Sacral agenesis: A whole exome sequencing and copy number study

Authors: Robert M. PORSCH[1], Elisa MERELLO[3], Patrizia DE MARCO[3], Guo CHENG[2], Laura RODRIGUEZ[4], Paul K. TAM[2,5], Stacey S. CHERNY[1,6,7], Pak C. SHAM[1,5,6,7], Valeria CAPRA[3#], Maria-Mercè GARCIA-BARCELO[2,5#],Desmond D. CAMPBELL[1,6#]

[#]:co-corresponding authors

Affiliations:

[1]Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

[2]Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

[3]Istituto Giannina Gaslini, Genoa, Italy

[4]AbaCid-Genética. Grupo HM Hospitales. Madrid, Spain

[5]Centre for Reproduction, Development, and Growth, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

[6]Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

[7]State Key Laboratory of Brain and Cognitive Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

Email Addresses:

Robert M. PORSCH (robert.porsch@hku.hk), Elisa MERELLO (ElisaMerello@ospedale-gaslini.ge.it) , Patrizia DE MARCO (PatriziaDeMarco@gaslini.org), Guo CHENG (guochenghnny@gmail.com), Laura RODRIGUEZ (laurarodriguez@abacid.es), Paul K. TAM (paultam@hku.hk_, Stacey S. CHERNY (cherny@hku.hk), Pak C. SHAM (pcsham@hku.hk), Valeria CAPRA (valeriacapra@ospedale-gaslini.ge.it), Maria-Mercè GARCIA-BARCELO (mmgarcia@hku.hk), Desmond D. CAMPBELL (ddc123@hku.hk)

Submitting Author:

Maria-Mercè GARCIA-BARCELO / Room 1-05D, 1/F

The Hong Kong Jockey Club Building for Interdisciplinary Research

5 Sassoon Road, Pokfulam, Hong Kong SAR

ABSTRACT

Background

Caudal regression syndrome (CRS) or sacral agenesis is a rare congenital disorder characterized by a constellation of congenital caudal anomalies affecting the caudal spine and spinal cord, the hindgut, the urogenital system, and the lower limbs. CRS is a complex condition, attributed to an abnormal development of the caudal mesoderm, likely caused by the effect of interacting genetic and environmental factors. A well-known risk factor is maternal-insulin-dependent diabetes mellitus.

Results

In this pilot study, exome sequencing and copy number variation (CNV) analyses of 5 CRS trios implicate diabetes related genes, including *MORN1, ZNF330, CLTCL1 and PDZD2*. *De novo* mutations were identified in *SPTBN5, MORN1 and ZNF330* and inherited damaging mutations in *PDZD2* (homozygous) and *CLTCL1 (*compound heterozygous). In addition, a compound heterozygous mutation in *GLTSCR2*, a direct regulator of the CRS-related gene *PTEN,* was identified.Pathway based tests suggested the involvement of both pancreatic cancer ($\boldsymbol{p} < 1 \times \mathbf{10}^{-4}$), and an immunity-related ($\boldsymbol{p} < 1 \times \mathbf{10}^{-4}$) KEGG pathways.

Two CNV deletions, one *de novo* (3q13.13) and one homozygous (8p23.2), were detected in one of our CRS patients. These deletions overlapped with CNVs previously reported in patients with similar phenotype.

Conclusion

Despite the genetic diversity and the complexity of the phenotype, this pilot study identified genetic features common across CRS patients including mutations in genes associated with diabetes.

Keywords: Sacral agenesis, Caudal regression, Copy-number variation, Whole exome sequencing

BACKGROUND

Caudal Regression Syndrome (CRS; Caudal Dysgenesis Syndrome, Caudal Dysplasia Sequence, Congenital Sacral Agenesis; OMIM 600145) is a rare (1 in 7,500-100,000 births[1, 2]) congenital disorder characterized by varying degrees of spinal column agenesis associated with anomalies of central nervous, genito-urinary, cardiac, respiratory and gastro-intestinal systems [3] with anorectal malformations (ARMs) being the most common.

CRS has been attributed to an abnormal fetal development of the caudal mesoderm that occurs before the fourth week of gestation [4]. During the abnormal gastrulation, prospective notochordal cells that are wrongly specified in terms of their rostrocaudal positional encoding are eliminated. Eventually, fewer cells or even no cells will be available to form the notochord at a given abnormal segmental level. The consequences of such a segmental notochordal paucity are manifold and affect the development of the spinal column and spinal cord as well as of other organs that rely on the notochord as their inductor. If the prospective notochord is depleted a wide array of segmental vertebral malformations develop including segmentation defects, indeterminate or block vertebrae, or absence of several vertebrae. Because of lack of neural induction and absence of a floor plate, fewer prospective neuroectodermal cells will be induced to form the neural tube. The resulting malformation essentially depends on the segmental level and the extent of the abnormality along the longitudinal embryonic axis, with subsequent interference on the processes of primary and/or secondary neurulation [5]. However, what triggers such abnormal events is not known.

The caudal spinal abnormalities are considered the defining characteristics of CRS. Cama et al. [6] and Pang et al. [7] classified the disorder into 5 categories according to the degree of caudal spine involvement: Type I**)** total sacral agenesis with normal or short

transverse pelvic diameter and some lumbar vertebrae possibly missing; Type II**)** total sacral agenesis without involvement of lumbar vertebrae; Type III**)** subtotal sacral agenesis or sacral hypodevelopment; Type IV hemisacrum and Type V**)** coccygeal agenesis.

As with many congenital disorders, a well-known risk factor for CRS is maternal insulin-dependent diabetes mellitus (type 1 diabetes)[8]. CRS has thye highest known relative risk (252) of any congenital diosorder [9]. While animal studies have shown that embryos exposed to higher levels of glucose develop growth anomalies, hyperglycemia has not been associated with abnormal fetal development in humans [10]. The exact mechanism by which maternal diabetes affects fetal development in humans remains unclear [11]. During normal pregnancies insulin sensitivity is reduced at the start of the third trimester in order to provide metabolic fuel for both mother and the developing fetus. However, since insulin is unable to cross the placenta, the fetus starts producing its own insulin in order to metabolize nutrition. It has been suggested that insulin, antibodies to insulin, or some other abnormality of carbohydrate metabolism could affect the development of a genetically susceptible fetus.

Evidence for a genetic cause is provided by the existence of familial segregation as well as animal models. Indeed, while the most severe forms of CRS present sporadically, milder CRS forms can be transmitted within families in a dominant manner with reduced penetrance and phenotypic variability [4]. Thus far, patients presenting association of sacral agenesis type IV , presacral mass, and ARM , a status known as Currarino syndrome, have been associated with mutations in *MNX1* gene [12–17]. Yet *MNX1* mutations account for only 50% of sporadic and 90% of familial cases [17]. Although private mutations in genes such as *VANGL1* [18], *HOXD13*[19] and *PTEN* [20] have been described in sporadic cases with caudal dysgenesis and/or vertebrae anomalies, no firm genetic association has been established.

