersed appearance of endocrine hormone expressing cells and their convergence to form the definitive islet. *Developmental Biology*, *230*(2), 189–203. https://doi.org/10.1006/DBIO.2000.0103

Boch, J. (2011). TALEs of genome targeting. In *Nature Biotechnology* (Vol. 29, Issue 2). https://doi.org/10.1038/nbt.1767

Bult, C., Blake, J., Smith, C., Kadin, J., & Richardson, J. (2019). *the Mouse Genome Database Group*. Mouse Genome Database (MGD) 2019. Nucleic Acids Res. 2019 Jan. 8;47 (D1): D801–D806.

Cai, C. L., Liang, X., Shi, Y., Chu, P. H., Pfaff, S. L., Chen, J., & Evans, S. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Developmental Cell*, *5*(6), 877–889. https://doi.org/10.1016/S1534-5807(03)00363-0

Cedar, H., & Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics 2009 10:5*, *10*(5), 295–304. https://doi.org/10.1038/nrg2540

Ching, S. T., Infante, C. R., Du, W., Sharir, A., Park, S., Menke, D. B., & Klein, O. D. (2018). Isl1 mediates mesenchymal expansion in the developing external genitalia via regulation of Bmp4, Fgf10 and Wnt5a. *Human Molecular Genetics*, *27*(1). https://doi.org/10.1093/hmg/ddx388

Choi, T. Y., Choi, T. I., Lee, Y. R., Choe, S. K., & Kim, C. H. (2021). Zebrafish as an animal model for biomedical research. *Experimental & Molecular Medicine 2021 53:3*, *53*(3), 310–317. https://doi.org/10.1038/s12276-021-00571-5

Dahm, P., & Gschwend, J. E. (2003). Malignant Non-Urothelial Neoplasms of the Urinary Bladder: A Review. In *European Urology* (Vol. 44, Issue 6). https://doi.org/10.1016/S0302-2838(03)00416-0

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*(10). https://doi.org/10.1038/ng.3656

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1). https://doi.org/10.1093/bioinformatics/bts635

Domcke, S., Hill, A. J., Daza, R. M., Cao, J., O'Day, D. R., Pliner, H. A., Aldinger, K. A., Pokholok, D., Zhang, F., Milbank, J. H., Zager, M. A., Glass, I. A., Steemers, F. J., Doherty, D., Trapnel, C., Cusanovich, D. A., & Shendure, J. (2020). A human cell atlas of fetal chromatin accessibility. *Science*, *370*(6518). https://doi.org/10.1126/science.aba7612

Draaken, M., Knapp, M., Pennimpede, T., Schmidt, J. M., Ebert, A. K., Rösch, W., Stein, R., Utsch, B., Hirsch, K., Boemers, T. M., Mangold, E., Heilmann, S., Ludwig, K. U., Jenetzky, E., Zwink, N., Moebus, S., Herrmann, B. G., Mattheisen, M., Nöthen, M. M., … Reutter, H. (2015a). Genome-wide Association Study and Meta-Analysis Identify ISL1 as Genome-wide Significant Susceptibility Gene for Bladder Exstrophy. *PLoS Genetics*, *11*(3). https://doi.org/10.1371/journal.pgen.1005024

Draaken, M., Knapp, M., Pennimpede, T., Schmidt, J. M., Ebert, A. K., Rösch, W., Stein, R., Utsch, B., Hirsch, K., Boemers, T. M., Mangold, E., Heilmann, S., Ludwig, K. U., Jenetzky, E., Zwink, N., Moebus, S., Herrmann, B. G., Mattheisen, M., Nöthen, M. M., … Reutter, H. (2015b). Genome-wide Association Study and Meta-Analysis Identify ISL1 as Genome-wide Significant Susceptibility Gene for Bladder Exstrophy. *PLOS Genetics*, *11*(3), e1005024. https://doi.org/10.1371/JOURNAL.PGEN.1005024

Draaken, M., Reutter, H., Schramm, C., Bartels, E., Boemers, T. M., Ebert, A. K., Rösch, W., Schröder, A., Stein, R., Moebus, S., Stienen, D., Hoffmann, P., Nöthen, M. M., & Ludwig, M. (2010). Microduplications at 22q11.21 are associated with non-syndromic classic bladder exstrophy. *European Journal of Medical Genetics*, *53*(2). https://doi.org/10.1016/j.ejmg.2009.12.005

Driever, W., Stemple, D., Schier, A., & Solnica-Krezel, L. (1994). Zebrafish: genetic tools for studying vertebrate development. In *Trends in Genetics* (Vol. 10, Issue 5). https://doi.org/10.1016/0168-9525(94)90091-4

Dupret, B., Völkel, P., Vennin, C., Toillon, R. A., Le Bourhis, X., & Angrand, P. O. (2017). The histone lysine methyltransferase Ezh2 is required for maintenance of the intestine integrity and for caudal fin regeneration in zebrafish. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1860*(10). https://doi.org/10.1016/j.bbagrm.2017.08.011

Duttke, S. H. C., Lacadie, S. A., Ibrahim, M. M., Glass, C. K., Corcoran, D. L., Benner, C., Heinz, S., Kadonaga, J. T., & Ohler, U. (2015). Human Promoters Are Intrinsically Directional. *Molecular Cell*, *57*(4), 674. https://doi.org/10.1016/J.MOLCEL.2014.12.029

Ebert, A. K., Reutter, H., Ludwig, M., & Rösch, W. H. (2009). The exstrophy-epispadias complex. *Orphanet Journal of Rare Diseases*, *4*(1). https://doi.org/10.1186/1750-1172-4-23

Ebert, A. K., Zwink, N., Reutter, H. M., & Jenetzky, E. (2021). A Prevalence Estimation of Exstrophy and Epispadias in Germany From Public Health Insurance Data. *Frontiers in Pediatrics*, *9*. https://doi.org/10.3389/fped.2021.648414

Eeles, R. A., Olama, A. A. Al, Benlloch, S., Saunders, E. J., Leongamornlert, D. A., Tymrakiewicz, M., Ghoussaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S., Dadaev, T., Neal, D. E., Hamdy, F. C.,

Donovan, J. L., Muir, K., Giles, G. G., Severi, G., Wiklund, F., Gronberg, H., … Easton, D. F. (2013). Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nature Genetics*, *45*(4). https://doi.org/10.1038/ng.2560

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjostedt, E., Lundberg, E., Szigyarto, C. A. K., Skogs, M., Ottosson Takanen, J., Berling, H., Tegel, H., Mulder, J., … Uhlen, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics : MCP*, *13*(2), 397–406. https://doi.org/10.1074/MCP.M113.035600

Fernandes, J. C. R., Acuña, S. M., Aoki, J. I., Floeter-Winter, L. M., & Muxel, S. M. (2019). Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Non-Coding RNA 2019, Vol. 5, Page 17*, *5*(1), 17. https://doi.org/10.3390/NCRNA5010017

Gao, R., Liang, X., Cheedipudi, S., Cordero, J., Jiang, X., Zhang, Q., Caputo, L., Günther, S., Kuenne, C., Ren, Y., Bhattacharya, S., Yuan, X., Barreto, G., Chen, Y., Braun, T., Evans, S. M., Sun, Y., & Dobreva, G. (2019). Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate. *Cell Research*, *29*(6). https://doi.org/10.1038/s41422-019-0168-1

Garieri, M., Delaneau, O., Santoni, F., Fish, R. J., Mull, D., Carninci, P., Dermitzakis, E. T., Antonarakis, S. E., & Fort, A. (2017). The effect of genetic variation on promoter usage and enhancer activity. *Nature Communications*, *8*(1). https://doi.org/10.1038/S41467-017-01467-7

Gearhart, J. P., Ben-Chaim, J., Jeffs, R. D., & Sanders, R. C. (1995). Criteria for the prenatal diagnosis of classic bladder exstrophy. *Obstetrics and Gynecology*, *85*(6). https://doi.org/10.1016/0029-7844(95)00069-4

Gurung, P. M. S., Barnett, A. R., Wilson, J. S., Hudson, J., Ward, D. G., Messing, E. M., & Bryan, R. T. (2020). Prognostic DNA Methylation Biomarkers in High-risk Non–muscle-invasive Bladder Cancer: A Systematic Review to Identify Loci for Prospective Validation. In *European Urology Focus* (Vol. 6, Issue 4). https://doi.org/10.1016/j.euf.2019.02.012

Hao, Y., & Li, G. (2020). Role of EFNA1 in tumorigenesis and prospects for cancer therapy. In *Biomedicine and Pharmacotherapy* (Vol. 130). https://doi.org/10.1016/j.biopha.2020.110567

Hawe, J. S., Wilson, R., Schmid, K. T., Zhou, L., Lakshmanan, L. N., Lehne, B. C., Kühnel, B., Scott, W. R., Wielscher, M., Yew, Y. W., Baumbach, C., Lee, D. P., Marouli, E., Bernard, M., Pfeiffer, L., Matías-García, P. R., Autio, M. I., Bourgeois, S., Herder, C., … Chambers, J. C. (2022). Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. *Nature Genetics*, *54*(1). https://doi.org/10.1038/s41588-021-00969-x

Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science*, *187*(4173). https://doi.org/10.1126/science.1111098

Hoshi, M., Reginensi, A., Joens, M. S., Fitzpatrick, J. A. J., McNeill, H., & Jain, S. (2018). Reciprocal spatiotemporally controlled apoptosis regulates Wolffian duct cloaca fusion. *Journal of the American Society of Nephrology*, *29*(3), 775–783. https://doi.org/10.1681/ASN.2017040380/-/DCSUPPLEMENTAL

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C.,

Koch, R., Rauch, G. J., White, S., … Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, *496*(7446). https://doi.org/10.1038/nature12111

Itou, J., Kawakami, H., Quach, T., Osterwalder, M., Evans, S. M., Zeller, R., & Kawakami, Y. (2012). Islet1 regulates establishment of the posterior hindlimb field upstream of the Hand2-Shh morphoregulatory gene network in mouse embryos. *Development*, *139*(9). https://doi.org/10.1242/dev.073056

Johnsson, P., Lipovich, L., Grandér, D., & Morris, K. V. (2014). Evolutionary conservation of long non-coding RNAs; Sequence, structure, function. In *Biochimica et Biophysica Acta - General Subjects* (Vol. 1840, Issue 3). https://doi.org/10.1016/j.bbagen.2013.10.035

Kim, J., Lee, Y., Lu, X., Song, B., Fong, K. W., Cao, Q., Licht, J. D., Zhao, J. C., & Yu, J. (2018). Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. *Cell Reports*, *25*(10). https://doi.org/10.1016/j.celrep.2018.11.035

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics*, *203*(3). https://doi.org/10.1002/aja.1002030302

Kitchen, M. O., Bryan, R. T., Haworth, K. E., Emes, R. D., Luscombe, C., Gommersall, L., Cheng, K. K., Zeegers, M. P., James, N. D., Devall, A. J., Fryer, A. A., & Farrell, W. E. (2015). Methylation of HOXA9 and ISL1 predicts patient outcome in high-grade non-invasive bladder cancer. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0137003

Kolvenbach, C. M., Dworschak, G. C., Frese, S., Japp, A. S., Schuster, P., Wenzlitschke, N., Yilmaz, Ö., Lopes, F. M., Pryalukhin, A., Schierbaum, L., van der Zanden, L. F. M., Kause, F., Schneider, R., Taranta-Janusz, K., Szczepańska, M., Pawlaczyk, K., Newman, W. G., Beaman, G. M., Stuart, H. M., … Hilger, A. C. (2019). Rare Variants in BNC2 Are Implicated in Autosomal-Dominant Congenital Lower Urinary-Tract Obstruction. *The American Journal of Human Genetics*, *104*(5), 994–1006. https://doi.org/10.1016/J.AJHG.2019.03.023

Kulkarni, B., & Chaudhari, N. (2008). Embryogenesis of bladder exstrophy: A new hypothesis. *Journal of Indian Association of Pediatric Surgeons*, *13*(2). https://doi.org/10.4103/0971-9261.43017

Lek M. et al. (2016). Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616).

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., & Dewey, C. N. (2009). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, *26*(4). https://doi.org/10.1093/bioinformatics/btp692

Liang, X., Song, M. R., Xu, Z. G., Lanuza, G. M., Liu, Y., Zhuang, T., Chen, Y., Pfaff, S. L., Evans, S. M., & Sun, Y. (2011). Isl1 Is required for multiple aspects of motor neuron development. *Molecular and Cellular Neuroscience*, *47*(3). https://doi.org/10.1016/j.mcn.2011.04.007

Liaw, A., Cunha, G. R., Shen, J., Cao, M., Liu, G., Sinclair, A., & Baskin, L. (2018). Development of the human bladder and ureterovesical junction. *Differentiation; Research in Biological Diversity*, *103*, 66–73. https://doi.org/10.1016/J.DIFF.2018.08.004

Lin, L., Bu, L., Cai, C. L., Zhang, X., & Evans, S. (2006). Isl1 is upstream of sonic hedgehog in a pathway required for cardiac morphogenesis. *Developmental Biology*, *295*(2). https://doi.org/10.1016/j.ydbio.2006.03.053

Lin, Y. C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van De Peer, Y., Tavernier, J., & Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications 2014 5:1*, *5*(1), 1–12. https://doi.org/10.1038/ncomms5767

Liu, J., Welm, B., Boucher, K. M., Ebbert, M. T. W., & Bernard, P. S. (2012). TRIM29 functions as a tumor suppressor in nontumorigenic breast cells and invasive ER+ breast cancer. *American Journal of Pathology*, *180*(2). https://doi.org/10.1016/j.ajpath.2011.10.020

Lu, H., Hao, L., Yang, H., Chen, J., & Liu, J. (2019). miRNA-34a suppresses colon carcinoma proliferation and induces cell apoptosis by targeting SYT1. *International Journal of Clinical and Experimental Pathology*, *12*(8).

Manichaikul, A., Palmer, A. A., Sen, Ś., & Broman, K. W. (2007). Significance thresholds for quantitative trait locus mapping under selective genotyping. *Genetics*, *177*(3). https://doi.org/10.1534/genetics.107.080093

Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies via imputation of genotypes : Supplementary Methods Imputation of missing genotypes Imputation of completely missing SNPs. *Nature Genetics*, *39*(7).

Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. In *Nature* (Vol. 469, Issue 7330). https://doi.org/10.1038/nature09784

Marshall, V. F., & Muecke, E. C. (1962). Variations in Exstrophy of the Bladder. *Journal of Urology*, *88*(6). https://doi.org/10.1016/s0022-5347(17)64883-3

Masden, T. B., Taela, A., Rocha, M. da, Moores, D. C., & Radulescu, A. (2019). Isolated thoracoschisis with rib agenesis and liver herniation: A case report. *American Journal of Case Reports*, *20*. https://doi.org/10.12659/AJCR.919125

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., Dsouza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., … Costello, J. F. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, *466*(7303). https://doi.org/10.1038/nature09165

Meeker, N. D., Hutchinson, S. A., Ho, L., & Trede, N. S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *BioTechniques*, *43*(5), 610–614. https://doi.org/10.2144/000112619/ASSET/IMAGES/LARGE/FIGURE2.JPEG

Mingardo, E., Beaman, G., Grote, P., Nordenskjöld, A., Newman, W., Woolf, A. S., Eckstein, M., Hilger, A. C., Dworschak, G. C., Rösch, W., Ebert, A. K., Stein, R., Brusco, A., Di Grazia, M., Tamer, A., Torres, F. M., Hernandez, J. L., Erben, P., Maj, C., … Reutter, H. (2022). A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy. *Communications Biology 2022 5:1*, *5*(1), 1–11. https://doi.org/10.1038/s42003-022-04092-3

Narkis, G., Tzchori, I., Cohen, T., Holtz, A., Wier, E., & Westphal, H. (2012). Isl1 and Ldb Co-regulators of transcription are essential early determinants of mouse limb development. *Developmental Dynamics*, *241*(4). https://doi.org/10.1002/dvdy.23761

Nordenskjöld, A., Arkani, S., Pettersson, M., Winberg, J., Cao, J., Fossum, M., Anderberg, M., Barker, G., Holmdahl, G., & Lundin, J. (2023). Copy number variants suggest different molecular pathways for

the pathogenesis of bladder exstrophy. *American Journal of Medical Genetics Part A*, *191*(2), 378–390. https://doi.org/10.1002/AJMG.A.63031

Palmbos, P. L., Wang, L., Yang, H., Wang, Y., Leflein, J., Ahmet, M. L., Wilkinson, J. E., Kumar-Sinha, C., Ney, G. M., Tomlins, S. A., Daignault, S., Kunju, L. P., Wu, X. R., Lotan, Y., Liebert, M., Ljungman, M. E., & Simeone, D. M. (2015). ATDC/TRIM29 drives invasive bladder cancer formation through miRNA-mediated and epigenetic mechanisms. *Cancer Research*, *75*(23). https://doi.org/10.1158/0008-5472.CAN-15-0603

Perner, B., Englert, C., & Bollig, F. (2007). The Wilms tumor genes wt1a and wt1b control different steps during formation of the zebrafish pronephros. *Developmental Biology*, *309*(1), 87–96. https://doi.org/10.1016/J.YDBIO.2007.06.022

Plaisance, I., Chouvardas, P., Sun, Y., Nemir, M., Aghagolzadeh, P., Aminfar, F., Shen, S., Shim, W. J., Rochais, F., Johnson, R., Palpant, N., & Pedrazzini, T. (2022). A transposable element into the human long noncoding RNA CARMEN is a switch for cardiac precursor cell specification. *Cardiovascular Research*, *00*, 1–16. https://doi.org/10.1093/CVR/CVAC191

Popiołek, A., Brzoszczyk, B., Jarzemski, P., Chyrek-Tomaszewska, A., Wieczór, R., Borkowska, A., & Bieliński, M. (2021). Prostate-specific antigen and testosterone levels as biochemical indicators of cognitive function in prostate cancer survivors and the role of diabetes. *Journal of Clinical Medicine*, *10*(22). https://doi.org/10.3390/jcm10225307

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., Willer, C. J., & Frishman, D. (2011). LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, *27*(13). https://doi.org/10.1093/bioinformatics/btq419

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3). https://doi.org/10.1086/519795

Qi, B. Q., Beasley, S. W., Williams, A. K., & Frizelle, F. (2000). Apoptosis during regression of the tailgut and septation of the cloaca. *Journal of Pediatric Surgery*, *35*(11), 1556–1561. https://doi.org/10.1053/jpsu.2000.18309

Reutter, H., Draaken, M., Pennimpede, T., Wittler, L., Brockschmidt, F. F., Ebert, A. K., Bartels, E., Rösch, W., Boemers, T. M., Hirsch, K., Schmiedeke, E., Meesters, C., Becker, T., Stein, R., Utsch, B., Mangold, E., Nordenskjöld, A., Barker, G., Kockum, C. C. lementsson, … Mattheisen, M. (2014). Genome-wide association study and mouse expression data identify a highly conserved 32 kb intergenic region between WNT3 and WNT9b as possible susceptibility locus for isolated classic exstrophy of the bladder. *Human Molecular Genetics*, *23*(20). https://doi.org/10.1093/hmg/ddu259

Reutter, H., Shapiro, E., & Gruen, J. R. (2003). Seven new cases of familial isolated bladder exstrophy and epispadias complex (BEEC) and review of the literature. *American Journal of Medical Genetics. Part A*, *120A*(2), 215–221. https://doi.org/10.1002/AJMG.A.20057

Rieke, J. M., Zhang, R., Braun, D., Yilmaz, Ö., Japp, A. S., Lopes, F. M., Pleschka, M., Hilger, A. C., Schneider, S., Newman, W. G., Beaman, G. M., Nordenskjöld, A., Ebert, A. K., Promm, M., Rösch, W. H., Stein, R., Hirsch, K., Schäfer, F. M., Schmiedeke, E., … Reutter, H. (2020). SLC20A1 Is Involved in Urinary Tract and Urorectal Development. *Frontiers in Cell and Developmental Biology*, *8*. https://doi.org/10.3389/fcell.2020.00567

Satish Kumar, K. V., Mammen, A., & Varma, K. K. (2015). Pathogenesis of bladder exstrophy: A new hypothesis. In *Journal of Pediatric Urology* (Vol. 11, Issue 6). https://doi.org/10.1016/j.jpurol.2015.05.030

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., & Sharp, P. A. (2008). Divergent transcription from active promoters. *Science (New York, N.Y.)*, *322*(5909), 1849. https://doi.org/10.1126/SCIENCE.1162253

Sigova, A. A., Mullen, A. C., Molinie, B., Gupta, S., Orlando, D. A., Guenther, M. G., Almada, A. E., Lin, C., Sharp, P. A., Giallourakis, C. C., & Young, R. A. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(8), 2876–2881. https://doi.org/10.1073/PNAS.1221904110/SUPPL_FILE/SD05.DOC

Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2020). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology 2020 22:2*, *22*(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Stepanenko, A. A., & Dmitrenko, V. V. (2015). HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene*, *569*(2), 182–190. https://doi.org/10.1016/J.GENE.2015.05.065

Su, T., Liu, H., Zhang, D., Xu, G., Liu, J., Evans, S. M., Pan, J., & Cui, S. (2019). LIM homeodomain transcription factor Isl1 affects urethral epithelium differentiation and apoptosis via Shh. *Cell Death and Disease*, *10*(10). https://doi.org/10.1038/s41419-019-1952-z

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1). https://doi.org/10.1093/nar/gku1003

Tasian, G., Cunha, G., & Baskin, L. (2010). Smooth Muscle Differentiation and Patterning in the Urinary Bladder. *Differentiation; Research in Biological Diversity*, *80*(0), 106. https://doi.org/10.1016/J.DIFF.2010.05.004

Toma, M. I., Erdmann, K., Diezel, M., Meinhardt, M., Zastrow, S., Fuessel, S., Wirth, M. P., & Baretton, G. B. (2014). Lack of ephrin receptor A1 is a favorable independent prognostic factor in clear cell renal cell carcinoma. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0102262

Villicaña, S., & Bell, J. T. (2021). Genetic impacts on DNA methylation: research findings and future perspectives. In *Genome Biology* (Vol. 22, Issue 1). https://doi.org/10.1186/s13059-021-02347-6

Warot, X., Fromental-Ramain, C., Fraulob, V., Chambon, P., & Dollé, P. (1997). Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development (Cambridge, England)*, *124*(23), 4781–4791. https://doi.org/10.1242/DEV.124.23.4781

Warrington, N. M., Shevroja, E., Hemani, G., Hysi, P. G., Jiang, Y., Auton, A., Boer, C. G., Mangino, M., Wang, C. A., Kemp, J. P., McMahon, G., Medina-Gomez, C., Hickey, M., Trajanoska, K., Wolke, D., Ikram, M. A., Montgomery, G. W., Felix, J. F., Wright, M. J., … The 23and Me Research Team. (2018). Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Human Molecular Genetics*, *27*(11). https://doi.org/10.1093/hmg/ddy121

Weiss, D. A., Oliver, E. R., Borer, J. G., Kryger, J. v., Roth, E. B., Groth, T. W., Shukla, A. R., Mitchell, M. E., Canning, D. A., & Victoria, T. (2020). Key anatomic findings on fetal ultrasound and MRI in the prenatal diagnosis of bladder and cloacal exstrophy. *Journal of Pediatric Urology*, *16*(5). https://doi.org/10.1016/j.jpurol.2020.07.024

Wilfinger, A., Arkhipova, V., & Meyer, D. (2013). Cell type and tissue specific function of islet genes in zebrafish pancreas development. *Developmental Biology*, *378*(1). https://doi.org/10.1016/j.ydbio.2013.03.009

Williamson, S. R., Lopez-Beltran, A., Montironi, R., & Cheng, L. (2011). Glandular lesions of the urinary bladder: Clinical significance and differential diagnosis. In *Histopathology* (Vol. 58, Issue 6). https://doi.org/10.1111/j.1365-2559.2010.03651.x

Wullweber, A., Strick, R., Lange, F., Sikic, D., Taubert, H., Wach, S., Wullich, B., Bertz, S., Weyerer, V., Stoehr, R., Breyer, J., Burger, M., Hartmann, A., Strissel, P. L., & Eckstein, M. (2021). Bladder tumor subtype commitment occurs in carcinoma in situ driven by key signaling pathways including ECM Remodeling. *Cancer Research*, *81*(6). https://doi.org/10.1158/0008-5472.CAN-20-2336

Yang, Z., Li, C., Fan, Z., Liu, H., Zhang, X., Cai, Z., Xu, L., Luo, J., Huang, Y., He, L., Liu, C., & Wu, S. (2017). Single-cell Sequencing Reveals Variants in ARID1A, GPRC5A and MLL2 Driving Self-renewal of Human Bladder Cancer Stem Cells. *European Urology*, *71*(1). https://doi.org/10.1016/j.eururo.2016.06.025

Zhang, R., Knapp, M., Suzuki, K., Kajioka, D., Schmidt, J. M., Winkler, J., Yilmaz, Ö., Pleschka, M., Cao, J., Kockum, C. C., Barker, G., Holmdahl, G., Beaman, G., Keene, D., Woolf, A. S., Cervellione, R. M., Cheng, W., Wilkins, S., Gearhart, J. P., … Reutter, H. (2017). ISL1 is a major susceptibility gene for classic bladder exstrophy and a regulator of urinary tract development. *Scientific Reports*, *7*. https://doi.org/10.1038/srep42170

Zhang, S., Liu, Y., Liu, Z., Zhang, C., Cao, H., Ye, Y., Wang, S., Zhang, Y., Xiao, S., Yang, P., Li, J., & Bai, Z. (2014). Transcriptome profiling of a multiple recurrent muscle-invasive urothelial carcinoma of the bladder by deep sequencing. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091466

9. List of publications

Dworschak, G. C., Punetha, J., Kalanithy, J. C., **Mingardo, E.**, Erdem, H. B., Akdemir, Z. C., Karaca, E., Mitani, T., Marafi, D., Fatih, J. M., Jhangiani, S. N., Hunter, J. V., Dakal, T. C., Dhabhai, B., Dabbagh, O., Alsaif, H. S., Alkuraya, F. S., Maroofian, R., Houlden, H., … Reutter, H. (2021). Biallelic and monoallelic variants in PLXNA1 are implicated in a novel neurodevelopmental disorder with variable cerebral and eye anomalies. *Genetics in Medicine 2021 23:9*, *23*(9), 1715–1725. https://doi.org/10.1038/s41436-021-01196-9

Enders, M., Weier, A., Chunder, R., An, Y., Bremm, F., Feigenspan, A., Buettner, C., Ekici, A. B., **Mingardo, E**., Odermatt, B., & Kuerten, S. (2023). Impact of the Voltage-Gated Calcium Channel Antagonist Nimodipine on the Development of Oligodendrocyte Precursor Cells. *International Journal of Molecular Sciences 2023, Vol. 24, Page 3716*, *24*(4), 3716. https://doi.org/10.3390/IJMS24043716

Häberlein, F., **Mingardo, E.**, Merten, N., Schulze Köhling, N. K., Reinoß, P., Simon, K., Japp, A., Nagarajan, B., Schrage, R., Pegurier, C., Gillard, M., Monk, K. R., Odermatt, B., Kostenis, E., & Gomeza, J. (2022). Humanized zebrafish as a tractable tool for in vivo evaluation of pro-myelinating drugs. *Cell Chemical Biology*, *29*(10), 1541–1555. https://doi.org/10.1016/j.chembiol.2022.08.007

Mingardo, E., Beaman, G., Grote, P., Nordenskjöld, A., Newman, W., Woolf, A. S., Eckstein, M., Hilger, A. C., Dworschak, G. C., Rösch, W., Ebert, A. K., Stein, R., Brusco, A., Di Grazia, M., Tamer, A., Torres, F. M., Hernandez, J. L., Erben, P., Maj, C., … Reutter, H. (2022). A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy. *Communications Biology 2022 5:1*, *5*(1), 1–11. https://doi.org/10.1038/s42003-022-04092-3

Mingardo, E., Odermatt, B., Haeberlein, F., Gomeza, J., Schmitt, N.-K., Kostenis, E., & Garbow, N. (n.d.). *Introduction High-Throughput Cell Analysis of a Fluorescent Zebrafish Reporter Line Using the PerkinElmer EnSight Multimode Plate Reader Multimode Detection A P P L I C A T I O N N O T E:* https://resources.perkinelmer.com/lab-solutions/resources/docs/app_mmd_ensight-zebrafish-cellcounting_014340_01.pdf .

