Heliyon 11 (2025) e42079

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Insufficient gene expression and lost gene regulatory network may underlie the mechanism of Hirschsprung Disease in 5p–syndrome

Yizhao Luan [a,1], Peng Li [b,1], Yuanyuan Luo [c], Hong Zhang [c], Xiaochun Zhu [b], Yan Zhang [a], Aihua Yin [a,***], Qiang Wu [c,**], Chengwei Chai [b,*]

[a] *Prenatal Diagnosis Centre, Maternal and Children Metabolic-Genetic Key Laboratory, Guangzhou Key Laboratory of Prenatal Screening and Diagnosis, Guangdong Women and Children Hospital, Guangzhou, 511442, China*
[b] *Department of Pediatric General Surgery, Guangdong Women and Children Hospital, Guangzhou, 511442, China*
[c] *Department of Pediatric Surgery, Guangdong Provincial Key Laboratory of Research in Structural Birth Defect Disease, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, 510623, China*

A B S T R A C T

Cri-du-chat syndrome (CDC, OMIM 123450) is a rare chromosomal syndrome that results from partial deletions on the short arm of chromosome 5, known as 5p minus. Substantial clinical and genetic heterogeneity were observed in CDC patients. Large efforts have been dedicated to correlating the deleted regions on 5p arm with observed symptoms in CDC patients. However, the genetic basis of many specific phenotypes, including the co-occurrence of Hirschsprung Disease (HSCR), have yet been clarified. Here, we conducted a study on two patients with CDC and HSCR using whole genome sequencing (WGS) analyses. Our WGS data confirmed the deletion regions on 5p associated with CDC and indicated potential unknown genetic mechanisms underlying HSCR. On the one hand, leveraging human single-cell atlas for developing enteric nervous system, we demonstrated that some affected genes in these two patients overlapped with those showing expression changes along the development pseudotime of enteric nervous cells (ENC) and overlapped with known HSCR genes including *RET*, *NRG1*, *ERBB (ERBB2 and ERBB3)*, *ITGB (ITGB1)*. On the other hand, integrating gene regulatory relationship estimated from single cell chromatin accessibility omics of enteric neurons, we found that the 5p deletion regions contained key cis-regulatory regions for HSCR-related gene *GDNF*. Taken together, our study reveals the genetic basis of HSCR or intestinal phenotypes in 5p minus patients, highlighting the importance of studying gene regulatory relationships to explain phenotypic heterogeneity.

1. Introduction

Cri-du-chat (CDC, OMIM 123450) is a chromosomal syndrome that results from partial deletions on the short arm of chromosome 5, known as 5p deletion or 5p minus. According to a survey of the birth prevalence of congenital anomalies among 21,472 consecutive newborn babies in Tokyo, the incidence of CDC was approximately 1 in 14,000 [1]. To date, several relatively large clinical cohorts have been reported, with a total number of CDC patients exceeding 1000 [2–13]. Substantial clinical and genetic heterogeneity were observed in these patients, posing challenges in elucidating phenotype-genotype correlations, understanding pathological mechanisms, and developing therapeutic drugs.

A wide spectrum of clinical phenotypes was observed in 5p minus individuals. The classical features have been characterized by a high-pitched cry that sounds like that of a cat, intellectual disability and delayed development including small head size, low birth

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: yinaihua@gdwch.com.cn (A. Yin), wuqiangll@hotmail.com (Q. Wu), 2018760140@gzhmu.edu.cn (C. Chai).
[1] Yizhao Luan and Peng Li contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2025.e42079
Received 24 August 2024; Received in revised form 14 December 2024; Accepted 16 January 2025
Available online 23 January 2025

weight, and hypotonia. Indeed, the impact of this condition on patients' bodies and lives is multidimensional and enduring beyond development delay, including dysmorphology, multiorgan dysfunction, and social behavior disorders. For dysmorphology, the most frequently observed are distinctive facial features, including widely spaced eyes, low-set ears, a small jaw, and a rounded face. Multiorgan dysfunction can be found including conductive hearing loss and abnormalities of brain [14], vision, renal, genital, anal, limb, and gastrointestinal. As of social behavior disorders, the patients can present hyperactivity, fluctuating emotion, and aggressive behaviors; A significant proportion of patients will exhibit delayed speech development, or can communicate only in simple ways, such as short understandable sentences, and less than 10 words. Notably, some 5p minus individuals don't exhibit any obvious symptoms and appear to be normal healthy.

The development of genomic diagnostic techniques has facilitated research on both the chromosomal and molecular levels of this disease. Ninety percent of cases are *de novo*, and 10 % are inherited, due to a rearrangement in the parents [4]. 5p deletions can be interstitial or terminal and range from 560 kb to 40 Mb in size [15]. Hence, the variability seen among individuals may be attributed to the differences in their genotypes. Although many individuals have similar breakpoints, there is no common recurring breakpoint. Studies differ on the correlation between deletion size and severity of clinical features. In addition, there is also phenotypic heterogeneity among family members with the same deletion. Recent studies suggest that epigenetic mechanisms may also be involved in the pathogenesis of CDC [7].

