

Hedgehog-induced-oxidative-phosphorylation rescues neuronal differentiation defect of human enteric neural crest cells underlying Hirschsprung disease

Zhixin Li

University of Hong Kong <https://orcid.org/0000-0001-9962-3942>

Kathy Nga-Chu Lui

The University of Hong Kong

Sin-Ting Lau

University of Hong Kong

Frank Pui-Ling Lai

University of Hong Kong

Peng Li

The Seventh Affiliated Hospital of Sun Yat-sen University <https://orcid.org/0000-0001-7226-4214>

Patrick CHUNG

The University of HK

Kenneth Wong

The University of Hong Kong <https://orcid.org/0000-0001-7371-503X>

Paul TAM

The University of Hong Kong

Maria-Mercè Gracia-Barcelo

University of Hong Kong

Chi-chung Hui

The Hospital for Sick Children

Pak Sham

University of Hong Kong

Elly Ngan (✉ engan@hku.hk)

University of Hong Kong <https://orcid.org/0000-0002-0022-7999>

Article

Keywords:

Posted Date: April 8th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1433241/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Hirschsprung disease (HSCR) is characterized by the absence of enteric neurons caused by the defects of enteric neural crest cells (ENCCs), leading to intestinal obstruction. Here, using iPSC-based models of HSCR and single-cell transcriptomic analysis, we identified a core gene set of 118 genes commonly dysregulated in all HSCR-ENCCs, with *HDAC1* found to be a master regulator of these genes. Furthermore, upregulation of RNA splicing mediators and enhanced alternative splicing events were associated with severe form of HSCR. In particular, the higher inclusion rate of exon 9 in *PTBP1* and the perturbed expression of a PTBP1-target, *PKM*, were significantly enriched in these HSCR-ENCCs, and associated with the defective oxidative-phosphorylation (OXPHOS) and impaired neurogenesis. Hedgehog-induced OXPHOS significantly enhanced the survival and differentiation capacity of HSCR-ENCCs. In sum, we have defined the core factors underpinning HSCR disease and demonstrated the implications of OXPHOS in ENCC development and HSCR pathogenesis.

Introduction

Hirschsprung disease (HSCR), the most common congenital enteric neuropathy, is characterized by the absence of nerve cells along a variable length of the colon (short segment, S-HSCR; long segment, L-HSCR; total colonic aganglionosis, TCA). It is caused by the problems of enteric neural crest cells (ENCCs) in growth, migration and/or differentiation, leading to intestinal obstruction.

Glial cell-derived neurotrophic factor and RET tyrosine receptor kinase constitute the key signaling pathways for the development of the enteric nervous system (ENS)¹. *RET* variants and mutations that are not necessarily deleterious have been found to be associated with most HSCR cases. Specifically, the risk allele “T” in *RET* rs2435357 is most frequently associated with S-HSCR². The *BACE2/APP/BACE1*³, *NRG1/ERBB*^{4,5} and *Semaphorin C/3D*⁶ signaling pathways have also been identified through genome-wide genetic screens in various subgroups of HSCR patients. Nevertheless, the underlying causes of the phenotypic heterogeneity are still not well-understood.

It is generally believed that the majority of HSCR cases are caused by cumulative actions of multiple genetic variants, each with minor effects. While each individual may carry different constellations of genetic variants, they likely converge on shared pathogenic pathways, leading to HSCR disease. On the other hand, additional modifiers likely determine disease severity and account for the disease spectrum. Since ENCC samples from patients cannot be collected at or before disease onset, most transcriptome analyses were performed on resected diseased bowels. Thus, these transcriptomic data mainly reflect end-stage disease status, and molecular events occurring prior to disease onset remain unclear. To date, various directed differentiation protocols have been established for the derivation of neural crest (NC) cells from human induced pluripotent stem cells (iPSC) and to model ENS development and HSCR disease states^{3,7,8,9,10}. More importantly, subsequent scRNA-seq analysis further revealed that these iPSC-derived NC cells and their neuronal derivatives highly resemble the developing ENS in E13.5 mouse embryonic guts⁸. It opens new opportunities for understanding this complex congenital disease. The use

of a single-cell approach exploits cell heterogeneity and provides insight into the involved cellular changes of gene expression across the disease severity axis, allowing exploration of the molecular pathways, genes and gene regulatory networks involved in specific disease states and cellular phenotypes.

Increasing evidence suggest that metabolic rewiring plays an active role in the regulation of migration and development of neural crest cells (NCCs)^{11,12}. Hedgehog-mediated metabolic reset has been observed in cancers, muscle, adipose tissue and during stem cell renewal and cell fate transition. It is achieved mainly through the activation of GLI-independent non-canonical Hedgehog signaling, which can rapidly activate the machineries of primary energy metabolism through activation of AMP kinase (AMPK)¹³. The non-canonical pathway driving the metabolic rewiring is SMO- and cilium-dependent, and that can be uncoupled with the canonical Hedgehog signaling cascades at the level of SMO by selective partial agonism with a low dosage of the Hedgehog antagonist (such as cyclopamine) that allows SMO translocation to the cilium. Hedgehog signaling has been implicated in various stages of ENS development¹⁴⁻¹⁹. Nevertheless, till now, how the non-canonical arm of Hedgehog pathway influences ENS development remains largely unknown.

In the present study, we established seven iPSC lines from HSCR patients demonstrating the full disease spectrum and recapitulating HSCR-associated phenotypes *in vitro*. We then performed a high-resolution transcriptomic analysis of control- and HSCR-ENCCs to identify major changes in molecular pathways and networks associated with disease severity. Defective oxidative phosphorylation (OXPHOS) was found to be highly associated with neuronal differentiation defect of HSCR-ENCCs. Activation of fatty acid oxidation (FAO) pathway and Hedgehog-induced metabolic rewiring restored the differentiation capacity of HSCR-ENCCs, highlighting the relevance of OXPHOS in HSCR pathogenesis.

Results

Establishment and characterization of iPSC-based model of HSCR

Seven patients with HSCR from different subgroups, including 4 S-HSCR, 2-L-HSCR and 1 TCA, were recruited for this study. Each patient was given a disease score according to the extent of the aganglionosis (0–5; control=0 and TCA=5). The *RET* alleles in rs2435357 and variants in the coding region of *RET* gene were identified through sequencing the genomic DNA from the patients' blood. The clinical and genetic features of each patient are summarized in Fig. 1a & b. Briefly, among the four S-HSCR patients, HSCR#5, #10, and #20 showed relatively mild phenotypes with disease scores of 1 or 2. In contrast, HSCR#1 demonstrated a longer aganglionic segment up to the lower sigmoid, representing the intermediate phenotype between S- and L-HSCR.

Skin fibroblasts from these HSCR patients were used to generate iPSC lines and the corresponding genotype of *RET* alleles in each hiPSC line was confirmed by Sanger sequencing. Two control iPSC lines which carry the non-risk allele "C/C" in rs2435357 were also included in this study. NCCs carrying the

exact genetic makeup of the patients were derived from iPSC lines using an *in vitro* differentiation protocol⁷ (Fig. 1c). The migration and neuronal differentiation capabilities of the control- and HSCR-NCCs were assessed using scratch (Fig. 1d) and *in vitro* differentiation (Fig. 1f) assays, respectively, as previously described^{7,8,20}. The control- and HSCR-iPSC lines generated comparable yields of NCCs (Supplementary Fig. 1a and 1b). However, the HSCR-NCCs exhibited lower expression levels of *RET* (Supplementary Fig. 1c) and were less competent to migrate, and to generate neurons than those of the control NCs (Fig. 1d-g). Interestingly, among these S-HSCR-NCCs, HSCR#1-NCCs exhibited the most severe defects in migration (Fig. 1e), which may have contributed to their higher disease score. Furthermore, the neuronal differentiation defects of the L- and TCA-HSCR NCs were found to be more severe than those of the S-HSCR NCs, as observed in the formation of neuron-like processes and the expression levels of pan-neuronal markers (neurofilament [NF] and protein gene product 9.5 [PGP9.5]) (Fig. 1g). These data showed that the disease phenotypes and severities are highly associated with the combined effects of migration and differentiation defects of iPSC-derived NCs.

We then sequenced the control and HSCR-NCCs at high resolutions using Smart-Seq. A total of 3,342 individual cells were sequenced, and 6,344 average genes as well as 770,870 mean confidently mapped reads per cell were identified (Supplementary Data 1 and Supplementary Fig. 2a-c). We projected all single cells on a *t*-distributed stochastic neighbor embedding (*t*-SNE) plot (Fig. 1h) and identified five transcriptionally distinct clusters. Cluster 1 was identified to be the major cluster showing an expression profile highly resemble the major population of ENS cells identified in E12.5 mouse embryos, which express glial and progenitor marker genes, such as *Sox10*, *ErbB3*, and *Plp1*, without any neuronal gene expression²¹ (Fig. 1i & j). We also compared this major cluster (Cluster 1) of our iPSC-derived cells to mouse NCCs on enteric and non-enteric lineages at E13.5 obtained from three independent datasets. They included our own dataset where two subpopulations of vagal NCCs: mouse enteric (mENCCs)¹⁹ and cardiac NCCs (mCNCCs), were sequenced at single cell level using 10^x Genomics; and another set of published scRNA-seq sequence data of mouse NCCs on enteric (mENCCs) and peripheral nervous system (mPNS) lineages derived from sciRNA-seq3 platform²². All NCCs from different platforms and lineages were integrated together by Seurat software²³ to remove technical and batch effects. As demonstrated by the correlation analysis and principal component analysis (PCA), the iPSC-derived cells showed the highest similarity to the mouse ENCCs when compared to the NCCs on other lineages (Fig. 1k and Supplementary Fig. 2f). Therefore, hereafter, the NCCs derived from the iPSC lines will be termed ENCCs. RNA velocity was proven to be a powerful method to predict cell state based on unspliced and spliced mRNAs in single cell²⁴. Based on the RNA velocity analysis, Cluster 1 was predicted as the progenitor cells which had great differentiation potential towards clusters 2-5 (Supplementary Fig. 2d), further suggesting that the majority of iPSC-derived cells are at the progenitor state. It is also noteworthy that the five clusters demonstrated different, uneven proportions of cells in the control and HSCR samples, reflecting the heterogeneity between samples (Supplementary Fig. 2e) and that may be also associated with the disease phenotypes.

Stratification of HSCR patients and reconstruction of the disease severity axis based on single-cell transcriptomes

Cluster 1 represents the major population of cells found in all iPSC-derived ENCCs and highly resembled the major progenitor population of ENS as identified in E12.5 and E13.5 mouse embryos, so cells in Cluster 1 were used for the subsequent single-cell transcriptome analyses. We first examined whether the transcriptome profiles of HSCR-ENCCs are associated with the disease severity. Genes that were uniquely expressed in one sample were excluded for downstream analysis to remove individual effects. A cladogram was constructed to estimate the overall transcriptional similarities between cells in Cluster 1 for all samples. Samples with similar clinical severity scores tended to be clustered together (Fig. 2a). HSCR#6 was located at a unique branch in the cladogram, suggesting that the pathogenic pathways involved in TCA are different from those in S- and L-HSCR. Although HSCR#1 was classified as S-HSCR based on clinical stratification, it showed similarities with the L-HSCR specimens (HSCR#17 and #23). Next, we compared each HSCR-ENCCs with those from the two controls. At a false discovery rate (FDR) of 1%, we identified 879, 909, 2,301, 3,635, 6,445, 1,790, and 818 differentially expressed genes (DEGs) in HSCR#5, #10, #20, #1, #17, #23, and #6, respectively. Except for those in HSCR#6, which was derived from a patient with TCA carrying a deletion mutation in the *RET* gene (*RET* G731del), the ENCCs with higher clinical scores demonstrated significantly higher numbers of DEGs in both threshold-free (Supplementary Fig. 2g) and fixed-threshold DEG analysis and comparison ($R^2=0.464$, $P<0.01$) (Supplementary Fig. 2h), implying that more genes and pathways are disrupted in more advanced disease states. TCA-ENCCs exhibited unique transcription profiles with enriched molecular pathways associated with GDNF/*RET* signaling as described previously⁷ and in Supplementary Fig. 3. To identify the *RET*-independent pathways implicated in HSCR pathogenesis, we excluded HSCR#6 from subsequent analyses.

Since the S- and L-HSCR patients show different extent of aganglionosis, we then determined whether there is a pseudo-axis that can reflect the clinical phenotypic heterogeneity and/or the extent of aganglionosis based on single-cell transcriptomes from patients and control iPSC-derived ENCCs. To this end, we employed PCA to re-cluster the control- and HSCR- ENCCs. A disease severity axis was observed along the PC2 axis, with the ENCCs distributed according to the extent of aganglionosis (Fig. 2b). In particular, S-HSCR#1, which had a relatively longer aganglionic segment, ENCCs derived from this iPSC-line fell between the S-HSCR- and L-HSCR-ENCCs in the PC2 axis (Fig. 2c). Therefore, the stratification of disease severity based on single-cell profiling of iPSC-derived ENCCs was consistent with clinical stratification. Furthermore, the case-control divergence in this component may reflect cells at varying severity states along a shared disease process. The ENCCs of S-HSCR#5, #10, and #20 were more similar to the control group, demonstrating they were less severe than the S-HSCR#1 ENCCs which was at the intermediate state, and the L-HSCR ENCCs (Fig. 2c).

To further delineate the genes implicated in the disease severity, the differential expression analysis was performed with iPSC-derived ENCCs from two control and six S/L-HSCR-iPSC lines using Monocle³⁷, and 833 genes were identified as the significantly DEGs (q-value < 0.01) along the disease severity axis.