Nagarajan, B., Harder, A., Japp, A., Häberlein, F., **Mingardo, E.**, Kleinert, H., Yilmaz, Ö., Zoons, A., Rau, B., Christ, A., Kubitscheck, U., Eiberger, B., Sandhoff, R., Eckhardt, M., Hartmann, D., & Odermatt, B. (2020). CNS myelin protein 36K regulates oligodendrocyte differentiation through Notch. *Glia*, *68*(3), 509–527. https://doi.org/10.1002/GLIA.23732

10. Appendix: A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy

communications biology

ARTICLE

https://doi.org/10.1038/s42003-022-04092-3 OPEN

A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy

Enrico Mingardo[1,2,3,59✉], Glenda Beaman[4,59], Philip Grote[5,6,59], Agneta Nordenskjöld[7,8], William Newman[4], Adrian S. Woolf[9,10], Markus Eckstein[11,12,13], Alina C. Hilger[3,14], Gabriel C. Dworschak[1,3], Wolfgang Rösch[15], Anne-Karolin Ebert[16], Raimund Stein[17], Alfredo Brusco[18], Massimo Di Grazia[19], Ali Tamer[5], Federico M. Torres[20], Jose L. Hernandez[21], Philipp Erben[13,22], Carlo Maj[23], Jose M. Olmos[24], Jose A. Riancho[24], Carmen Valero[24], Isabel C. Hostettler[25,26,27], Henry Houlden[28], David J. Werring[29], Johannes Schumacher[30], Jan Gehlen[30], Ann-Sophie Giel[30], Benedikt C. Buerfent[30], Samara Arkani[31,32], Elisabeth Åkesson[33,34], Emilia Rotstein[35], Michael Ludwig[36], Gundela Holmdahl[37], Elisa Giorgio[38,39], Alfredo Berettini[40], David Keene[41], Raimondo M. Cervellione[41], Nina Younsi[17], Melissa Ortlieb[16], Josef Oswald[42], Bernhard Haid[17,42], Martin Promm[15], Claudia Neissner[15], Karin Hirsch[43], Maximilian Stehr[44], Frank-Mattias Schäfer[44,45], Eberhard Schmiedeke[46], Thomas M. Boemers[47], Iris A. L. M. van Rooij[48], Wouter F. J. Feitz[49], Carlo L. M. Marcelis[50], Martin Lacher[51], Jana Nelson[51], Benno Ure[52], Caroline Fortmann[52], Daniel P. Gale[53], Melanie M. Y. Chan[53], Kerstin U. Ludwig[3], Markus M. Nöthen[3], Stefanie Heilmann[3,54], Nadine Zwink[55], Ekkehart Jenetzky[55,56], Benjamin Odermatt[1,2], Michael Knapp[57,59] & Heiko Reutter[3,58,59✉]

Classic bladder exstrophy represents the most severe end of all human congenital anomalies of the kidney and urinary tract and is associated with bladder cancer susceptibility. Previous genetic studies identified one locus to be involved in classic bladder exstrophy, but were limited to a restrict number of cohort. Here we show the largest classic bladder exstrophy genome-wide association analysis to date where we identify eight genome-wide significant loci, seven of which are novel. In these regions reside ten coding and four non-coding genes. Among the coding genes is EFNA1, strongly expressed in mouse embryonic genital tubercle, urethra, and primitive bladder. Re-sequence of EFNA1 in the investigated classic bladder exstrophy cohort of our study displays an enrichment of rare protein altering variants. We show that all coding genes are expressed and/or significantly regulated in both mouse and human embryonic developmental bladder stages. Furthermore, nine of the coding genes residing in the regions of genome-wide significance are differentially expressed in bladder cancers. Our data suggest genetic drivers for classic bladder exstrophy, as well as a possible role for these drivers to relevant bladder cancer susceptibility.

A full list of author affiliations appears at the end of the paper.

The bladder exstrophy-epispadias complex (BEEC) is a spectrum of congenital abnormalities which involves the abdominal wall, bony pelvis, the urinary tract, the external genitalia, and in the worse cases also the gastrointestinal tract. The BEEC represents the severe end of all human congenital anomalies of the kidney and urinary tract. The most common defect form, classic bladder exstrophy (CBE), is characterized by pubic diastasis, the evaginated bladder plate template, and an epispadic urethra. At birth, the visible bladder mucosa appears reddish and mucosal polyps may be seen on the surface. CBE is associated with kidney and other upper urinary tract anomalies with a higher occurrence in males compared to females[1]. Associated long-term complications include malignancies of the bladder comprising mainly urothelial cell carcinoma and adenocarcinoma[2,3]. Recently, the CBE live prevalence for Germany has been estimated to be ~1:30,700[4]. Given the overall European population of ~450,000,000 (https://ec.europa.eu/) citizens, presumptively ~15,000 CBE patients live in Europe. State-of-the-art health care for this population should take the genetic and bladder cancer disposition into account.

To determine the genetic contribution to CBE, we previously performed two genome-wide association studies (GWAS) with subsequent meta-analysis and identified a susceptibility locus on chromosome 5q11.1[5,6]. The present study aimed to identify further risk loci. Furthermore, we investigated if the identified genetic risk loci might be involved in the associated bladder cancer susceptibility. For this purpose, we performed the largest GWAS for CBE to date comprising 628 patients and 7352 ethnically matched controls. In detail, the present meta-analysis included seven independent discovery samples (Supplementary Information: Supplementary Table 1) comprising: 98 patients of Central European origin and 526 ethnically matched controls[5], 110 patients of Central European origin and 1,177 ethnically matched controls[6], 172 patients of Central European origin and 2588 ethnically matched controls, 57 patients of Italian origin and 1,325 ethnically matched controls, 62 patients of Spanish origin and 279 ethnically matched controls, 80 patients of Swedish origin and 238 ethnically matched controls, and 49 patients of UK origin and 1,219 ethnically matched controls, identifying eight genome-wide significant risk loci, seven of which are novel. Within these loci reside 10 coding genes (*LPHN2*, *EFNA1*, *SLC50A1*, *DPM3*, *KRTCAP2*, *ISL1*, *TRIM29*, *SYT1*, *PAWR*, *GOSR2*) and four non-coding genes (one pseudogene and three long non-coding RNA, respectively, HMGB1P47, ISL1-DT, LINC01974, and LINC01716). Among these coding genes, *EFNA1* has been previously shown to be strongly expressed in mouse embryonic genital tubercle, urethra, and primitive bladder prompting us to re-sequence this gene in our cohort. To assess their embryonic and fetal expression, we generated mouse embryonic bladder total RNA-seq at CBE-relevant developmental stages E10.5, E12.5, and E15.5, and human embryonic and fetal urinary bladder and genital tissues total RNA-seq at gestational week 7, 7 to 7.5, 7.5, 8, and 9. Finally, to evaluate their possible link in the overall CBE bladder cancer susceptibility, we analyzed the expression of these genes in urothelial carcinoma tissues and in different bladder cancer cell lines obtained from the Cancer Cell Line Encyclopedia (EMBL-EBI) compared to healthy bladder tissue transcriptomic (GEO).

Results

GWAS meta-analysis. The meta-analysis of 628 patients with CBE and 7,352 ethnically matched controls comprised seven independent GWAS. These seven GWAS included the first two GWAS cohorts[5,6], and five new CBE cohorts described above from Central European, Italy, Spain, Sweden, and the UK along with ethnically matched control samples. We used a total of 8,289,003 SNPs with info score >0.4 and mean dosage for the minor allele >1% in cases and controls in at least one sample, obtaining a genomic inflation factor λ of 1.068. The respective Q-Q plot is shown in Supplementary information (Supplementary Fig. 16). Single marker analysis identified eight genome-wide significant loci shown in the Manhattan plot in Supplementary information (Supplementary Fig. 17) and the strongest signal at rs6874700 $p = 5.58 \times 10^{-24}$ corresponds to the 5q11.1 previously reported locus (Fig. 1)[6]. Table 1 shows the relative risks in each sample and in the meta-analysis for the most strongly associated SNP (top SNP) from each locus. Notably, with the exception of the UK sample where the top associated SNP on chromosome 12 was not significant, the direction of effect was consistent between all studies for these top SNPs. A complete list of all genome-wide significant SNPs is given in Supplementary Data 2. Regional association results for all eight genome-wide significant loci are shown in Fig. 1 and in Supplementary Information (Supplementary Figs. 1–8). For conditional logistic regression analyses the regional association plots are presented in Supplementary Information (Supplementary Figs. 20–27). The results provide no evidence that secondary signals in any of the eight loci are present.

Re−sequencing of *EFNA1*. Among the most significant markers, marker rs4745 resides directly in *EFNA1*. Mouse *Efna1* has been shown to be expressed in CBE-relevant embryonic anatomical structures (https://www.gudmap.org/). This prompted us to re-sequence *EFNA1* in 580 CBE patients. We identified 14 rare variants in 14 independent patients (Supplementary information, Supplementary Table 4). Four of these variants residing in the coding region of *EFNA1* were found to be novel: two heterozygous missense variants c.116 T > C (p.Ile39Thr) and c.503 C > T (p.Ala168Val); one homozygous missense variant c.167 A > G (p.Asp56Gly), and a heterozygous loss of function (LoF) frameshift variant at c.341delT (p.Phe114Serfs*28). Parental samples were only available for the patient carrying variant c.116 T > C demonstrating paternal transmission. In silico prediction tools, Mutation Taster, Poly-Phen-2, and SIFT defined the missense variant c.116 T > C (p.Ile39Thr) as disease-causing, deleterious, and possibly damaging. The CADD score of 25.3 supports a functional implication of this variant on *EFNA1* regulation. None, of the other missense variants, were scored deleterious. The LoF variant c.341delT (p.Phe114Serfs*28) has a CADD score of 25.6. For the estimation of the enrichment of rare protein-altering variants in EFNA1 in our cohort compared to the general population resembled by gnomAD, we use a very conservative comparison. Hence, we only included the three novel coding variants with CADD score >20, identified here in our re-sequencing approach of *EFNA1*. We compared these to missense or LoF variants in gnomAD less or equal to 5 (≤5 in 250,000; MAF ≤ 0,00002) consistent with rare penetrant dominant phenotypes. These criteria identified 162 missense and LoF variants in gnomAD (baseline 250,000 alleles; https://gnomad.broadinstitute.org/gene/ENSG00000169242?dataset=gnomad_r2_1). Per se, it is possible that some of these variants are cis/trans in the same individuals but if we would be able to define this it would only make the association stronger. For comparison, we used Fisher's exact test. Taking this assumption, the chi-square statistic using Fisher´s exact test, yielded 18.0159, and the p value is 0.000022. We added this statistic to our results.

Analysis of mouse and human embryonic total RNA-sequencing data of the identified genes in mouse and human embryonic and fetal urinary bladder and genital tissues. In the linkage disequilibrium block of all eight top SNPs reside 10

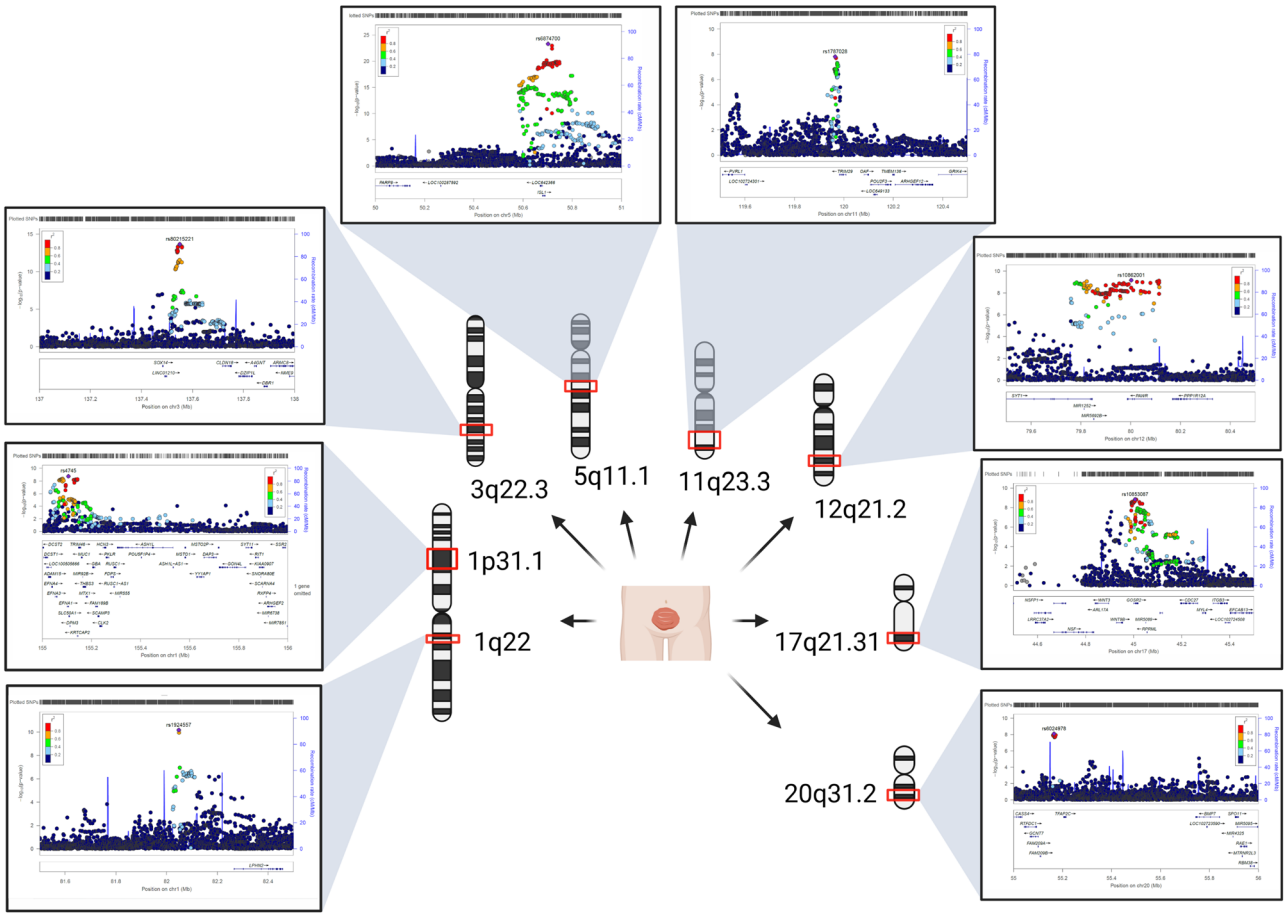


Fig. 1 Chromosome regional association loci of CBE. The eight CBE regional association loci (red boxes in chromosomes) reside in chromosome 1, 3, 5, 11, 12, 17, and 20. In the panels, details of the genome-wide association loci: every dot represents an SNP (*x* axes) plotted to the relative $-\log_{10}$(*P* value) (*y* axes). SNP are colored according to the relative r^2 value.

coding genes, four non-coding genes comprising one pseudogene, and three Long Intergenic Non-Protein Coding RNAs (Table 2). All 10 coding genes showed expression in mouse embryonic bladder at E10.5, E12.5, and E15.5. *Isl1*, *Trim29*, *Syt1*, and *Pawr* showed differential expression through different mouse embryonic stages (Table 3, Fig. 2a). As in mouse transcriptome, all 10 coding genes showed expression in the human embryonic bladder during different developmental stages. *DPM3*, *TRIM29*, *SYT1*, and *PAWR* showed differential expression at different human embryonic respectively fetal stages (Table 3, Fig. 2b). While the two *LINC01974* and *LINC01716* are not expressed in any of the bladder developmental stages, the pseudogene *HMGB1P45* and the long noncoding RNA *ISL1-DT* are strongly downregulated; the first from weeks 7 to 7.5 followed by gene silencing at week 8 and 9 and the latter shows a downregulation trend from weeks 7 to 9.

Comparison of RNA-sequencing data of the identified genes in healthy bladder tissue and different types of bladder cancer cell lines and muscular invasive urothelial carcinoma. Expression of non-coding genes in healthy bladder tissue and muscular invasive urothelial carcinoma could not be detected due to polyA-RNA-sequencing (Table 4). Despite this, we found all coding genes to be differentially expressed when compared to healthy bladder tissue (Table 5, Fig. 2c). In detail, *SLC50A1* and *SYT1* were significantly upregulated in all cancers compared to healthy bladder tissue. On contrary, *DPM3* and *KRTCAP2* were significantly downregulated. All other coding genes *LPHN2*, *EFNA1*, *ISL1*, *TRIM29*, *PAWR*, and *GOSR2* were differentially up- or downregulated in different cancers (Table 5, Fig. 2c).

Discussion

Recently, we described *SLC20A1*, encoding a sodium-phosphate symporter, as the putative monogenic dominant disease gene for isolated BEEC[7]. We were able to support our genetic data through functional studies in non-BEEC human embryos, mouse embryos, and zebrafish Morpholino knockdown experiments. To our knowledge, the present genetic study with a focus on the multifactorial genetic background of the BEEC is the largest study on CBE to date. We identified eight genome-wide significant risk loci. Within these loci, we determined possible CBE candidate genes using transcriptome datasets of CBE-relevant mouse embryonic, human embryonic, and fetal urogenital tissues at different developmental stages. Additionally, we provide a possible link between the identified putative candidate genes and CBE-associated bladder cancer susceptibility.

In detail, in direct proximity to the most significant markers of all eight risk loci reside nine coding genes that are expressed in CBE-relevant mouse and human urogenital tissues during different embryonic stages. Four of these candidate genes (*Isl1*, *Trim29*, *Syt1*, *Pawr*) showed differential expression in mouse embryonic urogenital tissues, five of these candidate genes (*DPM3*, *ISL1*, *TRIM29*, *SYT1*, and *PAWR*), and two of the non-coding genes (*HMGB1P47*, *ISL1-DT*) showed differential expression in human embryonic urogenital tissues. Previous reports of transgenic mouse lines of *Isl1* and *Syt1* revealed

Table 1 Top SNPs from genome-wide significant loci.

SNP	Chromosome	risk/ other allele	Info Score[1]	GWAS1 RR (95% CI)	GWAS2 RR (95% CI)	CE RR (95% CI)	Italy RR (95% CI)	Spain RR 95% CI	Sweden RR 95% CI	UK RR 95% CI	Meta-analysis RR 95% CI	*p* value	Corrected *p* value[2]
rs1924557	1:82047934	T/C	0.93	1.75 (0.68–4.47)	3.61 (1.37–9.49)	3.76 (2.16–6.53)	2.03 (0.98–4.19)	3.51 (1.33–9.26)	1.23 (0.68–2.22)	5.00 (1.59–15.7)	2.35 (1.82–3.03)	6.65×10^{-11}	2.67×10^{-10}
rs4745	1:155106227	T/A	0.96	1.19 (0.72–1.98)	1.40 (0.91–2.16)	1.47 (1.16–1.87)	1.80 (1.16–2.78)	1.17 (0.73–1.86)	1.63 (1.13–2.36)	1.16 (0.77–1.74)	1.50 (1.32–1.71)	4.44×10^{-10}	1.58×10^{-9}
rs80215221	3:137550420	G/A	0.99	2.29 (1.29–4.09)	1.95 (1.20–3.17)	1.39 (1.08–1.79)	1.89 (1.24–2.89)	2.82(1.76–4.52)	1.70 (1.16–2.51)	1.99 (1.31–3.03)	1.68 (1.47–1.92)	2.26×10^{-14}	1.49×10^{-13}
rs6874700	5:50701750	A/T	0.99	2.54 (1.49–4.33)	1.33 (0.87–2.02)	1.60 (1.28–2.01)	1.98 (1.33–2.95)	2.14 (1.33–3.45)	1.94 (1.32–2.83)	1.84 (1.20–2.80)	1.91 (1.68–2.16)	5.58×10^{-24}	1.48×10^{-22}
11:119964758:G:C	11:119964758	G/C	0.98	1.38 (0.81–2.36)	1.54 (0.98–2.42)	1.32 (1.03–1.69)	1.35 (0.90–2.04)	1.14(0.72–1.80)	1.68 (1.12–2.52)	2.13 (1.40–3.24)	1.47 (1.29–1.68)	8.89×10^{-9}	2.63×10^{-8}
rs10862001	12:80002106	A/G	0.83	2.02 (1.12–3.64)	1.48 (0.89–2.47)	1.77 (1.27–2.45)	1.97 (1.22–3.18)	1.03 (0.55–1.93)	2.11 (1.26–3.54)	0.83 (0.40–1.69)	1.68 (1.43–1.99)	6.84×10^{-10}	2.37×10^{-9}
rs10853087	17:45006112	C/G	0.97	1.18 (0.72–1.93)	1.44 (0.94–2.21)	1.63 (1.29–2.06)	1.60 (1.08–2.38)	1.46 (0.93–2.29)	1.26 (0.87–1.83)	1.53 (1.01–2.31)	1.47 (1.30–1.66)	1.41×10^{-9}	4.67×10^{-9}
20:55165923:G:A	20:55165923	G/A	0.97	2.30 (1.13–4.67)	1.72 (0.95–3.12)	1.46 (1.08–1.98)	2.57 (1.39–4.76)	1.00 (0.58–1.71)	2.08 (1.23–3.51)	1.89 (1.04–3.43)	1.65 (1.39–1.96)	7.31×10^{-9}	2.19×10^{-8}

[1]Mean imputation quality score (Info score)), Chromosomal positions are annotated according to Human Genome version 19 (hg19); *CE* Central Europe, *UK* United Kingdom
[2]*p* value corrected for genomic inflation $\lambda = 1.068$

phenotypic overlap to the human CBE phenotypic spectrum. The *Hoxb6Cre*;*Isl1* cKO hindlimb skeletons exhibited proximal defects in particular the os pubis and ischium, two posterior segments of the pelvic girdle, were missing, resembling pubic diastasis, a human BEEC-specific feature[1,8]. Transgenic Syt1$^{tm1a(EUCOMM)Wtsi}$/Syt1$^{tm1a(EUCOMM)Wtsi}$ mice among other features develop thoracoschisis[9], a rare congenital anomaly characterized by the evisceration of intra-abdominal organs through a thoracic wall defect[10] mirroring the BEEC associated infraumbilical abdominal wall defect[1].

One of the most significant markers identified in the present GWAS resides within *EFNA1*. In general, a probability of being LoF intolerant (pLi) score of 0.46 for *EFNA1* is suggestive of possessing LoF intolerance for this gene in the context of the CBE condition. Although the pLi of 0.46 is only suggestive of LoF we have to consider that bladder exstrophy is not a mortal condition at birth. Hence, we believe that a value of 0.46 is supportive for *EFNA1* to be implied in CBE[11]. Previously, in *EFNA1* only two LoF variants were observed in the entire gnomad (frequency of 0,000016). Here, we observed one in 580 (frequency of 0,0017). Furthermore, in *EFNA1* in the entire gnomad database, 96 missense variants were observed in 125,099 individuals (frequency of 0,00077). Here, we identified three in 580 (0,0052). Based on this observation, we performed a conservative estimation of whether LoF or missense variants might be enriched in *EFNA1* in our CBE cohort compared to the general population showing a significant difference between both cohorts (p 0.000022). This finding suggests a possible implication of these variants in CBE formation in a multifactorial inheritance model among the affected.