The challenge in interpreting the genetic heterogeneity of the disease lies in the fact that the existing disease-associated genes cannot explain the phenotypes of the patients. For example, Hirschsprung Disease (HSCR) has been identified to be combined with 5p deletion [16,17]. HSCR is characterized by the absence of enteric neurons caused by the defects of growth, migration and/or differentiation of enteric neural crest cells (ENCC), leading to intestinal obstruction [18]. It is known that the 5p arm contains one of HSCR-related genes, *GDNF*, located at 5p13.2 [19]. However, in the reported cases with HSCR and 5p deletion, the breakpoints were in 5p15.2 and 5p14.3–15.33. These studies suggested that there may be other genetic causes for HSCR in the 5p deletion region.

In human tissues, genes expressed following cell type-specific programs are regulated by *cis*-regulatory elements (CREs) such as enhancers and promoters. Many genetic variants associated with human disease and complex traits may affect these CREs [20]. Therefore, we reasonably infer that the 5p deletion leads to novel HSCR-related mutations or the loss of regulatory relationships of known disease-related genes, resulting in the corresponding phenotype and occurrence of the disease. In this study, we conducted a targeted study on two patients with CDC and HSCR using whole-genome sequencing (WGS) analyses and identified potential genetic and structural variations. By integrating single-cell transcriptomes, we analyzed the molecular and cellular functions associated with these genetic variations, including impact on cell lineage trajectory. Finally, we correlated these variations with gene regulatory relationships specific to enteric neuronal cells.

2. Materials and methods

2.1. Patient recruitment

Two probands were recruited. For Patient 1, we collected peripheral blood samples from the patient and her parents. For Patient 2, we collected the Paraffin section of the aganglionic segment of the colon. This research received approval from the Ethics Committee of Guangzhou Women and Children's Medical Center, Guangzhou Medical University (2021-065B00). Clinical data and biological specimens were obtained from the probands and families with written informed consent.

2.2. Sample preparation, library construction and sequencing

Blood DNA was extracted from QIAamp DNA Blood Maxi Kit (Qiagen, Germany), and intestinal DNA was extracted using QIAamp DNA Mini Kit (Qiagen, Germany) following the standard manufacturer's instructions. All samples were quantified using the Qubit dsDNA BR Assay Kit (Thermo Fisher, USA), and DNA integrity was evaluated with 1 % agarose gel electrophoresis.

Paired-end libraries with approximately 350bp insert sizes were constructed from 1 μg of genomic DNA from each accession using VAHTS Universal DNA Library Prep Kit for Illumina V3 (Vazyme, China) according to the manufacturer's specifications. Briefly, gDNA was fragmented using a biorupter ultrasonic processor (Diagenode, Belgium) to a size of 250–300 bp, then after purified, the fragmented DNA was end repaired, "A"-tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. Agilent Bioanalyzer 2100 system (Agilent, USA) was used to check the insert size of the libraries. The concentrations of the constructed libraries were initially measured and diluted to 1 ng/μl by Qubit®4.0 (Thermo Fisher, USA). Finally, these libraries were sequenced on the Illumina Novaseq6000 platform (Illumina, USA) by Genergy Biotechnology Co. Ltd. (Shanghai, China).

2.3. Whole-genome sequencing data processing

Raw data were processed using fastp (version 0.23.2) [21] to trim the adapter and low-quality bases, and data quality was checked with fastqc (v0.11.9) [22]. Clean reads were aligned to the Human genome hg19 using Sentieon (Version.202112.01) bwa mem, then sorted by sentieon sort. Duplicated reads were excluded using Sentieon Dedup. The single nucleotide variants (SNVs) and indels, were called by Sentieon DNAscope [23]. All genomic variants were annotated using ANNOVAR [24]. Variants with a frequency below 0.01 in 1000 Genomes Project (1000g2015aug_all) were retained.

Copy number variations was detected by Control-freeC (v11.6) [25], and Structure variation was identified by lumpy (v0.2.13) [26]. CNVs and SVs were firstly filtered by subtracting the overlaps with the unaffected parents and with the entries from the Database

of Genomic Variations (DGV), which contains genomic alterations that involve segments of DNA that are larger than 50bp, identified in healthy control samples [27]. DGV dataset were downloaded from the "dgvMerged" database from ANNOVAR owner [24]. The comparison and manipulation of genomic features were performed using BEDtools functions [28]. The invocation of BEDtools functions was performed with the R package 'bedtoolsr' [29]. Online UCSC Liftover was used to convert genome coordinates between different versions of genome assembly [30].

2.4. ScRNA-seq data

Single cell transcriptomes were downloaded from the Gut cell atlas portal (gutcellatlas.org), which contained 428K intestinal single cells from fetal, pediatric, adult donors, and up to 11 intestinal regions [31]. The normalized h5ad data (Full_obj_log_counts_soupx_v2.h5ad) and cell type identity information from the original study were used. Differential expression analyses were reproduced between cell groups and conditions of interest using the *FindMarkers*(.) function from the Seurat package [32]. Cell type-specific markers were defined using differential expression analysis with the genes with a higher expression (avg_log2FC > 1, adjusted $P < 0.05$) and pct.1 $\geq$ 30 % in the cell subtypes of interest.

2.5. CRE catalog and CRE-target link

Cell type specific CRE catalog was downloaded from one published work established by integrating single-cell chromatin accessibility assays on 30 adult human tissue types and 15 fetal tissue types [33,34]. The original coordinate for adult CREs is based on the hg38 genome assembly, and for fetal CREs is based on hg19 assembly. Corresponding chromosome coordinates of cytoband is obtained from UCSC database(https://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/) [35]. Chromosome coordinate conversion between hg38 and hg19 was conducted using UCSC LiftOver. CRE-gene map for adult tissues was downloaded from Ref. [34], and that for fetal tissues was downloaded from Ref. [33]. GTF files for hg38 and hg19 were downloaded from UCSC database.