When we examined their expression dynamics along this pseudo disease severity axis, we observed four major categories of transcriptional gene modules in characterized patterns (Fig. 2d-f; Supplementary Data 2). As revealed by over-representation analysis (ORA), module 1 consisted of genes that are up-regulated in HSCR-ENCCs with low disease severity, and are largely involved in “negative regulation of neuron differentiation and Notch signaling,” such as *NOTCH3* and *GLI3* (Fig. 2f). Module 2 consisted of genes that show rapid down-regulation and are enriched in “cell cycle and synapse organization,” such as *ORC3* and *GPC4*. On the other hand, module 3 consisted of genes that are activated at the higher disease severity state, during cell shift from moderate S-HSCR to L-HSCR, and are predominantly “RNA splicing” genes, such as *PTBP1* and *U2AF2*. Lastly, module 4 consisted of genes that are downregulated at the severe disease state and enriched in energy metabolism, such as *COX7B* and *MDH1*. Each disease-associated module was enriched with specific biological processes, suggesting that distinct cellular events are interrupted at different disease stages, accounting for the varying disease severities. In particular, perturbation of neurogenesis, proliferation, RNA splicing, and energy metabolism of ENCCs were found to be the key events associated with the disease severity of HSCR.

To complement with the ORA analysis, the expression of whole pathways was further scored using an additive model. Genes involved in neurogenesis (GO:0050771), proliferation (GO:0051726), RNA splicing (GO:0033120) and cellular respiration (GO:1901857) in GO database were used to estimate the overall changes of the key biological processes. From these analyses, we found genes implicated in neurogenesis (Fig. 2g) and energy metabolism genes (Fig. 2j) were mildly repressed in the S-HSCR-ENCCs, but severely downregulated in L-HSCR-ENCCs, leading to low pathway scores in these two categories. Genes implicated in cell proliferation (Fig. 2h) and RNA splicing (Fig. 2i) were also uniquely repressed and activated in S-HSCR and L-HSCR-ENCCs, respectively. HSCR#1, representing the intermediate case, exhibited interrupted RNA splicing and energy metabolism, while the neurogenesis pathway score was similar to that of S-HSCR-ENCCs. The expression levels of representative genes in these cellular processes are shown in Supplementary Fig. 4. Together, our results revealed that the enhanced RNA splicing accompanied by reduced energy metabolism are associated with severe form of HSCR.

Identification of *HDAC1* as a master regulator across HSCR severity axis

We next aimed to determine the pathway(s) that commonly affected across various patient bases and identify the potential “driver genes” of the disease. A total of 118 DEGs are shared between the S- and L-HSCR groups, including 49 downregulated and 69 upregulated genes. Among these, a core gene set of 20 genes was significantly perturbed along the disease severity axis (DEG across PC2, FDR<0.05) and functionally enriched in biological pathways. We then used a local linear regression model to identify the rate of expression changes of these core genes across severity axis. We found that 14 and 6 core genes were consistently upregulated and downregulated, respectively, in HSCR (Fig. 3a). Based on the STRING database²⁵, these genes encode proteins that show strong interactions and involvement in neurogenesis and NC development (Fig. 3b). Specifically, *HDAC1* was identified as a hub gene in the PPI network that bridging multiple crucial biological processes. Since most of the core genes were upregulated in the

HSCR-ENCCs and showed similar expression pattern with HDAC1 (Fig. 4d), we performed motif enrichment analysis to explore the potential transcriptional coactivator effects of HDAC1 on their expression. We found that 13 of the 20 core genes had at least two sites enriched with the HDAC1 DNA-binding motif (Supplementary Fig. 5). Furthermore, 6 of them were predicted to be “high-confidence” direct target genes of HDAC1 (Fig. 3c). Particularly, we deliberately used the non-DEGs as background genes to perform HDAC1 motif enrichment analysis and observed significantly more enrichment in the DEGs compared to the non-DEGs (OR = 3.41, χ^2 P -value = 1.60E-04). This suggested the binding of HDAC1 to the human genome was neither universal nor random and it was associated with the disease state. Using Single cell regulatory network inference and clustering (SCENIC) analysis²⁶, we established a gene regulatory network (GRN) of HDAC1. HDAC1 exhibited transcript enrichment (orange) paired with concomitant activation of downstream regulatory networks (blue) in various disease states (Fig. 3e). Intriguingly, the HDAC1 GRN was slightly activated in cells of low severity state (S-HSCR) but highly activated in severe disease state (L-HSCR) (Fig. 3e, boxplot). This implies that the upregulation of HDAC1 and activation of its GRN are highly associated with disease severity, highlighting its role in different forms of HSCR.

A whole-transcriptome correlation analysis was performed to identify HDAC1-mediated (activate or repress) genes and pathways. We first evaluated the correlation (co-expression) between HDAC1 and the other genes in the control- and HSCR-ENCCs to identify significant alterations in the regulatory architecture of HDAC1 among HSCR-ENCCs. To identify potential direct binding genes of HDAC1, we then performed a motif enrichment analysis on the differentially expressed HDAC1-associated genes between the control- and HSCR-ENCCs. From these analyses, we predicted 5,336 and 512 genes to be the HDAC1-activated and repressed genes, respectively, where the activated genes represent the direct targets of the HDAC1 with binding motif(s) in their promoter regions (Fig. 3f). We found that the putative HDAC1-activated genes are enriched in “RNA splicing and protein localization to nucleus” pathways, while the repressed target genes are enriched in “action potential, regulation of membrane and pH”, likely implicated in regulating the ion channels of the cellular membrane (Fig. 3g).

The transcriptomic data suggest that the elevated transactivation activity of HDAC1 would be an underlying cause of HSCR. As revealed by the scRNA-seq data, the *HDAC1* transcript levels were elevated in ENCCs derived from 2 S-HSCR (HSCR#20, #1) and 2 L-HSCR (HSCR #17 & #27) lines (Supplementary Fig. 6a). Therefore, we performed additional biological assays to delineate how HDAC1 may influence the development of hENCCs and determine how the aberrant HDAC1 activity may cause HSCR disease. We focused on two control- and 4 HSCR- lines: HSCR#5 and HSCR#20 were the two S-HSCR without and with an elevated expression of *HDAC1* transcripts, respectively; HSCR#1 exhibited an intermediate transcriptomic signature bridging S- and L-HSCR and had high expression of *HDAC1* transcripts; and a L-HSCR (HSCR#17) showed elevated expression of *HDAC1*. Western blot analysis revealed that the total protein expression of HDAC1 in control- and all HSCR-ENCCs are highly comparable (Supplementary Fig. 6b). We also examined the repressor function of HDAC1 based on its ability to catalyze the removal of acetyl groups from core histones (H3). As shown in Supplementary Fig. 6b, the H3 were highly acetylated

in both control and all HSCR-ENCCs as monitored based on the level of H3K27Ac (Supplementary Fig. 6b), so no obvious deacetylation was found in HSCR-ENCCs. HDAC1 forms Sin3, NuRD and CoREST complexes to repress gene expression through promoting histone deacetylation. Therefore, we also examined the potential repressor function of HDAC1 based on the combined expression levels of various subunits in the Sin3, NuRD and CoREST complexes (Table S2), the inferred Sin3, NuRD and CoREST complex activities were highly comparable between control and HSCR-ENCCs (Supplementary Fig. 6c). All these data imply that the aberrant HDAC1 activity as seen in HSCR-ENCCs is unlikely due to the enhanced expression of HDAC1 itself nor its repressor function. On the other hand, expression of HDAC1 was only found in the nuclei of HSCR-ENCCs (Fig. 3h), while both cytoplasmic and nuclear expression of HDAC1 were observed in control-ENCCs. Given the overall expression levels of HDAC1 was comparable between the control- and HSCR-ENCCs, the enhanced nuclear localization of HDAC1 may increase the transactivation activity of HDAC1 in HSCR-ENCCs, leading to the aberrant expressions of HDAC1 target genes. To directly demonstrate the impact of the elevated HDAC1 activity in the development of ENCCs, we performed the *in vitro* differentiation assays using control- and HSCR-ENCCs, where the control- and HSCR-ENCCs were directed to neuronal lineage to generate enteric neuronal progenitors (NPs) in absence or presence of HDAC1 inhibitor, Valporic acid (VPA, 0.4mM). Consistently, inhibition of HDAC1 activity, by addition of VPA, favored the neuronal lineage differentiation of all HSCR-ENCC, and significantly more NF expressing NPs were found at the day 5 of differentiation (Fig. 3i). In summary, aberrantly high HDAC1 transactivation activity represents a common cause of neuronal differentiation defect associated with HSCR disease.

Global analysis of splicing among different HSCR-ENCCs

Alterations in RNA splicing pathways were found to be highly associated with advanced disease states in HSCR (Fig. 2i). Consistent with this, we identified 37 RNA-splicing mediators, including *PTBP1*, *SRSF2*, *U2AF1*, *U2AF2*, specifically upregulated in severe HSCR cases (Fig. 4a). Subsequent global alternative splicing (AS) analysis also revealed that the sequencing reads in S-HSCR#1 and L-HSCR ENCCs show significantly higher splicing frequencies than the control ENCCs (Fig. 4b). The percentage of transcripts with sequences spliced in (percent spliced in [PSI]) was estimated for cassette exons, alternative 5'/3' splice sites, and various AS events. The most common AS events identified were cassette inclusion and alternative first exons, accounting for more than 60% of the total AS events (Fig. 4c). Except for intron retention, all the other six AS types had significantly higher numbers of events in ENCCs at high disease state than those in low disease states ($P < 0.05$, *t*-test). Additionally, the overall expression of the 37 RNA-splicing mediators significantly and positively correlated with AS frequency (Fig. 4d). L-HSCR-ENCCs exhibited distinct AS profiles, with a unique cluster in the PCA (Fig. 4e). In total, 4,735 AS events occurred at significantly higher rates in the L-HSCR-ENCCs than in control-ENCCs, affecting 2,063 genes (FDR < 0.05). The core AS genes that could contribute to the advanced disease state (PC2) are shown in Fig. 4f and many of them are involved in neuronal differentiation and cell metabolism. Polypyrimidine Tract Binding Protein 1 (*PTBP1*) showed the highest association with L-HSCR and were present in multiple AS forms resulting from the inclusion of exon 9 (*PTBP1* [1]) as well as alternative first exons *PTBP1* [2] and [4]) and retention of the intron (*PTBP1* [3]) (Fig. 4g, Supplementary Data 5). Different isoforms of *PTBP1*

have been implicated in neurogenesis^{27,28} or glycolysis²⁹. In addition, we found that many of other top differentially spliced genes are the putative targets of PTBP1 [1] (inclusion of exon 9) (Fig. 4f, highlighted in red). For instance, Adenylate Kinase 3 (*AK3*) encodes a GTP:ATP phosphotransferase which is involved in cell energy metabolism. Ubiquitin Specific Peptidase 14 (*USP14*) is a deubiquitinating enzyme and associated with neurodegeneration. The inclusion of exon 4 and exon 3 in *AK3* and *USP14* were found significantly enriched in all HSCR and L-HSCR, respectively. Intriguingly, the combination of transcriptional and post-transcriptional profiles explains the measured clinical scores of the HSCR cases, providing a comprehensive picture of the molecular constituents and mechanisms underlying the heterogeneity of HSCR (Fig. 4h).

Inclusion of *PTBP1* exon 9 was associated with the disrupted cellular metabolism in L-HSCR-ENCCs

Among all AS events, the inclusion of *PTBP1* exon 9 appears to contribute most significantly to the advanced disease state; *PTBP1* was overexpressed in the HSCR, particularly higher in the severe disease cases, with strong enrichment of exon 9 (Fig. 5a, Supplementary Fig. 7). Exon 9 in *PTBP1* encodes a linker between RNA recognition motif 2 (RRM2) and RRM3 in *PTBP1*. It also possesses splicing regulatory activity that are distinct from its RNA-binding activity²⁸. Therefore, we further analyzed the relationship between *PTBP1* exon 9 inclusion and other AS events found in HSCR-ENCCs. First, AS events significantly correlated with *PTBP1* exon 9 were selected. Then, the exons targeted by PTBP1 through motif recognition were used for further identification of HSCR-associated AS events. From this, approximately 2,000 and 1,000 AS events were predicted to be enhanced and repressed, respectively, by *PTBP1* exon 9 in HSCR-ENCCs (Fig. 5b). Remarkable overlaps were found between our predicted *PTBP1* exon 9 targets and public data²⁸ (Fig. 5c). Overall, the *PTBP1* exon 9-repressed exons represented genes that are related to “metabolic processes” and “biosynthetic process” pathways, while the genes with *PTBP1* exon 9-activated exons were involved in “protein binding”, “RNA localization”, “axonogenesis” and “axon guidance” pathways (Fig. 5d).

One of the known targets of *PTBP1* is *PKM*, which encodes two key enzymes (pyruvate kinase M [*PKM*] 1 and *PKM2*) involved in glycolysis, and they are responsible for converting phosphoenolpyruvate (PEP) to pyruvate as a substrate for oxidative phosphorylation (OXPHOS). Upregulation of *PTBP1* or inclusion of the *PTBP1* exon 9 enhances the binding of *PTBP1* to the splice sites flanking exon 9 in *PKM* transcripts, resulting in mutual exclusion (MX) of exon 9 and inclusion of exon 10 to generate *PKM2* (Fig. 5e). Significant RNA recognition motif (RRM) enrichment was found in the *PKM* gene bodies of the HSCR-ENCCs ($P = 4.9 \times 10^{-17}$), with *PTBP1* exon 9 inclusion inversely correlated with *PKM* exon 9 (*PKM1*) (Spearman $r = 0.536$). According to the correlation distribution of *PTBP1* exon 9 inclusion to other AS events (Fig. 5b), we found more than 95% of the correlation values located within -0.4 to 0.4. The relatively low correlation value may be attributed to the low RNA materials in single cells, and thus limits the accuracy and sensitivity of the AS event detection by the algorithms. However, the strong correlation between *PTBP1* exon 9 inclusion and MXE10 events in *PKM* suggests a highly confident causal regulation between them (Fig. 5e). In concordant with these findings, subsequent quantitative RT-PCR

analysis also showed a significantly high *PTBP1* exon 9 and a low *PKM1* transcript level in L-HSCR-ENCCs, while *PKM2* levels did not show association with any disease state (Fig. 5f).