A CRS-like phenotype can be induced by administration in animals of retinoic acid (RA), lithium, cadmium, sulphamide, or organic solvents [21, 22].Several mutated genes including *Cyp26a1*, *Hoxd13* [23], *Wnt-3a* [24], *Acd*, *Ptf1a*, and *Pcsk5* underlie a CRS-like phenotype in mice [11, 25], yet mutations in the human orthologs have never been identified in CRS patients. Interestingly, the reverse is also true,*Mnx1* (formerly *Hxlb9*)mutant mice do not present Currarino syndrome features[26]. These exceptions to the human-mouse phenotype correlation suggest differences in genetic etiology between humans and experimental organisms [11].

In order to search for genetic risk factors for CRS we have exome sequenced five sporadic CRS cases and their respective healthy parents. Due to the sporadic nature of the disease we have focused on *de novo* or recessive inherited damaging genetic variants. In addition, we investigated the relative burden of rare mutations in pathways. We also used a SNP chip assay in order to identify rare and de novo CNVs.

METHODS

Subjects

The records of patients treated between 1995-2010 at the Neurosurgery Department of Giannina (Genoa, Italy) and at the AbaCid-Genética, Grupo HM Hospitales (Madrid, Spain) for congenital anomalies of the spine were reviewed. For all patients family history, cardiac, respiratory and endocrine status were collected. Neurological, neurophysiological (Somatosensory evoked potential, SEP), radiological, neuroradiologic (MRI), orthopaedic, physical, urological (urodynamic, cystography) and surgical assessments were performed for each case . For this pilot study, we selected four Italian trios (CR5, CR17, CR41, CR46) as well as one trio from Spain (CURR20). Please note that we amended either an A, B or C to the

patient ID to indicated father, mother or child respectively. Selected cases had lower spine agenesis with additional anomalies of axial skeleton and internal organs in common. All cases were sporadic and only one children had a mother with diabetes type I. The local ethical committees approved the study and written consent was obtained from all patients and parents.

Bioinformatics Processing

Capture, alignment and base-calling

Whole exome sequencing (WES) was performed at the Centre of Genomic Sciences of the University of Hong Kong. All five trios were exome sequenced using Illumina HiSeq PE100 and captured with TrueSeq Exome Enrichment kit (Illumina Inc.). Alignment was done using BWA MEM [27] against Human Genome HG-19. Duplicated reads were flagged with Picard-tools [28]. The GATK tool set was used to realign indels, perform base recalibration, remove duplicates, perform indel and SNP calling, and was used for genotype refinement to improve accuracy of genotype calls [29]. A hard filter was used to remove variants with insufficient quality, the GATK recommended criteria were used (see supplementary methods). Relatedness of our participants was investigated with PLINK [30]. We then screened variants for their potential pathogenicity and frequency, retaining for further analysis only variants that were rare. We considered a variant to be rare if its minor allele frequency was ≤1% in each of several public databases (see supplementary methods). We considered a variant to be potentially deleterious according to a score obtained from KGGSeq [31]. KGGSeq's prediction algorithm makes use of available biological information (the mutation's effect on the gene, i.e. stop gain or

loss, frameshift, splice site, missense), as well as scoring from other publicly available prediction algorithms (PolyPhen-2, SIFT and others).

De novo and compound heterozygous

Single nucleotide variants (SNVs) and small indels

Subsequent analysis of *de novo* and compound heterozygous, as well as homozygous, mutations was performed using KGGSeq [31]. We defined a *de novo* mutation as a first time genetic alteration of a specific locus in a proband. Compound heterozygous mutations were defined as the condition in which two nonsynonymous recessive alleles (MAF≤1%) affect two different loci within a certain gene and were inherited from separate parents. The expected probability of *de novo* mutations was evaluated using the framework of Samocha et al. [32] who approximated *de novo* mutation probabilities for each gene. Since a similar framework was not available for compound heterozygous mutations we made use of the only large control trio dataset publicly available, the Genome of the Netherlands (GoNL) [33]. The GoNL is a population dataset containing 250 unaffected parents-offspring trios. On the basis of this data we prioritized genes with a higher than expected frequency of compound heterozygous mutations. Thus we classed as a candidate risk locus any gene for which a recessive or *de novo* model could be constructed in any of our trios using the set of rare potentially deleterious variants we had identified. Detected *de novo*, compound heterozygous and homozygous mutations were validated using Sanger sequencing.

Analysis of kinship revealed misattributed paternity within one family (CR46). Hence the family CR46 was excluded from all family based analyses (*de novo*, compound heterozygous and homozygous mutation analysis).

Copy number variation

Several groups have reported methodology and software for calling CNVs from exome sequencing data. We tried 3 programs (EXCAVATOR [34] , CoNIFER [35], , and CONTRA [36]) for this and found no consistency between used tools. Confirming previous studies demonstrating limited power of CNV detection tools from exome sequencing data [37].

Thus we investigated copy number variation (CNV) in the families CR5, CR17 and CR41 with Illumina's HumanCoreExome-24 beadchip. Quality control of the assayed genotypes was performed using GenomeStudio (Illumina Inc.) using the default settings. CNV calling and *de novo* CNV detection was performed with PennCNV [38]. We then screened CNVs for their potential pathogenicity as follows. We retained for further analysis CNVs which allowed construction of a recessive disease model for any gene in any of our trios. We also retained *de novo* CNVs and rare CNVs. We deemed a CNV to be rare if it did not overlap with any CNV detected in the 1000 Genome Project. *De novo* CNVs were validated by quantitative real-time PCR (ABI Prism 7900 Sequence Detection System; Applied Biosystems) using TaqMan® Copy Number Assay. Ensembl's genome browser was used investigate CNVs on their overlap with genes or regulatory elements [39].

Pathway based test

Analysis was extended by investigating the burden of rare (MAF≤1%) nonsynonymous variants in KEGG [40] defined pathways (174 different pathways). We compared the burden of rare potentially deleterious mutations in patients versus controls in each pathway (see formula below). Controls were acquired from the UK10K project [41]. This control cohort consisted of

929 individuals affected with psychiatric disorders. This dataset was used as control due to its good coverage, making it ideal for investigations of rare variants.

