CRISPR-Cas9 mediated correction of Hirschsprung-associated mutations in human iPSC restores neural crest cell functions

Frank Pui-Ling LAI[1,aξ], Sin-Ting LAU[1,aξ], Reeson Xu WANG[1,b], John Kwong-Leong WONG[2,b], Tingwen ZHOU[1,b], Hongsheng GUI[2,b], Wing Hon LAI[3,a], Hung-Fat TSE[3,c], Paul Kwong-Hang TAM[1,c], Maria-Mercedes GARCIA-BARCELO[1,d], Elly Sau-Wai NGAN[1,d]*

[1]Department of Surgery, [2]Department of Psychiatry, [3]Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong.

ξ Authors contribute equally

[a] Postdoctoral Fellow
[b] Postgraduate Research Student
[c] Clinical Chair Professor/Professor
[d] Non-clinical Associate Professor

*_Corresponding author_:
Dr. Elly SW Ngan
Associate Professor
Department of Surgery
The University of Hong Kong,
Faculty of Medicine Building,
21 Sassoon Road, Pokfulam, Hong Kong, SAR, China.
Tel: (852) 2831-5410
Fax: (852) 3917-9621
Email: engan@hku.hk

Short Title: iPSC for HSCR modeling

Abbreviations: hiPSC: human induced pluripotent stem cells; CRISPR: clustered regularly-interspaced short palindromic repeats

Grant Support: This work described in this paper was substantially supported by a HMRF grant (Project no.: 01121476) to E.S.W.N., and research grants from the Research Grants Council of Hong Kong Special Administrative Region, China Hong Kong (HKU17119514, T12C-714/14-R).

Disclosures: The authors declare no conflicts of interest.

Author contributions: S.T.L., K.L., H.F.T. and P.K.H.T. generated patient iPSC. F.P.L.L. performed genome-editing on hiPSC. S.T.L. and F.P.L.L. analyzed patient

iPSC and hPSC. R.X.W and T.Z. studied the functional defects of mutant VCL. J.K.L.W. and M.G. performed genetic and bioinformatic analysis. P.K.H.T. provided patient samples and clinical data of the patients. E.S.W.N. supervised the whole project and prepared the manuscript.

Abstract:

Background and aim: Hirschsprung (HSCR) disease is caused by the failure of enteric neural crest cells (ENCCs) to fully colonize the bowel, thus leading to bowel obstruction and megacolon. Heterozygous mutations in the coding region of the *RET* gene account for the severe form of HSCR (total colonic aganglionosis, TCA). However, 80% of HSCR patients have short segment HSCR (S-HSCR), whose genetic etiology remains unclear. Here, we aimed to establish a new strategy for the functional identification of novel mutation(s) in HSCR disease. ***Methods:*** Three HSCR-iPSC lines (from 1 TCA-HSCR patient with a *RET G731del* mutation and 2 S-HSCR patients without a *RET* mutation) and two *RET* isogenic mutant iPSC lines ($RET^{+/-}$ and $RET^{-/-}$) were established to generate disease-relevant ENCCs. The migration and differentiation capacities of the iPSC-derived ENCCs were analyzed with various *in vitro* differentiation and migration assays. Disease-associated mutation(s) were identified by combining genetic and transcriptomic data obtained from patient blood and iPSC-derived ENCCs, respectively. Mutations in HSCR-iPSCs were corrected with CRISPR/Cas9. ***Results:*** ENCCs derived from all diseased iPSC lines exhibited defects in migration and neuronal lineage differentiation. *RET* mutations were functionally associated with cellular phenotypes of ENCCs *in vitro*. Integration of genetic and transcriptomic data further identified a novel mutation in *Vinculin* (*VCL M209L*) associated with S-HSCR. Correcting the *RET G731del* and *VCL M209L* mutations in HSCR-iPSCs restored the ENCC functions. ***Conclusion:*** We identified *VCL* as a novel HSCR gene. Importantly, our results also illustrate the value of human iPSCs for defining mutations that functionally contribute to a complex disease. ***Key words:*** PSC-based model, RET, Vinculin

Introduction:

Hirschsprung (HSCR) disease, the most common neurocristopathy, affects ~1 to 1.3 in 5,000 newborns. It is a complex congenital disorder of the colon, in which certain nerve cells are absent, owing to incomplete colonization of the bowel by enteric neural crest cells (ENCCs), thus causing functional intestinal obstruction. Phenotypic severity is classified according to the length of the aganglionic segment, referred to as short segment (S-HSCR: 80% of the patients), long segment (L-HSCR: 15% of the patients) and total colonic aganglionosis (TCA: 5% of the patients)[1]. HSCR disease is caused by disruptions in the developmental pathways by which ENCCs colonize the developing gut and give rise to various subtypes of neurons and glia to form a functional enteric nervous system (ENS).