PKM1 efficiently converts PEP into pyruvate that is preferentially used for OXPHOS, favoring neurogenesis³⁰. Consistently, the overall expression of OXPHOS pathway genes were found to be decreased in the HSCR cases as shown in the scRNA-seq data, highly correlated with reduced *PKM1* levels (Fig. 5g). Concordantly, the cellular aerobic respiration of HSCR-ENCCs was found to be lower than that of the control cells as determined by the oxygen consumption rate (OCR) using Seahorse analyzer. The association of the OXPHOS with the OCR were further illustrated by inducing metabolic stress with addition of oligomycin, carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP) and antimycin A/rotenone in succession. In both control- and HSCR-ENCCs, treatment with oligomycin, an ATP synthase inhibitor, induced a loss of OCR, indicating that the mitochondrial respiration is coupled with ATP production in ENCCs (Fig. 5h). To measure reserve respiratory capacity, which is an indicative parameter of maximal respiratory efficiency of mitochondria, we next added FCCP, an ionophore that induces high proton conductance into the mitochondrial membrane with rapid acceleration of the mitochondrial electron transport chain (ETC). FCCP treatment resulted in a high OCR increase in control-ENCCs, but not in the HSCR-ENCCs (Fig. 5h). Both the basal OCR (Fig. 5i) and the maximal respiratory capacity of mitochondria (Fig. 5j) were reduced in HSCR-ENCCs. Therefore, the reduced *PKM1* activity in HSCR, at least partly due to the inclusion of *PTBP1* exon 9, may perturb OXPHOS, leading to the retarded neuronal differentiation of ENCCs.

Hedgehog-enhanced cellular metabolism promoted growth and neuronal differentiation of ENCCs

Our previous study demonstrated that activation of Hedgehog pathway by addition of a Smoothed agonist (SAG) primes ENCCs to a more advanced cell state and favors the neuronal lineage differentiation⁸. Subsequent analysis of the transcriptomes of these SAG-primed ENCCs at single cell level further revealed that the genes implicated in glycolytic and fatty acid oxidation (FAO) pathways are robustly upregulated (Fig. 6a & b). In concordance with these observations, a full activation of Hedgehog pathway by SAG or a partial agonism through applying a low dosage of cyclopamine (Cyc^{Low}, 1nM) significantly increased the cellular aerobic respiration of ENCCs as monitored based on the OCR (Fig. 6c), accompanied by significant increases in both OCR/ extracellular acidification rate (ECAR) ratio (Fig. 6d) and the maximum respiration (Fig. 6e). In addition, a specific activation of AMPK-PKM2-PDHA1 pathway was observed upon an addition of SAG or Cyc^{Low}, and this pathway was blocked by a high dosage of cyclopamine (Cyc^{High}, 100nM), which antagonizes the Hedgehog pathway (Fig. 6f). Activation of the AMPK-PKM2 pathway by the full (SAG) or partial (Cyc^{Low}) agonism of Hedgehog pathway led to an enhanced ATP production, which could be abolished by an addition of PKM2 inhibitor (PKM2i) (Fig. 6g). Both lactate (Fig. 6h) and pyruvate (Fig. 6i) levels were significantly increased upon activation of the Hedgehog pathway, and they served as substrates to support the anaerobic and aerobic cellular metabolism, respectively, resulting in a higher ATP yield. Importantly, the Hedgehog-enhanced metabolic rewiring greatly increased the survival and/or growth of ENCCs as revealed by the MTT assay (Fig. 6j).

The involvement of FAO pathway in Hedgehog-enhanced metabolic rewiring was also determined using a FAO-specific inhibitor, Etomoxir (ETO). Unlike glycolysis, a full agonism of the Hedgehog pathway was required for activation of the FAO pathway, where only SAG, but not Cyc^{Low}, could significantly increase the FAO-associated OCR (Fig. 6k) and maximum respiration (Fig. 6l). In addition, activation of FAO or Hedgehog pathway by GW9508 and SAG, respectively, led to a significant enhancement in the mitochondrial activity as monitored by the mitotracker (Fig. 6m, red), while the SAG-treated cells exhibited lower mitochondrial activity when ETO was added, suggesting the involvement of FAO in SAG-mediated mitochondrial activity. More importantly, activation of FAO pathway conferred ENCCs a greater neuronal differentiation capacity and significantly more neurons (TUJ1⁺, NF⁺) were found at day 5 of neuronal differentiation when FAO pathway was activated by the addition of GW9508 (Fig. 6n). Our results support an essential role of the Hedgehog/FAO/mitochondria axis in mediating the neuronal differentiation of ENCCs.

Hedgehog-induced OXPHOS restored the neuronal differentiation capacity of HSCR-ENCCs

We next sought to determine whether the Hedgehog-induced metabolic rewiring can rescue the differentiation defects of HSCR-ENCCs. The FAO-dependent mitochondria activity in HSCR-ENCCs was elevated upon SAG treatment which was blocked by ETO (Fig. 7a-c). Intriguingly, activation of the Hedgehog/FAO/mitochondria axis could also greatly improve the differentiation capacity of HSCR-ENCCs and significantly more TUJ1⁺ and NF⁺ neurons were detected when FAO pathway was activated by GW9508 (Fig. 7d & e).

Based on the result from the time-course experiment as described previously, addition of SAG during ENCC induction could maximize the yield and prime cells to the neurogenic lineage with enhanced cellular metabolic rate⁸. Therefore, we employed the same differentiation strategy to generate HSCR-ENCCs (Fig. 7f). Consistently, SAG-treated HSCR-ENCC-derived TUJ1⁺ neurons grew better, exhibited better differentiation competency and required a shorter differentiation time to generate neuronal subtypes, such as tyrosine hydroxylase (TH) (Fig. 7g & h). In sum, our data suggest that activation of the Hedgehog pathway promotes glycolysis and FAO to support the survival/growth and differentiation of ENCCs (Fig. 7i).

Discussion

In the present study, we depicted the cellular landscape of HSCR-ENCCs in mild to advanced disease stages by analyzing the transcriptomes of control- and HSCR-ENCCs at the single-cell level. Intriguingly, we identified common core regulatory machineries across different patient bases and built a model network of interacting disease-modifiable factors that detail the complex nature of HSCR.

Through high-resolution scRNA-seq and systemic multi-patient comparison analyses, we defined the core and disease-modifiable gene sets associated with HSCR pathogenesis. As supported by the clinical and molecular features of the intermediate case (HSCR#1), we observed a pseudo-disease severity axis that

can reflect the difference from mild to severe HSCR disease states (S-HSCR to L-HSCR). Additionally, our results suggest that disruption of RNA splicing and energy metabolism pathways was associated with the disease severity axis and would promote more extensive aganglionosis. Importantly, we identified a core gene set with 118 genes representing the shared molecular machinery underlying the pathogenesis of S- and L-HSCR. Specifically, HDAC1 was identified to be the overarching regulator of the core gene set, where transcriptional activation of HDAC1 targets were found elevated in HSCR-ENCCs.

Enhanced AS events were found in HSCR-ENCCs, particularly, in the more advanced disease states. It could be the result of the upregulation of RNA splicing mediating factors. Many of the top disease-associated AS genes, such as *PTBP1*, *AK3*, and *USP14*, are associated with neurogenesis and energy metabolism. In particular, the inclusion of exon 9 in *PTBP1* contributed to many HSCR-associated AS events. Inclusion of exon 9 in *PTBP1* was associated with the reduced MXE10 in *PKM* and a significantly lower abundances of *PKM1* in ENCCs in advanced disease states. These changes likely prohibit metabolic rewiring from glycolysis to OXPHOS which is required for supporting energy needs in cell state transition and differentiation^{31,32}. In concordant with the expression data, significantly lower OXPHOS was consistently detected in HSCR-ENCCs and that was highly associated with the poor survival and differentiation capacity of HSCR-ENCCs.

Glycolysis and FAO are essential metabolic pathways for producing pyruvate and acetyl-CoA, respectively, to fuel the tricarboxylic acid (TCA) cycle for ATP generation under an energy-demanding condition. The Hedgehog signaling is frequently coupled with bioenergetics of various cancers, and is associated with the insulin-independent glucose uptake in muscle and brown adipose tissue in obese and diabetic patients¹³. Here, we used Cyc^{Low} and SAG to achieve a partial agonism and a full activation of the Hedgehog pathway and revealed that the Hedgehog-mediated glycolytic and FAO metabolic pathways greatly enhance the survival and neuronal differentiation competency of ENCCs, respectively. These functions are conserved in HSCR-ENCCs, which usually have lower OXPHOS activity. According to the transcriptomic analysis as described previously, activation of the Hedgehog pathway primes ENCCs to neurogenic lineage without altering their neuronal differentiation trajectory⁸. It suggests that Hedgehog/FAO/OXPHOS axis would be the potential mechanism underlying this process, which is consistent with the preference for mitochondrial OXPHOS during the cell state transition and neuronal differentiation. The non-canonical metabolic routing represents a quick and flexible way to provide energy and prepare ENCCs for cell state transition/differentiation, but it likely requires the reinforcement by the canonical Hedgehog signaling, which further initiates a transcriptional glycolysis response and modulates the expression of the entire machinery of primary metabolism including the Hedgehog/FAO/OXPHOS axis. As supported by our own transcriptional profiling data of SAG-primed ENCCs and the data from the differentiation assays, the full activation of Hedgehog pathway by SAG treatment robustly upregulates the expression of genes implicated in various cellular metabolic pathways, and that leads to a sustainable and profound effect on cell state transition and differentiation. In particular, HSCR-ENCCs frequently show low OXPHOS activity, the prolonged activation of Hedgehog is required for improving their survival and differentiation capacity. Nevertheless, the aberrant activation of

the canonical pathway is associated with the pro-oncogenic nature of the Hedgehog pathway and HSCR disease. The use of FAO agonist, or in combination with a selective partial agonism of Hedgehog pathway, may represent an alternative approach to restore the functions of ENCCs in HSCR patients and warrants further studies.

In conclusion, our study provides a comprehensive biological picture of the core molecular mechanisms underlying HSCR heterogeneity. In particular, we show that the metabolic rewiring is critical for the development of ENCCs and the defective OXPHOS represents an underlying cause of L-HSCR.

Methods

Patients

Seven HSCR patients from different subgroups (4 S-HSCR, 2 L-HSCR and 1 TCA) were recruited at Queen Mary hospital, Hong Kong. Each patient was given a disease score according to the length of aganglionic segment (0-5; control= 0 and TCA=5) (Fig. 1A & B). The *RET* alleles in rs2435357 and variants in the coding region of *RET* gene were identified through sequencing the genomic DNA from the patients' blood. In brief, among the four S-HSCR patients, HSCR#5, #10, #20 showed relatively mild phenotypes with disease scores 1 or 2, while HSCR#1 had a longer aganglionic segment up to the lower sigmoid, representing the intermediate phenotype between S- and L-HSCR. Skin biopsies were obtained from these patients for generation of iPSC lines. The study was approved by the institutional review board of The University of Hong Kong together with the Hospital Authority (UW 13-419).

Human induced pluripotent stem cells (iPSCs)

A control hiPSC line (IMR90-iPSC) was obtained from WiCell Research Resources (WiCell, WI., RRID:CVCL C434). Another control hiPSC cell line (UE02302) was generated from urine derived cells of a male individual by episomal reprogramming vectors carrying the four reprogramming factors³³. All iPSCs used in this study were at the intermediate (35-65) passage numbers and maintained on Matrigel (Corning, 354277)-coated plate with mTeSR1 medium (StemCell Technologies, 85850).

Generation of iPS cell lines

Fibroblasts were derived from skin biopsies obtained from patients diagnosed with HSCR disease. Fibroblasts were then maintained and expanded in DMEM with 10% FBS. iPSC were generated from fibroblasts using episomal vectors carrying the four reprogramming factors (Oct4, Klf4, Sox2, and c-Myc)³⁴. In brief, 1×10^6 fibroblasts were transfected with episomal vectors using Nucleofector with transfection kit (Lonza, VPD-1001) and the transfected cells were then replated on a 35mm gelatin-coated well. The medium was changed 24 hours later and replaced daily from thereon. At day 7, cells were dissociated and replated on MEFs in ES medium (DMEM/F12 (Thermo Fisher Scientific,, 10565018) plus 10% KnockOut Replacement Serum (KSR) (Thermo Fisher Scientific,, 10828028), 0.5% L-glutamine (Thermo Fisher Scientific, 25030018), 1% NEAA (Thermo Fisher Scientific, 11140050), 0.1% β -

mercaptoethanol (Thermo Fisher Scientific, 21985023), 0.5% penicillin-streptomycin (Thermo Fisher Scientific, 15140122) and FGF2 (Peprotech, 100-18B). Cells were fed daily with ES medium. Following the appearance of hESC-like colonies after approximately 3–4 weeks, individual colonies were isolated and clonally expanded to establish iPSC lines.