Comparative analysis of control and bladder cancer tissues showed that all of the ten candidate genes were differentially expressed in bladder cancers. *SLC50A1* and *SYT1* were significantly upregulated in all cancers compared to healthy bladder tissue. On contrary, *DPM3* and *KRTCAP2* were significantly downregulated. *LPHN2* has been suggested to have a regulatory role in urothelial bladder cancer[12]. *EFNA1* plays a pivotal role in the pathogenesis of several tumors, including renal cell carcinoma, bladder, and prostate cancer[13,14]. Mapping all putative candidate genes prioritized in the present study to the search tool for retrieval of interacting genes (STRING), we found probable interaction of three proteins comprising: (i) gene fusions between EFNA1 and SLC50A1, and (ii) co-expression between EFNA1 and DPM3, (iii) and EFNA1 and SLC50A1 (Supplementary information: Supplementary Fig. 28). The PPI enrichment p-value was determined with 0.000205. Gene clustering analysis suggested clustering for *EFNA1*, *DPM3*, and *SLC50A1* (Supplementary information: Supplementary Fig. 28). All three genes *EFNA1*, *DPM3*, and *SLC50A1*, respectively the genomic region 1q21-q22, been previously associated with the 2D:4D ratio, a sexually dimorphic trait, that has been extensively used in adults as a biomarker for prenatal androgen exposure[15]. Markers in the region of *EFNA1*, *DPM3, SLC50A1* have previously been associated with prostate cancer risk[16]. Prostate cancer risk on the other side correlates with serum testosterone levels[17]. All of these observations suggest a possible gene-environmental interaction for this region. Adding to this hypothesis in the context of embryonic CBE formation, CBE presents with a higher occurrence rate in males compared to females[1], a skewed sex ratio that is so far not explained, but could be influenced by differences in intrauterine androgen exposure between males and females.

The tumor suppressor gene *TRIM29* is up regulated during early and late embryonic bladder development but is downregulated in three different bladder cancers[18]. More specific, TRIM29 protein has been shown to be a driver of invasive and non-invasive bladder cancer. Interestingly, TRIM29-driven bladder cancers in transgenic mice were indistinguishable from

Table 2 Coding and non-coding genes in the LD blocks of the most significant markers for CBE.

Gene name	Associated SNP	Protein encoded	Type	Cellular location	Function
LPHN2	rs1924557	Adhesion G protein-coupled receptor L2	G protein-coupled receptor	Transmembrane of plasma membrane	Exocytosis regulator
EFNA1	rs4745	Ephrin-A1	Tyrosine kinase receptor	Transmembrane of plasma membrane	GPI-bound ligand for Eph receptors, involved in cell migration, repulsion, and adhesion
SLC50A1	rs4745	Solute Carrier Family 50 Member 1	Glucose transporter	Transmembrane of plasma membrane	Sugar transport across membranes
DPM3	rs4745	Dolichol-phosphate mannosyltransferase subunit 3	Synthase of mannosyl residual	Endoplasmic reticulum membrane	Stabilizer subunit of DPM complex (DPM1, DPM2, and DPM3)
KRTCAP2	rs4745	Kerotinocyte-associated protein 2	Subunit of the oligosaccharyl transferase complex	Endoplasmic reticulum membrane	Protein N-glicosilation. Transfer of defined glycan (Glc(3)Man(9)GlcNAc(2).
ISL1	rs6874700	Insulin Gene Enhancer Protein ISL-1	Transcription Factor	Nucleus	DNA-binding transcriptional activator
TRIM29	11:119964758	Tripartite motif-containing protein 29	Zinc finger and Leucine zipper motif		Nucleic acid binding and macrophage activation
SYT1	rs10862001	Synaptotagmin-1	Ca(2+) sensor	Transmembrane of synaptic vesicles	Triggering neurotransmitter release
PAWR	rs10862001	Pro-Apoptotic WT1 Regulator	Apoptosis inducer	Nucleus and cytoplasm	Downregulation of BCL2 via its interaction with WT1
GOSR2	rs10853087	Golgi SNAP receptor complex member 2	Trafficking membrane protein	Golgi	Protein transport from the cis/medial-Golgi
LINC01974	rs10853087	/	Long ncRNA	Unknown	Unknown
LINC01716	20:55165923	/	Long ncRNA	Unknown	Unknown
HMGB1P47	rs6874700	/	Long ncRNA	Unknown	Unknown
ISL1-DT	rs6874700	/	Long ncRNA. ISL1 divergent transcript.	Unknown	Unknown

Table 3 RNA expression patterns of coding and non-coding genes in the LD blocks of the most significant GWAS markers in mouse and human embryonic urogenital tissue.

Mouse embryonic transcriptome data					Human embryonic transcriptome data						
Marker	**Gene**	**log2fc E10.5 vs. E12.5**	**log2fc E10.5 vs. 15.5**	**log2fc E12.5 vs. 15.5**	**Gene**	**log2fc 7 vs. 7-7.5**	**log2fc 7-7.5 vs. 7.5**	**log2fc 7 vs. 7.5**	**log2fc 7.5 vs. 8**	**log2fc 7 vs. 9**	**log2fc 8 vs. 9**
rs1924557	LPHN2	−0.21	0.57	−1.16	LPHN2	0.16	−0.54	−0.38	0.09	−0.10	0.19
rs4745	EFNA1	−0.12	−0.76	−0.64	EFNA1	−0.25	−0.34	−0.58	0.00	−0.30	0.28
rs4745	SLC50A1	0.80	1.34	0.54	SLC50A1	0.03	0.07	0.10	−0.28	−0.18	0.00
rs4745	DPM3	0.06	0.86	0.81	**DPM3**	0.00	**−1.72**	**−1.72**	0.19	−0.52	1.00
rs4745	KRTCAP2	0.27	0.31	−0.04	KRTCAP2	−0.58	0.00	−0.58	0.58	−0.58	0.58
rs6874700	**ISL1**	−0.26	**−5.01**	**−4.76**	ISL1	−1.26	0.88	−0.38	−1.37	−1.48	0.28
11:119964758	**TRIM29**	**1.90**	1.10	−0.80	**TRIM29**	**2.00**	−1.26	0.74	0.00	**1.87**	1.14
rs10862001	**SYT1**	**2.85**	**2.41**	−0.44	**SYT1**	−0.25	1.04	0.79	**−2.54**	0.08	**1.8**
rs10862001	**PAWR**	−0.32	1.37	**1.69**	**PAWR**	0.70	0.55	1.25	**−1.66**	0.32	0.74
rs10853087	GOSR2	−0.16	−0.44	−0.28	GOSR2	−0.14	0.34	0.19	−0.19	−0.14	−0.14
Marker	**Non-coding Gene**	**log2fc E10.5 vs. E12.5**	**log2fc E10.5 vs. 15.5**	**log2fc E12.5 vs. 15.5**	**Non-coding gene**	**log2fc 7 vs. 7-7.5**	**log2fc 7-7.5 vs. 7.5**	**log2fc 7 vs. 7.5**	**log2fc 7.5 vs. 8**	**log2fc 7 vs. 9**	**log2fc 8 vs. 9**
rs6874700	**HMGB1P47**	n.e.			**HMGB1P47**	**−2.00**	0.00	**−2.00**	activated	**−3.00**	suppressed
rs6874700	**ISL1−DT**	n.e.			**ISL1−DT**	−1.38	0.85	−0.53	−1.8	**−1.61**	0.00
rs10853087	LINC01974	n.e.			LINC01974			n.e.			
20:55165923	LINC01716	n.e.			LINC01716			n.e.			

Bold: differential expressed genes (log2fc <−1.5 or >1.5).
log2fc log2 fold change, *vs.* versus, *Chr.* chromosome, *n.e.* not expressed. Differential expression defined with log2fc <−1,5 or >1,5.

gene expression signatures of human bladder cancers[19]. *PAWR*, has been previously shown to be a key altered gene in human bladder cancer stem cells[20]. *SYT1* has been reported as a possible oncogene in colon cancer[21]. The knockdown of *SYT1* markedly inhibits colon cancer cell proliferation, migration, and invasion, and induces cell apoptosis, indicating that *SYT1* may function as an oncogene in colon cancer[21]. *ISL1* has been associated with high-risk non-muscle-invasive bladder cancer in several studies[22,23]. Here we found downregulation of *ISL1* in embryonic stages of mouse and human CBE urogenital tissues. Vice versa we found dysregulation of *ISL1* expression in three bladder cancers. Hence, dysregulation of *ISL1* expression in human embryonic and adult bladder tissues might contribute to the CBE and bladder malignancies vice versa.

To date, this is the largest genetic study on CBE. We have identified eight genome-wide significant risk loci. Our transcriptomic analysis of CBE-relevant mouse embryonic, human embryonic, and fetal urogenital tissues suggests candidate genes within these loci. Bladder cancer transcriptomic suggests these candidate genes play a possible role in the CBE-associated bladder cancer susceptibility. Identification of the different expressions to turn these developmental genes on later in life might ultimately lead to preventive strategies for bladder cancer per se.

Methods

Patients and recruitment. This study was approved by the institutional ethics committee of each participating center. All experimental protocols were approved by the institutional committee of the University of Bonn (Lfd.Nr.031/19). The study was conducted according to the Declaration of Helsinki principles. Written informed consent was obtained from all patients, guardians, and healthy controls. We included 420 newly recruited isolated CBE patients and 5,649 healthy controls of European origin. Details can be found in the Supplementary information (Supplementary Table 1). Details about the 208 CBE patients and 1,703 ethnically matched controls of our previous studies, included in the present meta-analysis, are described in [5,6], in summary, CBE patients were recruited under written informed consent by BEEC expert physicians.

Sample description. In addition to the two previously described samples GWAS1 and GWAS2[6], five new samples of patients with bladder exstrophy and representative controls were obtained from Central Europe, the United Kingdom, Italy, Spain, and Sweden. The number of cases and controls used in this study are shown in Supplementary Information (Supplementary Table 1).

Genotyping. All samples, cases, and controls, were genotyped on Illumina human genotyping arrays. In GWAS 1 (Reutter et al. 2014), cases and controls were genotyped in two batches. Due to the discontinuation of the genotyping arrays utilized for earlier batches, different arrays were used comprising Illumina's Human610-Quad (H610Q) and Human660W-Quad Bead Chips and the Illumina HumanOmni1-Quad-v1 Bead Chip. In GWAS 2[6], all cases and controls were genotyped using the Illumina BeadChip HumanOmniExpress. The five novel GWAS case samples were newly genotyped simultaneously using the Illumina "Infinium Global Screening Array-24 v2.0". The five novels ethnically matched control samples were also genotyped using the Illumina "Infinium Global Screening Array-24 v2.0". However, the five novel control samples were not genotyped together with the five novel case samples but independently of each other.

Quality control of individuals. An individual was excluded if (i) the call rate was <97%; (ii) the rate of autosomal heterozygosity deviates more than six standard deviations from the mean; (iii) the rate of X-chromosomal heterozygous genotypes was >2% for a supposed male individual or <10% for a supposed female individual. PLINK version 1.9 and KING were used to detect pairs of closely related individuals within and between samples[24,25]. From each pair of individuals with an estimated identity by descending probability >0.2 or kinship coefficient >0.0884, the individual with a higher rate of missing genotypes was discarded. Individuals being outliers in a multidimensional scaling analysis (MDS) were removed. The post-quality control sample sizes are presented in Supplementary Table 1.

Principal component analysis. Individuals with:

$$\left(\frac{c_1 - m_1}{s_1}\right)^2 + \left(\frac{c_2 - m_2}{s_2}\right)^2 \geq 81 \quad (1)$$

were considered to be outliers. c_1 and c_2 denote the first two MDS coordinates of the individual and m_1, m_2 and s_1 and s_2 denote the mean and standard deviation, respectively, of the first two MDS coordinates in European HapMap individuals. For the five new samples (Central Europe, Italy, Spain, Sweden, and the UK), Supplementary Information (Supplementary Figs. 9–13) show the first two MDS coordinates for all genotyped individuals together with Asian and African HapMap individual. For the samples of GWAS1 and GWAS2, the MDS coordinates used in our previous studies are shown in Supplementary Information (Supplementary Figs. 14, 15).

Quality control of variants and imputation. Separately in each of the different ethnicity samples, SNPs were removed if (i) the minor allele frequency was <1% in

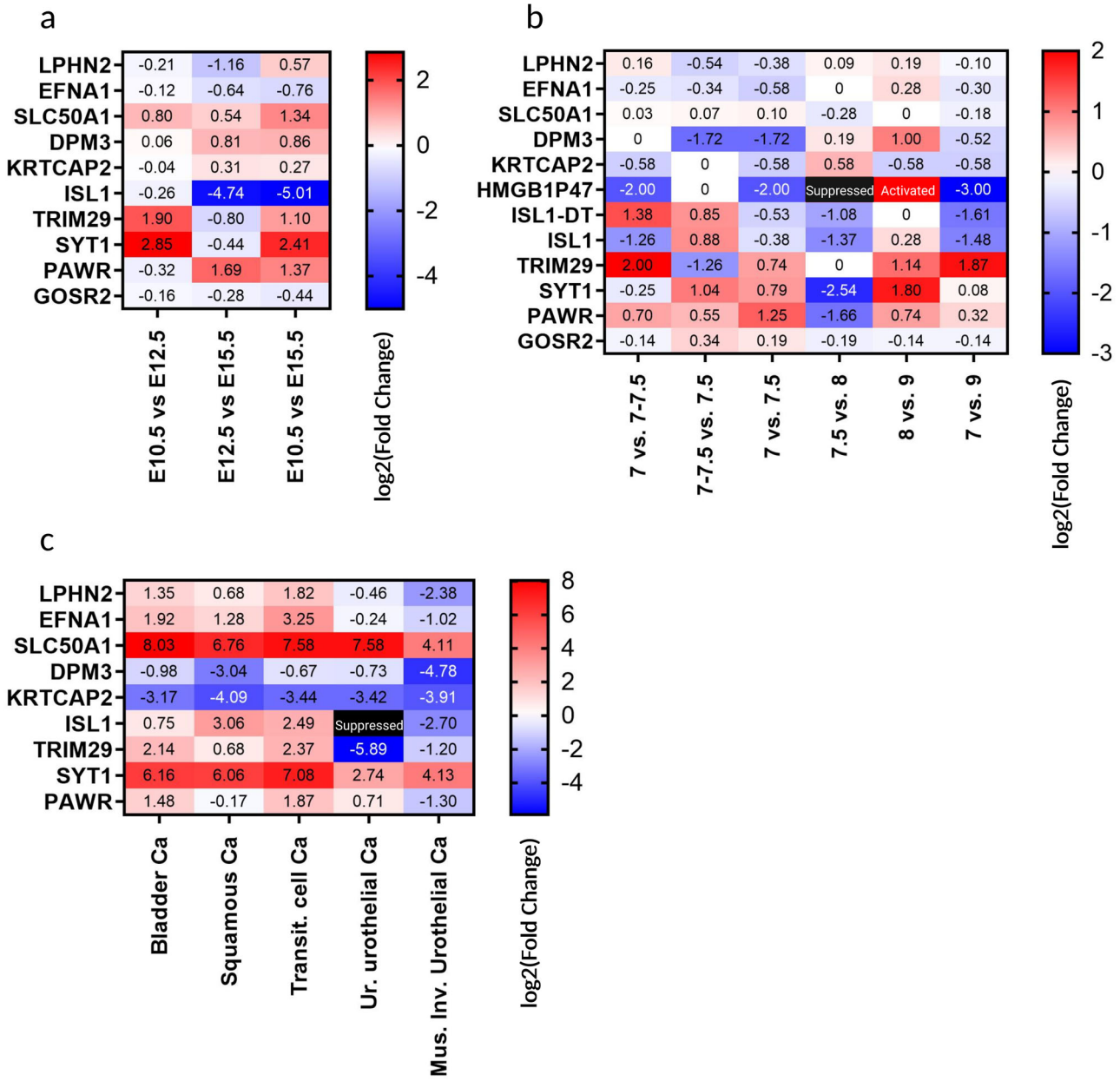


Fig. 2 Expression heatmaps of genes that reside in the LD blocks of the eight significant genetic markers. a Genes expression pattern in mouse embryonic bladder from embryonic day E10.5 to E12.5, E12.5 to E15.5, and E10.5 to E15.5. **b** Genes expression pattern in human embryonic and fetal bladder from week 7 to 7-7.5, 7-7.5 to 7.5, 7 to 7.5, 7.5 to 8, 8 to 9, and 7 to 9. **c** Genes expression pattern of 3-year-old control bladder tissue compared to Bladder carcinoma (Bladder Ca), Bladder squamous cell carcinoma (Squamous cell Ca), Bladder transitional cell carcinoma (Transit. cell Ca), Ureter urothelial carcinoma (Ur. Urothelial Ca), Muscle invasive urothelial cancer (Mus. Inv. Urothelial Ca). Legend: Suppressed = gene is silenced and no expression is detected. Activated = gene shows expression after a silenced state.

either cases or controls; (ii) the successful genotyping rate was >95% in either cases or controls; (ii) the p value for Hardy–Weinberg equilibrium was $<10^{-4}$ in controls or $<10^{-6}$ in cases. SNPs satisfying the quality filters were uploaded for imputation on the University of Michigan Imputation Server using the Haplotype Reference Consortium panel[26].

Statistical analysis for genome-wide association analysis. Association testing was performed by logistic regression using SNPTEST version 2.5.2 for the allele dosage and adjusted for the sample-specific top five MDS coordinates[27]. For each SNP, a meta-analysis with the fixed-effects inverse variance-weighting approach was conducted by including only those samples in which the info score was >0.4 and the mean dosage for the minor allele was >1% in cases and controls for the respective SNP. SNPs reaching a p value $<5*10^{-8}$ in the meta-analysis are considered to be genome-wide significant. Q-Q and Manhattan plots for the meta-analysis were created by SAS[28]. Regional association plots for genome-wide significant loci were generated with LocusZoom[29]. To look for secondary signals of association in loci of genome-wide significance, logistic regression using SNPTEST conditioned on the most associated SNP in the locus was carried out.

Protein–protein interaction networks analysis. Putative candidate genes within identified risk loci were mapped to the STRING to acquire protein–protein interaction (PPI) networks (https://string-db.org/). The search tool integrates both known and predicted PPIs. Here it was used to predict functional interactions of proteins[30,31]. Active interaction sources, including text mining, experiments, databases, and co-expression as well as species limited to "Homo sapiens" and an interaction score >0.4 were applied to construct the PPI networks. In the networks, the nodes correspond to the proteins and the edges represent the interactions. STRING was employed to seek potential interactions among putative candidate genes. Active interaction sources, including experimental repositories, computational prediction methods, and public text collections as well as species limited to "Homo sapiens" and a combined score >0.4, were applied.

Table 4 Average TPM of coding and non-coding genes in the LD blocks of the most significant GWAS markers calculated for each bladder cancer cell line.

Marker	Coding and non-coding genes	Control tissue	Bladder Ca	Transit. cell Ca	Squamous cell Ca	Ur. urothelial Ca	Mus. Inv. Urothelial Ca
rs1924557	LPHN2	6.87	17.47	24.25	11.00	5.00	1.32
rs4745	EFNA1	8.25	31.30	78.50	20.00	7.00	4.06
rs4745	SLC50A1	0.24	62.60	46.00	26.00	46.00	4.15
rs4745	DPM3	131.29	66.55	82.50	16.00	79.00	4.77
rs4745	KRTCAP2	170.89	18.95	15.75	10.00	16.00	11.37
rs6874700	ISL1	0.36	0.61	2.03	3.00	0.00	0.06
11:119964758	TRIM29	11.86	52.27	61.25	19.00	0.20	5.15
rs10862001	SYT1	0.06	4.30	8.13	4.00	0.40	1.05
rs10862001	PAWR	13.47	37.65	49.25	12.00	22.00	5.46
rs10853087	GOSR2	11.25	16.80	12.50	5.00	16.00	3.69
rs6874700	HMGB1P47	RNA-polyA-seq	0.00	0.00	0.00	0.00	RNA-polyA-seq
rs6874700	ISL1-DT	RNA-polyA-seq	0.11	0.20	0.5	0.00	RNA-polyA-seq
rs10853087	LINC01974	RNA-polyA-seq	0.04	0.00	0.00	0.00	RNA-polyA-seq
20:55165923	LINC01716	RNA-polyA-seq	0.01	0.00	0.00	0.00	RNA-polyA-seq

urothelial carcinoma tissues and control bladder tissue.
TPM for Non-coding genes in Control tissue and Muscolar invasive urothelial cancer do not show reads due to polyA-sequencing. Legend: Bladder tissue of a 3-year-old bladder donor (control tissue); Bladder carcinoma (Bladder Ca); Bladder transitional cell carcinoma (Transit. cell Ca); Bladder squamous cell carcinoma (Squamous Ca); Ureter urothelial carcinoma (Urothelial Ca); Bladder urothelial carcinoma (Bl. urothelial Ca); Muscle invasive urothelial cancer (Mus. Inv. Urothelial Ca); RNA-polyA-seq data do not include miRNAs or lincRNAs.

Table 5 log2-fold change of genes in the LD blocks of the most significant GWAS markers of bladder cancer cell types and muscle-invasive urothelial cancer over bladder control tissue.

Marker	Coding and non-coding genes	Control tissue vs. bladder Ca	Control tissue vs. transit. cell Ca	Control tissue vs. squamous cell Ca	Control tissue vs. Ur. urothelial Ca	Control tissue vs. Mus. Inv. urothelial Ca
rs1924557	**LPHN2**	1.35	**1.82**	0.68	0.46	**−2.38**
rs4745	**EFNA1**	**1.92**	**3.25**	1.28	−0.24	−1.02
rs4745	SLC50A1	**8.03**	**7.58**	**6.76**	**7.58**	**4.11**
rs4745	**DPM3**	−0.98	−0.67	**−3.04**	−0.73	**−4.78**
Rs4745	**KRTCAP2**	**−3.17**	**−3.44**	**−4.09**	**−3.42**	**−3.91**
rs6874700	**ISL1**	0.75	**2.49**	**3.06**	suppressed	**−2.70**
11:119964758	**TRIM29**	**2.14**	**2.37**	0.68	**−5.89**	−1.20
rs10862001	**SYT1**	**6.16**	**7.08**	**6.06**	**2.74**	**4.13**
rs10862001	**PAWR**	1.48	**1.87**	−0.17	0.71	−1.30
rs10853087	**GOSR2**	0.54	0.12	−1.21	−0.47	**−1.65**
rs6874700	HMGB1P47	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq
rs6874700	ISL1-DT	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq
rs10853087	LINC01974	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq
20:55165923	LINC01716	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq	RNA-polyA-seq

Bold: differential expressed genes (log2fc <−1.5 or >1.5)
Log2 fold change for non-coding genes of different bladder cancers and Muscolar Invasive Urothelial cancer over control tissue shows no value due to polyA-sequencing of control tissue. Differential Expression defined with log2fc <−1.5 or >1.5 (marker in bold italic letters). Legend: Bladder tissue of a 3-year-old bladder donor (control tissue); Bladder carcinoma (Bladder Ca); Bladder transitional cell carcinoma (Transit. cell Ca); Bladder squamous cell carcinoma (Squamous Ca); Ureter urothelial carcinoma (Urothelial Ca); Bladder urothelial carcinoma (Bl. urothelial Ca); Muscle invasive urothelial cancer (Mus. Inv. Urothelial Ca); RNA-polyA-seq data does not include miRNAs or lincRNAs). *log2fc* log2 fold change; differential expression defined with log2fc <−1.5 or >1.5 (marker in bold italic letters). *vs.* versus, *Chr.* chromosome.

Re-sequencing of *EFNA1*. Re-sequencing of all coding exons of *EFNA1* of transcript ENST00000368407.3 was performed in 580 CBE patients, all of which were included in the current GWAS. PCR conditions can be obtained upon request, primer sequences are shown in Supplementary information (Supplementary Table 2). Sequencing files for patient, parent, and control DNA were added to databases created using PreGap4 software, with control DNA processed as the reference sequence.

Genes prioritization. Lower p-value SNP of each associated region was imputed in LDproxy Tool (https://ldlink.nci.nih.gov/?tab=ldproxy) for European populations of CEU (Utah residents from north and west Europe); TSI Toscani in Italia; FIN Finnish in Finland; GBR British in England and Scotland; IBS Iberian population in Spain. Out of this, genes that reside in the linkage disequilibrium blocks defined from LD variants of r^2 above 0.8 to the top SNPs were taken into consideration for this study (Supplementary information, Supplementary Figs. 1–8). LD blocks coordinate regions imputed in hg19 are described in Supplementary Information (Supplementary Table 3).

No variants were significant LD associated with rs1924557 in chromosome 1 to determine an LD block region.

RNA isolation and mRNA library preparation of mouse embryonic urinary bladder and genital tissues. Animals were anesthetized with Isoflurane and sacrificed by cervical dislocation. Ethical consent is documented and approved by the local authorities of the Regierungspräsidium Darmstadt. Embryos from timed-pregnant females of the SWISS strain were harvested at embryonic days (E) E10.5, E12.5, and 15.5 (Supplementary information: Supplementary Fig. 18). The respective developmental Theiler stages were determined as 18 (TS18), TS21, and TS23. From E10.5 embryos, the urogenital ridge was dissected under an M205C stereo microscope (Leica Microsystems, Germany) surgically isolated, and transferred into QIAzol®. Embryos were pooled for each time point. For E10.5 stage biopsies from three embryos were pooled biopsies to prepare RNA, for E12.5 and E15.5 stages two embryos were pooled for RNA preparation. From E12.5 (primitive bladder) and E15.5 (bladder) embryos, the distinct structures of the developing and distinct visible bladder were surgically isolated (Supplementary information: Supplementary Fig. 18), combined, and transferred into QIAzol®.

Processing of total mouse embryonic RNA-sequencing data. About 50 million unique mapped reads per sample were obtained from each RNA-seq experiment. The reads were aligned using STAR aligner[32]. Read count was calculated with GenomicFeatures Bioconductor package. Calculation and normalization of "transcripts per kilobase million (TPM)" accounting for reads per kilobase (RPK) was performed as described elsewhere[33]. The fold change was calculated by dividing the subsequent stage by the preceding one and the log2 function was applied to the division as following: log2 (FoldChange) = log2 (subsequent embryonic stage/preceding embryonic stage). Differentially expressed genes were identified with values less than or equal to −1.5 or ≥1.5, respectively. The same algorithm was applied for the calculation of TPM of already deposited human embryonic and fetal RNA-seq data at EMBL-EBI expression atlas (accession code: E-MTAB-6592).

The raw RNA-sequencing data of mouse embryonic urinary bladder are deposited at GEO with the accession id: GSE190641.