The receptor tyrosine kinase RET and glial cell-derived neurotrophic factor (GDNF) constitute the important signaling pathway in ENS development[2-4]. GDNF (from the gut mesenchyme) binds to the RET receptor (expressed in ENCCs) and activates effector signaling molecules, which in turn mediate proliferation, migration and differentiation of enteric ENCCs. Mutations in the *GDNF* gene are very rare, whereas *RET* variants/mutations are associated with most HSCR cases. Heterozygous mutations in coding (CD) sequences of the *RET* gene account for only 5-20% of all HSCR cases and are primarily associated with TCA, whereas non-coding (NCD) variants/single nucleotide polymorphisms spanning the *RET* gene are frequently detected in S-HSCR cases. Both CD and NCD variants/mutations may lead to aberrant *RET* expression. In particular, most of the CD mutations found in the extracellular domain of RET disrupt the protein maturation of RET, thus leading to decreased *RET* expression[5] and suggesting that haploinsufficiency of *RET* may

represent a cause of HSCR. *Ret*$^{-/-}$ mutant mice exhibit complete intestinal aganglionosis, owing to a combination of cell death and failure to migrate, accompanied by kidney defects. Intriguingly, whereas most of the *RET* CD mutations found in patients are heterozygous, heterozygous deletion or mutations of *Ret* in mice are not sufficient to cause an HSCR-like phenotype[2-4]. Only when *Ret* expression is decreased to approximately one-third of its normal levels do *Ret* hypomorphic mice recapitulate the ENS phenotypes of HSCR patients, including colonic aganglionosis with incomplete penetrance, a tendency for males to be affected more frequently than females and an absence of kidney defects[6]. These data suggest that decreased *RET* expression is the underlying cause of HSCR. The biological implications of the NCD variants of *RET*, however, are less defined. The risk allele "T" at rs2435357 in the intronic enhancer of the *RET* gene is the strongest risk allele associated with S-HSCR. This genetic variant has been shown to decrease *RET* expression in a reporter assay[7]. Nevertheless, this risk allele "T" is frequently found in healthy individuals, thus suggesting that interplay between and/or accumulation of additional CD or NCD functional variants are likely involved in S-HSCR pathogenesis.

In this study, we firstly made use of three induced pluripotent stem cell (HSCR-iPSC) lines carrying *RET* CD mutation, a control iPSC line (IMR90) and a "*RET*-corrected" HSCR-iPSC line to perform a proof-of-concept study and to establish a causal relationship between the genetic mutation with cellular phenotypes "in the dish". We then generated additional HSCR-iPSC lines from two S-HSCR patients without *RET* CD mutation to further demonstrate how a human iPSC model can be used to generate disease-relevant cells for the identification of new susceptibility

genes as well as for a better understanding of disease etiology (Figure 1 and Supplementary Table 1).

Materials and Methods:

Human iPS cells

A control human iPS cell line (IMR90-iPSC) was obtained from WiCell Research Institute (WiCell, WI). Patient-iPS cell lines were generated from fibroblasts of HSCR patients by episomal reprogramming vectors or retroviruses carrying the four reprogramming factors and were maintained on Matrigel-coated plates in mTeSR1 medium. The study was approved by the institutional review board of the University of Hong Kong together with the Hospital Authority (UW 13-419).

Plasmid construction

The human codon-optimized Cas9 expression plasmid was obtained from Addgene (44720). Two guide RNA (gRNA) plasmids were prepared using Gibson assembly, and the corresponding sequences are listed in Supplementary Table 2. The donor plasmid containing the *VCL* gene spanning the region chr10:74072141-74078191, a 2229-bp neomycin selection cassette (loxP-neo-Tk) flanked by two loxP sites and an HSV-TK promoter was prepared. VCL expression constructs carrying the wild-type or HSCR-associated mutation were prepared.