We applied an adjusted burden test to identify pathways with significantly higher burden of rare potentially deleterious mutations in patients versus controls. Due to the differences in population [42] and calling platforms, we modified the classical burden test by standardizing the test statistic *U* by the mean number of rare nonsynonymous mutations:

$$U = \frac{\sum_{i=1}^{n} \sum_{j=1}^{g} x_{ij}}{\bar{x}}$$

where n is the number of subjects in the sample, $\bar{x}$is the sample mean of the number of rare mutations an individual carries across the genome, and $x_{ij} = \{0,1,2\}$ is the genotype call for whether the i^{th} subject carries no rare variants (0), a rare variant from one parent only (1), or a rare variant from both parents (2) in the j^{th} position of the pathway. We estimated the null distribution of *U* via bootstrapping with 10,000 iterations. Hence we randomly choose a set of *n*=5 controls and computed the test statistic *U* 10,000 times to obtain a sample distribution of *U,*against which we compared *U* obtained from our cases to obtain a p-value. P-values were then controlled for multiple testing by reporting the Benjamini & Hochberg False Discovery Rate (FDR) [43]. Furthermore all HLA related genes were dropped from the pathways prior to the analysis. This was necessary because for our sample we used a different aligner software to that used for the UK10K variant calling. The complexity of the HLA region can present difficulties to aligners thus inducing unwanted technical bias [44]. In addition, we only considered those pathways in which patients and controls had at least three or more rare variants[45].

RESULTS

After extensive quality control and MAF (MAF≤1%) filtering we retained 127,344 variants of which 91.5% were known in dbSNP137. Out of these, 25,487 missense, 365 frameshift, 498 nonframe-shift, 54 splicing, 273 stop-gain and 52 stop-loss variants in 18,212 different genes were analyzed in regards to *de novo*, compound heterozygous and homozygous mutations.

De novo variants

We identified two missense and one frameshift *de novo* mutations in three different genes: *MORN1*, *SPTBN5,* and *ZNF330* in patients CR41C, CR5C, and CURR20 respectively (Table 2). *MORN1* encodes MORN (membrane occupation and recognition nexus) repeats [46]. The exact function of this gene is not known, however, in *Toxoplasma gondii* it is known to be involved in nuclear cell division [47]. Furthermore MORN repeats are known to be part of a number of genes, including junctophilins [48] which are involved in cardiomyopathy [49]. Notably, *MORN1* was reported to be produced by insulin producing cells (IPCs) derived from pancreatic stem cells [50]. The estimated probability for a *de novo* mutation to occur in *MORN1* is 0.8%. The probability of having a damaging *de novo* mutation, is lower than that of *MORN1* for 59% of the genes analyzed within the *de novo* framework, is lower than that of *MORN1* [32]. Pathogenicity analysis by KGGSeq suggests that the *de novo* mutation is damaging.

SPTBN5 (OMIM: 605916) is a beta-spectrin encoding protein. It plays an important role in linking proteins, lipids, and cytosolic factors of the cell membrane to the cytoskeletal filament systems of the cell [51]. *SPTBN5* is expressed in the cerebellum, pancreas, kidney, and bladder, as well as in a number of others systems. The gene has not been associated with any disease or disorder. The gene-specific probability of a *de novo* mutation is 1.8% and 99% of

genes have a lower probability making the *de novo* mutation less likely to be causally related. Further, KGGSeq's pathogenic prediction algorithm suggests that this variant is benign.

ZNF330 (OMIM: 609550) is a zinc finger protein with no known disease association and is mainly present in the nucleus during interphase as well as at the centromeres during mitosis [52]. Interestingly, this gene is differentially expressed in pancreatic the Islets of Langerhans and in peripheral blood mononuclear cells [53]. The approximated gene-based *de novo* mutation probability is 0.6%, relatively low but still within the 28th percentile of all genes.

Thus, given the pathogenic nature of the two *de novo* variants and their expression pattern in pancreatic cells, *MORN1* and *ZNF330* are potential candidate genes for CRS.

We detected one *de novo* CNV deletion on 3q13.13 in CR5C (Table 3). The deletion does not seem to encompass any gene or functional element, yet it overlaps with CNVs previously reported in patients with a similar phenotype. In particular, a documented *de novo* deletion in a Japanese patient with OEIS (omphalocele, exstrophy of the cloaca, imperforate anus, spinal defects) complex who also had also sacrum malformation (DECIPHER: 971) overlaps with the *de novo* CNV identified in CR5C.

Homozygous and Compound Heterozygous Mutations

In total we identified 8 compound missense heterozygous and two homozygous (one missense and one non-frameshift) mutations (*PDZD2, SYNGR1)* which passed the described filtering criteria (detailed in Table 2). Strikingly, and as in the *de novo* variants, mutations in genes related to diabetes were detected in two patients. None of the affected genes were recurrent. The two genes associated with diabetes were *PDZD2* and *CLTCL1* and were found mutated in CR5C and CR17C respectively.

PDZD2 has been shown to be an important promoter of fetal pancreatic progenitor cell proliferation [54, 55]. Ma et al. [56] showed that expression of *PDZD2* is specific to pancreatic beta cells. Furthermore, higher concentrations of secreted *PDZD2* in rat insulinoma cell lines were correlated with higher rate of cell proliferation and inhibited transcription of *INS*, an insulin promoter.

CLTCL1 is involved in the intracellular trafficking of glucose transporter *GLUT4* Intracellular trafficking of the glucose transporter GLUT4 from storage compartments to the plasma membrane is triggered in muscle and fat during the body's response to insulin [57].

Noteworthy, a compound heterozygous mutation in *GLTSCR2* (Glioma Tumor Suppressor Candidate Region Gene 2) was identified in patient CR5C. *GLTSCR2*, is expressed at high levels in pancreas and heart, is a tumor suppressor gene and a direct regulator of *PTEN*. Mutations of *PTEN* have been previously identified in a patient affected with *VACTERL* (Vertebral anomalies, Anal atresia, Cardiac defects, Tracheoesophageal fistula and/or Esophageal atresia, Renal & Radial anomalies and Limb defects) [20] which has commonalities with CRS [25].

Additional compound heterozygous mutations were identified in *DNAH10 (CR41),* an inner arm dynein heavy chain. Dynein proteins are implicated in many disorders such as motor neuropathies, cortical development diseases, as well as congenital malformation such as heterotaxia, situs inversus. Moreover, cytoplasmic Dyneins have been reported to interact with Kinesin (KIF1A, mutated in patient CR17) for Interkinetic nuclear migration in neural stem cells [58].

We detected a homozygous CNV deletion encompassing part of 8p23.2 in patient CR5C (Table 3). The CNV does not overlap with known genes but is contained within a duplication found in a patient with abnormal sacrum (DECIPHER: 271204). This documented patient, while

also harboring another deletion (7q34-7q36.3), displayed a great variety of phenotypes including central hypertonia, hypermetropia, long thorax, narrow mouth, seizures, and strabismus, as well as deeply set eyes. Additional detected rare CNVs (see supplementary Material) were overlapping with a number of other genes. However, none were directly related to CRS.

Pathway-based tests

KEGG pathways were tested for an excess of rare nonsynonymous mutations in cases versus controls. Table 4 lists the 10 most significantly burdened pathways. After adjusting for multiple testing by controlling for a False discovery Rate (FDR) of 0.05, two pathways remained significant; the KEGG pathway for pancreatic cancer ($p < 1 \times 10^{-4}$) and graft versus host disease ($p < 1 \times 10^{-4}$). Inspection of the QQ-plot (Figure 1) indicates overall robustness of the test.