Processing of bladder cancer RNA-sequencing data. Total RNA was purified using the QuantSeq library (Lexogen) with 500 ng RNA input. QuantSeq polyA RNA-tail libraries were sequenced (Single end 1 × 75 bp) on an Illumina Hiseq platform and generated data were further processed according to the GRCh38, TPM transformed, and further normalized. Sequencing, aligning and TPM calculation was performed by ImmunityBio™. Visualization of results in heatmaps was performed using graphpad PRISM 9.0.0.

RNA isolation and mRNA library preparation of human embryonic and fetal urinary bladder and genital tissues. Embryonic and fetal bladders and genital tissues were obtained by surgeons from terminated pregnancies after informed consent was obtained and with ethics approval. Pregnancies were terminated for social indications and the respective fetuses and embryos were healthy. The embryonic tissues comprised 7-week embryos, 7–7.5-week embryos, 7.5-week embryos, late 8-week embryos, and late 9-week embryos (Supplementary information: Supplementary Fig. 19). Samples comprised week 7 ($n = 2$), 7.5 ($n = 1$), 8 ($n = 3$), 9 ($n = 4$) for the bladder tissues and for the genital tissues from week 7 ($n = 3$), 8 ($n = 3$), 9 ($n = 3$), and 10 ($n = 4$). Gene expression data were extracted and analyzed after high throughput sequencing of paired-end mRNA libraries (Illumina). Data were deposited at EMBL-EBI expression atlas (accession code: E-MTAB-6592). Calculation of fold change of already deposited human embryonic RNA-seq data was carried out accordingly to our calculation of mouse embryonic data (see in Methods: Processing of total mouse embryonic RNA-sequencing data).

Processing of bladder cancer RNA-sequencing data. Cancer RNA-seq data were obtained from already deposited data at EMBL-EBI expression atlas (Cancer Cell Line Encyclopedia, experiment E-MTAB-2770). The deposited data does not include samples derived from CBE patients. Out of 1019 different cancer cell lines, the following cell line sample has been analyzed: 20 cell lines of bladder carcinoma, one cell line of bladder squamous cell carcinoma, four cell lines of bladder transitional cell carcinoma, one cell line of ureter urothelial carcinoma. TPM average was then calculated for each carcinoma cell type and data were compared with fold change to TPM of deposited mature urinary bladder polyA RNA-seq data (GEO accession: GSM1067793). In addition, RNA-polyA-seq data available from 38 cases of the CCC-EMN bladder cancer cohort [12] were generated from FFPE tissue all classified with Muscular invasive urothelial carcinoma. Demographic data is found in Supplementary Data 1.

Statistics and reproducibility. Quality control of individuals, principal component analysis, Quality control of variants and imputation, Statistical analysis for GWAS, and Genes prioritization is meticulously described in the methods above.

Average of TPM was calculated in R from biological replicates. Log2FC of human and mouse bladder was calculated as following: $\log_2(\text{AVERAGE\_TPM}_{\text{next\_stage}}/\text{AVERAGE\_TPM}_{\text{previous\_stage}})$. Log2FC of cancer cells was calculated as following: $\log_2(\text{AVERAGE\_TPM}_{\text{cancer\_line}}/\text{AVERAGE\_TPM}_{\text{control\_bladder\_tissue}})$. Sample size of mouse consists in $n = 3$ for embryo bladder at stage E10.5, and $n = 2$ for E12.5 and E15.5. Human fetal bladder samples comprised week 7 ($n = 2$), 7.5 ($n = 1$), 8 ($n = 3$), 9 ($n = 4$). Human cancer cells liens comprised $n = 20$ of bladder carcinoma cell, $n = 1$ of bladder squamous cell carcinoma cell line, $n = 4$ of bladder transitional cell carcinoma cell lines, $n = 1$ of ureter urothelial carcinoma cell line, $n = 38$ of muscular invasive urothelial carcinoma. Replicates are defined as a minimum of three technical replicates per sample size.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

GWAS generated and analyzed data during this study are included in this article and its supplementary information files. GWAS data are deposited at NHGRI-EBI GWAS Catalog with accession ID: GCST90132313. EFNA1 DNA sequencing data are deposited in GeneBank (BankIt) with the following accession numbers: OP312051; OP312052; OP312053; OP312054; OP312055; OP312056; OP312057; OP312058; OP312059; OP312060; OP312061; OP312062; OP312063. The raw RNA-sequencing data of the 38 Muscular Invasive Urothelial carcinomas are deposited at NCBI in Sequence Read Archive (SRA) with the following BioProject accession: PRJNA882449. The raw RNA-sequencing data of mouse embryonic urinary bladder are deposited at GEO with the accession id: GSE190641. The raw RNA-sequencing data of human embryonic and fetal urinary bladder and genital tissue are deposited at EMBL-EBI expression atlas with the following accession id: E-MTAB-6592. The raw RNA-sequencing data of cancer cell lines are obtained from EMBL-EBI expression atlas with the following accession id: E-MTAB-2770, PolyA RNA-sequencing of the mature urinary bladder is obtained from GEO with the following accession id: GSM1067793.

Received: 23 February 2022; Accepted: 11 October 2022;
Published online: 09 November 2022

References

1. Ebert, A.-K., Reutter, H., Ludwig, M. & Rösch, W. H. The exstrophy-epispadias complex. *Orphanet J. Rare Dis.* **4**, 23 (2009).
2. Dahm, P. & Gschwend, J. E. Malignant non-urothelial neoplasms of the urinary bladder: a review. *Eur. Urol.* **44**, 672–681 (2003).
3. Williamson, S. R., Lopez-Beltran, A., Montironi, R. & Cheng, L. Glandular lesions of the urinary bladder: clinical significance and differential diagnosis. *Histopathology* **58**, 811–834 (2011).
4. Ebert, A. K., Zwink, N., Reutter, H. M. & Jenetzky, E. A prevalence estimation of exstrophy and epispadias in Germany from Public Health Insurance Data. *Front. Pediatr.* **9**, 648414 (2021).
5. Reutter, H. et al. Genome-wide association study and mouse expression data identify a highly conserved 32 kb intergenic region between WNT3 and WNT9b as possible susceptibility locus for isolated classic exstrophy of the bladder. *Hum. Mol. Genet.* **23**, 5536–5544 (2014).
6. Draaken, M. et al. Genome-wide association study and meta-analysis identify ISL1 as genome-wide significant susceptibility gene for bladder exstrophy. *PLoS Genet.* **11**, 1005024 (2015).
7. Rieke, J. M. et al. SLC20A1 is involved in urinary tract and urorectal development. *Front. Cell Dev. Biol.* **8**, 567 (2020).
8. Itou, J. et al. Islet1 regulates establishment of the posterior hindlimb field upstream of the Hand2-Shh morphoregulatory gene network in mouse embryos. *Development* **139**, 1620–1629 (2012).
9. Bult, C., Blake, J., Smith, C., Kadin, J. & Richardson, J. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
10. Masden, T. B., Taela, A., Rocha, M., da. Moores, D. C. & Radulescu, A. Isolated thoracoschisis with rib agenesis and liver herniation: a case report. *Am. J. Case Rep.* **20**, 1915–1919 (2019).
11. Lek, M. et al. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60.706 humans. *Nature* **536**, 285–291 (2016).
12. Zhang, S. et al. Transcriptome profiling of a multiple recurrent muscle-invasive urothelial carcinoma of the bladder by deep sequencing. *PLoS One* **9**, 91466 (2014).
13. Hao, Y. & Li, G. Role of EFNA1 in tumorigenesis and prospects for cancer therapy. *Biomed. Pharmacother.* **130**, 110567 (2020).
14. Toma, M. I. et al. Lack of ephrin receptor A1 is a favorable independent prognostic factor in clear cell renal cell carcinoma. *PLoS One* **9**, 102262 (2014).
15. Warrington, N. M. et al. Genome-wide association study identifies nine novel loci for 2D:4D finger ratio. a putative retrospective biomarker of testosterone exposure in utero. *Hum. Mol. Genet.* **27**, 2025–2038 (2018).
16. Eeles, R. A. et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.* **45**, 385–391 (2013).
17. Popiołek. A. et al. Prostate-specific antigen and testosterone levels as biochemical indicators of cognitive function in prostate cancer survivors and the role of diabetes. *J. Clin. Med.* **10**, 5307 (2021).
18. Liu, J., Welm, B., Boucher, K. M., Ebbert, M. T. W. & Bernard, P. S. TRIM29 functions as a tumor suppressor in nontumorigenic breast cells and invasive ER+ breast cancer. *Am. J. Pathol.* **180**, 839–847 (2012).
19. Palmbos, P. L. et al. ATDC/TRIM29 drives invasive bladder cancer formation through miRNA-mediated and epigenetic mechanisms. *Cancer Res.* **75**, 5155–5166 (2015).
20. Yang, Z. et al. Single-cell sequencing reveals variants in ARID1A. GPRC5A and MLL2 driving self-renewal of human bladder cancer stem cells. *Eur. Urol.* **71**, 8–12 (2017).
21. Lu, H., Hao, L., Yang, H., Chen, J. & Liu, J. miRNA-34a suppresses colon carcinoma proliferation and induces cell apoptosis by targeting SYT1. *Int J. Clin. Exp. Pathol.* **12**, 2887–2897 (2019).

22. Gurung, P. et al. Prognostic DNA methylation biomarkers in high-risk non–muscle-invasive bladder cancer: a systematic review to identify loci for prospective validation. *Eur. Urol. Focus* **6**, 683–697 (2020).
23. Kitchen, M. O. et al. Methylation of HOXA9 and ISL1 predicts patient outcome in high-grade non-invasive bladder cancer. *PLoS One* **10**, 0137003 (2015).
24. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
25. Manichaikul, A., Palmer, A. A., Sen, Ś. & Broman, K. W. Significance thresholds for quantitative trait locus mapping under selective genotyping. *Genetics* **177**, 1963–1966 (2007).
26. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
27. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies via imputation of genotypes: Supplementary Methods Imputation of missing genotypes Imputation of completely missing SNPs. *Nat. Genet.* **39**, 906–913 (2007).
28. SAS Institute Inc. 2013. SAS 9.4 Language Reference. Cary. NC: SAS Institute.
29. Pruim. R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2011).
30. Szklarczyk, D. et al. STRING v10: protein-protein interaction networks. integrated over the tree of life. *Nucleic Acids Res.* **43**, 447–452 (2015).
31. Wullweber, A. et al. Bladder tumor subtype commitment occurs in carcinoma in situ driven by key signaling pathways including ECM Remodeling. *Cancer Res.* **81**, 1552–1566 (2021).
32. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
33. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2009).

Acknowledgements

We acknowledge the following institutions and organizations for founding and/or contributing to the project: BONFOR; Instituto de Salud Carlos III; Medical Research Council; Horizon 2020 Marie Skłodowska-Curie Actions Initial Training Network; RENALTRACT; Alzheimer Research UK; Stroke Foundation/British Heart Foundation. We acknowledge financial support by Deutsche Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg within the funding programme "Open Access Publication Funding".

Author contributions

E.M., B.O., P.G., M.K., W.N., G.B., A.N., A.S.W., and H.R. conceptualized and supervised the research. E.M., M.K., G.C.D., and M.E. C.M., were engaged in biostatistics and bioinformatics analysis. E.M., P.G., M.E. B.O., P.G., H.R., A.S.W., A.C.H., and G.B. designed and performed most experiments. W.R., A.-K.E., R.S., A.B., M.D.G., A.T., F.M.T., J.L.H., P.E., J.M.O., J.A.R., C.V., I.C.H., H.H., D.J.W., J.S., J.G., A.-S.G., B.C.B., S.A., E.Å., E.R., M.L., G.H., E.G., A.B., D.K., R.M.C., N.Y., M.O., J.O., B.H., M.P., C.N., K.H., M.S., F.-M.S., E.S., T.M.B.., I.A.L.M.v.R., W.F.J.F., C.L.M.M., M.L., J.N., B.U., C.F., D.P.G., M.M.Y.C., K.U.L., M.M.N., S.H., N.Z., E.J. were responsible for patient recruitment and biospecimen and clinical data collection. E.M., G.B., M.K., A.S.W., P.G., and H.R. drafted the manuscript. E.M. designed the graphic illustrations. All authors reviewed and edited the manuscript. All author(s) read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-04092-3.

Correspondence and requests for materials should be addressed to Enrico Mingardo or Heiko Reutter.

Peer review information *Communications Biology* thanks Jack Weaver and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Chiea Chuen Khor and George Inglis.

Reprints and permission information is available at http://www.nature.com/reprints

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

[1]Institute for Anatomy and Cell Biology, University Hospital Bonn, University of Bonn, Bonn, Germany. [2]Institute for Neuroanatomy, University Hospital Bonn, University of Bonn, Bonn, Germany. [3]Institute of Human Genetics, School of Medicine & University Hospital Bonn, University of Bonn, Bonn, Germany. [4]Centre for Genomic Medicine, University of Manchester, Manchester, UK. [5]Institute of Cardiovascular Regeneration, Centre for Molecular Medicine, Goethe University, Frankfurt am Main, Germany. [6]Georg-Speyer-Haus, Frankfurt am Main, Germany. [7]Department of Women´s and Children´s Health and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. [8]Pediatric Surgery, Astrid Lindgren Children Hospital, Karolinska University Hospital, Stockholm, Sweden. [9]Division of Cell Matrix Biology and Regenerative Medicine, Faculty of Biology Medicine and Health, School of Biological Sciences, University of Manchester, Manchester, UK. [10]Royal Manchester Children's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. [11]Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany. [12]Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), Erlangen, Germany. [13]BRIDGE-Consortium Germany e.V., Mannheim, Germany. [14]Department of Pediatrics and Adolescent Medicine, University Hospital Erlangen, Erlangen, Germany. [15]Department of Pediatric Urology, Clinic St. Hedwig, University Medical Center of Regensburg, Regensburg, Germany. [16]Department of Urology and Pediatric Urology, University Hospital of Ulm, Ulm, Germany. [17]Center for Pediatric, Adolescent and Reconstructive Urology, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany. [18]Department of Medical Sciences and Medical Genetics Unit, Città della Salute e della Scienza University Hospital, University of Torino, Torino, Italy. [19]Institute for Maternal and Child Health, IRCCS Burlo Garofalo, Trieste, Italy. [20]Translational Pediatrics and Infectious Diseases, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain. [21]Department of Internal Medicine, Hospital U M Valdecilla, University of Cantabria, IDIVAL, Santander, Spain. [22]Department of Urology and Urosurgery, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. [23]Institute of Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany. [24]Department of Internal Medicine. Hospital U M Valdecilla, University of Cantabria, IDIVAL, Santander, Spain. [25]Stroke Research Centre, University College London, Institute of Neurology, London, UK. [26]Neurogenetics Laboratory, The National Hospital of Neurology and Neurosurgery, London, UK. [27]Department of Neurosurgery, Klinikum rechts der Isar, Technical University Munich, Munich, Germany. [28]Neurogenetics Laboratory, The National Hospital of Neurology and Neurosurgery, London, UK. [29]Stroke Research Center, Department of Brain Repair and Rehabilitation, UCL

Institute of Neurology and The National Hospital for Neurology and Neurosurgery, London, UK. [30]Institute for Human Genetics, University of Marburg, Marburg, Germany. [31]Department of Women's and Children's Health and Center for Molecular Medicine, Bioclinicum, Karolinska Institutet, Stockholm, Sweden. [32]Department of Urology, Danderyds Hospital, Danderyd, Sweden. [33]Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. [34]R&D Unit, Stockholms Sjukhem, Stockholm, Sweden. [35]ME Gynecology and Reproduction Medicine, Karolinska University Hospital, and Dept of Clintec, Karolinska Institutet, Stockholm, Sweden. [36]Department of Clinical Chemistry and Clinical Pharmacology, University of Bonn, Bonn, Germany. [37]Department of Pediatric Surgery, Queen Silvia Children's Hospital, Gothenburg, Sweden. [38]Department of Molecular Medicine, University of Pavia, Pavia, Italy. [39]Laboratory of Molecular Medicine and Cytogenetics, IRCCS Mondino Foundation, Pavia, Italy. [40]Pediatric Urology Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [41]Paediatric Urology, Royal Manchester Children's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK. [42]Department for Pediatric Urology, Ordensklinikum Linz, Hospital of the Sisters of Charity, Linz, Austria. [43]Division of Pediatric Urology, Department of Urology, University of Erlangen-Nürnberg, Erlangen, Germany. [44]Department of Pediatric Surgery and Urology, Klinik Hallerwiese-Cnopfsche Kinderklinik, Nürnberg, Germany. [45]Department of Urology and Pediatric Urology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany. [46]Clinic for Paediatric Surgery and Paediatric Urology, Klinikum Bremen-Mitte, Bremen, Germany. [47]Department of Pediatric Surgery and Urology, University Hospital Cologne, Cologne, Germany. [48]Department of Pediatric Surgery, University of Leipzig, Leipzig, Germany. [49]Department for Health Evidence, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands. [50]Department of Urology, Pediatric Urology Center, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. [51]Department of Genetics, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. [52]Center of Pediatric Surgery Hannover, Hannover Medical School, Hannover, Germany. [53]Department of Renal Medicine, University College London, London, UK. [54]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. [55]Department of Child and Adolescent Psychiatry, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany. [56]Faculty of Health, School of Medicine, University of Witten/Herdecke, Witten, Germany. [57]Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany. [58]Division of Neonatology and Pediatric Intensive Care Medicine, Department of Pediatric and Adolescent Medicine, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany. [59]These authors contributed equally: Enrico Mingardo, Glenda Beaman, Philip Grote, Michael Knapp, Heiko Reutter. ✉email: enming@uni-bonn.de; Heiko.Reutter@uk-erlangen.de

Supplementary Information

A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy

Authors Enrico Mingardo, Glenda Beaman, Philip Grote, Agneta Nordenskjöld, William Newman, Adrian S. Woolf, Markus Eckstein, Alina C. Hilger, Gabriel C. Dworschak,Wolfgang Rösch, Anne-Karolin Ebert, Raimund Stein, Alfredo Brusco, Massimo Di Grazia, Ali Tamer, Federico M. Torres, Jose L. Hernandez, Philipp Erben, Carlo Maj, Jose M. Olmos, Jose A. Riancho, Carmen Valero, Isabel C. Hostettler, Henry Houlden, David J. Werring, Johannes Schumacher, Jan Gehlen, Ann-Sophie Giel, Benedikt C. Buerfent, Samara Arkani, Elisabeth Åkesson, Emilia Rotstein, Michael Ludwig, Gundela Holmdahl, Elisa Giorgio, Alfredo Berettini, David Keene, Raimondo M. Cervellione, Nina Younsi, Melissa Ortlieb, Josef Oswald, Bernhard Haid, Martin Promm, Claudia Neissner, Karin Hirsch, Maximilian Stehr, Frank-Mattias Schäfer, Eberhard Schmiedeke, Thomas M. Boemers, Iris A. L. M. van Rooij, Wouter F. J. Feitz, Carlo L. M. Marcelis, Martin Lacher, Jana Nelson, Benno Ure, Caroline Fortmann, Daniel P. Gale, Melanie M. Y. Chan, Kerstin U. Ludwig, Markus M. Nöthen, Stefanie Heilmann, Nadine Zwink, Ekkehart Jenetzky, Benjamin Odermatt, Michael Knapp, Heiko Reutter

Supplementary Table 1. Number of cases and controls for the seven independent samples

Sample	No of Cases	No of Controls
GWAS1	98	526
GWAS2	110	1.177
Central	172	2.588
Italy	57	1.325
Spain	62	279
Sweden	80	238
UK	49	1.219

Supplementary Table 2. Primer sequences used for *EFAN1* re-sequencing

EFNA1 EX1 FWD	AAAGGCGGAGTCGCTAGG
EFNA1 EX1 REV	GGGGTGCTCCCAGATATGAC
EFNA1 EX2 FWD	CTTGGGGTCCAGTGTGAAAT
EFNA1 EX2 REV	GCTAAACAGAGTGCCCAGCA
EFNA1 EX3-4 FWD	GAGTAGGGAGCTGAGAAAGCA
EFNA1 EX3-4 REV	CTCTCAGCCCAACAGGATTC
EFNA1 EX5 FWD	AAGGGGTCTGCTTGAAGAGG
EFNA1 EX5 REV	CGTTTTGAGGCTGCTAGGTG

Legend: exon (EX); Forward (FWD); Reverse (REV).

Supplementary Table 3. Linkage Disequilibrium (LD) blocks coordinates (hg19) for CBE associated top variants.

Marker	Chromosome (chr)	chr Start	chr End
rs1924557	chr1	n.a.	n.a.
rs4745	chr1	155089883	155142927
rs80215221	chr3	137538620	137558922
rs6874700	chr5	50659788	50748173
rs1790471	chr11	119964758	119968219
rs10862001	chr12	79847040	80114135
rs10853087	chr17	44989888	45046865
rs6024978	chr20	55161209	55175996

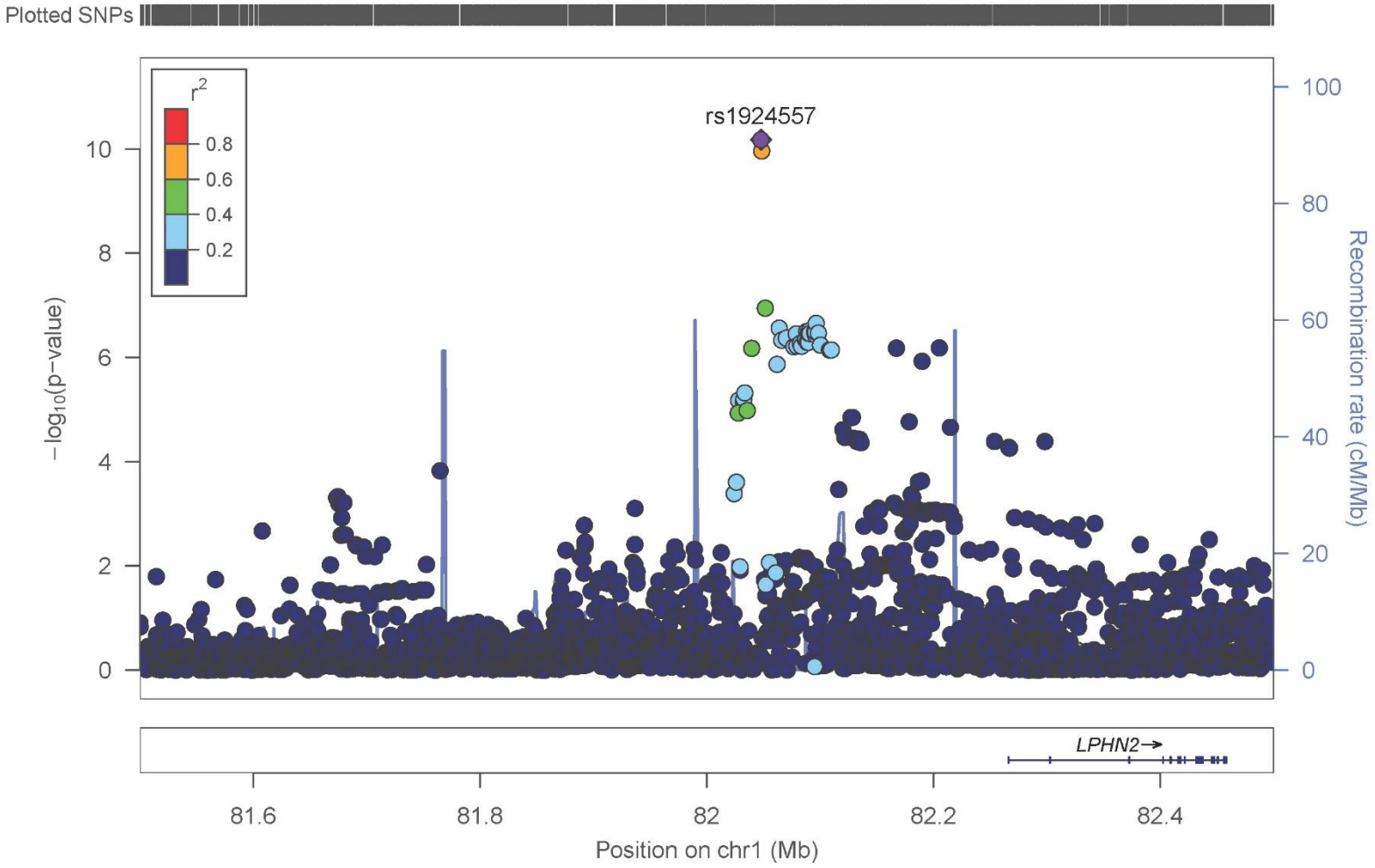
Legend: n.a. = no LD association.