Derivation of neural crest (NC) from iPSC lines

At day 0, Control or HSCR-iPSCs were seeded on Matrigel-coated plate (10^5 cells cm^{-2}) in iPS cell medium containing 10 ng ml^{-1} FGF2 and $10 \text{ }\mu\text{M}$ Y-27632 (Tocris Bioscience, 1254). Differentiation was then initiated by replacing iPSC medium with KSR medium, containing KnockOut DMEM (Thermo Fisher Scientific, 10829018) plus 15% KSR, 1% NEAA, 1% L-glutamine, 1X β -mercaptoethanol, 1% penicillin-streptomycin, LDN193189 (100 nM , Reprocell, 04-0074) and SB431542 ($10 \text{ }\mu\text{M}$, Abcam, ab120163). The dual SMAD inhibitors and a potent GSK inhibitor were added at different time frame during the NC induction, including LDN193189 (from day 0 to day 3), SB431542 (from day 0 to day 4), $3 \text{ }\mu\text{M}$ CHIR99021 (from day 2 to day 10, Reprocell, 04-0004). The NC cells were finally caudalized with $1 \text{ }\mu\text{M}$ retinoic acid (Abcam, ab120728) (from day 6 to day 9). The KSR medium was gradually changed to N2 medium at day 4 by increasing N2 from 25% to 75% from day 4 to 9 as described previously⁷. The N2 medium contained neural basal medium (Thermo Fisher Scientific, 22103049) and DMEM/F12 in 1:1 ratio supplemented with 0.5% N2 supplement (Thermo Fisher Scientific, 17502048), 1% B27 supplement (Thermo Fisher Scientific, 17504044), 5 mg ml^{-1} insulin (Thermo Fisher Scientific, 12585014) and 1% penicillin-streptomycin. The ENCCs were enriched by FACS with antibodies against p75^{NTR} and HNK-1 at day 10 of the differentiation as described^{7,8,20}. FACS-sorted ENCCs were collected for single-cell RNA sequencing.

***In vitro* differentiation of ENCCs to neuronal progenitors (NPs)**

Around 40 thousand FACS-enriched ENCCs were seeded as droplets on poly-ornithine/laminin/fibronectin (PO/LM/FN)-coated 24 well plate in N2 medium containing 10 ng ml^{-1} FGF2, $3 \text{ }\mu\text{M}$ CHIR99021 and $10 \text{ }\mu\text{M}$ Y-27632. The neuronal differentiation started 48 hours later and the attached ENCCs were then cultured with N2 medium containing BDNF (10 ng ml^{-1} Peprotech, 450-01), GDNF (10 ng ml^{-1} , Peprotech, 450-10) and ascorbic acid ($200 \text{ }\mu\text{M}$, Sigma, A4034-100G), NT-3 (10 ng ml^{-1} , Peprotech, 450-03), NGF (10 ng ml^{-1} , Peprotech, 450-01) and cAMP ($1 \text{ }\mu\text{M}$, Sigma, D0260). For valproic acid (VPA) treatment, 1 mM VPA (Sigma, P4543) was added during neuronal differentiation. The culture medium was changed every 2 days. ENCC-derived neurons at differentiation day 5 or 9 were fixed for immunocytochemistry analyses.

Fluorescence activated cell sorting (FACS) and flow cytometric analysis

For quantifying the ENCCs, the 10 day-differentiated cells were dissociated with Accutase and then incubated with anti-human antibodies including APC-HNK-1 (BD Pharmingen, 560845) and FITC-p75^{NTR} (Miltenyi Biotec, 130-091-917) for 30-45 minutes on ice. To stain for PE-RET (Neuromics, FC15018), the cells were fixed in 4% paraformaldehyde for 10 minutes at room temperature and

permeabilized using 0.1% (w/v) Saponin solution, then washed and blocked in PBS with 2% FBS. The cells are then stained with antibodies for 30-45 minutes on ice. Approximately 10^6 cells were stained and labeled cells were detected using FACS Aria III (Becton Dickinson Immunocytometry Systems, San Jose, CA, USA). Isotype-matched antibodies were used as controls. FlowJo version 8.2 (Tree Star, Inc.) was used to analyze flow data.

For cell sorting, HNK-1/p75^{NTR} stained cells were washed and resuspended in PBS with 2% FBS. The HNK-1 and p75^{NTR} double positive cells were enriched using fluorescence activated cell sorting (FACS) with BD FACS Aria III Cell Sorter. The HNK-1 and p75^{NTR} double positive cells were gated and sorted using the four-way purity mode and the purity of sorted cells was >96% and evaluated by flow cytometry. The sorted neural crest cells were collected for immunostaining or subsequent experiments. A list of primary antibodies and the working dilutions are provided in Supplementary Table 1.

Immunofluorescence analysis

For immunocytochemistry, the cells were fixed with 4% paraformaldehyde in PBS at room temperature for 30 min, followed by blocking with 1% bovine serum albumin (BSA) (Thermo Scientific, 23209) with or without 0.1% Triton X-100 (Sigma, T8787) in PBS buffer. Cells were then incubated in primary antibody overnight at 4°C and host-appropriate Alexa Fluor -488 or 594 secondary antibody (Molecular Probes, Invitrogen) (Supplementary Tables 1 & 2) for 1 h at room temperature. Cells were then counterstained with mounting medium with DAPI (Thermo Scientific, P36931) to detect nuclei. Cells were photographed using Carl Zeiss confocal microscope (LSM 810). Quantitative image analysis of differentiated neuronal cultures was performed with ImageJ plug-in tool. The total signal of neuronal marker was normalized with DAPI signal and the values reported in bar charts represent the mean \pm SEM. The same approach was used to quantitate the expression of neuronal marker in various groups. A minimum of 4,000 cells were analyzed per sample.

Migration assay

FACS-enriched ENCCs were plated on human fibronectin coated 12-well culture plates ($30,000$ cells cm^{-2}). After 24 hours, cells were treated with mitomycin ($10\mu\text{g ml}^{-1}$) to stop the cell proliferation. A wound was created in the center of each well by scratching with a pipette tip. Cells were allowed to migrate for 18 hours. The images of the initial wound and final wound were captured immediately and at 18 hours after scratching. The migration distance was obtained by comparing the width of the initial wound created and wound closure during 18 hours.

Reverse Transcription-PCR (qRT-PCR)

Total RNA was isolated from iPSC-derived ENCCs using RNeasy Mini Kit (Qiagen); and reverse transcribed in 20 μl reaction system using HiScript II Q RT SuperMix (Vazyme, R223-01), in accordance with the manufacturer's instructions. diluted cDNA samples were amplified by Luna® Universal qPCR Master Mix (New England BioLabs, M3003) with forward and reverse primers specific for *HDAC1*, *PKMEx2-4*, *PKM1*

(*PKM-Ex9*) and *PKM2* (*PKM-Ex10*). Fluorescence signal was measured by ViiA 7 Real-Time PCR System (Thermo Fisher Scientific) at the end of each cycle. Each individual sample was assayed in triplicate and gene expression was normalized with *GAPDH* expression. The primer sequences are listed in Supplementary Table 3.

Western blotting

FACS-sorted ENCCs were collected and lysed using protein lysis buffer containing 50mM Tris-HCl, pH7.5, 100mM NaCl, 1% Triton X-100, 0.1mM EDTA, 0.5mM MgCl₂, 10% glycerol, protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche). After incubation on ice for 15 minutes and the total proteins were collected by centrifuge for 10 minutes at 12000rpm, 4 °C. 20µg of total protein from cell lysates was separated on 12% SDS-polyacrylamide gels and blotted with the corresponding primary antibodies. A list of primary antibodies and working dilutions is provided in Supplementary Table 1. The same membranes were stripped and hybridized with anti-β-actin monoclonal antibody (Millipore, MAB 1501) as a protein-loading control. All blots were incubated with secondary horseradish peroxidase-conjugated anti-mouse or anti-rabbit or anti-goat antibody (1:2500, DAKO).

Seahorse Analysis

The Agilent Seahorse XFe96 Analyzer (Seahorse Bioscience, Billerica, MA) was used to measure basal, ATP-linked, and maximal oxygen consumption rate (OCR); reserve capacity; and extracellular acidification rate (ECAR). Cellular bioenergetics was performed using XF Cell Mito Stress Test Kit (Agilent Technologies, 103010-100), XF Glycolysis Stress Test Kit (Agilent Technologies, 103020-100), and XF Long Chain Fatty Acid Oxidation Stress Test Kit (Agilent Technologies, 103672-100). 5 x 10⁴ of FACS-enriched ENCCs (Passage 1 -2) were seeded onto fibronectin-coated Seahorse XF96 V3 PS Culture Microplate (Seahorse Bioscience Inc, Billerica, MA, USA, 101085-004) in N2 medium containing 10ng ml⁻¹ FGF2 and 3 µM CHIR99021. After a 24-hr incubation, N2 medium from each well were removed, and washed once with pre-warmed assay medium (Seahorse XF DMEM Medium, Seahorse Bioscience Inc, 103575-10) supplemented with 25 mM glucose, 2 mM L-glutamine, 1 mM sodium pyruvate, 0.5% N2 supplement, 1% B27 supplement, 1% penicillin-streptomycin, 10ng ml⁻¹ FGF2 and 3 µM CHIR99021, then 180µl of assay medium was added. Cells were incubated in 37°C incubator without CO₂ for 1hr to allow to pre-equilibrate with the assay medium. Cells were treated with/without PKM2 inhibitor (compound 3k, 1µM, Selleckchem, D8375), SAG (1µM) or cyclopamine-KAAD (1nM) for 1hr.

For Mito Stress Test, ten-fold concentrated compounds in the kit of oligomycin (Complex V inhibitor), carbonyl cyanide-4 (trifluoromethoxy) phenylhydrazone (FCCP, mitochondrial uncoupler), and a mixture of rotenone (complex I inhibitor) and antimycin A (complex III inhibitor) were loaded into the cartridge to produce final concentrations of 1.2µM, 2µM, and 1µM respectively.

For the Glycolysis stress test, ten-fold concentrations of compounds from the kit containing glucose (fuel for glycolysis), oligomycin and 2-Deoxy-D-Glucose (2-DG, competitive inhibitor of hexokinase) were loaded into the cartridge to produce final concentrations of 10 mM, 1.2 µM and 50 mM, respectively.

For Long Chain Fatty Acid Oxidation Stress Test, , ten-fold concentrations of compounds from the kit containing etomoxir (ETO, fatty acid oxidation inhibitor), oligomycin, FCCP and a mixture of rotenone and antimycin were loaded into the cartridge to produce final concentrations of 5 μ M 1.2 μ M, 2 μ M, and 1 μ M respectively.

Data were analysed using Agilent Seahorse analytics and normalized with number of cells seeded in each well.

Lactate and Pyruvate Assay

Lactate and pyruvate levels in medium was measured using Lactate Colorimetric Assay Kit (BioVision Incorporated, #K607) and Pyruvate Colorimetric/Fluorometric Assay Kit (BioVision Incorporated, #K609) according to the manufacturer's instructions, respectively. In brief, FACS-enriched ENCCs (Passage 1 -2) were seeded at 5 x 10⁵/well onto a fibronectin-coated 12-well plate and incubate overnight. On the next day, cells were treated with/without PKM2 inhibitor (1 μ M), SAG (1 μ M) or cyclopamine-KAAD (1nM) for 1hr. The medium were then collected and the levels of lactate or pyruvate were measured simultaneously. The levels were normalized to DNA concentration. The experiments were performed in triplicate.

ATPlite Luminescence Assay

ATP was detected using the ATPlite Luminescence ATP Detection Assay System (PerkinElmer, #6016943) according to the manufacturer's protocol. In brief, FACS-enriched ENCCs (Passage 1 -2) were seeded at 5 x 10³/well onto a fibronectin-coated 96-well plate and incubate overnight. On the next day, cells were treated with PKM2 inhibitor (1 μ M), SAG (1 μ M) or cyclopamine-KAAD (1nM) for 1hr. After incubation, 50 μ L lysis buffer (from the ATPlite kit) was added and shaken for 5min. Then, 50 μ L substrate (from the ATPlite kit) was added and incubated for 15min in dark. The ATP content was measured with VICTOR Nivo Multimode Microplate Reader (PerkinElmer).

Mitochondria activity

To measure the mitochondrial activity, FACS-enriched ENCCs (Passage 1 -2) were seeded at 1.5 x 10⁵/well onto a fibronectin-coated 24-well plate and incubate overnight. On the next day, cells were washed once with PBS, and starved in 2mM glucose NC medium for 2hrs. Cells were then treated with GW9508 (10 μ M), SAG (1 μ M) or ETO (5 μ M, 1hr) + SAG (1 μ M) for 1hr. 250 nM MitoTracker Red (Life Technologies, M7514) were then added, and incubated for 30 min at 37 °C, 5% CO₂. After incubation, cells were washed twice with PBS, and then fixed with 4% paraformaldehyde for 20min at room temperature. After fixation, cells were stained with DAPI, and imaged using a 63x objective on Carl Zeiss confocal microscope (LSM 800).

MTT assay

FACS-enriched ENCCs (Passage 1 -2) were seeded at 5×10^3 /well onto a fibronectin-coated 96-well plate and incubate overnight. On the next day, cells were treated with PKM2 inhibitor ($1\mu\text{M}$), SAG ($1\mu\text{M}$) or cycloplamine-KAAD (1nM) for 1hr. For pyruvate or lactate treatment, cells were washed once with PBS, and starved in 2mM glucose NC medium for 2hrs. Then 2mM pyruvate, 2mM pyruvate + $1\mu\text{M}$ SAG, 4mM lactate or 4mM lactate + $1\mu\text{M}$ SAG were added to the cells, and incubated for 3hr or 6hr. $10\mu\text{l}$ of MTT labelling reagent (final concentration 0.5mg/ml , Sigma, M2128) was added to each well for 3hr. Medium with labelling reagent were then removed, and $200\mu\text{l}$ DMSO was added to solubilize the purple formazan crystals. The absorbance was measured at 570nm using VICTOR Nivo Multimode Microplate Reader (PerkinElmer).