CRE-gene link data from adult cells was downloaded from the original study, and that from fetal cells was processed according to the following workflow. We downloaded the co-accessibility scores by cell type from GEO study: GSE149683 [33]. The peaks identified in 5p arm and in fetal enteric neurons and glia cells were retained. These peak pairs were annotated to promoters, genes and other intergenic regions using *annotatePeakInBatch*(.) command from the ChIPpeakAnno package [36]. The open regions of interest (ROIs) were grouped into 'upstream', 'downstream', 'overlapStart', 'overlapEnd', 'inside', and 'includeFeature' groups according to the relative positions of these ROIs and target genes.

3. Results

3.1. Clinical features

Two patients with CDC and HSCR were recruited in our study. Patient 1 was a four-year-old girl with medical history including CDC, congenital heart disease, and congenital tracheal stenosis. The patient exhibited the typical phenotype associated with CDC, including: (1) facial features characterized by ocular hypertelorism; (2) developmental delays, presenting as growth retardation, inability to perform daily activities independently, inability to speak, and inability to feed herself. Regarding the phenotypes associated with HSCR, the patient exhibited abdominal distention and difficulty in defecation immediately after birth. These symptoms had persisted to the present. At our hospital, the patient underwent a laparoscopically assisted transabdominal pull-through surgery, and pathological examination of the resected tissue confirmed the diagnosis of HSCR.

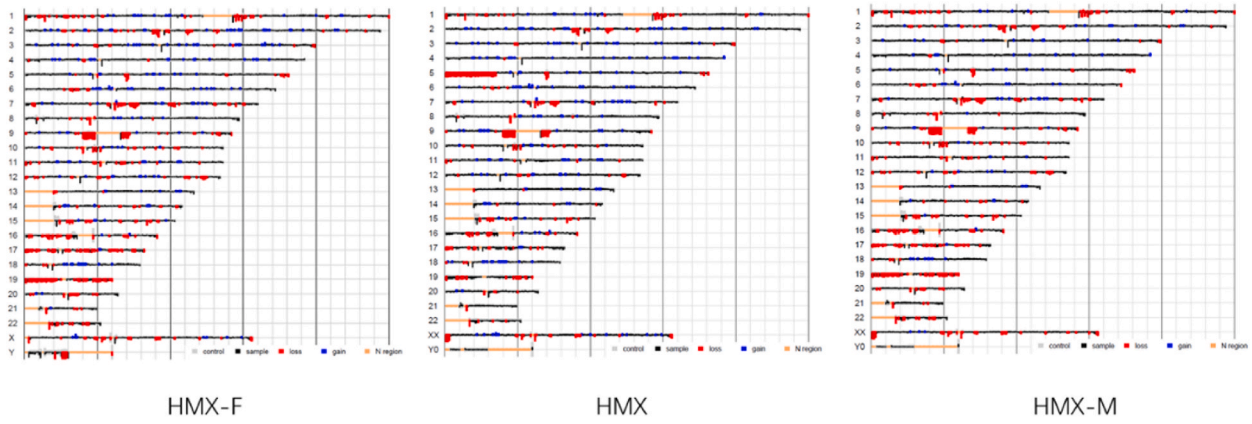

Fig. 1. Genome-wide copy number alterations in Patient 1 family.
Different CNV events with different sizes were identified in each member of patient 1 family. HMX is the proband, HMX-F and HMX-M indicate father and mother of the proband respectively. Genomic gain and loss regions are colored in red and blue, respectively.

Patient 2 was an 11-day-old boy who presented at birth with a high-pitched cry resembling a cat's meow. Chromosomal karyotype analysis indicated 46, XY, del 5p (13). He also exhibited distinct facial features associated with CDC, including small palpebral fissures, hypertelorism, micrognathia, a prominent philtrum, and poorly developed intranasal groove. After birth, the patient experienced drooling, abdominal distention, and an absence of bowel movements. Chest X-ray showed left lower lobe pneumonia, and abdominal X-ray indicated partial intestinal obstruction. Histopathological examination of tissues obtained by exploratory laparotomy revealed underdeveloped ganglion cells in the myenteric plexus of the distal ileum and transverse colon; the myenteric plexus in the upper rectum was wavy and lacked ganglion cells. Therefore, HSCR was diagnosed.

3.2. Identification of pathogenic mutations

To investigate the genetic basis of these diseases, we pursued WGS study. For Patient 1, patient-parent trio WGS on peripheral blood samples were conducted. For Patient 2, due to a long time elapsed since the initial medical record in our department (approximately four years), we performed WGS on the surgically excised tissue.