Supplementary table 4. Re-sequencing of EFNA1

Patients (n)	cDNA / protein change	exon / intron	SNP ID	gnomAD allele Frequency	Mutation Taster	PolyPhen-2	SIFT	CADD
n=1	93-36T>C	intron1	rs369393260	0.000256				
n=1	c.92+31G>A	intron1	rs372698388	0.00024				
n=1	c.92+117C>A	intron1	not reported	0				
n=1	c.92+147C>G	intron 1	not reported	0				
n=1	c.116T>C p.Ile39Thr	exon 2	not reported	0	disease causing	deleterious	possibly damaging	25.3
n=1	c.156G>A p.Pro52=	exon 2	rs376532577	0.000028	=	=	=	21.5
n=1	c.167A>G p.Asp56Gly	exon 2	not reported	0	disease causing	tolerated	benign	22.0
n=1	c.341delT p.Phe114Serfs*28	exon 2	not reported	0	LoF	LoF	LoF	25.6
n=1	c.455-13T>C	intron3	not reported	0				
n=1	c.454+75A>G	intron3	rs1033536381	6.6×10^{-06}				
n=1	c.454+61A>T	intron3	not reported	0				
n=1	c.455-52C>T	intron 3	not reported	0				
n=1	c.503C>T p.Ala168Val	exon 4	not reported	0	disease causing	tolerated	benign	6.97
n=1	c.521G>A p.Arg174Gln	exon 5	rs139969988	0.00059	benign	tolerated	benign	5.03

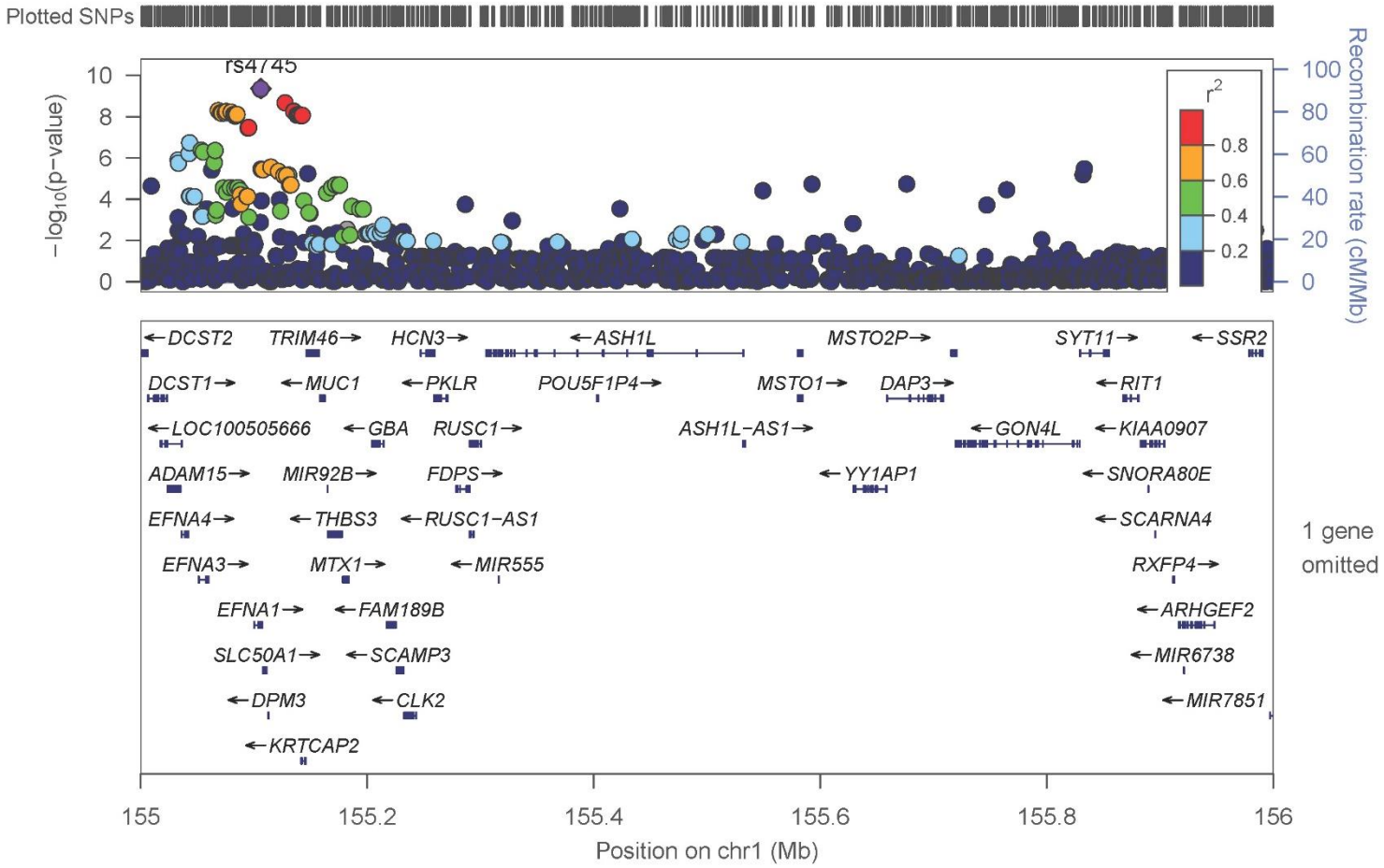
Legend: LoF = Loss-of-function

Supplementary Figure 1: Regional association result of genome-wide significant locus for chromosome 1, region 1.



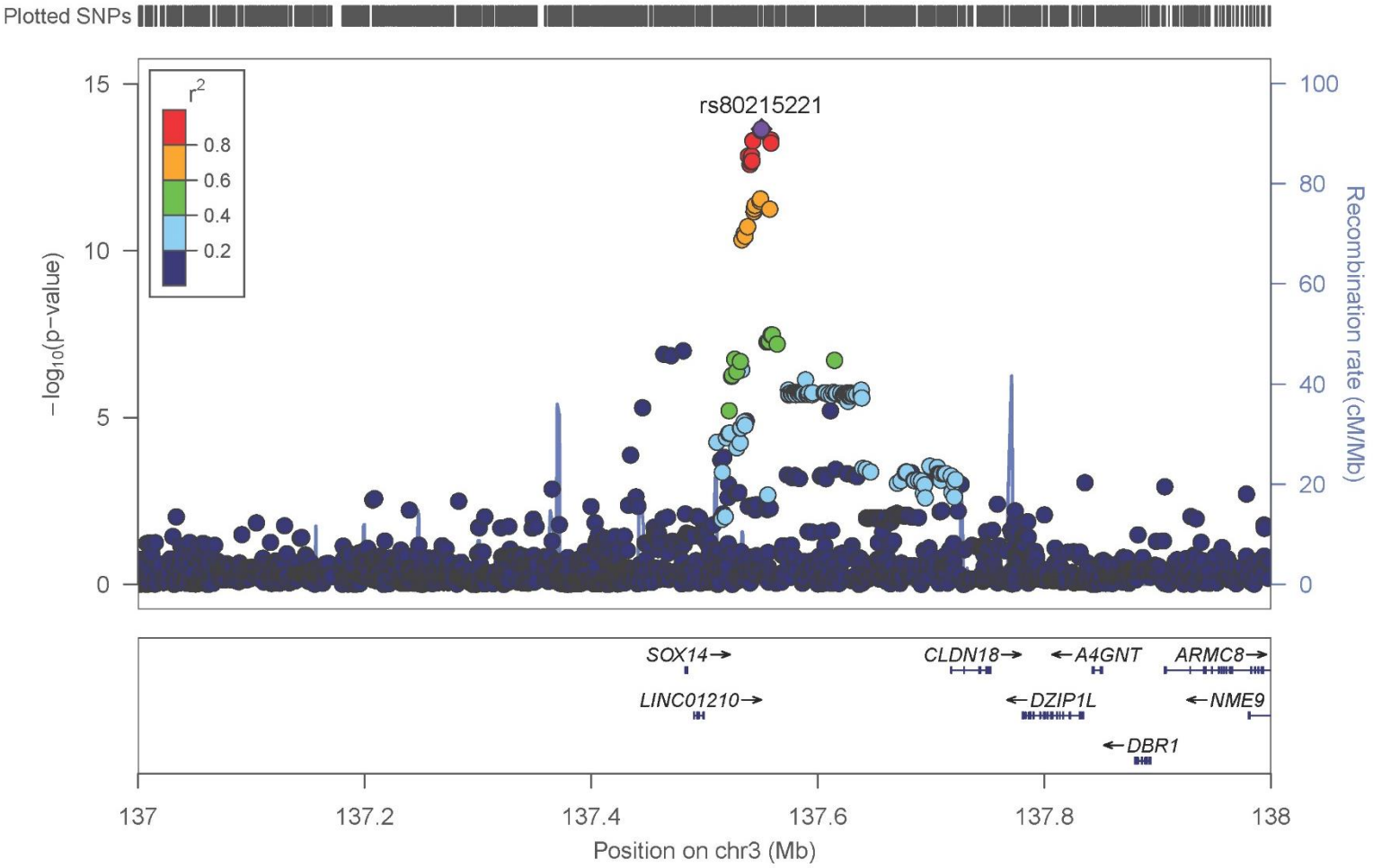
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 2: Regional association result of genome-wide significant locus for chromosome 1, region 2.



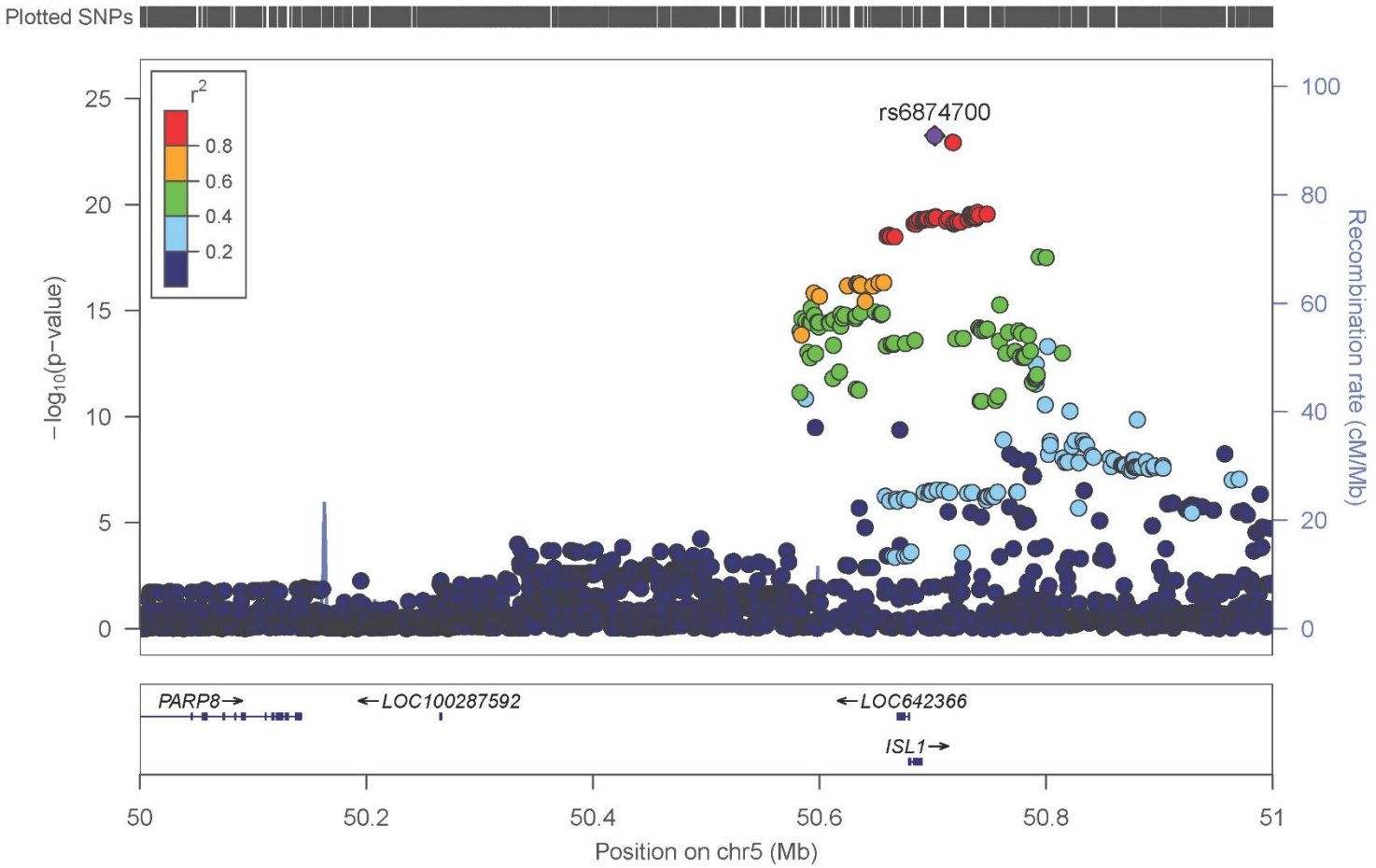
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 3: Regional association result of genome-wide significant locus for chromosome 3.



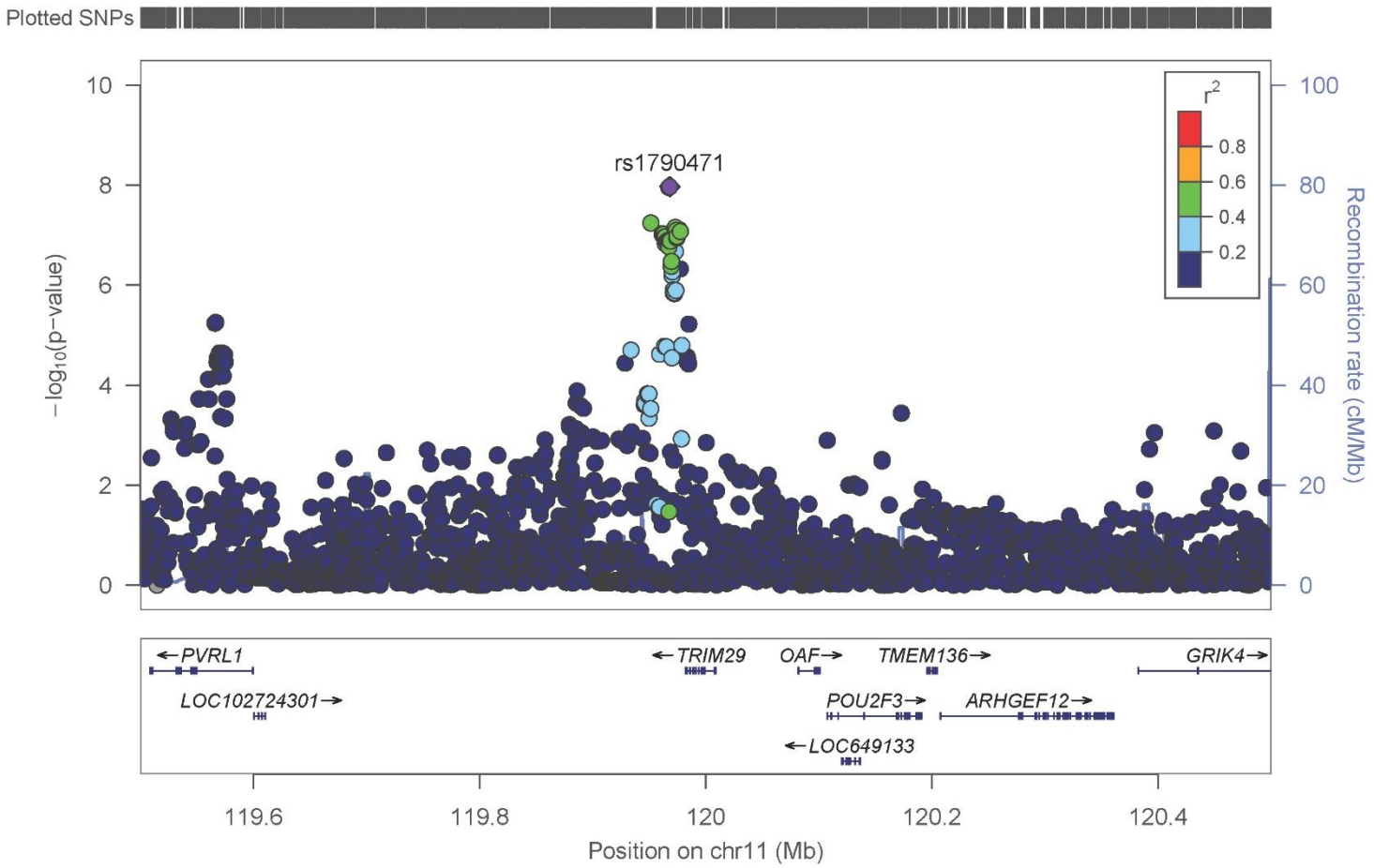
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 4: Regional association result of genome-wide significant locus for chromosome 5.



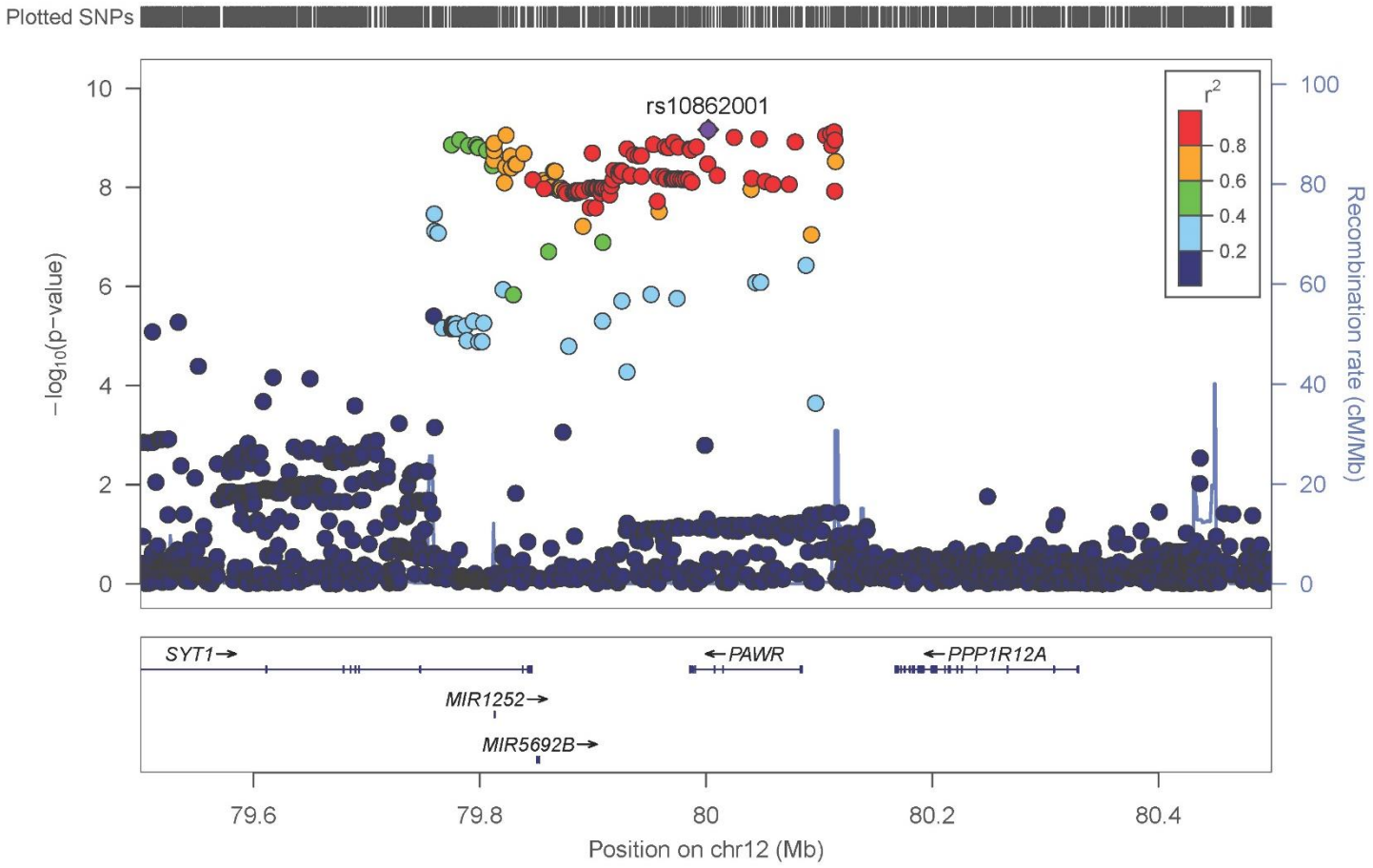
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 5: Regional association result of genome-wide significant locus for chromosome 11



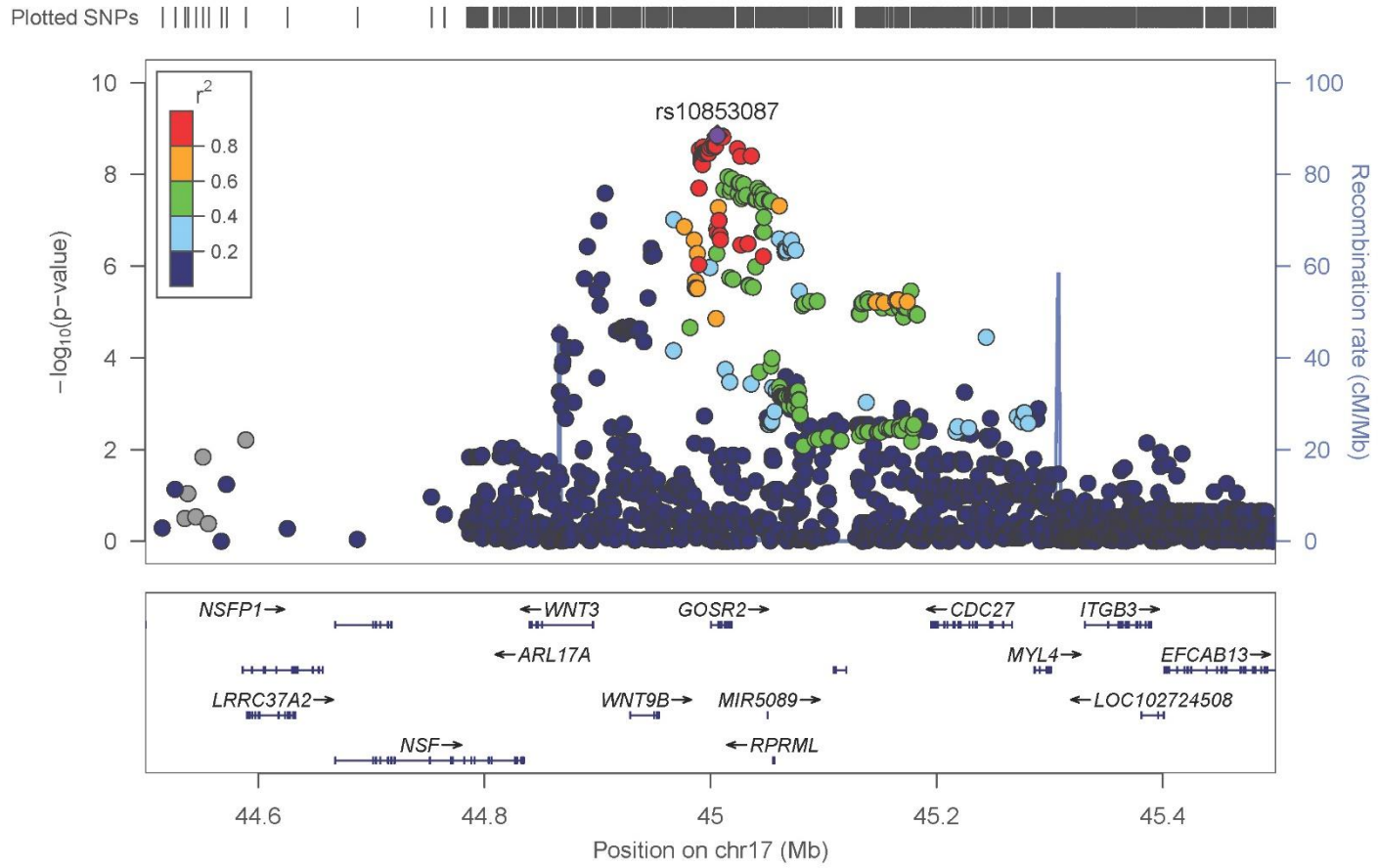
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 6: Regional association result of genome-wide significant locus for chromosome 12



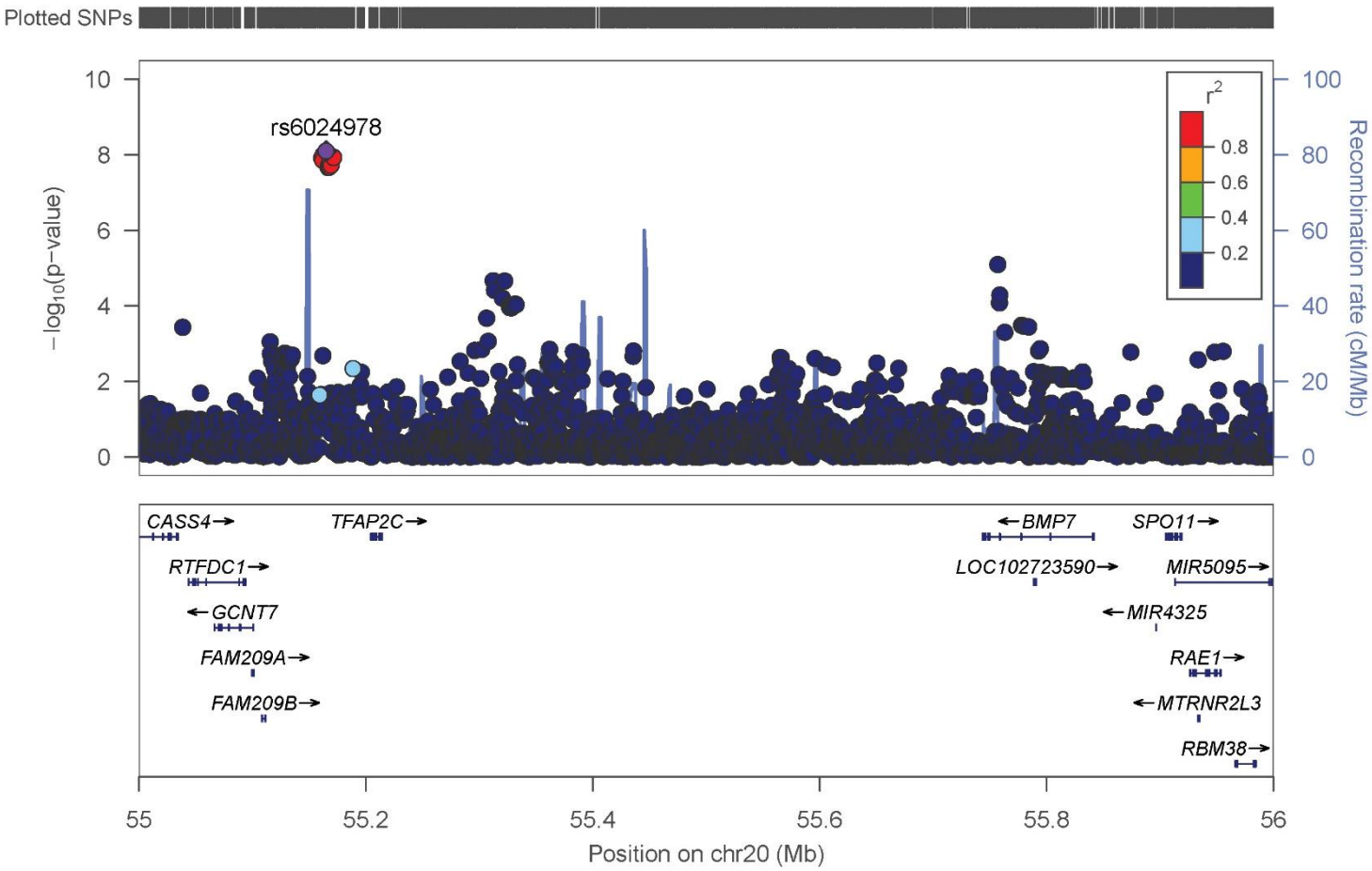
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 7: Regional association result of genome-wide significant locus for chromosome 17