Plate-based scRNA-seq

Plate-based single-cell (sc) RNA sequencing were performed at the Centre of Genomic Science, The University of Hong Kong. For scRNA-seq of human cells, the Smart-seq® v4 (Clontech) kit was used for first-strand synthesis. Single cells were directly sorted into 4 l of lysis buffer in a 384-well plate using a FACSaria III flow cytometer (BD Biosciences). First-strand DNA was synthesized within 16 cycles of amplification according to the manufacturer's instructions. cDNA was purified on Agencourt AMPureXP magnetic beads, washed twice with fresh 80% ethanol and eluted in 17 l of elution buffer. Then, 1 l of cDNA was checked and quantified on an Agilent Bioanalyzer high-sensitivity DNA chip. Sequencing libraries were produced using Illumina Nextera XT tagmentation according to the manufacturer's instructions except using 150 pg of input cDNA, 5 min of tagmentation and 12 cycles of amplification using the Illumina XT 24 index primer kit. Libraries were cleaned using an equal volume (50 ml) of Agencourt AMPureXP magnetic beads and re-suspended in 20l of elution buffer. Libraries were checked and quantified on an Agilent Bioanalyzer high-sensitivity DNA chip (size range 150-2000 bp) and by Qubit dsDNA BR (Molecular Probes). Libraries were pooled to a normalized concentration of 1.5 nM and sequenced on an Illumina™ NextSeq 500 using the 150 bp paired-end kit as per the manufacturer's instructions.

Preprocessing and quality control of Smart-Seq Single-Cell RNA Sequencing Data

Fastq files with paired-end reads were trimmed by Cutadapt (version 1.7) to remove read adapters or low-quality tail bases and then aligned to the ENSEMBL GRCh38 (release 90, <https://www.ensembl.org/>) human transcriptome by using Bowtie2 (version 2.3.4.1)³⁵ with options “--sensitive -mp 1,1 -np 1 -score-min L,0,-0.1 -l 1 -X 2000 -no-mixed -no-discordant -N 1 -L 25 -k 200,” which allows 1 mismatch during sequence alignment. Quality control of the reads of each cell was assessed by using FastQC (version 0.11.6). Gene expression level was quantified by using raw read count number and Transcript per million (TPM) values generated by RSEM (version 1.3.0)³⁶. For the quality control of the genes and cells, genes that were detected in less than 5 cells and cells that expressed with either fewer than 3000 genes or more than 9000 genes, or greater than 20% mitochondrial genes were excluded from the expression matrix. Finally, our study included 2,985 high-quality cells for downstream analysis. The distributions of uniquely

mapped reads count and detected gene number in the cells are shown in Supplementary Data 1 and Supplementary Fig. 2A-C.

Dimensionality reduction and cell clustering

The R package Seurat (version 3.1.4)²³ implemented in R (version 3.6.2) were used to perform dimensional reduction analysis on iPSC-derived ENCCs RNA-seq data. The “NormalizeData()” function from Seurat was used to normalize the raw counts, and the scale factor was set to 200,000, then followed by “FindVariableFeatures()” with default parameters to calculate highly variable genes for each sample. After performing “JackStraw()”, which returned the statistical significance of PCA (Principal component analysis) scores, we selected 15 significant PCs to conduct dimension reduction and cell clustering. Then, cells were projected in 2D space using *t*-SNE (*t*-distributed Stochastic Neighbor Embedding) with default parameters. Clustering on individual iPSC-derived ENCC used different resolutions and was finally integrated into 5 main clusters by the expression of canonical markers.

Single cell data integration and batch effect correction

To account for batch effect among different samples, we used “FindIntegrationAnchors()” in the Seurat package to remove batch effect and merge samples to one object. In detail, the top 4,500 genes with the highest expression and dispersion from each sample were used to find the integration anchors, and then the computed anchor set was applied to perform dataset integration.

Identification of differentially expressed genes and pathway enrichment analysis

To identify unique differentially expressed genes (DEGs) among each cluster, the “FindAllMarkers()” function from Seurat was used and non-parametric Wilcoxon rank sum tests were set to evaluate the significance of each individual DEG. The DEG analysis between HSCR and control-iPSC-derived ENCCs based on single-cell expression data was performed using monocle (version 2.14.0) R package³⁷. The DEGs with adjusted *P*-value less than 0.05 and Log₂ Fold Change (log₂FC) larger than 0.5 were thought to be significant and used in downstream analysis. Gene ontology (GO) term enrichment analyses were performed using clusterProfiler (version 3.14.3) R package³⁸. Terms that had an adjusted *P*-value < 0.05 was defined as significantly enriched.

Pathway score analysis

In order to measure the activity of whole pathways between samples at transcriptome level, we applied an simple additive model that ignore the interactions between genes to estimate the overall expression level of the pathways. The genes involved in neurogenesis (GO:0050771), proliferation (GO:0051726), RNA splicing (GO:0033120) and cellular respiration (GO:1901857) in GO database were used to evaluate the overall changes of the key biological processes.

Histone deacetylase activity inference

Key components of Sin3, NuRD and CoREST complexes were obtained from public database. A geometric average model was applied to estimate the overall activity of histone deacetylase complexes based on the average expression of their subunits. Geometric average was used to set the complexes activity to zero if any of the subunit was not expressed in single cells.

Transcriptomic timing analysis

A Bayesian linear regression model was applied to detect “switch-like” upregulation or downregulation of genes along the pseudotime axis. Crucially, the model probabilistically assigns a region along the axis associated with the positive or negative activation of each of each gene in the core gene set.

Gene regulatory network (GRN) analysis

Single cell regulatory network inference and clustering (SCENIC, version 1.1.2)²⁶ was used to infer transcription factor networks active using scRNA-Seq data. Analysis was performed using default and recommended parameters as directed on the SCENIC vignette (<https://github.com/aertslab/SCENIC>) using the hg19 RcisTarget database. Kernel density line histograms showing differential AUC score distribution across conditions were plotted with ggplot2 v.3.1.1 using the regulon activity matrix ('3.4_regulonAUC.Rds', an output of the SCENIC workflow) in which columns represent cells and rows the AUC regulon activity. Fold-change (FC) difference between median AUC values was calculated and the highest changed TFs were plotted.

Differential co-expression analysis

To predict HDAC1 activated and repressed target genes, we applied a three-step strategy to obtain the target genes of HDAC1 during HSCR pathogenesis. In the first step, we identified the significantly correlated genes to HDAC1 which represented the genes that had strong regulatory relationships with HDAC1. In the second step, we utilized the differential expression profile between control and HSCR-ENCCs to filter out the genes whose expression levels were not significantly disrupted by HDAC1. Finally, to confine the genes to be the actual targets of HDAC1, a public ChIP-seq dataset designed for HDAC1 from a similar cell state (neuronal progenitor) in Cistrome database³⁹ was utilized to obtain the direct/indirect binding genes of HDAC1. Activated and repressed target genes were determined by the positive/negative correlation and up/down-regulation of the differential expression, respectively. In addition, TF motif enrichment analysis by FIMO tool⁴⁰ of MEME suite⁴¹ was used to cross-validate the predicted activated targets of HDAC1. Similar method was applied to identify the potential target genes/exons of PTBP1 exon9 based on RNA binding motif analysis. Firstly, AS events that significantly correlated with PTBP1 exon9 were selected. Secondly, exons targeted by PTBP1 through motif recognition were kept. Finally, we focused on the AS events that significantly associated with HSCR disease.

Alternative splicing analysis

SUPPA (version 2.3)⁴² was used to perform the alternative splicing analysis. Firstly, transcript events and local alternative splicing (AS) events were generated from ENSEMBL GRCh38 genome annotation (gtf) file. Secondly, the transcript and local AS event inclusion levels (PSI, Percent-Spliced-In) were quantified from multiple samples. Lastly, differential splicing for AS events and differential transcript usage between HSCR and control iPSC-derived ENCCs were calculated based on single-cell level PSI value. Clustering of cells based on PSI value was also performed by Seurat package.

Protein-protein interaction (PPI) network analysis

The PPI among core gene set and HSCR associated DEGs were obtained from STRING (v11)²⁵. We included only the interactions based on manual curation or experimental evidence with a combined score >0.4. Cytoscape⁴³ was used to plot the network.

Statistical analysis

The differences among multiple treatment groups were analyzed with a two-sided unpaired Student's t test or one-way analysis of variance followed by Tukey post-test using GraphPad Prism 7 (GraphPad Software). A P value less than 0.05 was interpreted to represent a statistically significant difference. All experiments were replicated at least three times, and data are shown as means with SEM. No specific randomization or blinding protocols were used.

Reporting summary

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

Data availability

Raw sequencing data are available in the Sequence Read Archive (SRA) at the NCBI Center with the accession number PRJNA784249. The processed scRNA-seq data sets are available at <https://doi.org/10.5281/zenodo.6104610>. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file. Source data are provided with this paper.

Abbreviations

HSCR

Hirschsprung

iPSCs

induced pluripotent stem cells

ENCC

enteric neural crest cell

scRNA-seq
single-cell RNA sequencing
AS
alternative splicing
FAO
fatty acid oxidation.

Declarations

Grant Support: This work described in this paper was substantially supported by the HMRF grants (Project no.: 06173306, 08192786) from the Health Department, and the General Research Fund (GRF: HKU 17108019, 17101320) from the Research Grant Council of Hong Kong Special Administrative Region, China Hong Kong and a seed funding from Li Dak Sum Research Centre, University of Hong Kong to E.S.W.N., and a research grant from the Innovative Technology Fund (UIM/299).

Disclosures: The authors declare no conflicts of interest.

Author contributions: K.N.C. L., S S.T.L., F.P.L.L. and P.L. established iPSC lines and performed *in vitro* functional assays. Z.L. and P.C.S. performed and supervised the bioinformatics analyses. E.S.W.N., C.C.H. and P.C.S. supervised the project and prepared the manuscript. P.K.H.T., M.Y., K.K.Y.W., and P.H.Y.C. contributed to the recruitment of patients and provided clinical information. M.M.G.B. performed the genetic characterization of the recruited patients and provided genetic data.

Acknowledgements

Confocal imaging and RNA sequencing were performed using equipment maintained by the Li Ka Shing Faculty of Medicine Faculty Core and the Centre for PanorOmic Sciences, the University of Hong Kong, respectively. This work described in this paper was substantially supported by the HMRF grants (Project no.: 06173306, 08192786) from the Health Department, and the General Research Fund (GRF: HKU 17108019, 17101320) from the Research Grant Council of Hong Kong Special Administrative Region, China Hong Kong and a seed funding from Li Dak Sum Research Centre, University of Hong Kong to E.S.W.N., and a research grant from the Innovative Technology Fund (UIM/299).

Author information

These authors contributed equally: Zhixin LI, Kathy Nga-Chu LUI, Sin-Ting LAU.

The authors declare no conflicts of interest.

References

1 Schuchardt, A., D'Agati, V., Larsson-Blomberg, L., Costantini, F. & Pachnis, V. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. *Nature* **367**, 380-383 (1994).

- 2 Emison, E. S. *et al.* Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. *Am J Hum Genet* **87**, 60-74, doi:10.1016/j.ajhg.2010.06.007 (2010).
- 3 Tang, C. S. *et al.* Identification of Genes Associated With Hirschsprung Disease, Based on Whole-Genome Sequence Analysis, and Potential Effects on Enteric Nervous System Development. *Gastroenterology* **155**, 1908-1922 e1905, doi:10.1053/j.gastro.2018.09.012 (2018).
- 4 Garcia-Barcelo, M. M. *et al.* Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc Natl Acad Sci U S A* **106**, 2694-2699, doi:10.1073/pnas.0809630105 (2009).
- 5 Le, T. L. *et al.* Dysregulation of the NRG1/ERBB pathway causes a developmental disorder with gastrointestinal dysmotility in humans. *J Clin Invest* **131**, doi:10.1172/jci145837 (2021).
- 6 Jiang, Q. *et al.* Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with ret are critical to Hirschsprung disease liability. *Am J Hum Genet* **96**, 581-596, doi:10.1016/j.ajhg.2015.02.014 (2015).
- 7 Lai, F. P., Lau, S.T., Wong, J.W.L., Gui, H., Wang, X.R., Zhou, T., Lai, W.H., Tse, H.F., Tam, P.K., Garcia-Barcelo, M.M., Ngan, E.S.W. Correction of Hirschsprung-associated mutations in human induced pluripotent stem cells, via CRISPR/Cas9, restores neural crest cell function. *Gastroenterology* **153**, 139-153, doi:10.1053/j.gastro.2017.03.014 (2017).
- 8 Lau, S. T. *et al.* Activation of Hedgehog Signaling Promotes Development of Mouse and Human Enteric Neural Crest Cells, Based on Single-Cell Transcriptome Analyses. *Gastroenterology* **157**, 1556-1571 e1555, doi:10.1053/j.gastro.2019.08.019 (2019).
- 9 Barber, K., Studer, L. & Fattahi, F. Derivation of enteric neuron lineages from human pluripotent stem cells. *Nat Protoc* **14**, 1261-1279, doi:10.1038/s41596-019-0141-y (2019).
- 10 Lee, G., Chambers, S. M., Tomishima, M. J. & Studer, L. Derivation of neural crest cells from human pluripotent stem cells. *Nat Protoc* **5**, 688-701, doi:10.1038/nprot.2010.35 (2010).
- 11 Bhattacharya, D., Azambuja, A. P. & Simoes-Costa, M. Metabolic Reprogramming Promotes Neural Crest Migration via Yap/Tead Signaling. *Dev Cell* **53**, 199-211 e196, doi:10.1016/j.devcel.2020.03.005 (2020).
- 12 Varum, S. *et al.* Yin Yang 1 Orchestrates a Metabolic Program Required for Both Neural Crest Development and Melanoma Formation. *Cell Stem Cell* **24**, 637-653 e639, doi:10.1016/j.stem.2019.03.011 (2019).
- 13 Teperino, R. *et al.* Hedgehog partial agonism drives Warburg-like metabolism in muscle and brown fat. *Cell* **151**, 414-426, doi:10.1016/j.cell.2012.09.021 (2012).