We began by separately analyzing the genomic variations of each patient. From Patient 1 and her parents' genomes, we identified a total of 881 rare SNVs and indels and 2092 unique structural variation including CNVs (Fig. 1). Genomic sites with point mutations and CNV events that exhibited the same genotype in both parents and proband were further excluded. Additionally, CNV segments present in healthy control samples were also excluded, as they were not associated with the clinical phenotypes of interest. We finally retained 18 nonsynonymous SNVs and 471 unique CNV events (DEL + DUP) (Supplementary file 1). In patient 2, we identified 27,386 SNVs and indels, and 63,278 CNV events, out of which 8490 SNVs and indels, 692 CNVs and 25,548 SVs located in or overlapped with the exonic

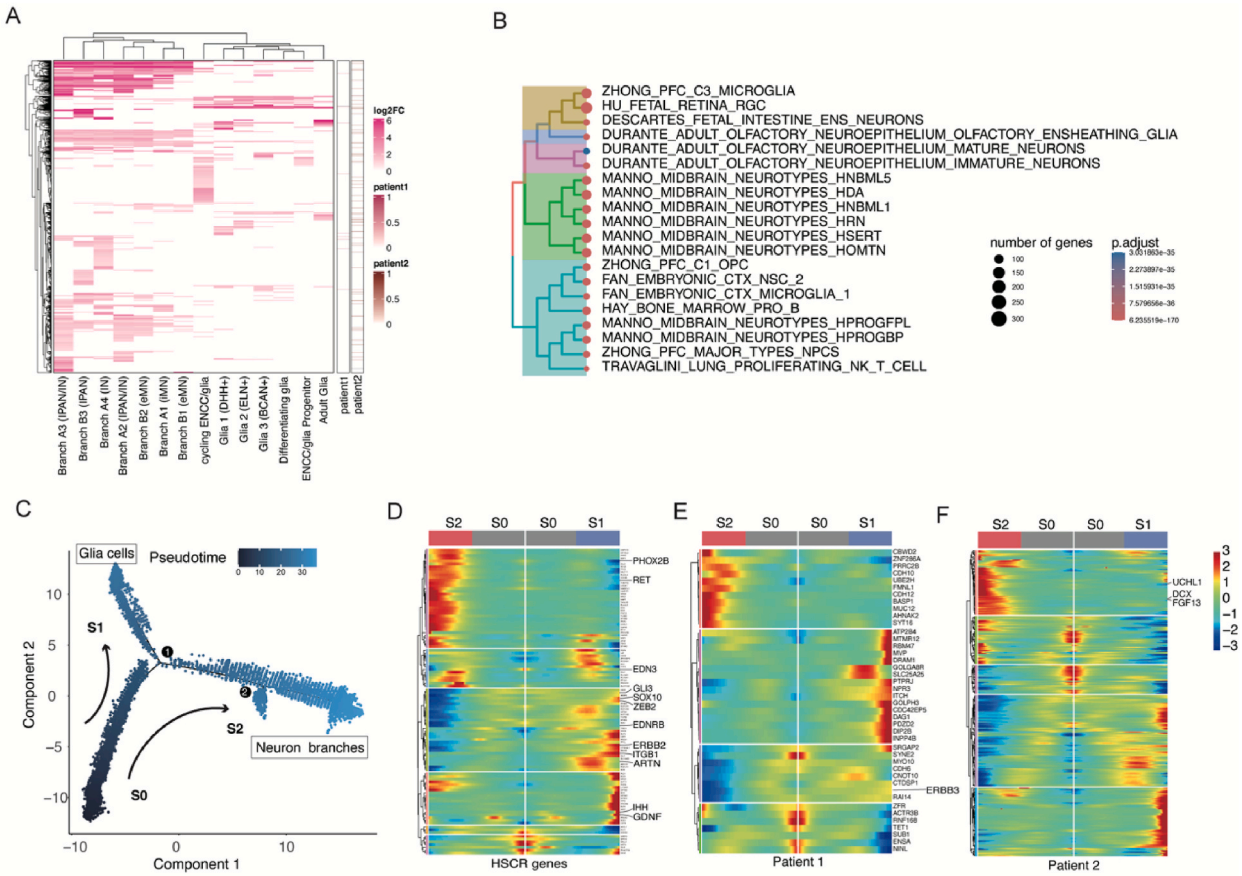

Fig. 2. Cell type-specific and cell trajectory-dependent expression of GOIs. GOIs were defined with the genes containing non-synonymous, stop-gain, stop-loss, and frameshift SNVs, as well as CNVs that overlap exonic, intronic, and splicing sites identified throughout the whole genome.
A, Heatmap showing the cell-type specific expression pattern of GOIs in ENC subtypes including enteric neuron and glial subtypes. Log2 fold change values were used and defined by comparing ENC subtypes to all the other cell types in human gut. Barplots on the right side indicated GOIs in Patient 1 and Patient 2. In the heatmap panel, the redder the color, the greater the magnitude of change.
B, Tree plot of top significant functional terms enriched by highly expressed genes in enteric neuron and glial cells.
C, Single-cell trajectory along the pseudotime predicted by Monocle2.
D-F, Heatmaps showing gene expression changes along the pseudotime during ENC development. HSCR-genes (D), GOIs from Patent 1 (E) and Patient 2 (F) were shown, respectively.

regions of protein-coding genes. Due to the paraffin embedding and prolonged storage of this sample, we applied a stringent cutoff for the sequencing depth of variations in this sample (DP and SU > 20). 325 SNVs and 956 CNV events were identified finally (Supplementary file 2).