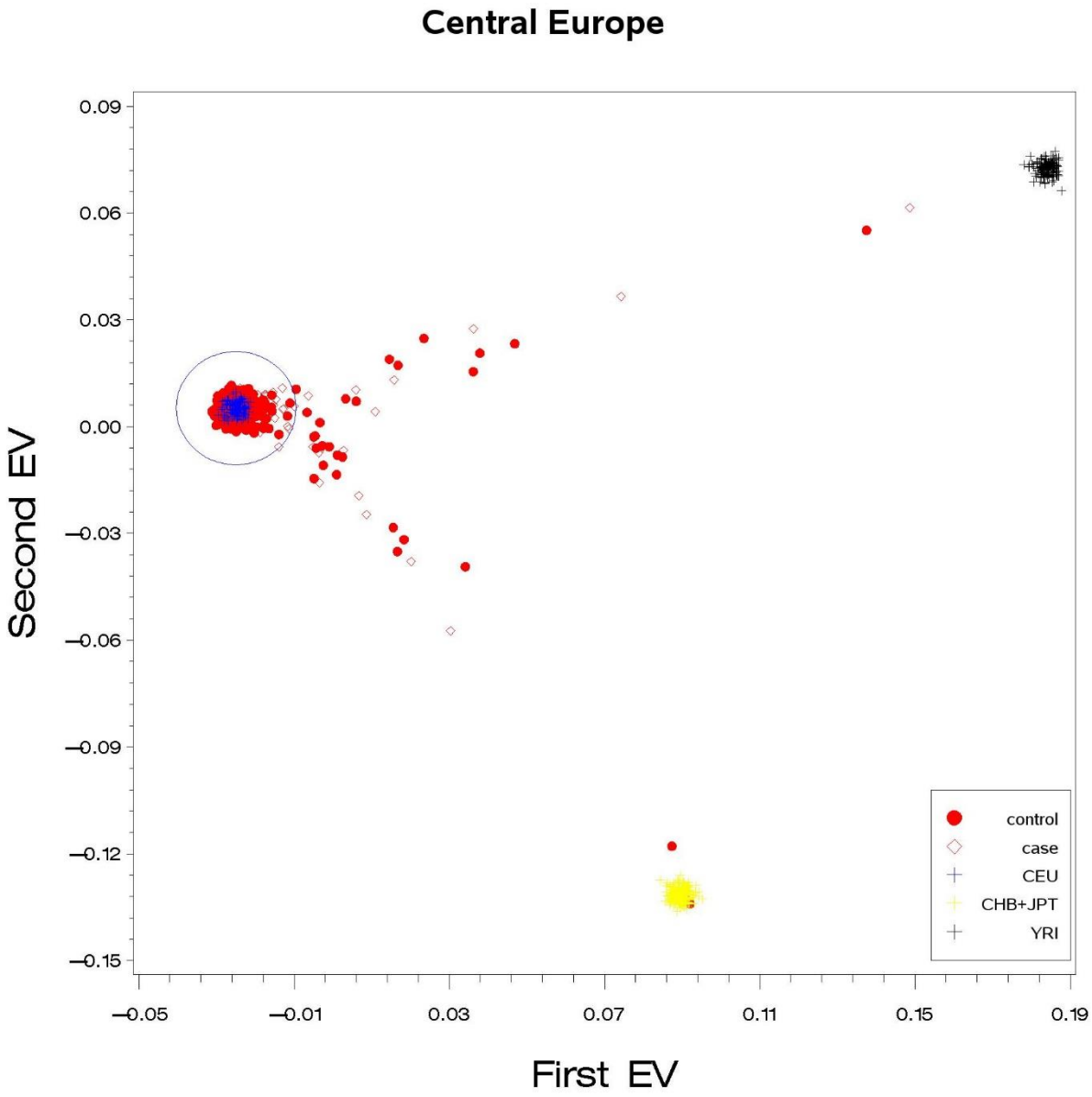
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 8: Regional association result of genome-wide significant locus for chromosome 20



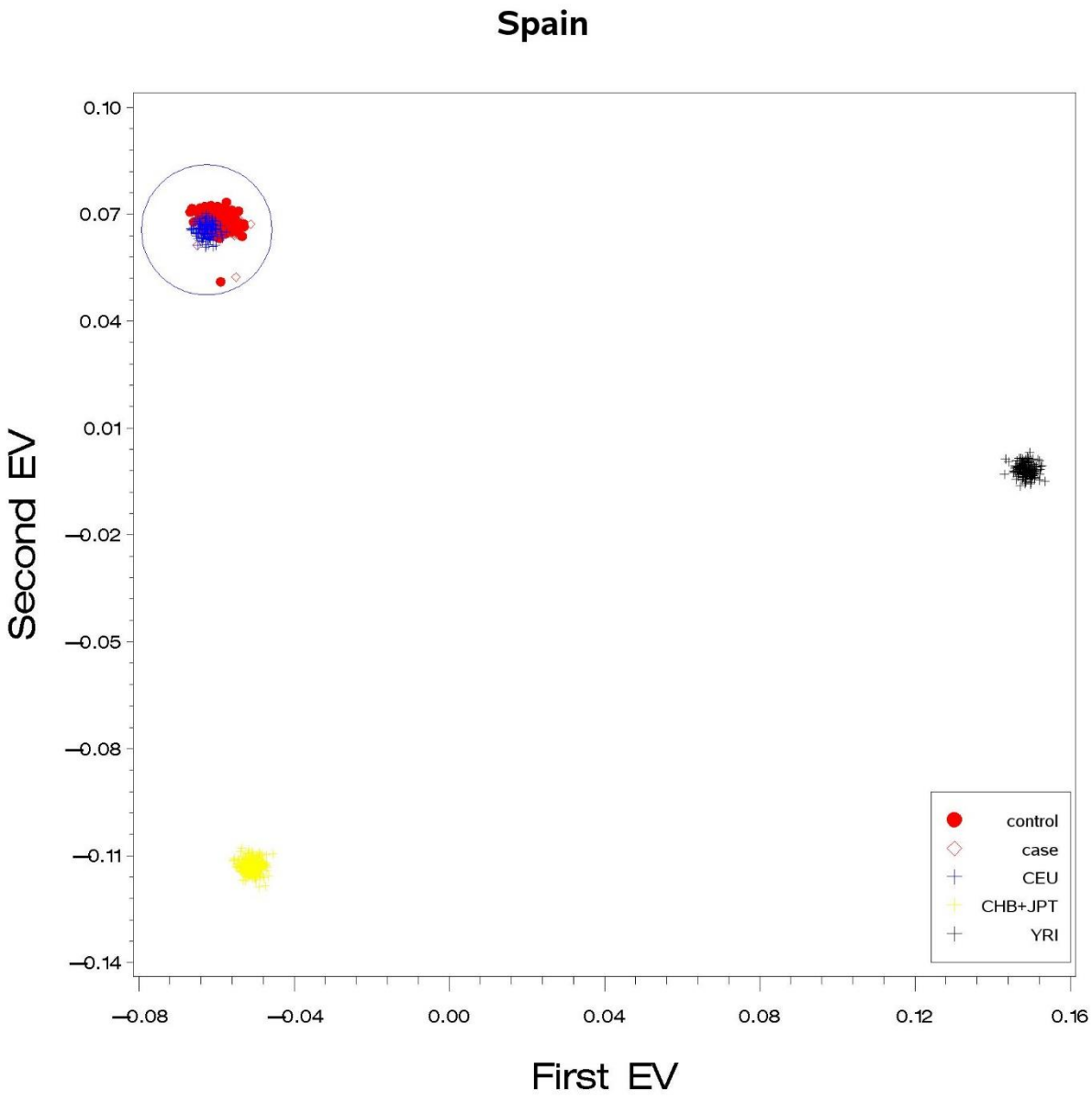
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). Colours of SNP represent the r^2 LD-block value. The most significant SNP is labelled in purple.

Supplementary Figure 9: Principal component analysis of Central Europe cohort.



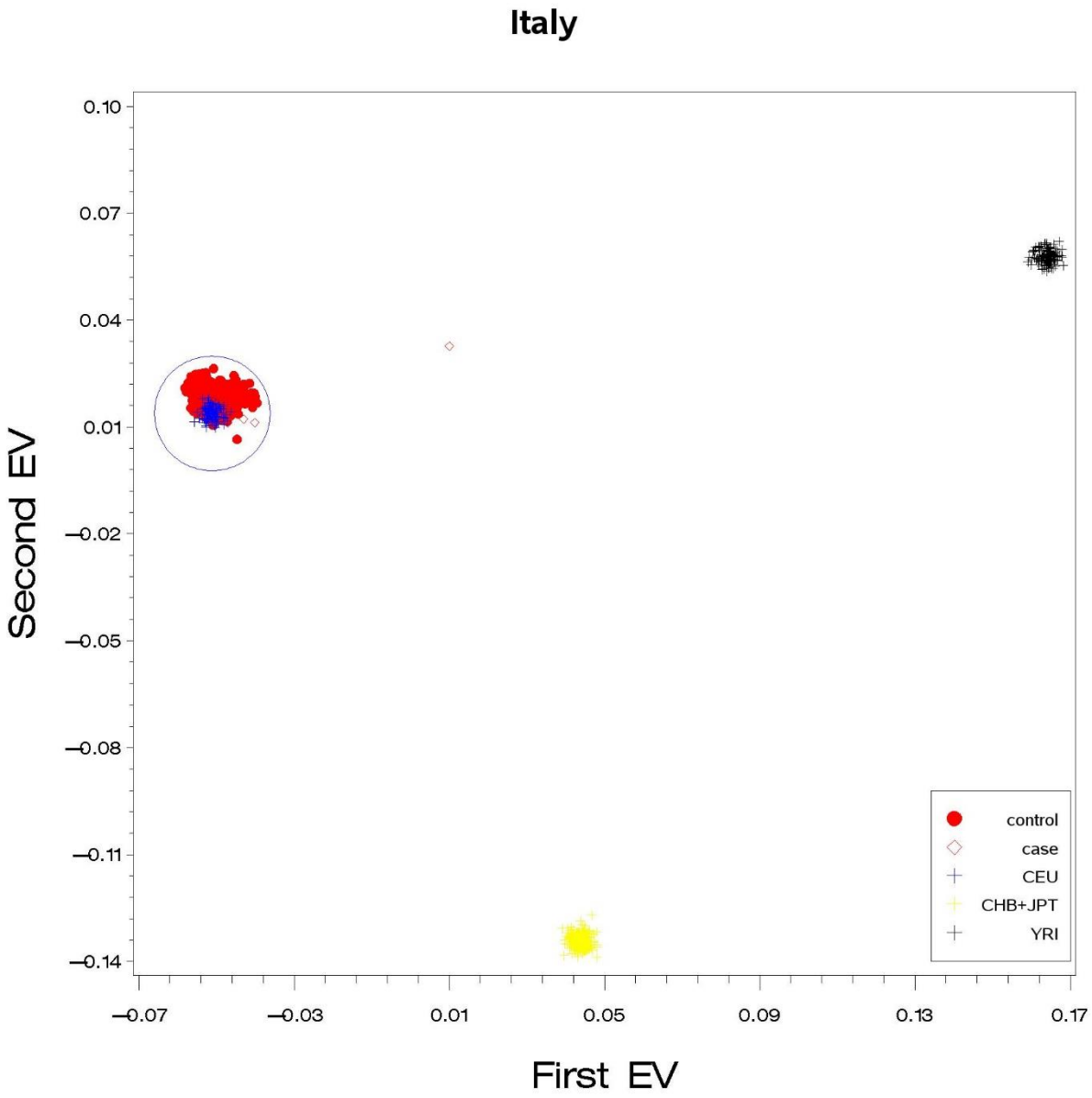
Only individuals within the red ellipses defined by condition (1) were used for the statistical analysis, individuals outside the red ellipsis were discarded. (CEU) Utah Residents from North and West Europe; (CHB) Han Chinese in Beijing, China; (JPT) Japanese in Tokyo, Japan; (YRI) Yoruba in Ibadan, Nigeria.

Supplementary Figure 10: Principal component analysis of Spain cohort.



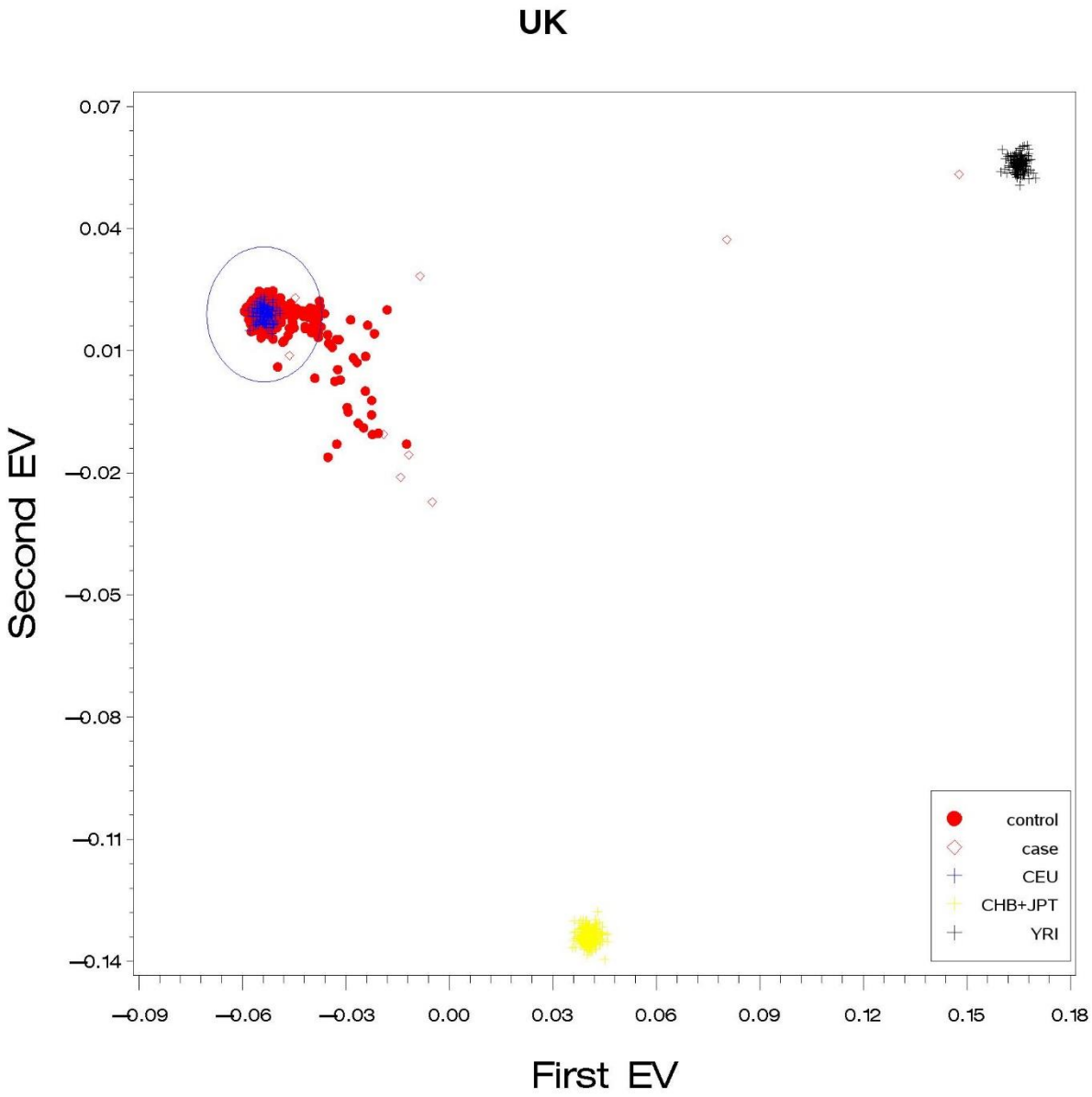
Only individuals within the red ellipses defined by condition (1) were used for the statistical analysis, individuals outside the red ellipsis were discarded. (CEU) Utah Residents from North and West Europe; (CHB) Han Chinese in Beijing, China; (JPT) Japanese in Tokyo, Japan; (YRI) Yoruba in Ibadan, Nigeria.

Supplementary Figure 11: Principal component analysis of Italy cohort.



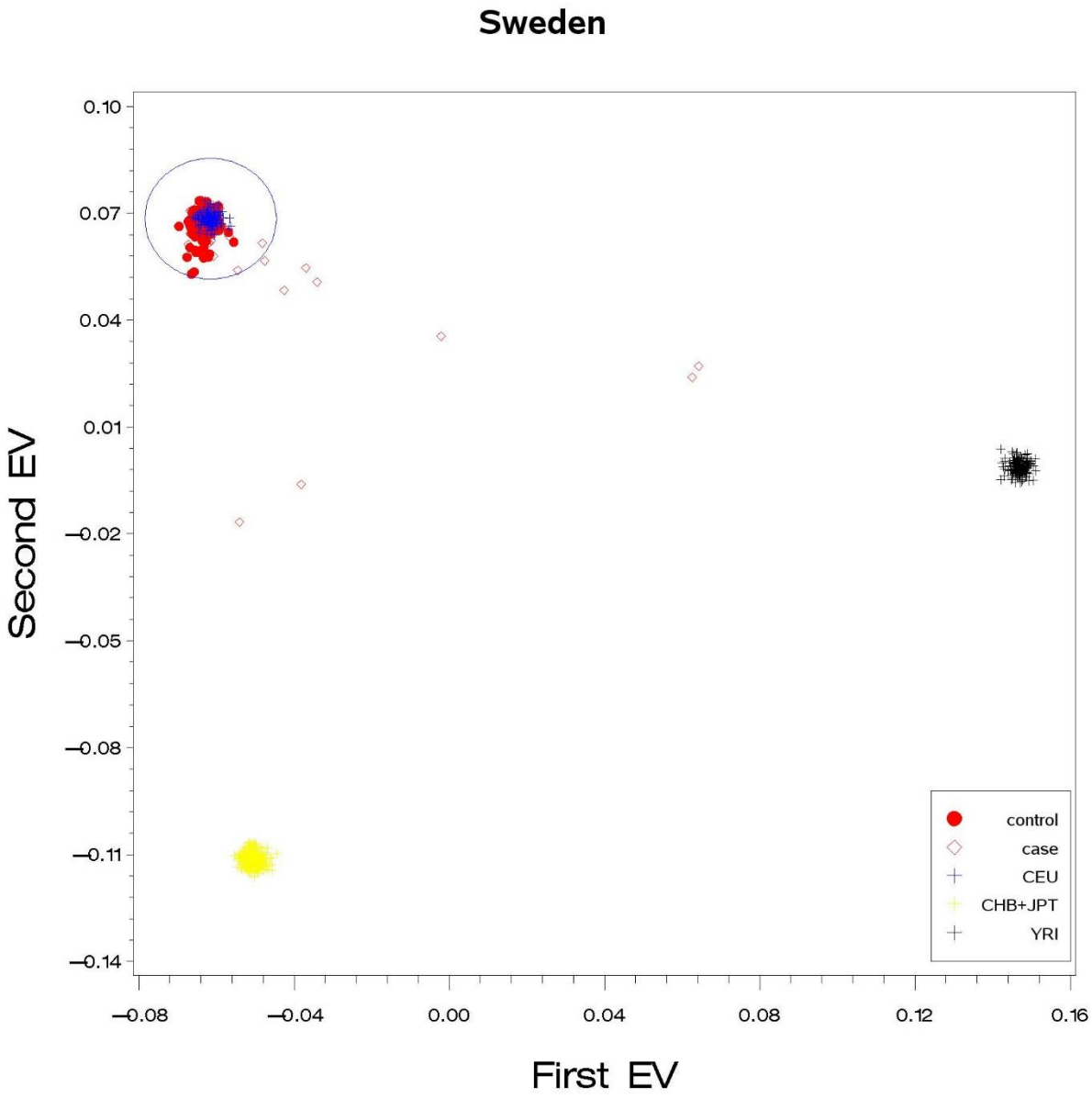
Only individuals within the red ellipses defined by condition (1) were used for the statistical analysis, individuals outside the red ellipsis were discarded. (CEU) Utah Residents from North and West Europe; (CHB) Han Chinese in Beijing, China; (JPT) Japanese in Tokyo, Japan; (YRI) Yoruba in Ibadan, Nigeria.

Supplementary Figure 12: Principal component analysis of UK cohort



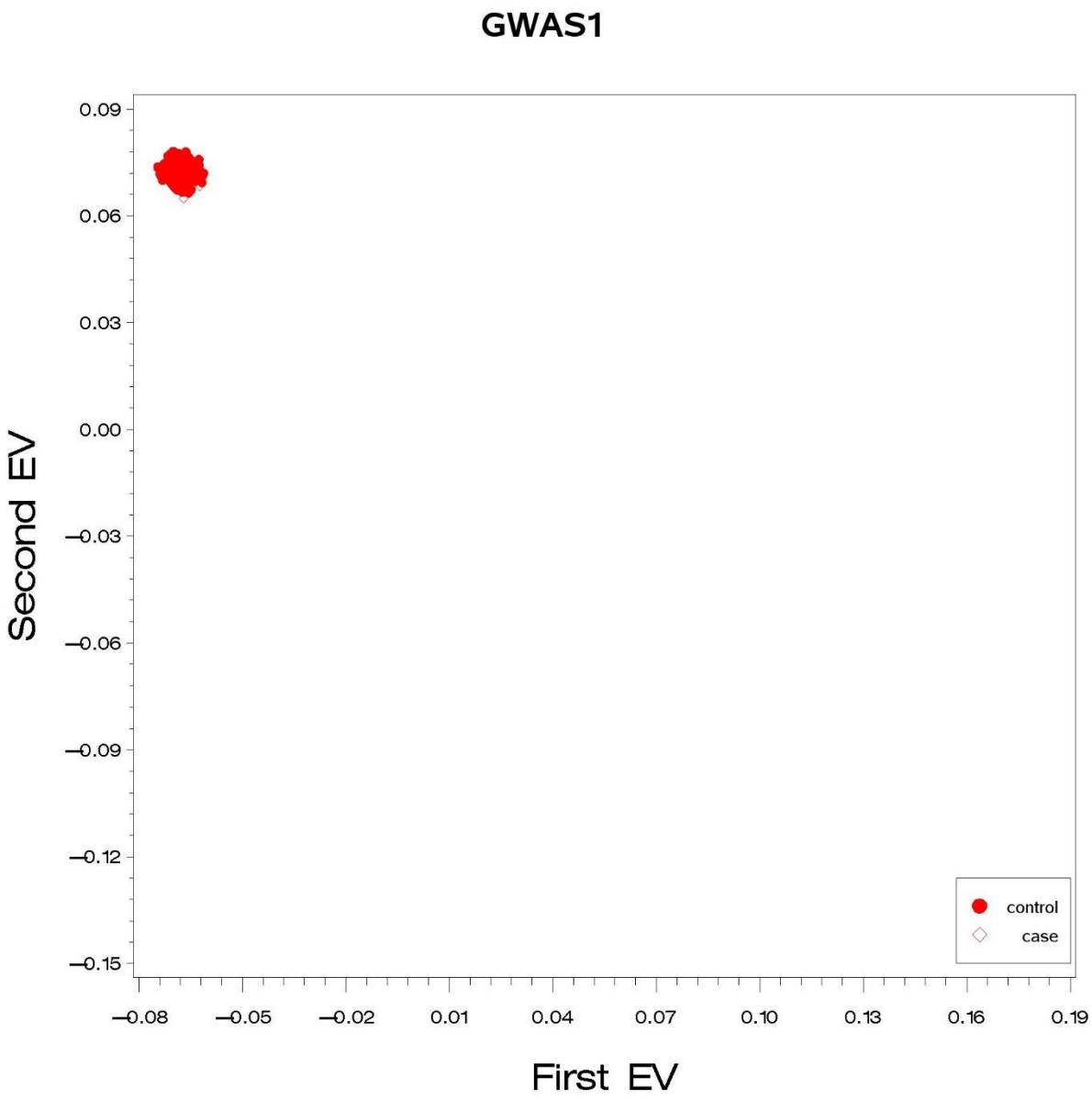
Only individuals within the red ellipses defined by condition (1) were used for the statistical analysis, individuals outside the red ellipsis were discarded. (CEU) Utah Residents from North and West Europe; (CHB) Han Chinese in Beijing, China; (JPT) Japanese in Tokyo, Japan; (YRI) Yoruba in Ibadan, Nigeria.