- 14 Liu, J. A. *et al.* Identification of GLI Mutations in Patients with Hirschsprung Disease That Disrupt Enteric Nervous System Development in Mice. *Gastroenterology*, doi:10.1053/j.gastro.2015.07.060 (2015).
- 15 Mao, J., Kim, B. M., Rajurkar, M., Shivdasani, R. A. & McMahon, A. P. Hedgehog signaling controls mesenchymal growth in the developing mammalian digestive tract. *Development (Cambridge, England)* **137**, 1721-1729, doi:10.1242/dev.044586 (2010).
- 16 Martinelli, D. C. & Fan, C. M. Gas1 extends the range of Hedgehog action by facilitating its signaling. *Genes Dev* **21**, 1231-1243, doi:10.1101/gad.1546307 (2007).
- 17 Ngan, E. S. *et al.* Hedgehog/Notch-induced premature gliogenesis represents a new disease mechanism for Hirschsprung disease in mice and humans. *J Clin Invest* **121**, 3467-3478, doi:10.1172/JCI43737 (2011).
- 18 Ramalho-Santos, M., Melton, D. A. & McMahon, A. P. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development (Cambridge, England)* **127**, 2763-2772 (2000).
- 19 Lai, F. P.-L. *et al.* Ciliary protein Kif7 regulates Gli and Ezh2 for initiating the neuronal differentiation of enteric neural crest cells during development. *Science Advances* **7**, eabf7472, doi:doi:10.1126/sciadv.abf7472 (2021).
- 20 Lai, F. P. *et al.* Ciliary protein Kif7 regulates Gli and Ezh2 for initiating the neuronal differentiation of enteric neural crest cells during development. *Sci Adv* **7**, eabf7472, doi:10.1126/sciadv.abf7472 (2021).
- 21 Lasrado, R. *et al.* Lineage-dependent spatial and functional organization of the mammalian enteric nervous system. *Science* **356**, 722-726, doi:10.1126/science.aam7511 (2017).
- 22 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).
- 23 Hao, Y. H. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573+, doi:10.1016/j.cell.2021.04.048 (2021).
- 24 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-018-0414-6 (2018).
- 25 Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**, D607-D613, doi:10.1093/nar/gky1131 (2019).
- 26 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083+, doi:10.1038/nmeth.4463 (2017).

- 27 Linares, A. J. *et al.* The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation. *Elife* **4**, e09268, doi:10.7554/eLife.09268 (2015).
- 28 Gueroussov, S. *et al.* An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349**, 868-873, doi:10.1126/science.aaa8381 (2015).
- 29 Zhu, W. *et al.* Roles of PTBP1 in alternative splicing, glycolysis, and oncogenesis. *J Zhejiang Univ Sci B* **21**, 122-136, doi:10.1631/jzus.B1900422 (2020).
- 30 Zheng, X. *et al.* Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation. *Elife* **5**, e13374, doi:10.7554/eLife.13374 (2016).
- 31 Shyh-Chang, N. & Ng, H. H. The metabolic programming of stem cells. *Genes Dev* **31**, 336-346, doi:10.1101/gad.293167.116 (2017).
- 32 Tarazona, O. A. & Pourquie, O. Exploring the Influence of Cell Metabolism on Cell Fate through Protein Post-translational Modifications. *Dev Cell* **54**, 282-292, doi:10.1016/j.devcel.2020.06.035 (2020).
- 33 Xue, Y. *et al.* Generating a non-integrating human induced pluripotent stem cell bank from urine-derived cells. *PLoS One* **8**, e70573, doi:10.1371/journal.pone.0070573 (2013).
- 34 Takahashi, K., Okita, K., Nakagawa, M. & Yamanaka, S. Induction of pluripotent stem cells from fibroblast cultures. *Nat Protoc* **2**, 3081-3089, doi:nprot.2007.418 [pii] 10.1038/nprot.2007.418 (2007).
- 35 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-U354, doi:10.1038/nmeth.1923 (2012).
- 36 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
- 37 Qiu, X. J. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979+, doi:10.1038/nmeth.4402 (2017).
- 38 Yu, G. C., Wang, L. G., Han, Y. Y. & He, Q. Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omicron-a Journal of Integrative Biology* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
- 39 Liu, T. *et al.* Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biology* **12**, R83, doi:10.1186/gb-2011-12-8-r83 (2011).
- 40 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).

- 41 Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Research* **43**, W39-W49, doi:10.1093/nar/gkv416 (2015).
- 42 Trincado, J. L. *et al.* SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biology* **19**, 40, doi:10.1186/s13059-018-1417-1 (2018).
- 43 Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

Figures

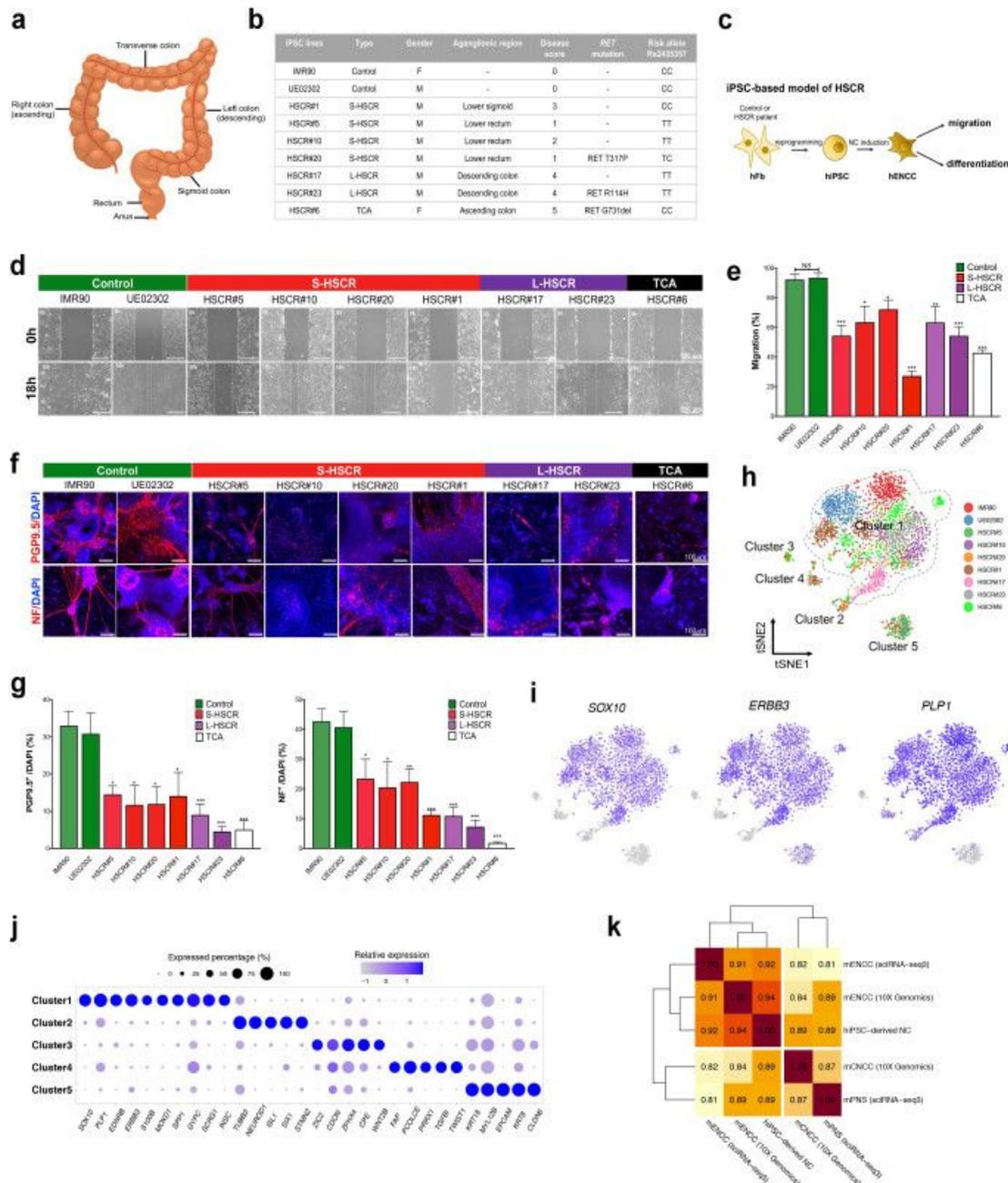


Fig. 1 Functional and molecular characterization of iPSC-based model of HSCR. (a) Anatomy of the human colon. (b) An overview of healthy and HSCR-iPSC lines used in this study. (c) Schematic shows the recapitulation of disease phenotypes using HSCR-iPSC lines. (d) *In vitro* scratch assay. (e) Bar chart shows the quantitative data of migration. (f) *In vitro* differentiation assay with two neuronal markers (PGP9.5 and NF) and (g) the corresponding quantitative analyses. (h) *t*-SNE projection of all 3,342 individual cells, colored by iPSC lines. 5 main clusters are labeled and Cluster 1 is marked with the dotted line. (i) Canonical markers expressed in Cluster 1. (j) Dot-plot shows top key markers expressed in each cluster. Color of dot indicates the relative expression and size of dot indicates the expressed percentage. (k) Correlation heatmap shows the similarities between human iPSC-derived NC cells and mouse *in vivo* enteric/non-enteric NC cells at E13.5 from two independent datasets. Sequencing platforms of datasets are noted.

Figure 1

See image above for figure legend.

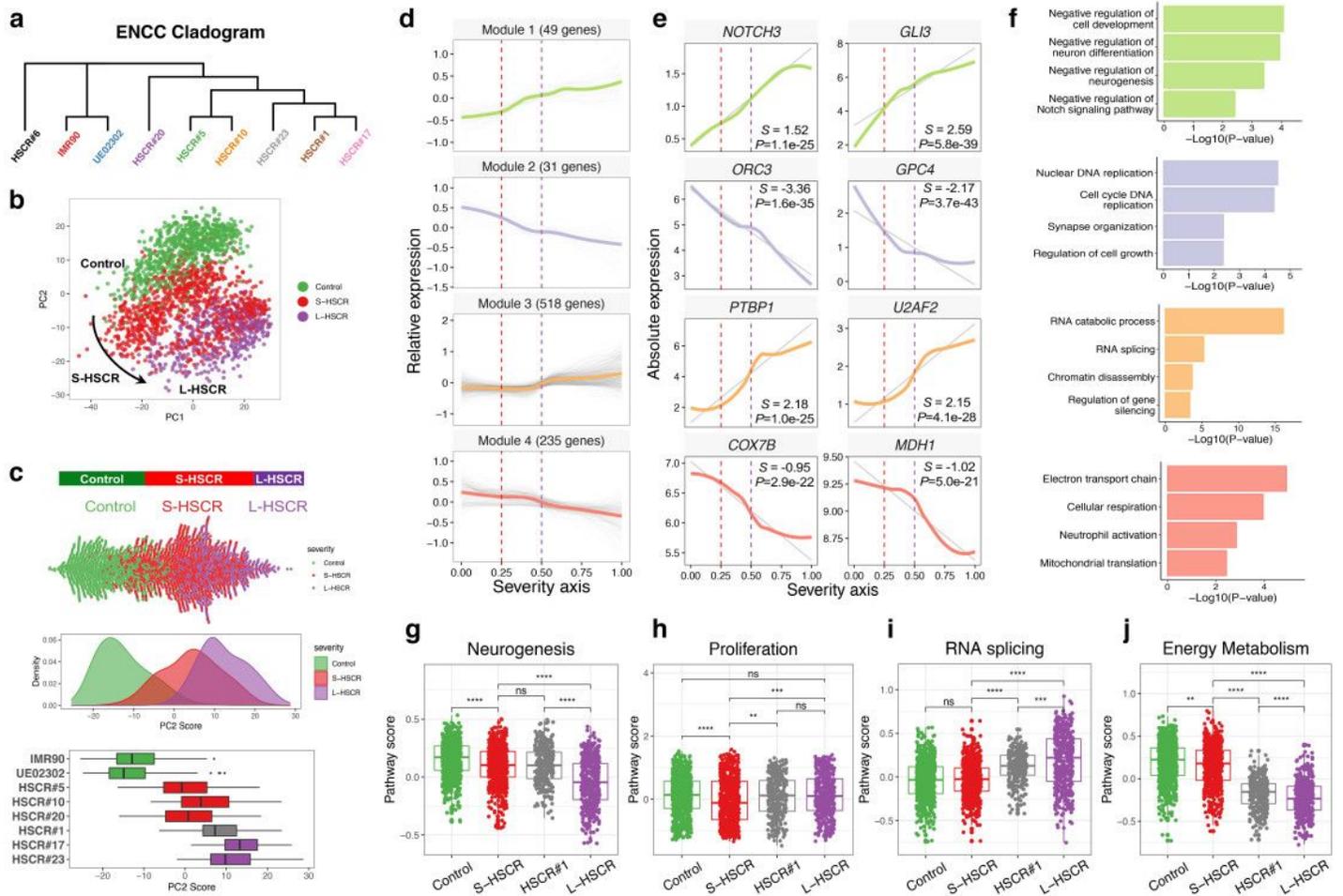


Fig. 2 Reconstruction of disease severity axis integrating different HSCR states. (a) Hierarchical clustering by average gene expression. **(b)** PCA projection of cells in Cluster 1 based on gene expression profile after excluding TCA HSCR-iPSC line (HSCR#6). **(c)** Distribution of cells along PC2 is highly associated with disease severity as shown in beeswarm, density and box plots. Colored by type of HSCR. **(d)** The expression dynamics of 833 top DEGs were cataloged into 4 major modules, colored by modules. Thick lines indicate the average gene expression patterns of each module. **(e)** Expression dynamics of representative genes in each gene module are shown as a LOESS smooth fit line (colored line) and the best fit lines (gray lines) in linear regression. S : slope, P : P -value of linear regression. **(f)** Gene ontology analyses of each gene module. **(g)-(j)** Overall pathway scores of neurogenesis, proliferation, RNA splicing and energy metabolism.