Because these two patients shared clinical symptoms of CDC and HSCR, we reasonably speculated that they have shared genomic alterations. We directly compared the SNVs and CNVs between these two patients. We found 3 common novel SNVs from three genes (*OR9G1*:NM_001005213:exon1:c.G82T:p.V28L, *OR9G1*:NM_001005213:exon1:c.G158A:p.C53Y, *OR9G9*:NM_001013358:exon1:c.G82T:p.V28L, *OR9G9*:NM_001005213:exon1:c.G158A:p.C53Y, and *AHNAK2*:NM_001350929:exon7:c.A3144C:p.E1048D), and common genomic deletion regions with a combined length of 3,957,110 nt. Notably, 3,793,729 nt deletion was found on the 5p arm, ranging from 5p15.1 to 5p13.2. A total of 11 protein-coding genes (*MYO10, CDH12, CDH10, CDH9, CDH6, PDZD2, MTMR12, NPR3, TARS1, ADAMTS12*, and *RAI14*) and 6 ncRNAs (*LINC02239*, LINC02228, LINC02211, LINC02064, *MIR4279*, and LINC02120) were found in this region. No genes implicated by these SNVs and CNVs overlapped with known HSCR-genes.

Subsequently, we investigated whether known genetic variants associated with CDC and HSCR were present in the individual genomes of both patients. We collected known genetic variants associated with CDC and HSCR after extensive literature search. Nearly all the CDC-related chromosomal deletions are in 5p region; The development of HSCR may be associated with variants in more than 120 genes (HSCR-genes; Table S1). Chromosomal deletions in different sub-regions on the 5p arm were identified in both patients, which could explain CDC occurrence: Patient 1 exhibited deletions scattered in the region of 5p15.3–13.2, while Patient 2 exhibited deletions scattered in the region of 5p15.3–14.2. Regarding the genetic basis of HSCR, we identified one CNV event in patent 1 resulting in one novel heterozygous deletion of *ERBB3* gene (chr12:56470600–56483912), of which the homozygous and compound heterozygous mutants have been considered pathogenic in HSCR [37]. In patient 2, we found two rare heterozygous nonsynonymous SNVs in HSCR-gene *UCHL1* with a high read depth (DP > 20)(chr4:41259692) and heterozygous CNV loss events in two HSCR-genes *DCX* and *FGF13*. Among these genes, *UCHL1* is associated with enteric neuron homeostasis [38,39], *DCX* may function in cell migration and axon outgrowth of enteric nervous system (ENS), and its low expression was enriched in HSCR patients [40,41], and *FGF13* is thought to be a marker of the developing ENS [40,42].

Taken together, our WGS data validated the critical chromosomal regions associated with CDC. While no common genomic variants associated with HSCR were found in these two patients, the WGS data revealed genomic mutations in known HSCR-related genes, which could lead to insufficient expression. However, the pathogenic significance of these mutations cannot be fully determined at this stage. Considering other reported cases of patients with both CDC and HSCR, it remains possible that additional genetic regulatory mechanisms contribute to the occurrence of HSCR in CDC patients.

3.3. Mapping genomic variants to cell lineage development associated with HSCR

Studying genes related to ENCC development and normal cellular function holds significant implications for unraveling HSCR. Utilizing genes with non-synonymous, stop-gain, stop-loss, and frameshift SNVs, as well as CNVs that overlap exonic, intronic, and splicing sites identified throughout the whole genome (Genes of interest, GOI), we analyzed the potential impact of these GOIs on the development of enteric neuronal cells (ENCs) in the normal human gut. In total, 100 GOIs in Patient 1 and 806 GOIs in Patient 2 were included (Supplementary file 3).

We then compiled a gene list specifically expressed in ENCs using a scRNA-seq dataset derived from fetal and adult intestinal tissues, which were supposed to be important for the development and cellular homeostasis of enteric nervous system (see Methods). A total of 2587 genes exhibited high expression levels in ENC subtypes by comparing to all the other cell types in the gut (log2FC > 1, adjusted $P < 0.05$), and were significantly overrepresented in cell type signature gene sets related to nervous systems, as expected (Fig. 2A and B). These findings demonstrated that ENCs exhibit specific regulation of gene expression during intestinal development. Among these genes, 57 are known to be associated with HSCR (Supplementary file 4). Notably, we found that 9 GOIs in Patient 1 and 68 GOIs in Patient 2 overlapped with these genes (Fig. 2A). Genomic mutations and deleted CNVs in these GOIs may influence cellular fate determination of enteric neurons and glial cells, that was involved in the HSCR pathogenesis.

To explore cell lineage-specific gene expression dynamics during ENS development, we analyzed single cell trajectory branch-dependent gene expression changes along the pseudotime predicted by Monocle2 using the same single-cell RNA-seq dataset (See Methods; Fig. 2C). Overall, individual cells were categorized based on age groups: fetal, pediatric, and adult cells (Fig. S1). Progenitor cells with high differentiation potential are predominantly found at earlier pseudotime stages; Most mature glial cells are located along one branch, whereas lineages formed during fetal development are situated along a different branch (Fig. 2C; Fig. S2). We first clustered known HSCR-genes according to their expression changes along the branches of the pseudotime. Some well-known causal genes were classified into sub-groups with different expression patterns, where *RET* and *PHOX2B* are highly expressed in neuron-enriched cell branches and *SOX10*, *ERBB2*, *ITGB1*, *IHH* and *GDNF* are highly expressed in glia-enriched cell branch (Fig. 2D). This result suggests that HSCR genes may contribute to HSCR phenotype by influencing spatiotemporally distinct cell types during development. The GOIs exhibiting significant expression changes (adjusted $P < 0.05$) along the pseudotime branches in both Patient 1 and Patient 2 were also clustered into sub-groups. These sub-groups included major clusters that corresponded to over-expression in distinct stages: the pre-branch stage, the glia-enriched branch, and the neuron-enriched branch (Fig. 2E and F), which was consistent with the expression profiles observed in known HSCR genes. Notably, we found some GOIs were identified to be cell markers and branch-dependent genes at the same time, including *ERBB3* in Patient 1 and *UCHL1*, *DCX* and *FGF13* in patient 2. These GOIs were supposed to be highly expressed following cell type specific mode during ENS development. However, in the genomes of our patients, genomic variants were found in these genes, which may lead to insufficient expression levels of these genes. Taken together, these results suggested that insufficient expression of a series of genes represent a predisposing risk for HSCR disease.