Although the above reported candidate risk loci are not part of either of these two pathways, the enrichment of mutations in genes of the pancreatic cancer pathway is in line with our finding of mutations in diabetes-related genes and with previously reported results.

DISCUSSSION

We have identified a number of potential genetic causal mechanisms for CRS. Here we discuss the relevance and implications of these findings, and will outline possible directions for future studies. In addition to the already known genetic risk factors [11], we have been able to identify a number of novel risk loci potentially connected to CRS.

Foremost, all four patients were affected by a homozygous, compound heterozygous mutations or *de novo* variant in a diabetes-relevant (*CLTCL1* and *PDZD2)* or pancreatic expressed (*MORN1* and *ZNF330)* gene. This result is in line with the increased risk of CRS in children born to diabetic mothers. In addition, one *de novo* (3q13.13) and one homozygous CNV (8p23.2) overlap with CNVs reported in patients with similar phenotype. Identification of overlapping CNVs in patients with similar phenotype is the central aim of DECIPHER. Since many patients with rare diseases harbor novel or extremely rare variants, it is crucial to accumulate evidence across different patients in order to foster our genetic understanding of the disease and its mechanism. Furthermore, we identified a heterozygous mutation in *GLTSCR2*, a direct regulator of *PTEN*. *PTEN* has been previously associated with CRS-like phenotypes (VACTER) [11]. Finally, we would like to point out that our patients have a significantly greater number of rare deleterious mutations in the pancreatic cancer and graft-versus-host disease pathways compared to controls. The excess of mutations in pancreatic cancer genes is a result which is consistent with previous research showing similar alteration of the Hedgehog protein (Hh) in patients affected with CRS compared to those suffering from pancreatic cancer [59] and it is tempting to speculate that this may somehow relate to insulin production. Furthermore, *PTEN* is known for its involvement in pancreatic cancer [60], indicating a connection between identified heterozygous mutation in the *PTEN* regulator *GLTSCR2* and the higher burden of rare deleterious mutation within the pancreatic cancer pathway. Equally interesting but not so easily justifiable, is the involvement of the immunology-related graft-versus-host disease pathway, a rare disease characterized by an aggressive immune response of donated tissues against the host organism [61].

There are, however, a number of limitations to our study. First, the sample size is small, but to be expected given the rarity of the disease. Second, we were not able to identify recurrent

affected genes across different patients. Third replication of our pathway burden test results is desirable given the differences in QC pipeline between cases and controls.

The diversity of identified potential disease mechanisms matches that of previous studies [11, 62–64] and also reflects the phenotypic diversity associated with CRS [62]. Furthermore, we [65] have shown that given a complex genetic disorder, one should expect a large genetic heterogeneity across patients. Thus the number of candidate genes identified is not surprising and is similar to that reported for other complex rare genetic disorders [66]. Further the increasing amount of candidate genes as well as those reported by others [25, 67–71] suggests that CRS might be caused by a multitude of private genetic risk factors. This makes it difficult to identify a common underlying genetic architecture. Differences in the genetic etiology between humans and experimental organisms makes it challenging to investigate the exact causal mechanism. In addition, some principal aspects of the disease are still unknown. A challenge, for example, is the uncertainty of the overall frequency of the disease. While some studies have estimated that 1 in 7,700 children might be affected by CRS [1], others suggest it might be as rare as 1 in 100,000 births [2]. This further complicates estimation of the expected number of disease causing mechanisms [72].

Conclusion

Despite the complexity of the phenotype, we were able to identify common genetic characteristics across patients, potentially causally related to the present phenotypes. Furthermore our data, although limited to a small group of patients, support a multigenic model for CRS. Future studies should consider larger accumulated samples across multiple centers in order to identify common genetic characteristics via whole genome or whole exome sequencing.

List of abbreviations

CRS, Caudal Regression; CNV, Copy Number Variation; ARM, Anorectal Malformation; SEP, Somatosensory Evoked Potential; WES, Whole Exome Sequencing; SNP, Single Nucleotide Polymorphism; SNV, Single Nucleotide Variation; MAF, Minor Allele Frequency; FDR, False Discovery Rate; IPC, Insuline Producing Cells; Hh, Hedgehog protein; QC, Quality Control

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The corresponding sequencing data, on which this study is based on, can be accessed through the European Genome-phenome Archive (EGA).

Author's contributions

RMP, DDC, SSC, GC and MMGB analyzed the data and drafted the paper. VC, EM, PDM and LR played a major role in collecting the samples and phenotyping the patients. PCS and PKT reviewed the study proposal, provided feedback on the study progress and mansucript. MMGB proposed the study idea.

Acknowledgements

We thank all participants who made this study possible.

Seed Funding Programme for Basic Research, University of Hong Kong,Project Code:201410159002 to MMGB.

Small Project Funding, University of Hong Kong, Project Code 201209176012 to DC.

EM; PDM, VC would like to acknowledge Ricerca Corrente Ministero della Salute Italia 5X Mille, Aletti- Volpati Trust Onlus and private funding resources. We also thank A.S.B.I. (Associazione Spina Bifida Italia).

Tables and Figures

Table 1. Clinical characteristics of the patients included in this study

Subjects	Sacral agenesis[1] and vertebral malformations	Ribs/Limbs anomalies	Genitourinary	Neural tube	ARM	Cardiac	Other	Maternal phenotype
CR5/M	CRS[1]-Type II Hemivertebra T7-T8	Additional 13th right rib					CPT (II) carrier	CPT (II) deficiency
CR17/F	CRS Type II		Hydronephrosis Hydroureter Bladder-exstrophy	Lipoma Low-lying conus medullaris			Omphalocele Twisted teeth	Diabetes type I
CR41/F	CRS Type I Deformed T7-T8-T9	Fusion of 5th and 6th left ribs Additional 13th left rib Club feet	Incontinence	Lipoma Blunt ending conus medullaris (T11)		Pulmonary vein atresia Inter-ventricular septal defect	Congenital hip dislocation. Motor delay.	
CR46/M	CRS Type II Lumbar kyphosis	Club feet			Anal Stenosis	Intra-ventricular septal defect	Congenital bilateral hip dislocation. Short neck.	Hydrocephalus
CURR20/F	CRS Type V				Anal Stenosis		Teratoma	

[1]According to Cama et al. [6] and Pang et al. [7] classification of sacral agenesis;. M: male; F: female; T= thoracic vertebra; S: sacral vertebra; ARM: anorectal malformations; CPT: Carnitine palmitoyl transferase

Table 2. *De novo*, compound heterozygous and homozygous variants.