Figure 2

See image above for figure legend.

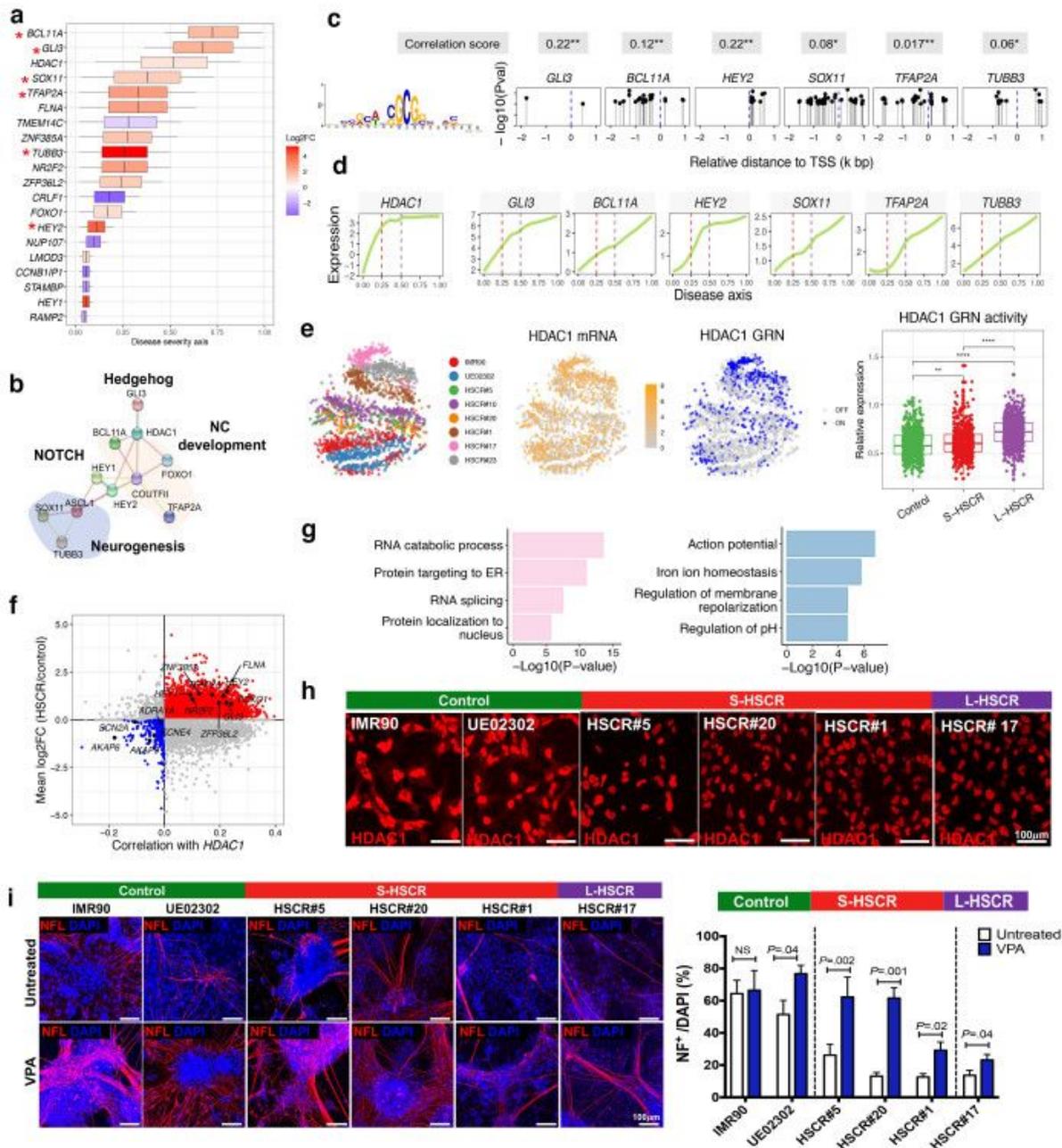


Fig. 3 Identification of a functionally enriched gene set and HDAC1 as a key regulator along HSCR severity axis. (a) Anchoring of the functionally enriched gene set to disease severity axis, colored by Log₂FC and HDAC1 targets are marked by asterisk. **(b)** Protein-protein interaction network of the core gene set. **(c)** Motif enrichment of HDAC1 at the promoter region (-2k bp to 1k bp relative to TSS) of its target genes. Expression correlations between HDAC1 and its targets are shown at the top panel. **(d)** Dynamic expression of HDAC1 and its activated target genes along disease severity axis. **(e)** *t*-SNE projection of cells and GRN analysis of HDAC1. mRNA (yellow) and binary regulator activity (active (blue) or inactive (grey)) inferred by SCENIC are shown. Overall activity of HDAC1 targeted GRN between control and S/L-HSCR is shown in boxplot. **(f)** Scatter plot shows putative HDAC1 activated and repressed targeted genes. Key genes are labeled. **(g)** Gene ontology analyses of HDAC1 activated and repressed targeted genes. **(h)** Immunocytochemistry shows the reduced cytoplasmic HDAC1 in HSCR hENCCs. **(i)** *In vitro* differentiation assays in absence or presence of HDAC inhibitor (Valproic acid, VPA, 0.4mM). Chart shows the quantitative data.

Figure 3

See image above for figure legend.

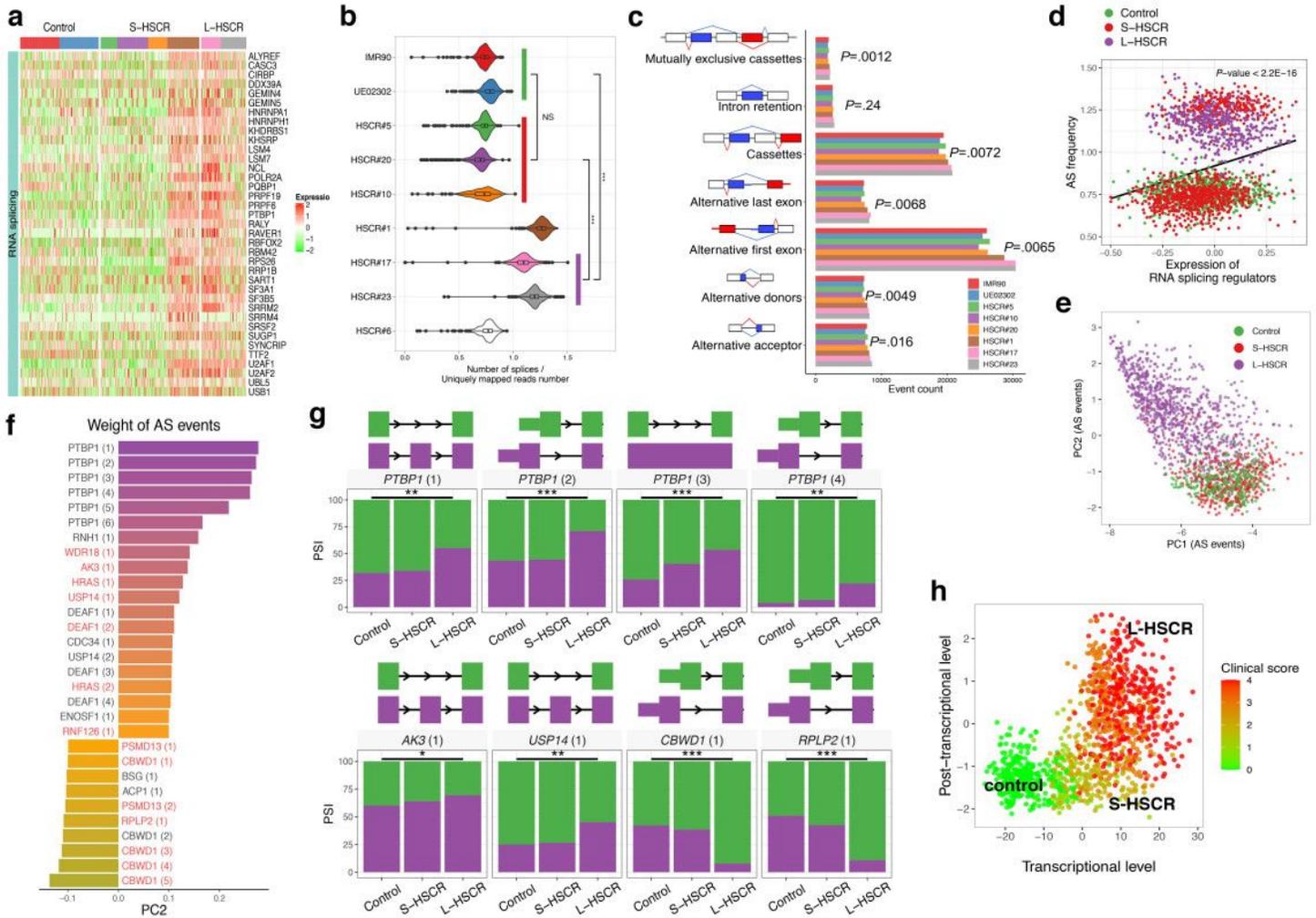


Fig. 4 Global analysis of RNA alternative splicing (AS) events in control- and HSCR-ENCCs. (a) Heatmap shows the top 37 up-regulated AS mediator genes enriched in HSCR#1 and L-HSCR. **(b)** Normalized number of splices in control- and HSCR-ENCCs. *P*-values (two-sided Mann-Whitney U test) from comparing ENCCs derived from the two controls and various HSCR lines are shown. The splicing frequency of each cell was estimated by the number of splices divided by uniquely mapped read number. **(c)** Number of different types of RNA splicing events in control- and HSCR-ENCCs. **(d)** The 37 RNA-splicing regulator genes are significantly up-regulated and positively correlated with AS frequency in HSCR#1- and two L-HSCR-ENCCs. **(e)** PCA projection of single cells based on AS events reveals that PC2 is highly associated with disease severity. **(f)** Top AS events positively and negatively contributed to PC2. Putative targets of PTBP1 (1) (Inclusion of exon 9) are labelled in red. **(g)** Significant differentially splicing events between control and S/L-HSCR. *P*-values <0.05 and <0.001 are marked by ** and ***. **(h)** Transcriptional and post-transcriptional levels represent two complementary dimensions describing the progression of HSCR disease.

Figure 4

See image above for figure legend.