3.4. Mapping genomic regulatory elements related to HSCR genes

It is known that the 5p arm contains one of HSCR-related genes, *GDNF*, located at 5p13.2 [19]. CREs are widely distributed in the human genome, which can control mRNA expression of distal genes [43]. To test whether 5p deletion resulted in loss of gene transcriptional regulation relationship of *GDNF* or other HSCR-related genes, we analyzed all the regulatory relationships between 5p CREs and their targets. Integrative analysis of single cell chromatin accessibility assays on 222 fetal and adult cell types collectively identified approximately 16,459 CREs within the 5p region, out of which 14,802 (89.93 %) can target genes by distal regulation, while only 1657 control target genes as promoters or by proximal regulation (See Methods). More than half of these CREs were uniquely identified in fetal tissues (3656; 22 %) or adult tissues (5950; 36 %) (Fig. S3), suggesting that gene expression regulation associated with the 5p region can be dynamic across development stages. In average, one CREs were shared by two cell types (median); only approximately 9 % of these CREs were found in more than 20 cell types, suggesting that the regulatory effects exhibited by the 5p regions were highly cell type-specific (Fig. S4).

In total, 3849 CREs in 5p regions were found in intestinal cells, out of which 1183 were found in enteric nerve system, including enteric neuron and glia cells from fetal and adult tissues (Fig. 3A). Linking CRE-target genes analyses revealed that there are numerous distal transcriptional regulation events occurring within 5p arm (Fig. 3B). In adult cells, the average distance between CREs and their target genes is 589,529 nt, with the maximum distance reaching up to 1,988,006 nt. In fetal cells, the average distance is 147,177 nt, with the maximum distance up to 484,321 nt. Interestingly, we found that many CREs and distal regulatory effects are distributed around the *GDNF* gene, which is associated with HSCR (Fig. 3B). Overall, our analysis indicated complex and cell type-specific gene regulatory networks in the 5p region, which may additionally contribute to the phenotypic heterogeneity observed in patients with 5p deletions.

4. Discussion

In this work, we carefully examined the point mutations and structural variants, mainly CNVs, in the genomes of two patients with 5p minus and HSCR disease. Specifically, we employed a core family trio WGS design for one of the patients. To identify phenotype-related genes, we compared the genomic variations carried by these probands with those reported in published studies. For genes not yet associated with the disease, we inferred their potential molecular and cellular functions by integrating publicly available multi-omics data from cell atlas-level resources. Single-cell RNA-seq data and single-cell ATAC-seq-derived data were combined.

By integrating cell lineage trajectory inferred using single-cell transcriptomes derived from multiple development stages of human intestinal tissues, we discovered numerous genes related to the development of enteric neurons; these genes showed similar gene expression dynamics with popularly known HSCR-genes including *RET*, *NRG1*, *ERBB (ERBB2 and ERBB3)*, *ITGB (ITGB1)*, yet have not yet been reported to be associated with HSCR (Supplementary file 5). The significance of these genes in enteric neuron development, homeostasis, and other cellular functions, as well as their specific regulatory pathways, will need to be validated through further functional experiments in future studies.

Until now, the detailed genetic causes of specific phenotype have not been definitively determined. Large efforts have been put into the genetic mapping of 5p deletion regions and correlating the deleted regions to symptoms in CDC patients, with the hope of finding dose-sensitive gene(s) that may cause one or several symptoms. Many issues remain unresolved including the relationship between the