Generation of RET mutant iPS cell lines

The CRISPR/Cas9D10A system[8] was used to target exon 10 or 12 of the *RET* locus in IMR90 human iPS cells. IMR90 iPS cells were transfected with gRNA constructs. The colonies were isolated and subjected to screening using PCR and Sanger

sequencing. The clones with bi-allelic nonsense mutations were expanded for follow-up assays.

Correction of the VCL mutation in HSCR-iPS cells

HSCR-iPSC#3 cells were transfected with the *VCL*-specific gRNA constructs, a Cas9-D10A nickase expression plasmid and a donor plasmid containing the wild-type *VCL* by using Nucleofector. The transfected cells were then plated on Matrigel-coated plate, and selection was initiated 24 h later with culture medium containing 100 ng/ml neomycin. Neomycin-resistant colonies were obtained in 2-3 weeks and isolated for screening.

Neural crest induction

Human PSCs were plated on Matrigel-coated plates (10^5 cells cm^{-2}) in ES cell medium containing 10 ng/ml FGF2 (PeproTech, 100-18B). Differentiation was initiated by replacing the ES medium with KSR medium and then gradually switching to N2 medium as described previously [9]. For vagal NCC induction, the cells were treated with LDN193189, SB431542, CHIR99021, and 1 μM retinoic acid. The differentiated cells were sorted for $p75^{NTR}$ and HNK1 or CD49d expression at day 10 as described [9-12]. CNS precursor control cells were generated by treatment with LDN193189 and SB431542 from day 0 to day 10 as previously described[13].

FACS and flow cytometry analysis

For flow cytometry analysis or cell sorting, the cells were dissociated with Accutase and labeled with anti-human antibodies. The labeled cells were detected using a FACSCalibur instrument. Isotype-matched antibodies were used as controls. FlowJo

version 8.2 (Tree Star, Inc.) was used to analyze the flow data. A list of primary antibodies and working dilutions is provided in Supplementary Table 3.

In vitro differentiation of ENCCs to enteric neurons

ENCCs ($3x10^4$) from the 10-day induction protocol were seeded as droplets on poly-ornithine/laminin/fibronectin (PO/LM/FN)-coated dishes in N2 medium containing 10 ng/ml FGF2 and 3μM CHIR99021. Neuronal differentiation was initiated by replacing the N2 medium with neuronal differentiation medium containing GDNF, ascorbic acid, brain-derived neurotrophic factor (BDNF), nerve growth factor (NGF), neurotrophin-3 (NT-3) and cyclic adenosine monophosphate (cAMP). The ENCCs differentiated into neurons in 1–2 weeks. The cells were then fixed for immunostaining.

In vitro differentiation of NCCs to mesenchymal lineages

To drive NCCs toward differentiation into the mesenchymal lineage, purified NCCs were incubated in αMEM containing 10% FBS for 2 weeks in uncoated tissue culture-grade dishes. hiPS cell–derived neural crest-mesenchymal precursor cells (hiPS cell–derived NCMP cells) were identified on the basis of the expression of CD73 and were expandable in αMEM containing 10% FBS. hiPS cell–derived NCMP cells were then subjected to mesenchymal lineage differentiation by culturing in Medium 231 containing a smooth muscle differentiation supplement or in Adipogenesis medium, Osteogenesis medium or Chondrogenesis medium for 2-4 weeks, according to the manufacturer's protocols.

Cell culture and transfection

The VCL expression construct was then transfected into HeLa cells using FuGene® HD Transfection Reagent. Cells overexpressing the wild-type or mutated GFP-VCL expression constructs were used for subsequent immunofluorescence and immunoprecipitation assays.

Immunofluorescence analysis

For immunofluorescence analysis, the cells were fixed, blocked and then incubated with a primary antibody listed in Supplementary Table 3. The cells were photographed using a Carl Zeiss confocal microscope (LSM 780). The quantitative image analysis of differentiated neuronal cultures was performed using ImageJ plugins. The percentage of neuronal cells was measured relative to the total number of cells (DAPI), and the values reported in bar charts represent the mean ± SEM.