Subjects	Genes	Nucleotide variants	RsID	Aminoacidic variants	OMIM Associated Disease	Functional Role	Mutation status
CR5C							
	SPTBN5	c.73G>A	-	p.(Glu25Lys)	-	interaction of cytoskeletal filament with other components of the cell[51]	*De novo*

	PDZD2	c.3317C>T	rs34748216	p.(Ser1106Phe)	-	insuline regulation[56]	H
	PKHD1L1	c.8291A>C c.11969G>A	rs118074609 / rs146831382	p.(Asn2764Thr)/p.(Gly3990Glu)	-	cellular immunity [73]	CH
	GLTSCR2	c.568C>T c.851C>T	rs34462252 / rs200463741	p.(Arg190Trp)/p.(Thr284Met)	-	PTEN regulation [74]	CH
CR17C							
	ARHGEF16	c.784A>G c.1477C>T	- / -	p.(Thr262Ala)/p.(Leu493Phe)	-	guanyl-nucleotide exchange factor	CH
	KIF1A	c.4781C>T c.2522A>T	rs201825284 / -	p.(Ser1594Leu)/p.(Asn841Leu)	mental retardation, spastic paraplegia-30, neuropathy	synaptic-vesicle transportation[75]	CH
	CLTCL1	c.4859G>A c.130G>T	rs5748024 / rs34869740	p.(Arg1620His)/p.(Val44Phe)		intracellular trafficking of the glucose transporter GLUT4 [76]	CH
CR41C							
	MORN1	c.319G>A	-	p.(Gly107Arg)	cardiomyopathy, hypertrophic-17	intracellular ion chanel communication [77]	*De novo*
	DNAH10	c.4846G>A c.10859C>T	- / rs202063832	p.(Ala1616Thr)/p.(Thr3620Leu)	primary ciliary dyskinesia	inner arm dynein heavy chain [78]	CH
CURR20C							
	ZNF330	c.6_7insT	-	p.(Lys3fs)	-	-	*De novo*
	VPS18	c.1697A>G c.1823G>A	- / -	p.(Tyr566Cys)/p.(Arg608His)	-	protein transportation to the vacuole [79]	CH
	PKD1L2[1]	c.6241_6242ins19nt[2] c.706_707delAA	- / -	p.(Thr2081fs)/p(Asn236fs)		-	CH
	SYNGR1	c.C606_607insCAA		p.(Pro202_Thr1insGln)		synaptic plasticity [80]	H

CH: compound heterozygous; H: homozygous; [1]mutatation information are given for the long form of the transcript (NCBI reference: NM_052892);
[2]19nt:GCTTTCCCCAGGCTTGGCAGTA

Table3. *De novo* and homozygous CNVs

Patients	Chromosome	Start Position	End Position	Length	Type	Genes or regulatory elements	Patients with related symptoms listed in DECIPHER (type of CNV, patient phenotype)
CR5C							
	3q13.13	109489534	109510473	20939	*De novo*/deletion	-	971, deletion; abnormality of the sacrum, Abnormality of the small intestine, Anal atresia, Cloacal exstrophy, Omphalocele, Spina bifida occulta.
	8p23.2	5599399	5605087	5688	homozygous/deletion	-	271204, duplication; Abnormality of the sacrum, Central hypotonia, Deeply set eye, Hypermetropia, Long thorax, Narrow mouth, Nasogastric tube feeding in infancy, Seizures, Strabismus.

Table 4. KEGG pathways most significantly enriched for potentially deleterious mutations

Pathway	P-value	FDR	Number of genes
Pancreatic cancer	$<1\times10^{-04}$	$<1\times10^{-04}$	70
Graft versus host disease	$<1\times10^{-04}$	$<1\times10^{-04}$	22
Antigen processing and presentation	0.002	0.081	69
Colorectal cancer	0.002	0.081	62
Folate biosynthesis	0.002	0.081	11
Neuroactive ligand receptor interaction	0.003	0.081	272
Glycosaminoglycan degradation	0.003	0.081	21
Endometrial cancer	0.010	0.190	52
Autoimmune thyroid disease	0.010	0.190	33
Intestinal immune network for IGA production	0.012	0.190	34

References

1. Unsinn KM, Geley T, Freund MC, Gassner I: **US of the spinal cord in newborns: spectrum of normal findings, variants, congenital anomalies, and acquired diseases.** *Radiographics* , **20**:923–38.

2. Entezami M: *Ultrasound Diagnosis of Fetal Anomalies*. Thieme; 2003.

3. Duhamel B: **From the Mermaid to Anal Imperforation: The Syndrome of Caudal Regression***. *Arch Dis Child* 1961, **36**:152–155.

4. Barkovich a J, Raghavan N, Chuang S, Peck WW: **The wedge-shaped cord terminus: a radiographic sign of caudal regression.** *AJNR Am J Neuroradiol* 1989, **10**:1223–31.

5. Tortori-Donati P, Fondelli MP, Rossi a, Raybaud C a, Cama a, Capra V: **Segmental spinal dysgenesis: neuroradiologic findings with clinical and embryologic correlation.** *AJNR Am J Neuroradiol* 1999, **20**:445–56.

6. Cama A, Palmieri A, Capra V, Piatelli GL, Ravegnani M, Fondelli P: **Multidisciplinary management of caudal regression syndrome (26 cases).** *Eur J Pediatr Surg Off J Austrian Assoc Pediatr Surg . [et al]* = *Zeitschrift fu□r Kinderchirurgie* 1996, **6 Suppl 1**:44–5.

7. Pang D, Hoffman HJ, Menezes AH: **Sacral agenesis and caudal spinal cord malformations**. *Neurosurgery* 1993, **32**:755–779.

8. Passarge E, Lenz W: **Syndrome of caudal regression in infants of diabetic mothers: observations of further cases.** *Pediatrics* 1966, **37**:672–5.

9. Reece EA, Hobbins JC: **Diabetic embryopathy: pathogenesis, prenatal diagnosis and prevention.** *Obstet Gynecol Surv* 1986, **41**:325–35.

10. Persaud O: **Maternal diabetes and the consequences for her offspring**. *J Devel Disabil*

2007, **13**:101 –133.

11. Semba K: **Etiology of Caudal Regression Syndrome**. *Hum Genet Embryol* 2013, **03**.

12. Hagan DM, Ross a J, Strachan T, Lynch S a, Ruiz-Perez V, Wang YM, Scambler P, Custard E, Reardon W, Hassan S, Nixon P, Papapetrou C, Winter RM, Edwards Y, Morrison K, Barrow M, Cordier-Alex MP, Correia P, Galvin-Parton P a, Gaskill S, Gaskin KJ, Garcia-Minaur S, Gereige R, Hayward R, Homfray T: **Mutation analysis and embryonic expression of the HLXB9 Currarino syndrome gene.** *Am J Hum Genet* 2000, **66**:1504–15.