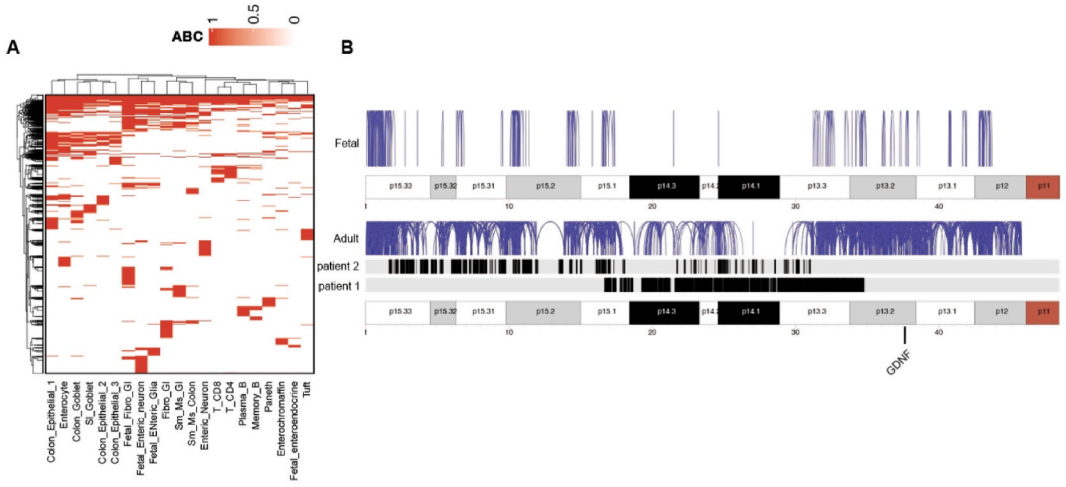

Fig. 3. CRE-target links between 5p regions.
A, CRE usage in intestinal cells.
B, 5p CRE-target links in enteric neurons derived from fetal and adult single cells. CNV deletion events of patient 1 and patient 2 were indicated in different tracks. The position of GDNF gene on the chromosome was indicated using the gene name and a segment.

size of the chromosomal deletion and the severity of the phenotype. In patients with simple 5p deletions or isolated 5p deletions, the severity of the phenotype may be correlated with the length of the chromosomal deletion [4,12]. Julián Nevado and partners have also suggested that severe phenotypes may be due to the deletion region containing functionally important genes. Our research indicates that the 5p deletion region contains many cis-regulatory elements, and the loss of these gene regulatory relationships may contribute to severe and diverse phenotypes. Previous studies have primarily relied on clinical phenotypes and genomic analyses of simple patient cohorts. Our research includes trio-based WGS, and during the analysis, we excluded background CNVs inherited from the parents and benign CNVs found in healthy individuals. This approach may be more effective in identifying genomic variants truly associated with phenotypes.

CRediT authorship contribution statement

Yizhao Luan: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Peng Li:** Writing – original draft, Investigation. **Yuanyuan Luo:** Investigation. **Hong Zhang:** Validation, Investigation. **Xiaochun Zhu:** Investigation. **Yan Zhang:** Methodology, Funding acquisition. **Aihua Yin:** Writing – review & editing, Supervision, Funding acquisition. **Qiang Wu:** Writing – review & editing, Funding acquisition. **Chengwei Chai:** Writing – review & editing, Supervision, Data curation, Conceptualization.

Data and code availability

Any information and request for the raw sequencing data that support the findings of this study and reagents should be directed to Chengwei Chai (2018760140@gzhmu.edu.cn). All the data set and code files used in this study are available at the Github repository at https://github.com/trumanLuan/5pminus.

Ethical approval

This research received approval from the Ethics Committee of Guangzhou Women and Children's Medical Center, Guangzhou Medical University (2021-065B00). Clinical data and biological specimens were obtained from the probands and families with written informed consent.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Chengwei Chai reports financial support was provided by Guangzhou Science and Technology Planning Project. Qiang Wu reports financial support was provided by Guangzhou Science and Technology Planning Project. Aihua Yin reports financial support was provided by Guangzhou Science and Technology Planning Project. Yan Zhang reports financial support was provided by Guangdong Basic and Applied Basic Research Foundation. Aihua Yin reports financial support was provided by National Key R&D Program of China. Aihua Yin reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the patients and the families for their participation and full cooperation during this study. This work was supported by Guangzhou Science and Technology Planning Project (grant no. 202201020594 to CWC, 202201020591 to QW, 202103000047 to AHY), Guangdong Basic and Applied Basic Research Foundation (2022A1515220097 to YZ), National Key R&D Program of China (grant no. 2023YFC2705605 to AHY) and National Nature Science Foundation of China (grant no. 82171856 to AHY).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2025.e42079.

References

[1] M. Higurashi, et al., Livebirth prevalence and follow-up of malformation syndromes in 27,472 newborns, Brain Dev. 12 (6) (1990) 770–773.
[2] R.C. Marinescu, et al., FISH analysis of terminal deletions in patients diagnosed with cri-du-chat syndrome, Clin. Genet. 56 (4) (1999) 282–288.
[3] E. Niebuhr, The Cri du Chat syndrome: epidemiology, cytogenetics, and clinical features, Hum. Genet. 44 (3) (1978) 227–275.
[4] J. Nevado, et al., Deep phenotyping and genetic Characterization of a cohort of 70 individuals with 5p minus syndrome, Front. Genet. 12 (2021) 645595.
[5] H.Y. Du, et al., Telomerase reverse transcriptase haploinsufficiency and telomere length in individuals with 5p- syndrome, Aging Cell 6 (5) (2007) 689–697.
[6] A. Zhang, et al., Deletion of the telomerase reverse transcriptase gene and haploinsufficiency of telomere maintenance in Cri du chat syndrome, Am. J. Hum. Genet. 72 (4) (2003) 940–948.