Supplementary Figure 13: Principal component analysis of Sweden cohort



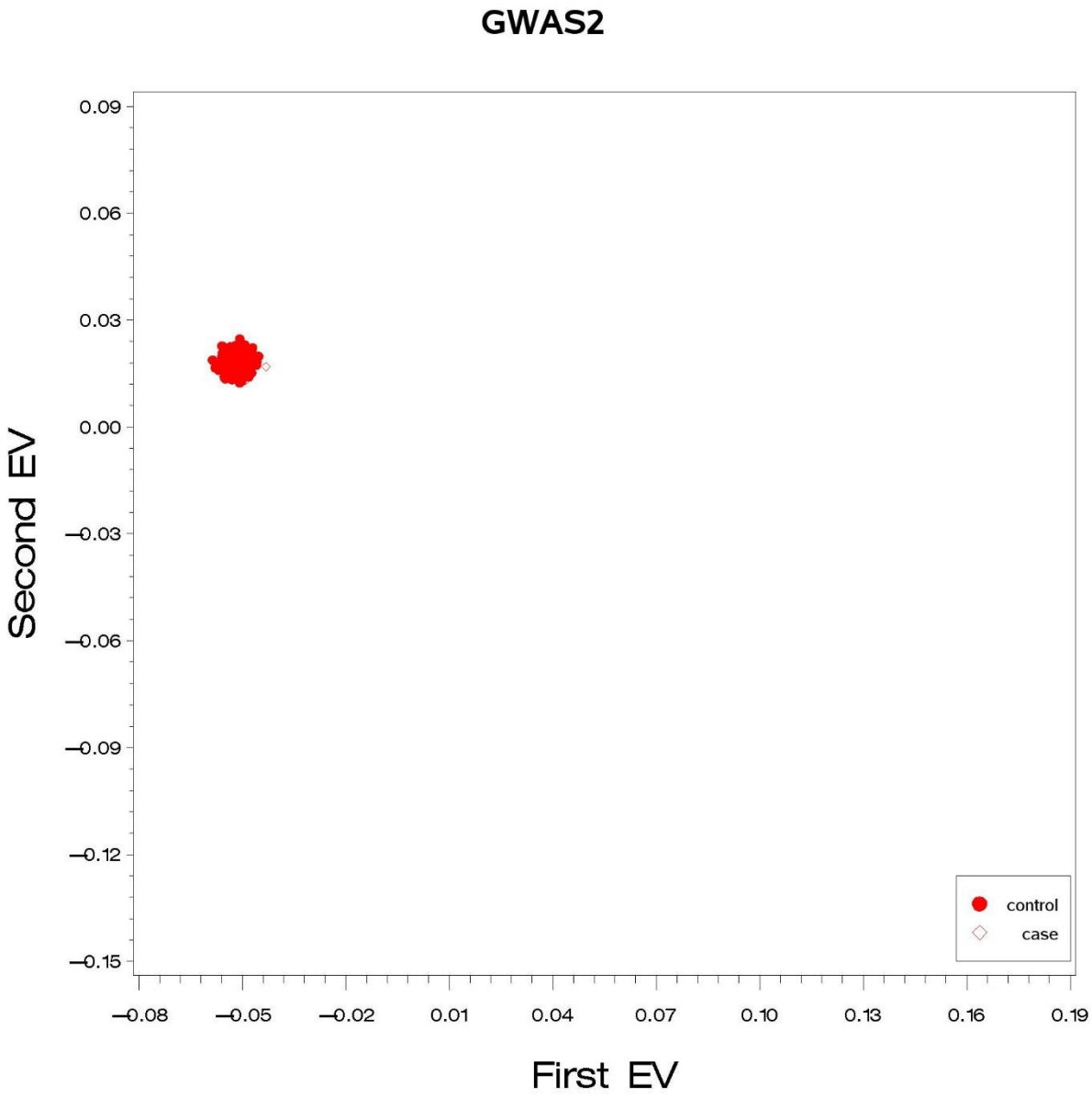
Only individuals within the red ellipses defined by condition (1) were used for the statistical analysis, individuals outside the red ellipsis were discarded. (CEU) Utah Residents from North and West Europe; (CHB) Han Chinese in Beijing, China; (JPT) Japanese in Tokyo, Japan; (YRI) Yoruba in Ibadan, Nigeria.

Supplementary Figure 14: Principal component analysis of GWAS 1 strata.



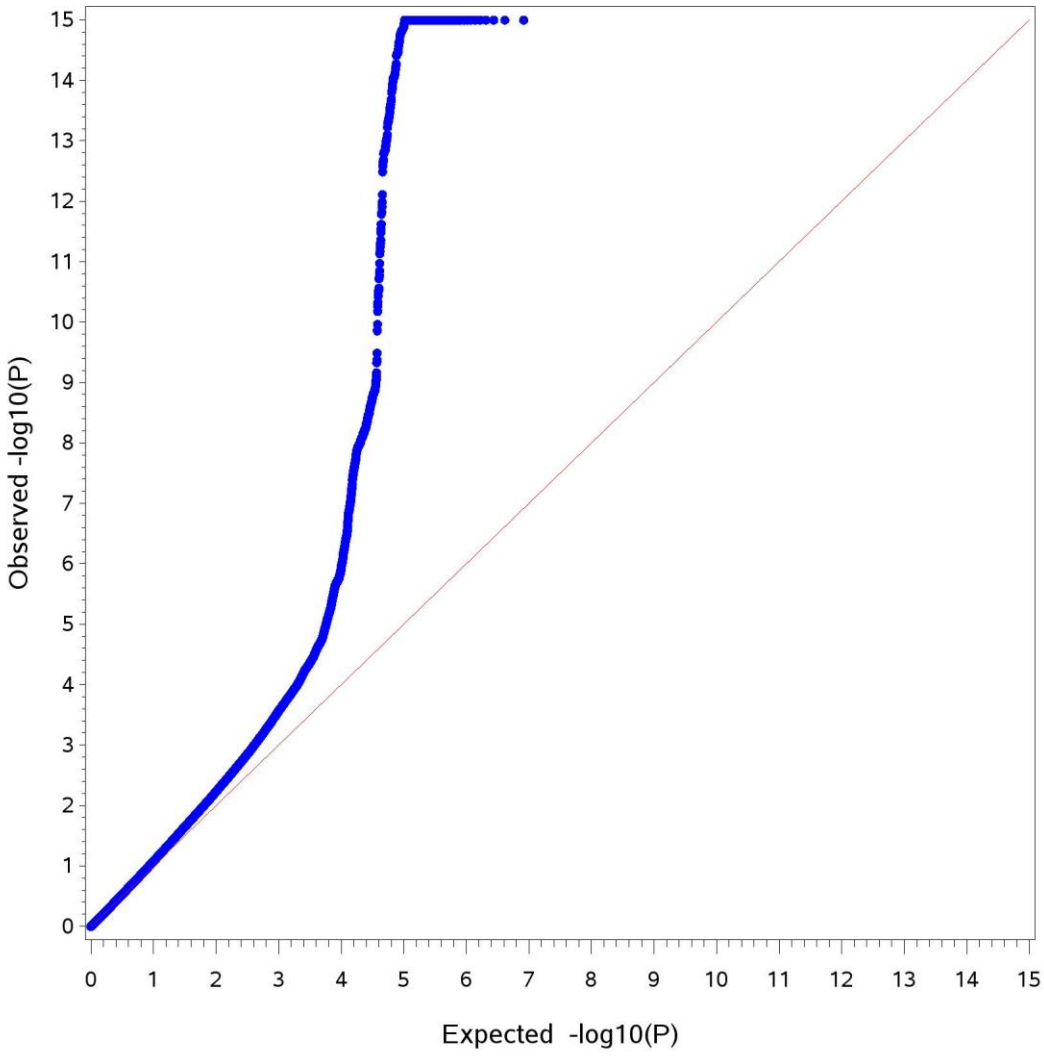
Only individuals within the red ellipses defined by condition (1) were used for the statistical analysis, individuals outside the red ellipsis were discarded. (CEU) Utah Residents from North and West Europe; (CHB) Han Chinese in Beijing, China; (JPT) Japanese in Tokyo, Japan; (YRI) Yoruba in Ibadan, Nigeria.

Supplementary Figure 15: Principal component analysis of GWAS 2 strata.



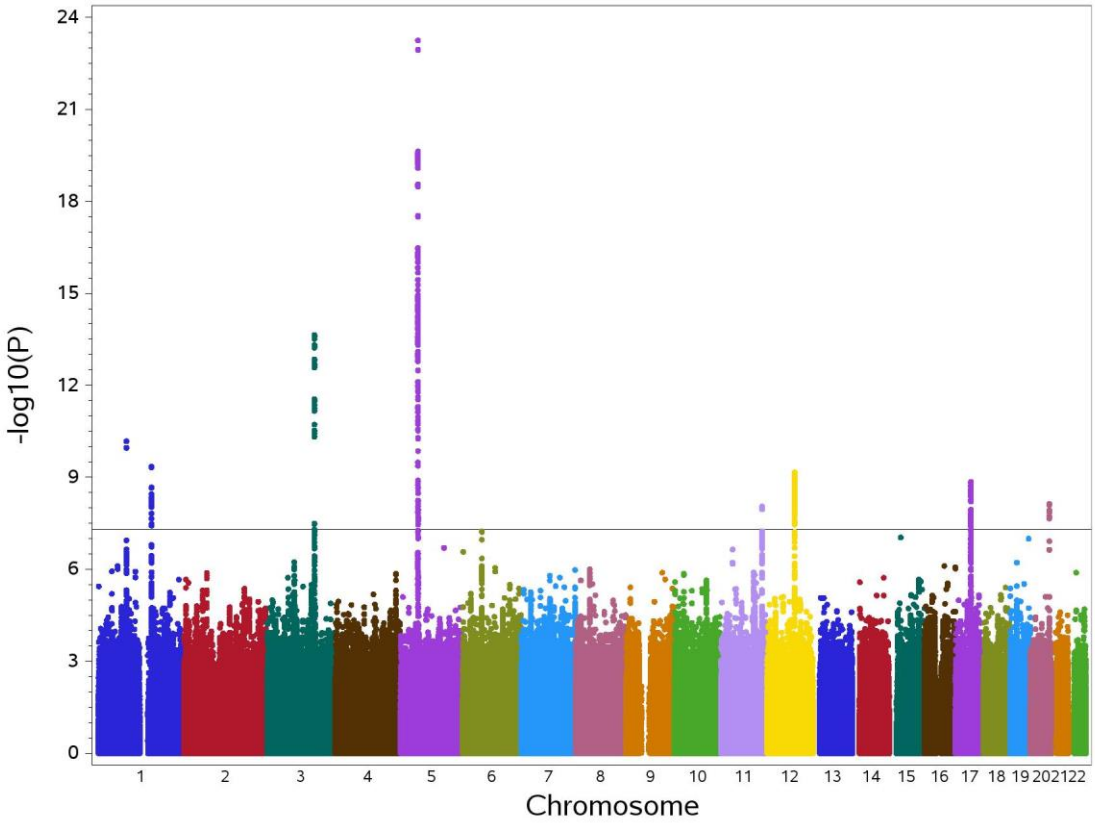
Only individuals within the red ellipses defined by condition (1) were used for the statistical analysis, individuals outside the red ellipsis were discarded. (CEU) Utah Residents from North and West Europe; (CHB) Han Chinese in Beijing, China; (JPT) Japanese in Tokyo, Japan; (YRI) Yoruba in Ibadan, Nigeria.

Supplementary Fig. 16: Meta-analysis Q-Q plot



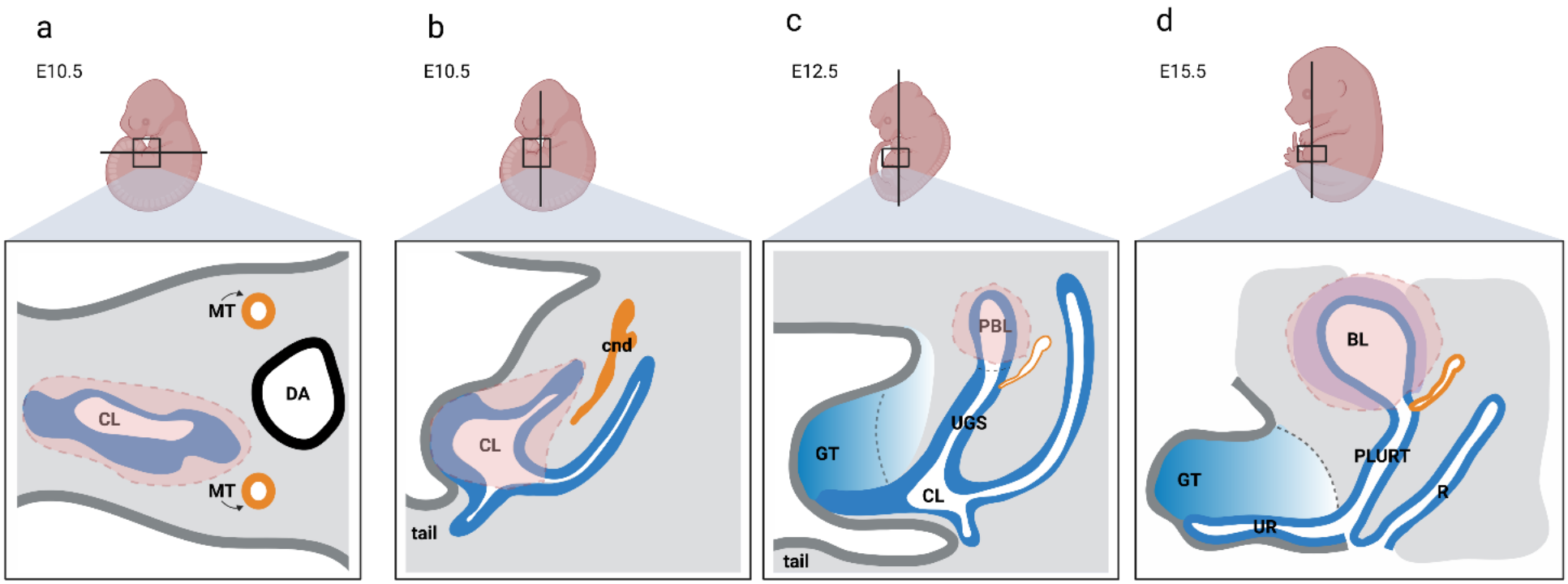
This Figure shows the Q-Q plots for the two-sided *P*-values obtained from the GWAS meta-analysis of BEEC. The X axis shows the expected distribution of -log10(*P*-values) under the null hypothesis of no association. The Y axis shows the distribution of the observed -log10(*P*-values) in the meta-analysis. The red line indicates where Y=X.

Supplementary Fig. 17: Manhattan plot of the Genome-Wide Association Studies for the 628 CBE patients and 7,352 ethnically matched controls.



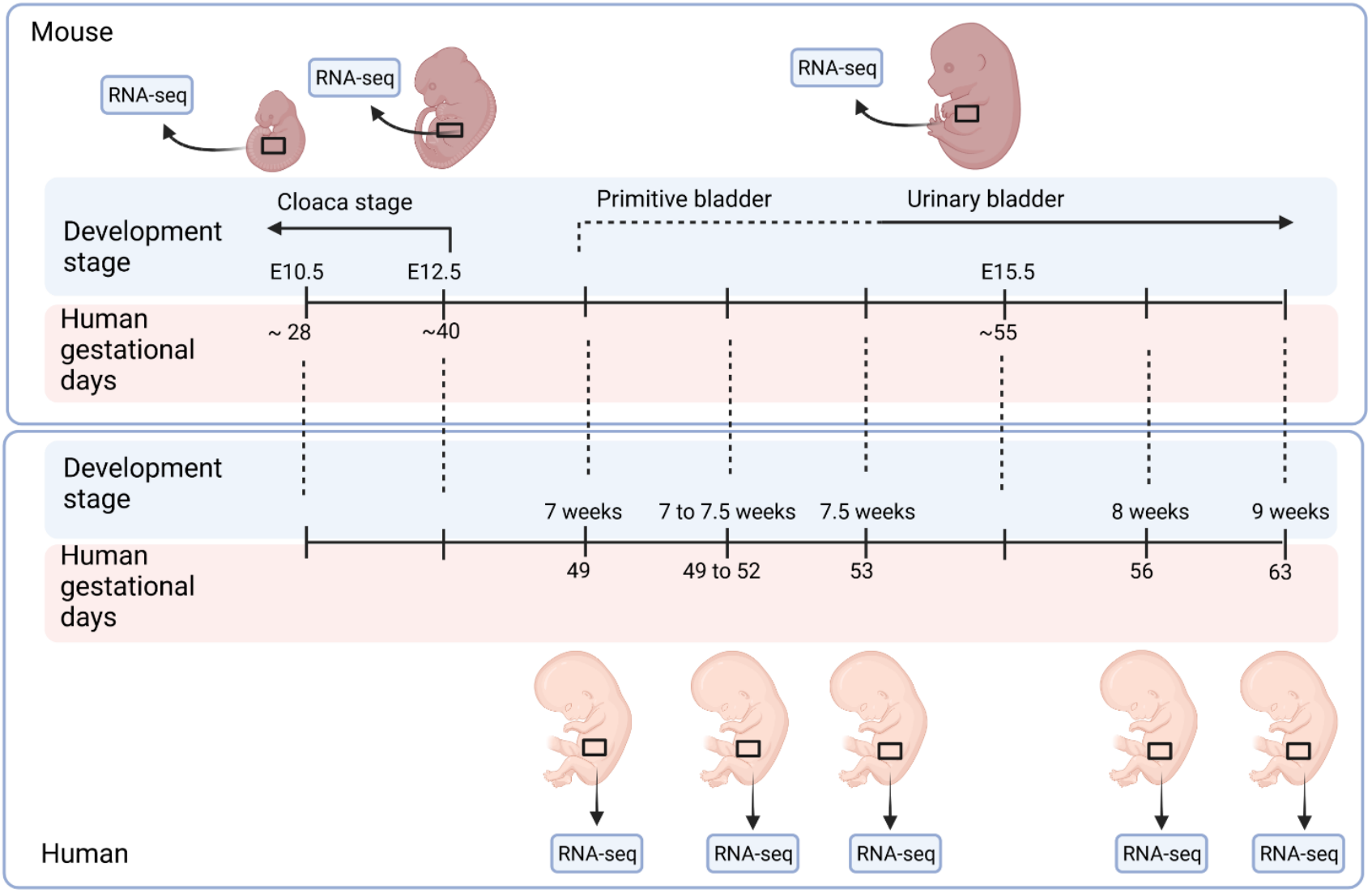
X-axis shows the chromosomes with each dot representing a SNP. *Y*-axis shows -$\log_{10}$(*P*-value) of the association of each SNP in the CBE cohort. Continue black horizontal line indicate the threshold of the genome-wide significance at *P*-value of $5*10^{-8}$.

Supplementary Fig. 18: Anatomical structures and details of mouse embryos dissected tissues for RNA-seq.

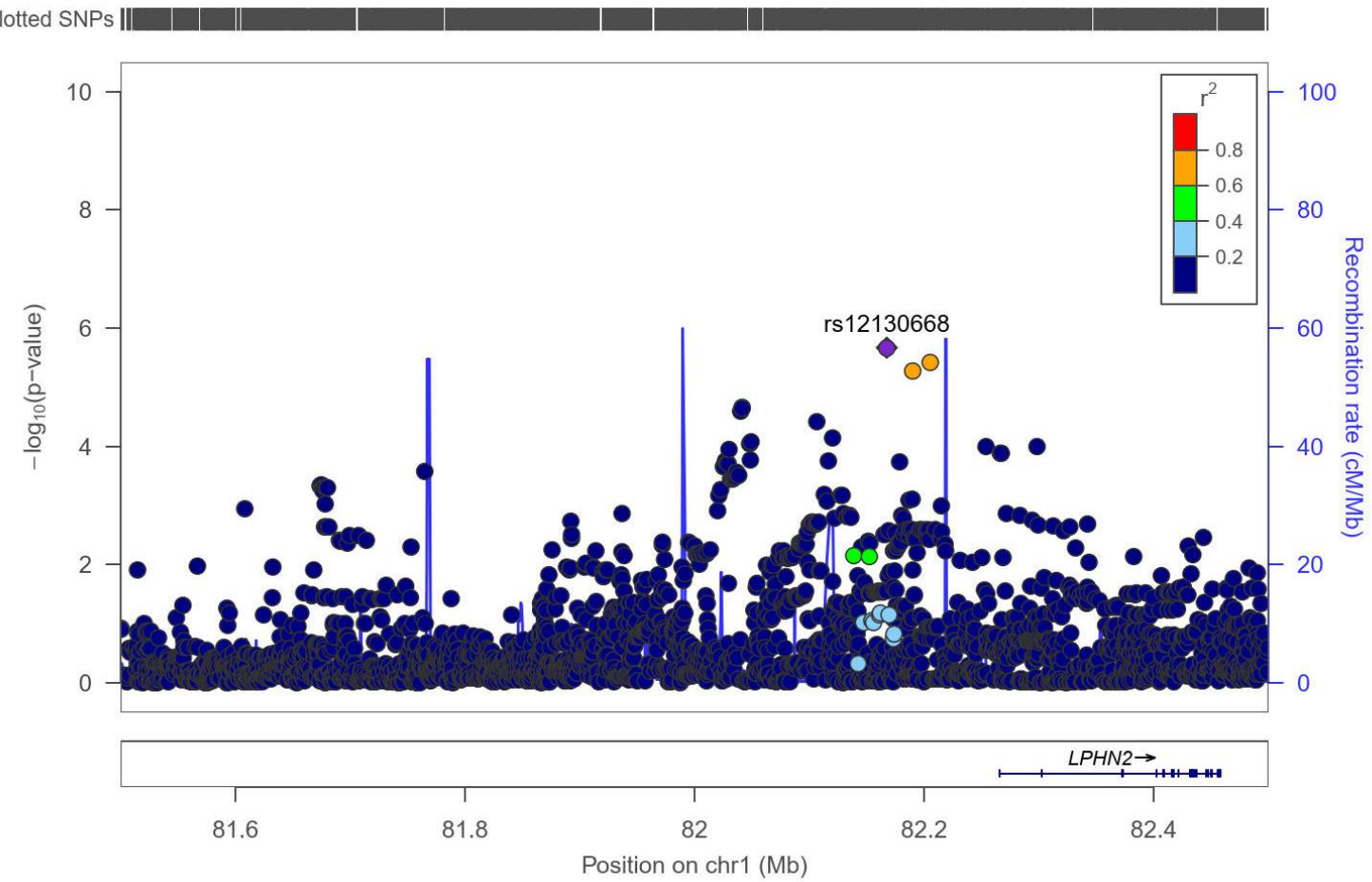


a) Embryonic day E10.5, transversal section. **b**) Embryonic day E10.5, sagittal section. **c**) Embryonic day E12.5, sagittal section. **d**) Embryonic day E15.5, sagittal section. Nomenclature: cloaca (CL); dorsal aorta (DA); mesonephric tubules (MT); common nephric duct (cnd); genital tubercle (GT); urogenital sinus (UGS); primitive bladder (PBL); urethra (UR); pelvic urethra (PLURT); rectum (R); bladder (BL).

Supplementary Fig. 19: Timelines of mouse embryonic and human embryonic and fetal bladder sampling for RNA-seq used in this study.

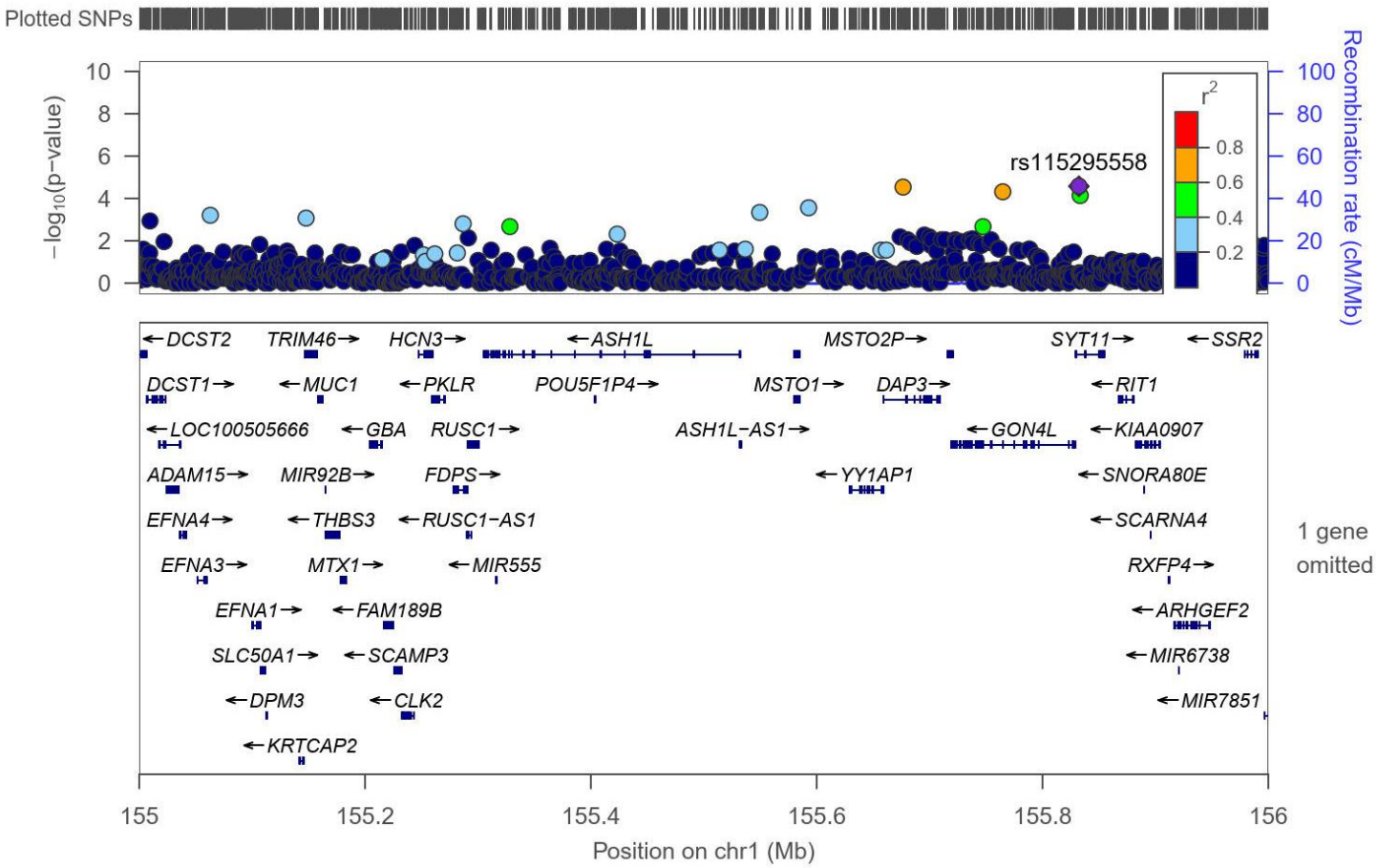


Supplementary Fig. 20: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 1, region 1.