Migration

In vitro migratory behaviors of ENCCs were measured using scratch assays.

Immunoblotting & immunoprecipitation

Total protein (20 μg) from cell lysates was separated on 8% SDS-polyacrylamide gels and blotted with specific antibodies listed in Supplementary Table 3. β-Actin expression was used as a protein-loading control.

To study the interaction of VCL and ACTIN, protein lysates were first immunoprecipitated with an anti-VCL antibody (Abcam, ab290) and protein A-agarose beads. The immunoprecipitates were separated by SDS-polyacrylamide gels and blotted with antibodies against ACTIN or VCL. All blots were incubated with a horseradish peroxidase-conjugated anti-mouse or anti-rabbit secondary antibody.

RNA and exome sequencing

RNA and exome sequencing were performed in the Centre of Genomic Science, The University of Hong Kong. The differential expression (DE) analyses were performed using DESeq2. The significance level was set at 0.05 for false discovery rate (FDR)-corrected p-values.

Network and functional enrichment analyses

A gene association network analysis was performed by GeneMANIA[14]. The resulting network bridged the genotyping and expression data. DAVID was used to detect enrichment of any particular similar type of genes[15].

Protein structure prediction

SWISS-MODEL[16] was used to predict the structure of the mutated VCL protein.

Experimental details are available in the supplementary methods.

Results:

Establishment of diseased and disease-corrected HSCR-iPSC lines

Because RET/GDNF signaling is crucial for ENCC development and *RET* mutations account for most cases of HSCR, here, we introduced a homozygous or heterozygous *RET* CD mutation in a control iPSC to test whether the HSCR-associated phenotypes and the disease severity can be captured in an iPSC-based model. We used the CRISPR/Cas9 system to specifically target the transmembrane or intracellular tyrosine kinase domain of RET and generated two isogenic iPSC lines

with compromised or reduced RET/GDNF signaling (Figure 2A). The heterozygous (IMR90$^{RET+/-}$) clone carries a 19 bp insertion in one allele that leads to a premature termination of the RET protein (RET$^{+/E734Kfs*\#40=>}$) and is equivalent to a heterozygous deletion of *RET* (IMR90$^{RET+/-}$) (Figure 2B). Another isogenic iPSC line carries two independent mutations in the two alleles of the *RET* gene, which cause a truncation deletion (allele 1: R606X) and a 2-amino acid deletion (allele 2: 601-602del) in the RET protein (Figure 2C).

An important aim of this study was to examine whether patient-specific iPSCs can be used to study the effects of genetic lesions and to model disease development. Many of the *RET* CD mutations interrupt ENCC development, thereby contributing to the severe form of HSCR (TCA). Therefore, a TCA-HSCR patient with a heterozygous CD mutation in exon 12 of the *RET* gene (G731del), which led to a single amino-acid deletion, was included in this study (Figure 2D). Skin fibroblasts were obtained from this patient and were reprogrammed into iPSCs with episomal vectors[17] (referred to as HSCR#6-iPSCs, Supplementary Figure 1). The presence of the RET G731del mutation in both the fibroblasts and iPSCs was confirmed by Sanger sequencing.

To determine how these mutations affect RET protein expression, we transfected plasmids carrying the wild-type or mutant *RET* cDNA into a human epithelial cell line (HeLa). A western blot analysis revealed that the E734Kfs*#40=> mutation in exon 12 of the *RET* gene causes premature termination leading to decreased total RET expression level in the IMR90$^{RET+/-}$ line (Figure 2E, Ex12). In the IMR90$^{RET-/-}$ line, the R606X (Ex10 A1) truncation mutation in allele 1 (A1) led to

premature termination of the RET protein, whereas the 601-602del mutation in the second allele (Ex10 A2) decreased the RET level. Therefore, we speculated that RET/GDNF signaling in the IMR90$^{RET-/-}$ line would be largely compromised. Therefore, comparing these two isogenic clones with their parental control line would allow us to establish a causal link between impaired RET/GDNF signaling and ENCC defects and, in turn, define the disease-relevant phenotypes in the dish.