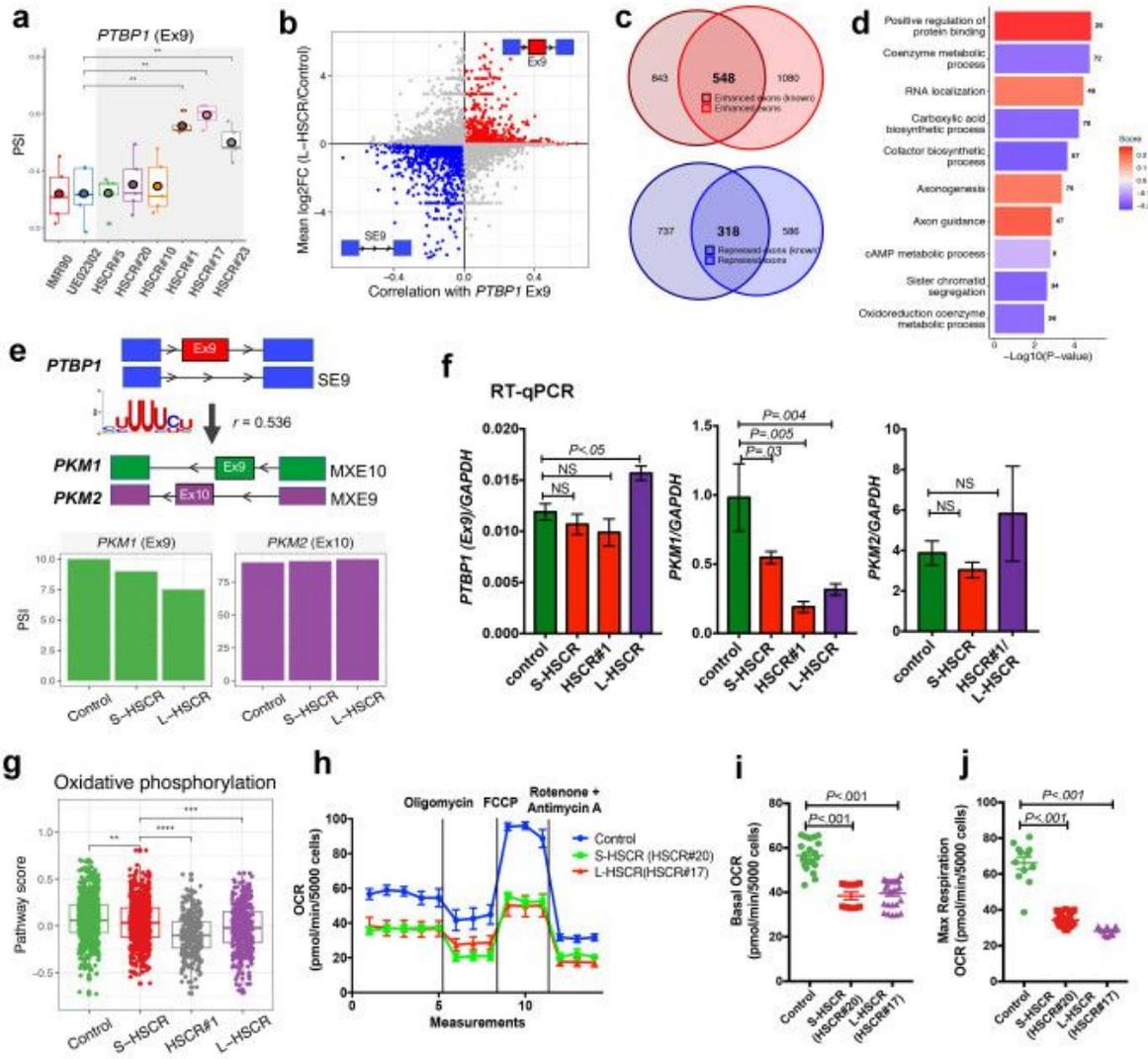


Fig. 5 Inclusion of *PTBP1* exon 9 is associated with dysregulation of cellular metabolic pathways in L-HSCR-ENCCs. (a) Boxplot shows Percent Spliced In (PSI) of exon9 (Ex9) inclusion rate in *PTBP1*. (b) Scatterplot shows the enhanced and repressed exons associated with *PTBP1* Ex9 enriched in L-HSCR. SE9: skipped exon 9. (c) Venn plots shows the *PTBP1* Ex9 -enhanced and -repressed exons enriched in L-HSCR overlapping with the known *PTBP1* target exons. (d) Gene ontology enrichment analyses of genes enriched with target exons of *PTBP1*-Ex9 in L-HSCR. (e) Schematic shows the regulation of *PKM1/2* splicing regulated by *PTBP1*-Ex9. SE: skipped exon; MXE: mutually exclusion of exon. Differential splicing of *PKM1* and *PKM2* between control, S-HSCR and L-HSCR. (f) RT-qPCR analyses of *PTBP1*, *PKM1* and *PKM2* expressions in ENCCs derived from the control- and various HSCR-iPSC lines. (g) Boxplot shows the reduced oxidative phosphorylation pathway scores in HSCR-ENCCs as inferred based on the expression of the genes implicated in these pathways. (h) Seahorse assays show reduced oxygen consumption rate (OCR) in HSCR-ENCCs. Quantitative analyses of the (i) basal OCR and (j) maximum respiratory rate in control and HSCR-ENCCs.

Figure 5

See image above for figure legend.

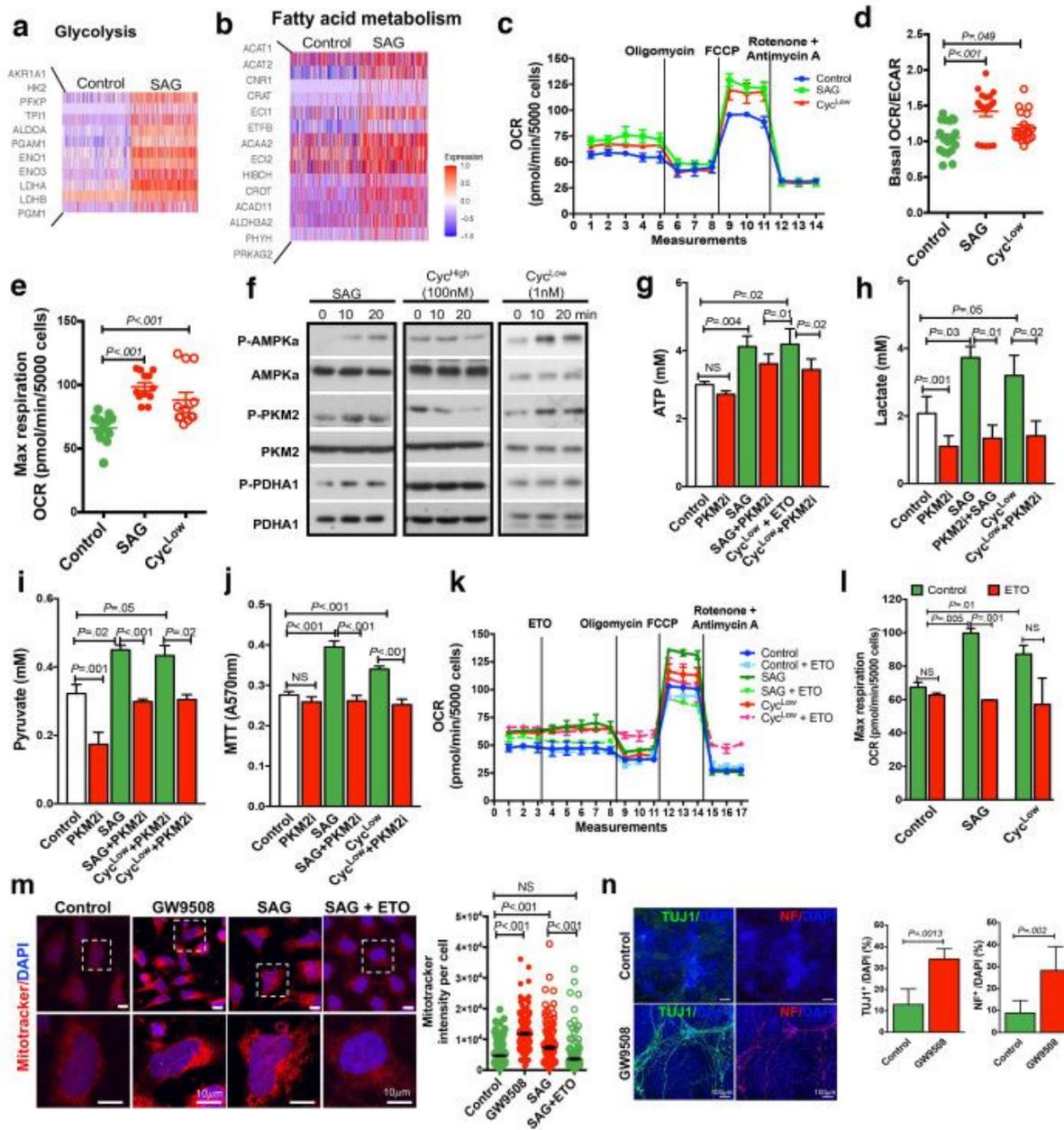


Fig. 6 Hedgehog-induced metabolic shift in ENCCs. Heatmaps show genes implicated in (a) glycolysis and (b) fatty acid metabolism pathways in ENCCs upon the SAG treatment. (c) Seahorse assays of control-ENCCs treated with SAG or cyclopamine (1nM, Cyc^{low}). Dot plots show the changes in (d) the basal oxygen consumption rate (OCR) to extracellular acidification rate (ECAR) ratios and (e) the maximum respiration rates upon different treatments. (f) Western blot shows the alternations of AMPK-PKM2-PDHA1 pathway in ENCCs subjected to SAG, cyclopamine (100nM, Cyc^{high}) & 1nM, Cyc^{low}) treatments. Bar charts show (g) ATP, (h) lactate, (i) pyruvate and (j) MTT assays in presence or absence of SAG or Cyc^{low}. (k) Seahorse assays of ENCCs treated with SAG and Cyc^{low} in absence or presence of FAO inhibitor (Etomoxir, ETO). (l) Bar chart shows the maximum respiration rates. (m) Mitochondria activities in ENCCs were monitored using the Mitotracker assay. (n) Immunocytochemistry of TUJ1⁺ and NF⁺ neurons in day 5 of neuronal differentiation. The quantitative data are shown in the bar charts.

Figure 6

See image above for figure legend.

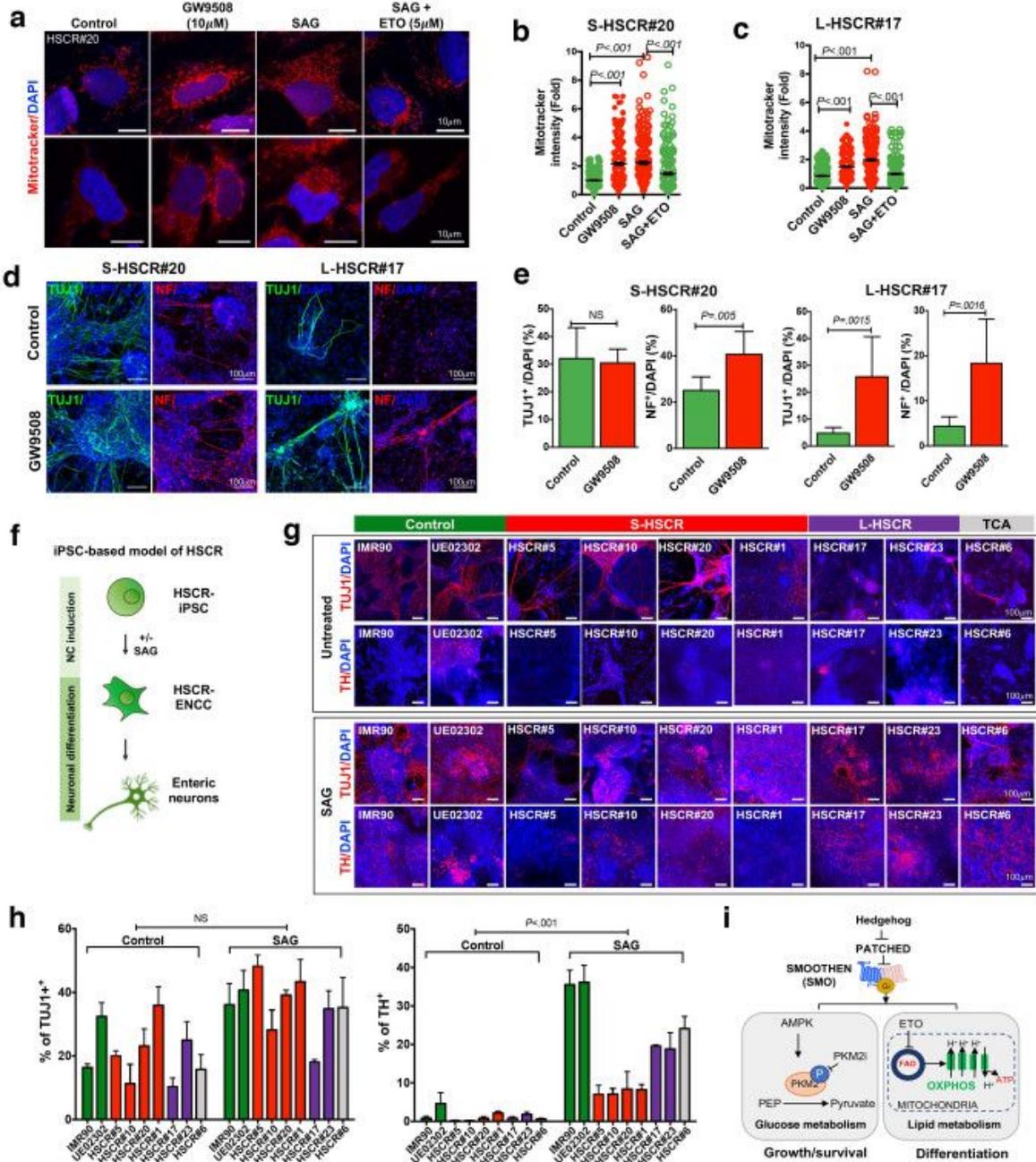


Fig. 7 Hedgehog-induced OXPPOS recues the differentiation defect of HSCR-ENCCs. (a) Mitochondria activities in HSCR-ENCCs treated with FAO agonist (GW9508), SAG and SAG in combination of ETO were monitored using the Mitotracker assay. The corresponding quantitative analyses of mitochondria activities in HSCR#20- and HSCR#17-ENCCs are shown in (b) and (c). (d) Immunocytochemistry of TUJ1⁺ and NF⁺ neurons in day 5 of neuronal differentiation. (e) The quantitative data of the *in vitro* differentiation assays are shown in the bar charts. (f) Schematic shows the differentiation strategy for the generation of HSCR-ENCCs. (g) Immunocytochemistry of TUJ1⁺ and TH⁺ neurons in day 9 of the neuronal differentiation of control and HSCR-ENCC with or without pretreatment with SAG. (h) The bar charts show the quantitative data. (i) Schematic shows the Hedgehog-mediated glycolysis and FAO regulate the survival and differentiation of ENCCs.

Figure 7

See image above for figure legend.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DescriptionofAdditionalSupplementaryFiles.docx](#)
- [SupplementaryTable13.reagentsscRNAseq.docx](#)
- [FigS17.pdf](#)
- [SupplementaryData1.xlsx](#)
- [SupplementaryData2.xlsx](#)
- [SupplementaryData3.xlsx](#)
- [SupplementaryData4.xlsx](#)
- [SupplementaryData5.xlsx](#)
- [EllyNganDataAccess.docx](#)