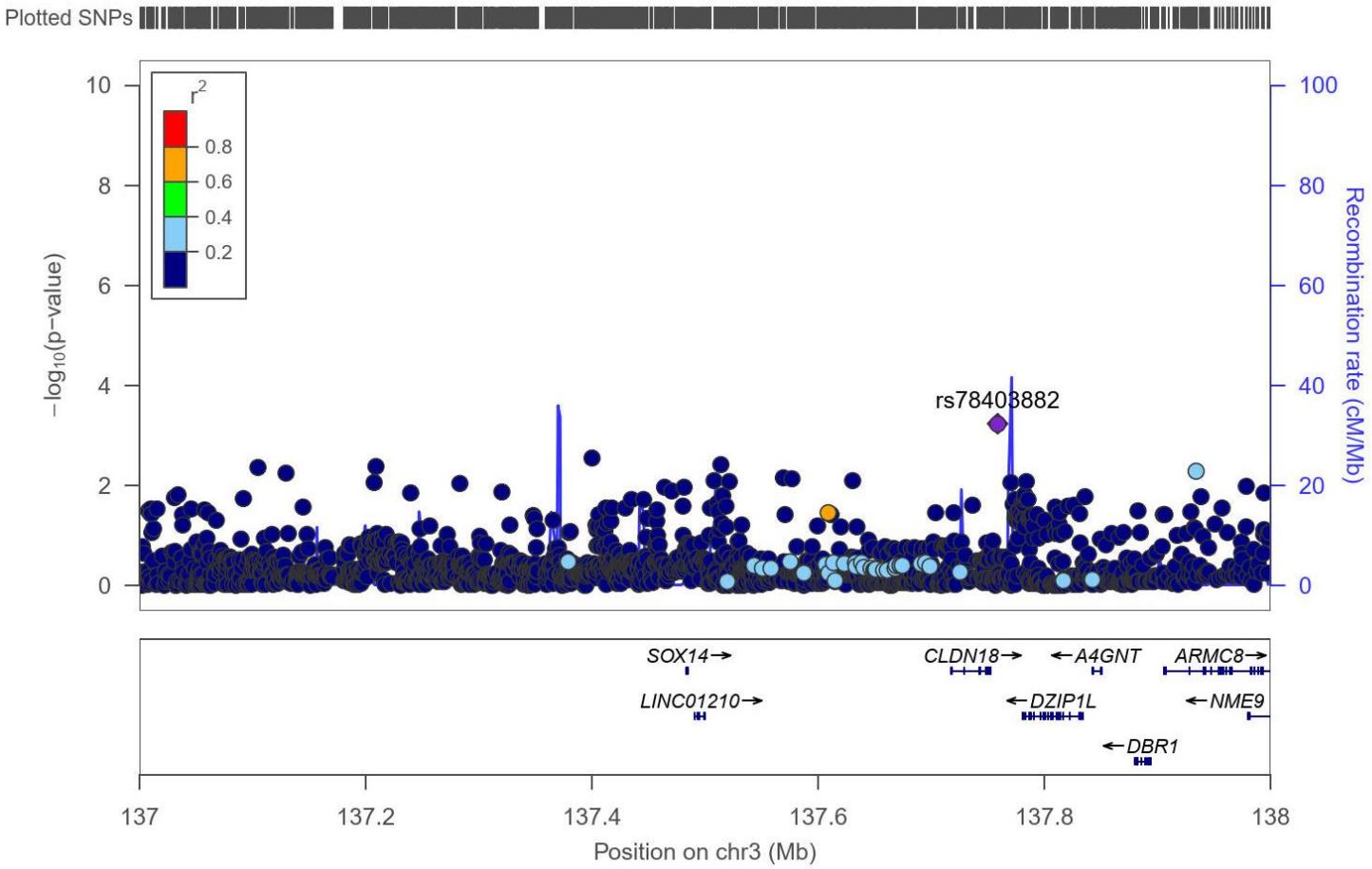
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

Supplementary Fig. 21: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 1, region 2.



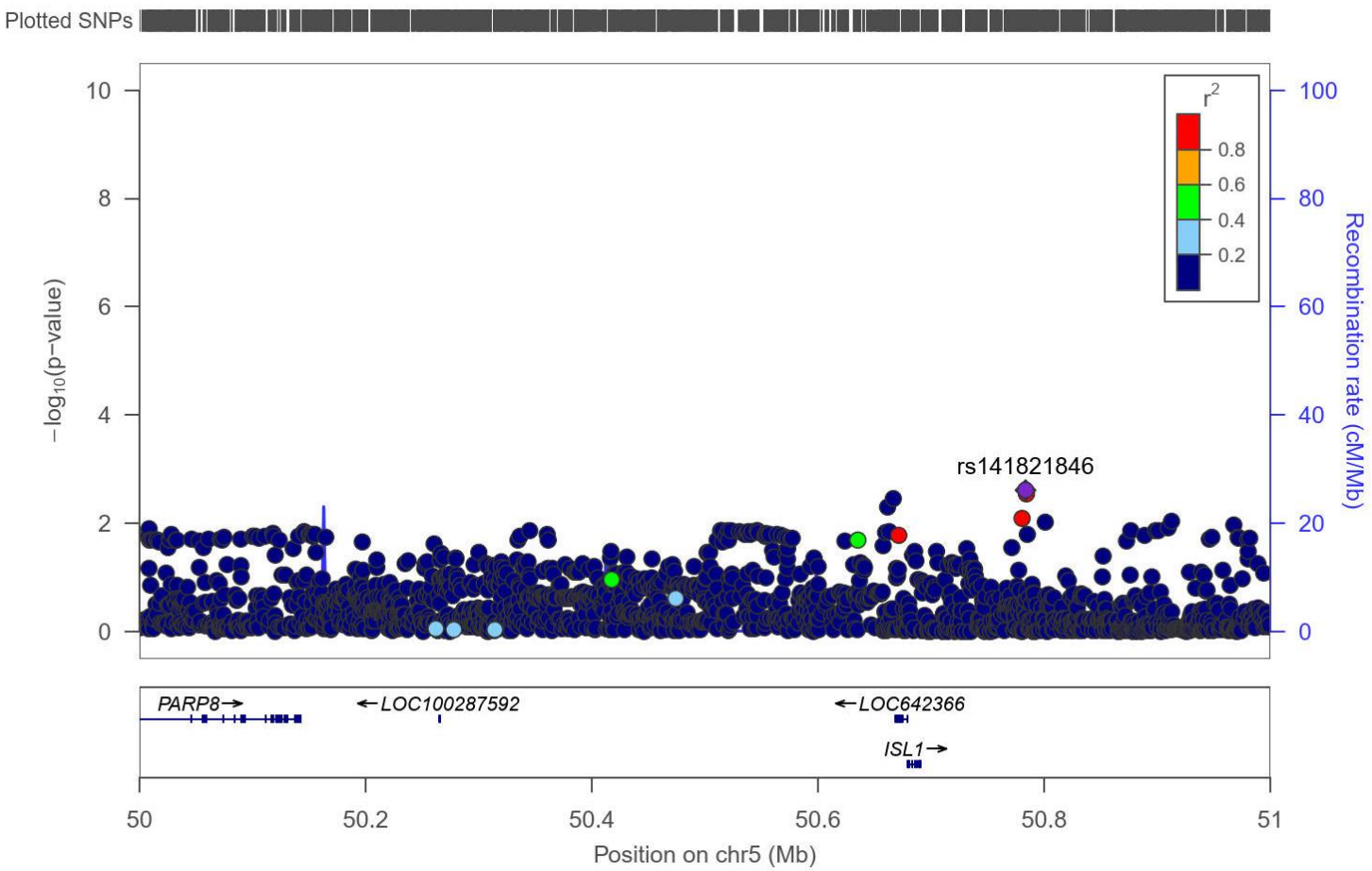
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

Supplementary Fig. 22: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 3.



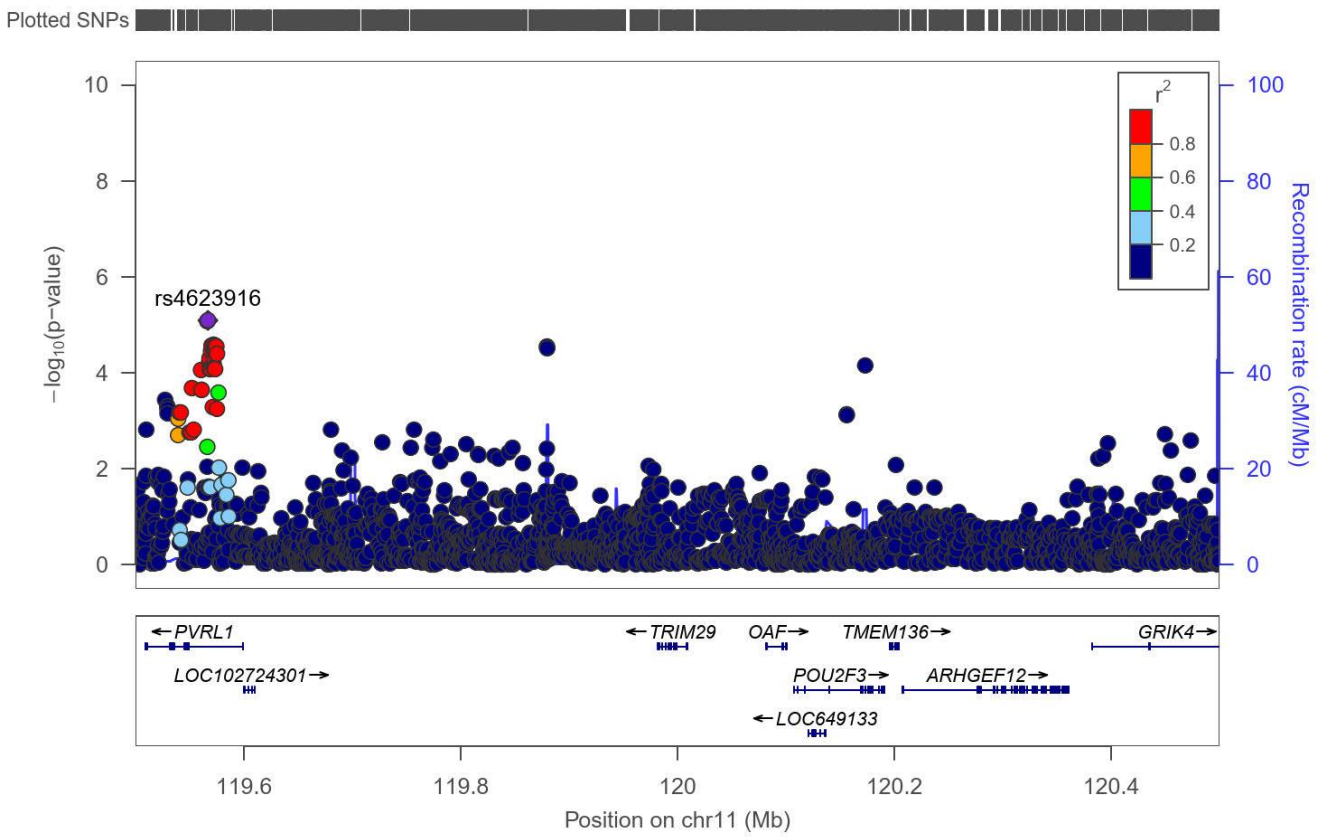
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

Supplementary Fig. 23: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 5.



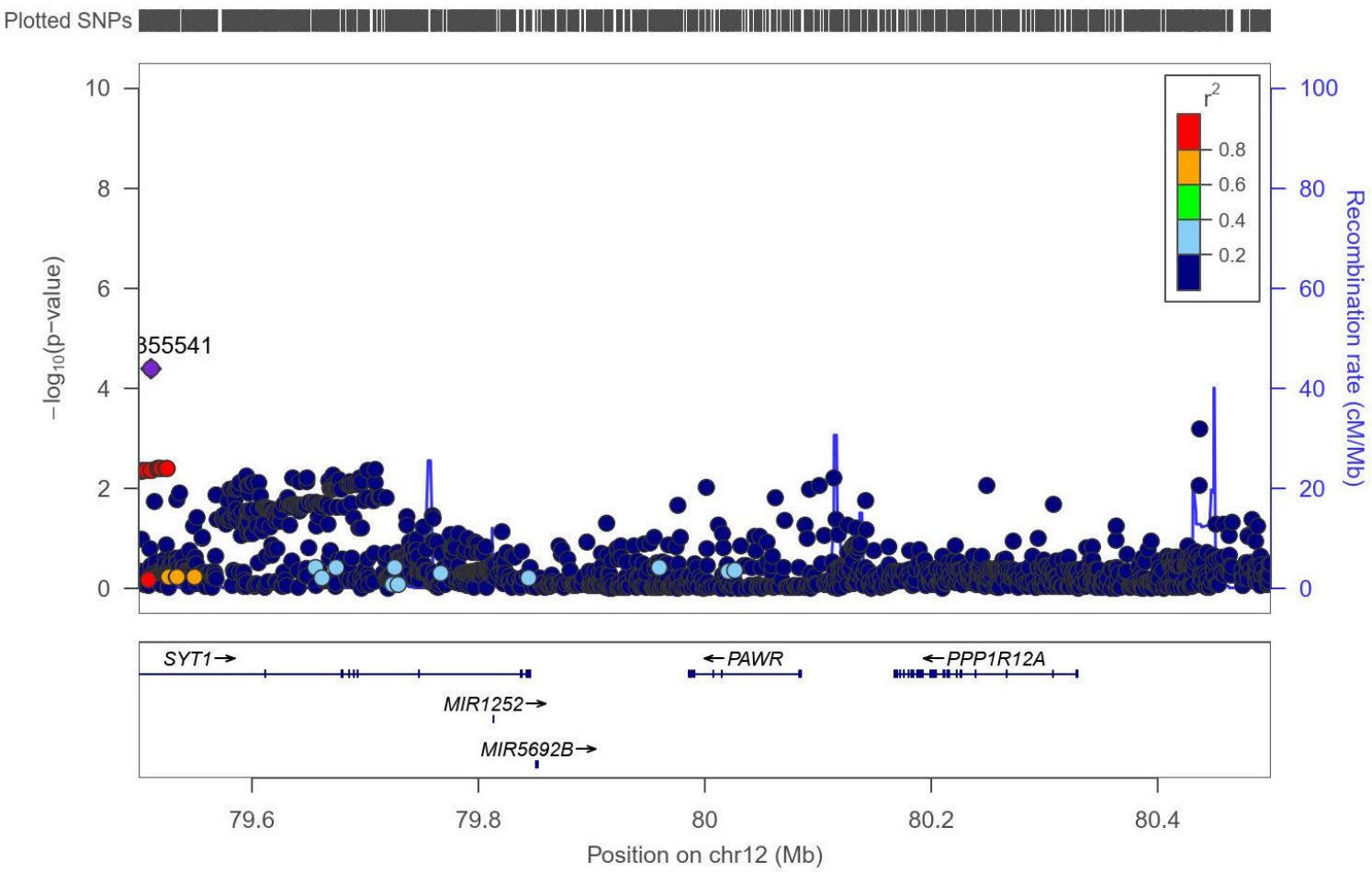
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

Supplementary Fig. 24: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 11.



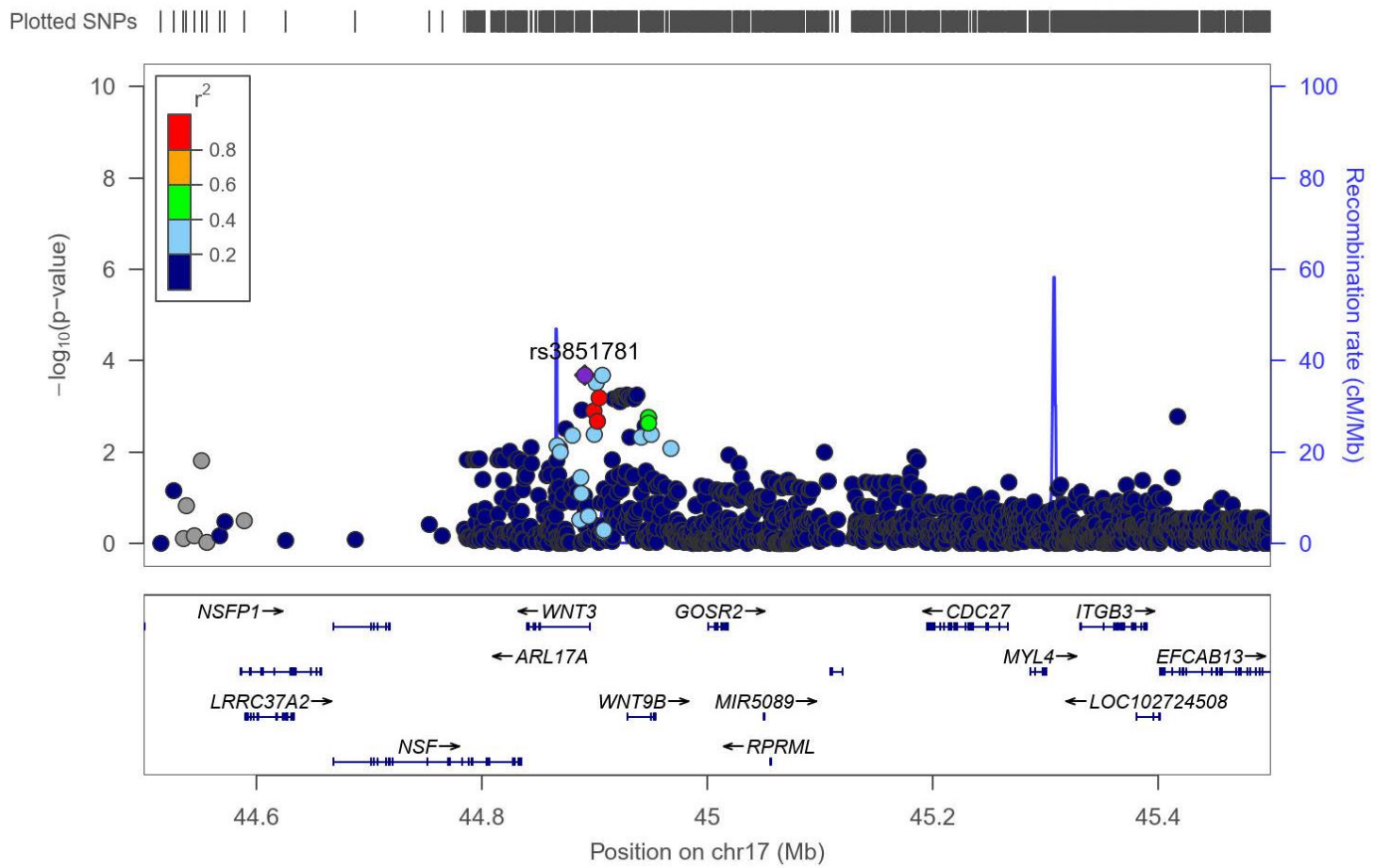
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

Supplementary Fig. 25: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 12.



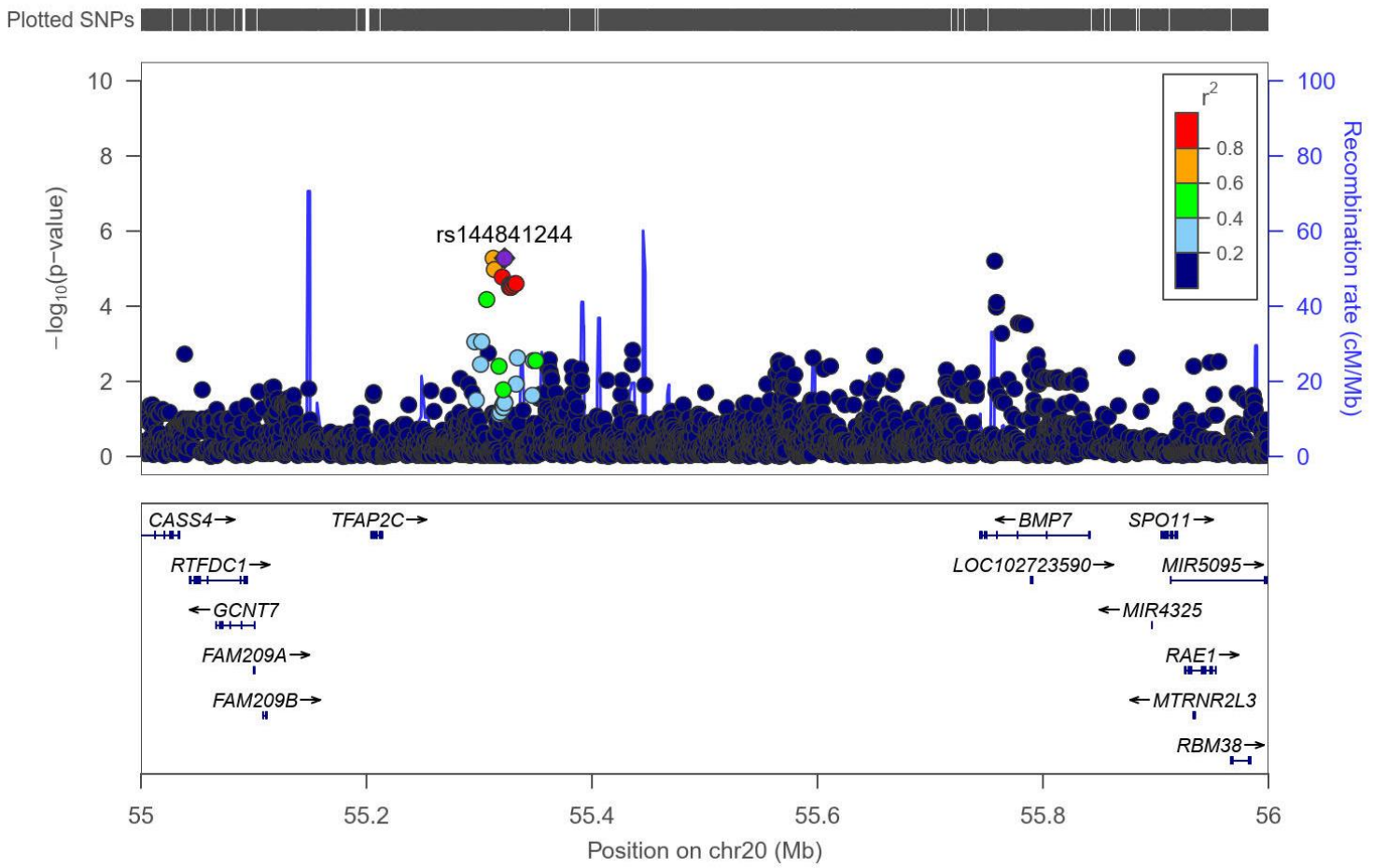
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

Supplementary Fig. 26: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 17.



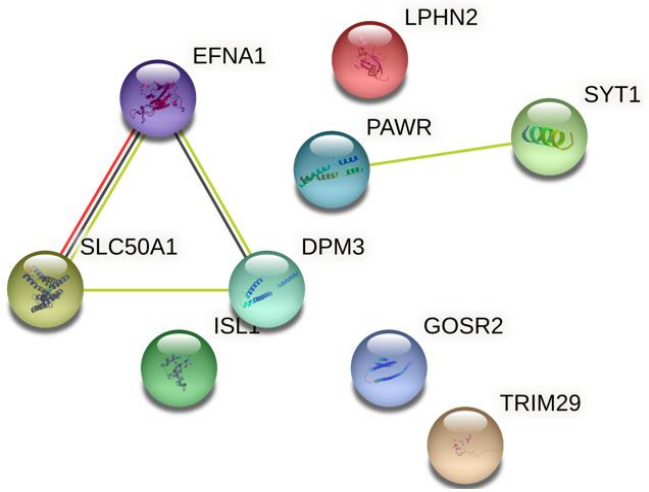
Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

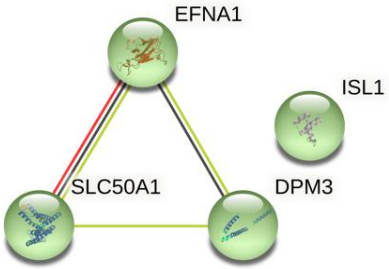
Supplementary Fig. 27: Regional association plots for conditional logistic regression in genome-wide significant locus in chromosome 20.

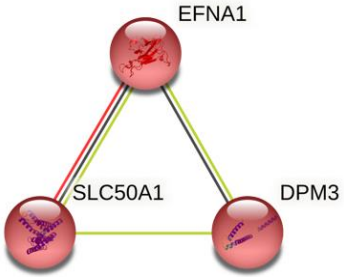


Every dot is a SNP in a 1 Megabase (Mb) window. *X*-axes represent chromosome position of the SNP and the relative genes position in the locus. *Y*-axes represent $-\log_{10}$(*P*-value) of each SNP. Blue peaks indicate the Recombination rate (cM/Mb). The most significant SNP is labelled in purple.

Supplementary Fig. 28. Network analysis of putative candidate genes.

<table>
<tr>
<td><u>Full STRING network</u>
(edges indicate both functional and physical protein associations)
 

Protein-protein interactions for three proteins/genes other than text mining: (i) gene fusions between EFNA1 and SLC50A1, (ii) co-expression between EFNA1 and DPM3, (iii) and EFNA1 and SLC50A1.</td>
<td><u>k- means clustering</u>
(network is clustered to a specified number of clusters)
 

<u>MCL clustering</u>
(network is clustered to a specified "MCL inflation parameter")
 
</td>
<td>Reference publication (PMID: 29659830): Warrington et al. identified the EFNA1-locus, with EFNA1 DPM3, EFNA3, KRTCAP2, and SLC50A1 as a locus that could implicate cholesterol metabolism in steroidogenesis as a link to testosterone exerting a role on 2D:4D ratio variation [5].</td>
</tr>
<tr>
<td>Number of nodes: 9; number of edges: 4; average node degree: 0.889; avg. local clustering coefficient: 0.556; expected number of edges: 0; PPI enrichment p-value: 0.000205</td>
<td colspan="2">Number of nodes: 4; number of edges: 3 (in both clusters); average node degree: 1.5; avg. local clustering coefficient: 0.75; expected number of edges: 0; PPI enrichment p-value: 7.38e-0 (based on the included number of genes/proteins = 4)</td>
</tr>
</table>

Edges:

Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other.

Known Interactions

from curated databases

experimentally determined

Predicted Interactions

gene neighborhood

gene fusions

gene co-occurrence

Others

textmining

co-expression

protein homology