To directly address the biological effect of this mutation, a genetically matched control line was established by correcting this *RET* mutation in HSCR#6-iPSCs by using the CRISPR/Cas9 genome-editing platform with a specific pair of gRNAs and a donor plasmid carrying the wild-type (WT) *RET*, as shown in Figure 2F. The corrected clone (HSCR#6-RET corrected-iPSC) had the same genetic background as HSCR#6-iPSCs but carried a WT *RET* gene, and the correction was confirmed by Sanger sequencing (Figure 2G). These two iPSC lines were used for subsequent functional analyses to directly address how *RET G731del* might interfere with human ENCC development.

S-HSCR is the most common form of HSCR, and many affected individuals do not carry any CD mutation in *RET* or in other known HSCR genes. Therefore, to generate iPSC lines, as described above, skin fibroblasts were obtained from two individual S-HSCR patients who did not carry any *RET* CD mutations (referred to as HSCR#3 and HSCR#5) (Supplementary Figures 2 & 3).

Derivation of ENCCs from diseased iPSC lines

To recapitulate the progressive differentiation process that occurs during normal ENCC development, we used a stepwise differentiation protocol[10]. Human iPSCs were first differentiated into multipotent NCCs before being further differentiated into neurons, including neurons expressing pan-neuronal or ENS-specific neuronal markers (Figure 3A). Cranial and vagal NCCs were derived from control and diseased iPSCs through previously described protocols in the absence and presence of retinoic acid (RA), respectively[9-11, 13, 18] (Figure 3B). To prospectively identify and isolate NCCs from the heterogeneous population, cell sorting was performed on day 10 with antibodies against HNK-1 and $p75^{NTR}$ (Figure 3B). Comparison of the transcription profiles of HNK-1 and $p75^{NTR}$ double-positive cranial and vagal NCCs derived from the control iPSC line (IMR90) showed that almost 3000 genes were commonly expressed in these two populations and that approximately 1,000 genes were uniquely expressed in each population. Additional comparisons made for the expression of NCC marker genes (*TWIST1*, *TFAP2A*, *SOX10*, *PAX3*, *NGFR*, *FOXD3* and *EDNRB*) showed that all these genes were expressed in HNK-$1^{+}p75^{NTR+}$ populations. In particular, HNK-$1^{+}p75^{NTR+}$ NCCs derived in the presence of RA showed high expression levels of these genes, which were accompanied by high levels of *HOXB* family genes (*HOXB2-5*, but not *HOXB9*). This gene expression pattern was consistent with that of enteric NCCs (Figure 3C). Most of the HNK-$1^{+}p75^{NTR+}$ NCCs also co-expressed CD49d, RET (Figure 3D) and SOX10 (Figure 3E). As early as day 9 in neuronal differentiation medium, the expression levels of various pan-neuronal markers (TUJ1, PGP9.5, and HU), PHOX2B and markers for enteric neuron subtypes such as TH (dopaminergic neuron), VIP (inhibitory neuron) and ChAT (excitatory neuron) were detectable (Figure 3F). These neuronal cells formed more mature neurons with extended

neuronal processes and became responsive to electric stimulation after they were cultured *in vitro* for another week (day 16) (Supplementary Figure 4). Both the gene expression patterns and cellular phenotypes of HNK-1^{+}p75^{NTR+} NCCs were consistent with those of mouse and human pluripotent stem cell-derived enteric NCCs, as defined in a recent paper by *Fattahi et. al.*[10]. Hereafter, the HNK-1^{+}p75^{NTR+} NCCs derived in the presence of RA will be termed "ENCCs". All diseased iPSC lines (IMR90$^{RET-/-}$, IMR90$^{RET+/-}$, HSCR#3, HSCR#5 and HSCR#6) and the control line (IMR90) generated ENCCs at comparable yields (Figure 3G).