[7] P. Holland, et al., Cri du chat syndrome patients have DNA methylation changes in genes linked to symptoms of the disease, Clin. Epigenet. 14 (1) (2022) 128.
[8] P.C. Mainardi, et al., Clinical and molecular characterisation of 80 patients with 5p deletion: genotype-phenotype correlation, J. Med. Genet. 38 (3) (2001) 151–158.
[9] S.N. Chehimi, et al., Breakpoint delineation in 5p- patients leads to new insights about microcephaly and the typical high-pitched cry, Mol Genet Genomic Med 8 (2) (2020) e957.
[10] L.E. Wilkins, et al., Clinical heterogeneity in 80 home-reared children with cri du chat syndrome, J. Pediatr. 102 (4) (1983) 528–533.
[11] S.N. Chehimi, et al., Novel rearrangements between different chromosomes with direct impact on the diagnosis of 5p- syndrome, Clinics 77 (2022) 100045.
[12] P.C. Mainardi, et al., The natural history of Cri du Chat Syndrome. A report from the Italian Register, Eur. J. Med. Genet. 49 (5) (2006) 363–383.
[13] C. Bel-Fenellos, et al., Cognitive-behavioral profile in pediatric patients with syndrome 5p-; genotype-phenotype correlationships, Genes 14 (8) (2023).
[14] R. Villa, et al., Structural brain anomalies in Cri-du-Chat syndrome: MRI findings in 14 patients and possible genotype-phenotype correlations, Eur. J. Paediatr. Neurol. 28 (2020) 110–119.
[15] J.M. Nguyen, et al., 5p deletions: current knowledge and future directions, Am J Med Genet C Semin Med Genet 169 (3) (2015) 224–238.
[16] I. Ullah, L. Mahajan, D. Magnuson, A newly recognized association of Hirschsprung disease with cri-du-chat syndrome, Am. J. Gastroenterol. 112 (1) (2017) 185–186.
[17] M. Aldaffaa, et al., Hirschsprung's disease in a genetically diagnosed Cri-du-chat syndrome baby, J. Pediatr. Surg. Case Rep. 91 (2023) 102600.
[18] Z. Li, et al., Transcriptomics of Hirschsprung disease patient-derived enteric neural crest cells reveals a role for oxidative phosphorylation, Nat. Commun. 14 (1) (2023) 2157.
[19] G. Martucciello, et al., GDNF deficit in Hirschsprung's disease, J. Pediatr. Surg. 33 (1) (1998) 99–102.
[20] J. Nasser, et al., Genome-wide enhancer maps link risk variants to disease genes, Nature 593 (7858) (2021) 238–243.
[21] S. Chen, et al., fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890.
[22] FastQC, 2015.
[23] D. Freed, et al., The Sentieon Genomics Tools - a fast and accurate solution to variant calling from next-generation sequence data, bioRxiv (2017).
[24] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (16) (2010) e164.
[25] V. Boeva, et al., Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data, Bioinformatics 28 (3) (2012) 423–425.
[26] R.M. Layer, et al., LUMPY: a probabilistic framework for structural variant discovery, Genome Biol. 15 (6) (2014) R84.
[27] J.R. MacDonald, et al., The Database of Genomic Variants: a curated collection of structural variation in the human genome, Nucleic Acids Res. 42 (Database issue) (2014) D986–D992.
[28] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (6) (2010) 841–842.
[29] M.N. Patwardhan, et al., Bedtoolsr: an R package for genomic data analysis and manipulation, J. Open Source Softw. 4 (44) (2019).
[30] W.J. Kent, et al., The human genome browser at UCSC, Genome Res. 12 (6) (2002) 996–1006.
[31] R. Elmentaite, et al., Cells of the human intestinal tract mapped across space and time, Nature 597 (7875) (2021) 250–255.
[32] Y. Hao, et al., Dictionary learning for integrative, multimodal and scalable single-cell analysis, Nat. Biotechnol. 42 (2) (2024) 293–304.
[33] S. Domcke, et al., A human cell atlas of fetal chromatin accessibility, Science 370 (6518) (2020).
[34] K. Zhang, et al., A single-cell atlas of chromatin accessibility in the human genome, Cell 184 (24) (2021) 5985–6001, e19.
[35] L.R. Nassar, et al., The UCSC Genome Browser database: 2023 update, Nucleic Acids Res. 51 (D1) (2023) D1188–D1195.
[36] L.J. Zhu, et al., ChIPpeakAnno: a bioconductor package to annotate ChIP-seq and ChIP-chip data, BMC Bioinf. 11 (2010) 237.
[37] T.L. Le, et al., Dysregulation of the NRG1/ERBB pathway causes a developmental disorder with gastrointestinal dysmotility in humans, J. Clin. Invest. 131 (6) (2021).
[38] M.F. Viola, et al., Dedicated macrophages organize and maintain the enteric nervous system, Nature 618 (7966) (2023) 818–826.
[39] S. Kulkarni, et al., Age-associated changes in lineage composition of the enteric nervous system regulate gut health and disease, Elife 12 (2023).
[40] T.A. Heanue, V. Pachnis, Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes, Proc. Natl. Acad. Sci. U. S. A. 103 (18) (2006) 6919–6924.
[41] Z. Wen, et al., Circular RNA CCDC66 targets DCX to regulate cell proliferation and migration by sponging miR-488-3p in Hirschsprung's disease, J. Cell. Physiol. 234 (7) (2019) 10576–10587.
[42] S. Roy-Carson, et al., Defining the transcriptomic landscape of the developing enteric nervous system and its cellular environment, BMC Genom. 18 (1) (2017) 290.
[43] I. Jung, et al., A compendium of promoter-centered long-range chromatin interactions in the human genome, Nat. Genet. 51 (10) (2019) 1442–1449.