13. Garcia-Barceló M, So M-T, Lau DK-C, Leon TY-Y, Yuan Z-W, Cai W-S, Lui VC-H, Fu M, Herbrick J-A, Gutter E, Proud V, Li L, Pierre-Louis J, Aleck K, van Heurn E, Belloni E, Scherer SW, Tam PK-H: **Population differences in the polyalanine domain and 6 new mutations in HLXB9 in patients with Currarino syndrome.** *Clin Chem* 2006, **52**:46–52.

14. Wang Y, Wu Y: **A novel HLXB9 mutation in a Chinese family with Currarino syndrome.** *Eur J Pediatr Surg* 2012, **22**:243–5.

15. Ross a J, Ruiz-Perez V, Wang Y, Hagan DM, Scherer S, Lynch S a, Lindsay S, Custard E, Belloni E, Wilson DI, Wadey R, Goodman F, Orstavik KH, Monclair T, Robson S, Reardon W, Burn J, Scambler P, Strachan T: **A homeobox gene, HLXB9, is the major locus for dominantly inherited sacral agenesis.** *Nat Genet* 1998, **20**:358–61.

16. Kim AY, Yoo S-Y, Kim JH, Eo H, Jeon TY: **Currarino syndrome: variable imaging features in three siblings with HLXB9 gene mutation.** *Clin Imaging* 2013, **37**:398–402.

17. Belloni E, Martucciello G, Verderio D, Ponti E, Seri M, Jasonni V, Torre M, Ferrari M, Tsui LC, Scherer SW: **Involvement of the HLXB9 homeobox gene in Currarino syndrome.** *Am J Hum Genet* 2000, **66**:312–9.

18. Kibar Z, Torban E, McDearmid JR, Reynolds A, Berghout J, Mathieu M, Kirillova I, De Marco P, Merello E, Hayes JM, Wallingford JB, Drapeau P, Capra V, Gros P: **Mutations in VANGL1 associated with neural-tube defects**. *N Engl J Med* 2007, **356**:1432–1437.

19. Garcia-Barceló M-M, Wong KK, Lui VC, Yuan Z, So M, Ngan ES, Miao X, Chung PH, Khong P, Tam PK: **Identification of a HOXD13 mutation in a VACTERL patient.** *Am J Med Genet A* 2008, **146A**:3181–5.

20. Reardon W, Zhou XP, Eng C: **A novel germline mutation of the PTEN gene in a patient with macrocephaly, ventricular dilatation, and features of VATER association**. *J Med Genet* 2001, **38**:820–823.

21. Padmanabhan R: **Retinoic acid-induced caudal regression syndrome in the mouse fetus.** *Reprod Toxicol* , **12**:139–51.

22. Rojansky N, Fasouliotis SJ, Ariel I, Nadjari M: **Extreme caudal agenesis. Possible drug-related etiology?** *J Reprod Med* 2002, **47**:241–5.

23. Young T, Rowland JE, van de Ven C, Bialecka M, Novoa A, Carapuco M, van Nes J, de Graaff W, Duluc I, Freund J-N, Beck F, Mallo M, Deschamps J: **Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos.** *Dev Cell* 2009, **17**:516–26.

24. Greco TL, Takada S, Newhouse MM, McMahon J a, McMahon a P, Camper S a: **Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development.** *Genes Dev* 1996, **10**:313–324.

25. Szumska D, Pieles G, Essalmani R, Bilski M, Mesnard D, Kaur K, Franklyn A, El Omari K, Jefferis J, Bentham J, Taylor JM, Schneider JE, Arnold SJ, Johnson P, Tymowska-Lalanne Z,

Stammers D, Clarke K, Neubauer S, Morris A, Brown SD, Shaw-Smith C, Cama A, Capra V, Ragoussis J, Constam D, Seidah NG, Prat A, Bhattacharya S: **VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5.** *Genes Dev* 2008, **22**:1465–77.

26. Catala M: **Genetic control of caudal development**. *Clin Genet* 2002, **61**:89–96.

27. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**. 2013, **00**:3.

28. **Picard** [http://broadinstitute.github.io/picard/]

29. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ: **A framework for variation discovery and genotyping using next-generation DNA sequencing data.** *Nat Genet* 2011, **43**:491–8.

30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559–75.

31. Li M-X, Gui H-S, Kwan JSH, Bao S-Y, Sham PC: **A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases.** *Nucleic Acids Res* 2012, **40**:e53.

32. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki J a, Rehnström K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook EH, Gibbs R a, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ: **A framework for the interpretation of de novo**

mutation in human disease. *Nat Genet* 2014(August).

33. The Genome of the Netherlands Consortium: **Whole-genome sequence variation, population structure and demographic history of the Dutch population.** *Nat Genet* 2014.

34. Magi A, Tattini L, Cifola I, D'Aurizio R, Benelli M, Mangano E, Battaglia C, Bonora E, Kurg A, Seri M, Magini P, Giusti B, Romeo G, Pippucci T, De Bellis G, Abbate R, Gensini GF: **EXCAVATOR: detecting copy number variants from whole-exome sequencing data.** *Genome Biol* 2013, **14**:R120.

35. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, Quinlan AR, Nickerson DA, Eichler EE: **Copy number variation detection and genotyping from exome sequence data**. *Genome Res* 2012, **22**:1525–1532.

36. Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, Tothill RW, Halgamuge SK, Campbell IG, Gorringe KL: **CONTRA: Copy number analysis for targeted resequencing**. *Bioinformatics* 2012, **28**:1307–1313.

37. Tan R, Wang Y, Kleinstein SE, Liu Y, Zhu X, Guo H, Jiang Q, Allen AS, Zhu M: **An evaluation of copy number variation detection tools from whole-exome sequencing data.** *Hum Mutat* 2014, **35**:899–907.

38. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF a, Hakonarson H, Bucan M: **PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data**. *Genome Res* 2007, **17**:1665–1674.

39. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt SE, Janacek SH,

Johnson N, Juettemann T, Kähäri AK, Keenan S, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Overduin B, Parker A, Patricio M, Perry E, Pignatelli M, et al.: **Ensembl 2015.** *Nucleic Acids Res* 2014, **43**:D662–669.

40. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information, knowledge and principle: Back to metabolism in KEGG**. *Nucleic Acids Res* 2014, **42**:199–205.

41. UK10K Consortium: **The UK10K project identifies rare variants in health and disease.** *Nature* 2015, **526**:82–90.

42. Zawistowski M, Reppell M, Wegmann D, St Jean PL, Ehm MG, Nelson MR, Novembre J, Zöllner S: **Analysis of rare variant population structure in Europeans explains differential stratification of gene-based tests.** *Eur J Hum Genet* 2014(August 2013):1–8.

43. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing**. *J R Stat Soc Ser B* 1995, **57**:289–300.

44. Huang Y, Yang J, Ying D, Zhang Y, Shotelersuk V, Hirankarn N, Sham PC, Lau YL, Yang W: **HLAreporter: a tool for HLA typing from NGS data**. *Genome Med* 2015, **7**:25.

45. Morris AP, Zeggini E: **An evaluation of statistical approaches to rare variant analysis in genetic association studies.** *Genet Epidemiol* 2010, **34**:188–93.

46. Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, Klein SL, Old S, Rasooly R, Good P, Guyer M, Peck AM, Derge JG, Lipman D, Collins FS, Jang W, Sherry S, Feolo M, Misquitta L, Lee E, Rotmistrovsky K, Greenhut SF, Schaefer CF, Buetow K, Bonner TI, Haussler D, Kent J, Kiekhaus M, Furey T, Brent M, et al.: **The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC).** *Genome Res*

2004, **14**:2121–7.

47. Ferguson DJP, Sahoo N, Pinches RA, Bumstead JM, Tomley FM, Gubbels M-J: **MORN1 has a conserved role in asexual and sexual development across the apicomplexa.** *Eukaryot Cell* 2008, **7**:698–711.

48. Takeshima H, Komazaki S, Nishi M, Iino M, Kangawa K: **Junctophilins: a novel family of junctional membrane complex proteins.** *Mol Cell* 2000, **6**:11–22.

49. Matsushita Y, Furukawa T, Kasanuki H, Nishibatake M, Kurihara Y, Ikeda A, Kamatani N, Takeshima H, Matsuoka R: **Mutation of junctophilin type 2 associated with hypertrophic cardiomyopathy.** *J Hum Genet* 2007, **52**:543–8.

50. Ebrahimie M, Esmaeili F, Cheraghi S, Houshmand F, Shabani L, Ebrahimie E: **Efficient and simple production of insulin-producing cells from embryonal carcinoma stem cells using mouse neonate pancreas extract, as a natural inducer**. *PLoS One* 2014, **9**.

51. Stabach PR, Morrow JS: **Identification and characterization of beta V spectrin, a mammalian ortholog of Drosophila beta H spectrin.** *J Biol Chem* 2000, **275**:21385–95.

52. Bolívar J, Díaz I, Iglesias C, Valdivia MM: **Molecular cloning of a zinc finger autoantigen transiently associated with interphase nucleolus and mitotic centromeres and midbodies. Orthologous proteins with nine CXXC motifs highly conserved from nematodes to humans.** *J Biol Chem* 1999, **274**:36456–64.

53. Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, Degenhardt J, Mayba O, Gnad F, Liu J, Pau G, Reeder J, Cao Y, Mukhyala K, Selvaraj SK, Yu M, Zynda GJ, Brauer MJ, Wu TD, Gentleman RC, Manning G, Yauch RL, Bourgon R, Stokoe D, Modrusan Z, Neve RM, de Sauvage FJ, Settleman J, Seshagiri S, Zhang Z: **A comprehensive transcriptional portrait of**

human cancer cell lines. *Nat Biotechnol* 2014, **33**:306–12.

54. Tsang SW, Shao D, Cheah KSE, Okuse K, Leung PS, Yao K-M: **Increased basal insulin secretion in Pdzd2-deficient mice.** *Mol Cell Endocrinol* 2010, **315**:263–70.

55. Suen PM, Zou C, Zhang YA, Lau TK, Chan J, Yao KM, Leung PS: **PDZ-domain containing-2 (PDZD2) is a novel factor that affects the growth and differentiation of human fetal pancreatic progenitor cells.** *Int J Biochem Cell Biol* 2008, **40**:789–803.

56. Ma RYM, Tam TSM, Suen APM, Yeung PML, Tsang SW, Chung SK, Thomas MK, Leung PS, Yao K-MM: **Secreted PDZD2 exerts concentration-dependent effects on the proliferation of INS-1E cells**. *Int J Biochem Cell Biol* 2006, **38**:1015–1022.

57. Vassilopoulos S, Esk C, Hoshino S, Funke BH, Chen C-Y, Plocik AM, Wright WE, Kucherlapati R, Brodsky FM: **A role for the CHC22 clathrin heavy-chain isoform in human glucose metabolism.** *Science* 2009, **324**:1192–6.

58. Tsai J-W, Lian W-N, Kemal S, Kriegstein AR, Vallee RB: **Kinesin 3 and cytoplasmic dynein mediate interkinetic nuclear migration in neural stem cells.** *Nat Neurosci* 2010, **13**:1463–1471.

59. Zavros Y: **The adventures of sonic hedgehog in development and repair. IV. Sonic hedgehog processing, secretion, and function in the stomach.** *Am J Physiol Gastrointest Liver Physiol* 2008, **294**:G1105–G1108.

60. Maitra A, Hruban RH: **A new mouse model of pancreatic cancer: PTEN gets its Akt together**. *Cancer Cell* 2005, **8**:171–172.

61. Martínez C, Urbano-Ispízua A: **Graft-versus-host disease therapy: something else beyond glucocorticoids?** *Haematologica* 2011, **96**:1249–51.

62. Singh SK, Singh RD, Sharma A: **Caudal regression syndrome--case report and review of literature.** *Pediatr Surg Int* 2005, **21**:578–81.

63. Kaciński M, Jaworek M, Skowronek-Bała B: **Caudal regression syndrome associated with the white matter lesions and chromosome 18p11.2 deletion.** *Brain Dev* 2007, **29**:164–6.

64. Chan BWH, Chan K-S, Koide T, Yeung S-M, Leung MBW, Copp AJ, Loeken MR, Shiroishi T, Shum ASW: **Maternal diabetes increases the risk of caudal regression caused by retinoic acid.** *Diabetes* 2002, **51**:2811–6.

65. Campbell DD, Porsch RM, Cherny SS, Capra V, Merello E, De Marco P, Sham PC, Garcia-Barceló M-M: **Cost effective assay choice for rare disease study designs**. *Orphanet J Rare Dis* 2015, **10**:10.

66. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE: **Rare-disease genetics in the era of next-generation sequencing: discovery to translation**. *Nat Rev Genet* 2013, **14**:681–691.

67. Adra A, Cordero D, Mejides A, Yasin S, Salman F, O'Sullivan MJ: **Caudal regression syndrome: etiopathogenesis, prenatal diagnosis, and perinatal management.** *Obstet Gynecol Surv* 1994, **49**:508–16.

68. Vlangos C, O'Connor B, Morley M: **Caudal regression in adrenocortical dysplasia (acd) mice is caused by telomere dysfunction with subsequent p53-dependent apoptosis**. *Dev …* 2009, **334**:418–428.

69. Padmanabhan R: **Retinoic acid-induced caudal regression syndrome in the mouse fetus**. *Reprod Toxicol* 1998, **12**:139–151.

70. Knight B: **Caudal Regression Syndrome: A Case Report.** *AANA J* 2011, **79**:281–282.

71. Evans J a, Vitez M, Czeizel a: **Congenital abnormalities associated with limb deficiency defects: a population study based on cases from the Hungarian Congenital Malformation Registry (1975-1984).** *Am J Med Genet* 1994, **49**:52–66.

72. Campbell DD, Porsch RM, Cherny SS, Capra V, Merello E, De Marco P, Sham PC, Garcia-Barceló M-M: **Cost effective assay choice for rare disease study designs**. *Orphanet J Rare Dis* 2015, **10**:10.

73. Hogan MC, Griffin MD, Rossetti S, Torres VE, Ward CJ, Harris PC: **PKHDL1 , a homolog of the autosomal recessive polycystic kidney disease gene , encodes a receptor with inducible T lymphocyte expression**. *Hum Mol Genet* 2003, **12**:685–698.

74. Yim J-H, Kim Y-J, Ko J-H, Cho Y-E, Kim S-M, Kim J-Y, Lee S, Park J-H: **The putative tumor suppressor gene GLTSCR2 induces PTEN-modulated cell death.** *Cell Death Differ* 2007, **14**:1872–9.

75. Rivire JB, Ramalingam S, Lavastre V, Shekarabi M, Holbert S, Lafontaine J, Srour M, Merner N, Rochefort D, Hince P, Gaudet R, Mes-Masson AM, Baets J, Houlden H, Brais B, Nicholson G a., Van Esch H, Nafissi S, De Jonghe P, Reilly MM, Timmerman V, Dion P a., Rouleau G a.: **KIF1A, an axonal transporter of synaptic vesicles, is mutated in hereditary sensory and autonomic neuropathy type 2**. *Am J Hum Genet* 2011, **89**:219–301.

76. Nahorski MS, Al-Gazali L, Hertecant J, Owen DJ, Borner GHH, Chen Y-C, Benn CL, Carvalho OP, Shaikh SS, Phelan A, Robinson MS, Royle SJ, Woods CG: **A novel disorder reveals clathrin heavy chain-22 is essential for human pain and touch development**. *Brain* 2015, **138**:2147–2160.

77. Landstrom AP, Weisleder N, Batalden KB, Bos JM, Tester DJ, Ommen SR, Wehrens XHT, Claycomb WC, Ko J-K, Hwang M, Pan Z, Ma J, Ackerman MJ: **Mutations in JPH2-encoded**

junctophilin-2 associated with hypertrophic cardiomyopathy in humans. *J Mol Cell Cardiol* 2007, **42**:1026–35.

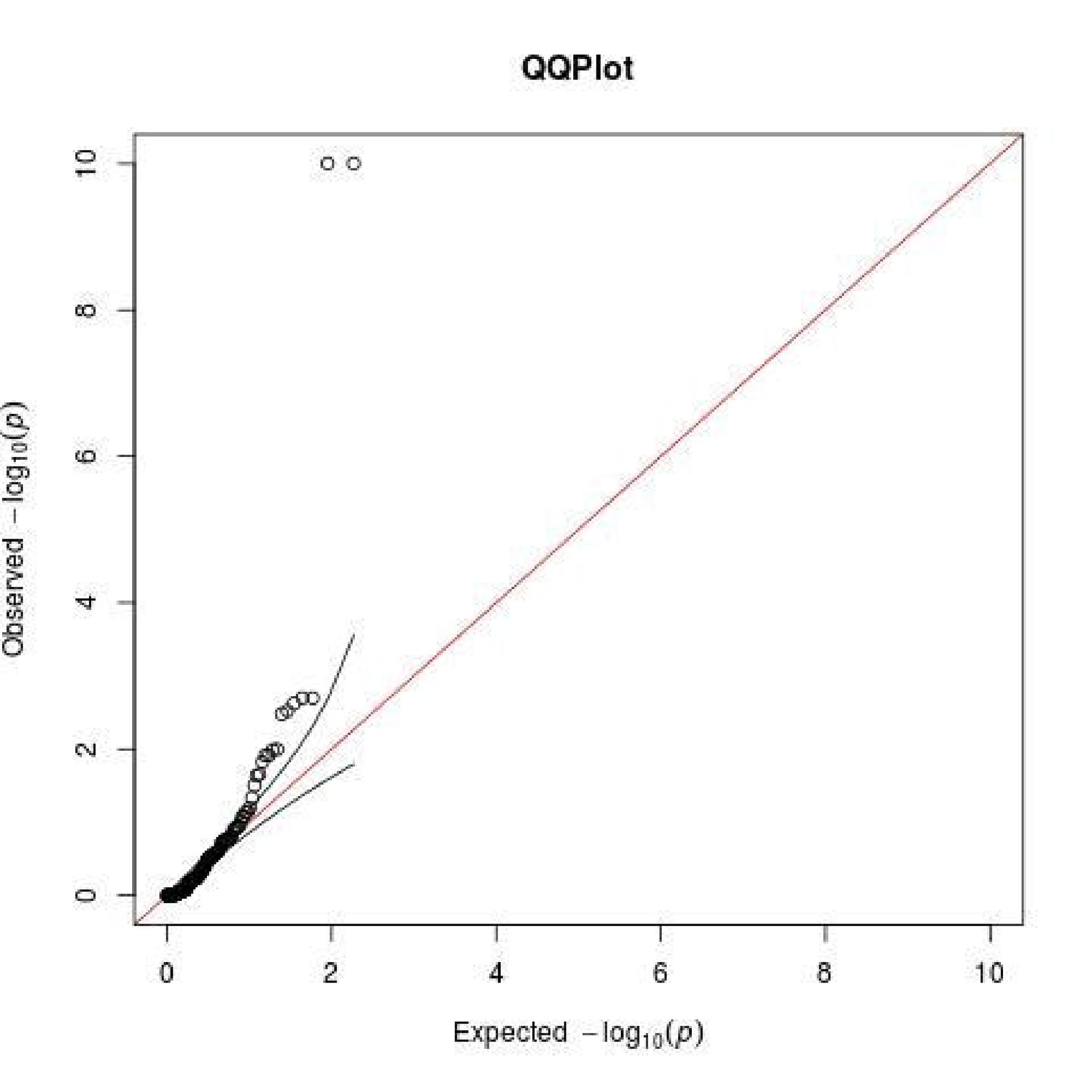
78. Maiti AK, Mattéi MG, Jorissen M, Volz A, Zeigler A, Bouvagnet P: **Identification, tissue specific expression, and chromosomal localisation of several human dynein heavy chain genes.** *Eur J Hum Genet* 2000, **8**:923–32.

79. Huizing M, Didier A, Walenta J, Anikster Y, Gahl W a., Krämer H: **Molecular cloning and characterization of human VPS18, VPS 11, VPS16, and VPS33**. *Gene* 2001, **264**:241–247.

80. Janz R, Südhof TC, Hammer RE, Unni V, Siegelbaum S a, Bolshakov VY: **Essential roles in synaptic plasticity for synaptogyrin I and synaptophysin I.** *Neuron* 1999, **24**:687–700.

Figure title and legend

Figure 1.
Title: QQ-Plot of adjusted Burden test.


QQPlot
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
2
4
6
8
10