HSCR-associated phenotypes were recapitulated in the dish with iPSCs

RET expression in ENCCs derived from the diseased iPSC lines carrying *RET CD* mutations were analyzed using flow cytometry with an antibody against the intracellular domain of RET. The percentages of RET^{+} cells obtained from the two isogenic iPSC lines (IMR90$^{RET-/-}$ and IMR90$^{RET+/-}$) were consistently decreased by ~40-50%, as compared with their parental line (IMR90). A similar observation was found in the HSCR#6-iPSC line, but significantly more RET-expressing ENCCs were obtained when the mutation was corrected (RET Corrected) (Figure 4A). We then examined how *RET* mutations interfere with the responsiveness of ENCCs to GDNF, on the basis of their neuronal differentiation capability, as monitored by the expression of a pan-neuronal marker (TUJ1). In the presence of GDNF, robust neuronal differentiation was observed in ENCCs carrying WT *RET*. IMR90$^{RET+/-}$ ENCCs differentiated into neurons less efficiently, whereas this process was almost completely abolished when both copies of the *RET* gene were interrupted in IMR90$^{RET-/-}$ ENCCs (Figure 4B). As demonstrated in scratch assay, retarded migration was also observed in these two *RET* mutant lines, as predicted

(Supplementary Figure 4). Similar defects were found in HSCR#6 ENCCs. Diseased ENCCs exhibited severe differentiation defects with very weak TUJ1 expression and extremely short neuronal fibers. Retarded migration of these cells was observed in the scratch assays, and the addition of GDNF did not promote their migration. Importantly, both the differentiation capacity (Figure 4C) and migratory behavior (Figure 4D) of ENCCs derived from HSCR#6-iPSCs was restored when the *RET* mutation was corrected. This proof-of-concept study indicated that HSCR-associated disease phenotypes and severity can be recapitulated in the dish using human iPSCs.

Similarly to the *RET* mutant ENCCs, ENCCs derived from the two S-HSCR iPSC lines (HSCR#3 and HSCR#5) also exhibited neuronal differentiation and migration defects, as monitored by *in vitro* differentiation and scratch assays (Figure 4E-F). The diseased ENCCs exhibited defects in the formation of axon outgrowth and various subtypes of neurons. The expression of pan-neuronal markers (e.g., TUJ1, PGP9.5) was dramatically decreased in diseased ENCCs. The percentages of RET expressing cells derived from these two S-HSCR-iPSC lines were only slightly lower than that of the healthy controls (IMR90-iPSC and RET-corrected HSCR#6-iPSC) (Supplementary Figure 6). Notably, the migration and differentiation defects in S-HSCR-ENCCs were relatively less severe than those of cells carrying *RET* CD mutations.

Disease-associated differentiation defects are lineage specific

The greatest concern regarding the analyses of S-HSCR iPSC lines is the differences in the genetic background among individuals, and comparisons were usually made between patients and an unmatched control. The different cellular

phenotypes observed in the control and patient cells may have been due to the significant biological heterogeneity of iPSC lines derived from different individuals. Therefore, we used a more rigorous approach to assess the differentiation defects of the control and S-HSCR-iPSC lines. First, CNS progenitors were generated from the control and diseased iPSC lines in a condition favoring the induction of CNS progenitors, in which the same neural induction medium was used, but cells were seeded at a higher cell density, as previously described[11] (Figure 5A). A comparable amount of CNS progenitors ($PAX6^{+}$) was obtained from the control and S-HSCR-iPSC lines (Figure 5B). Intriguingly, CNS progenitors from both the control and diseased cells efficiently differentiated further and generated various subtypes of neurons, as monitored by the expression of pan-neuronal markers (TUJ1, HU, PGP9.5) and subtype-specific neuronal markers (e.g., VIP, TH and nNOS) (Figure 5C).

We then tested the differentiation plasticity of the diseased NCCs toward non-neuronal NC lineages, such as adipogenic, osteogenic and chondrogenic lineages. Cranial NCCs were first generated in the absence of RA treatment through the same differentiation protocol as described above. FACS-sorted cranial NCCs derived from the control and diseased iPSCs were directed to a mesenchymal lineage, thus forming hiPSC–derived neural crest-mesenchymal precursor cells (hiPSC–derived NCMP cells). These hiPSC–derived NCMP cells were subsequently used to generate adipocytes, osteocytes or chondrocytes (Figure 5D). Comparable amounts of cranial NCCs were obtained from the control (IMR90) and S-HSCR-iPSC (HSCR#3 & HSCR#5) lines (Figure 5E), and 75-88% of these cells co-expressed SOX10 (Figure 5F). In addition, there were no obvious differences between the control and diseased

cells in their capacities to generate NCMP cells (Figure 5G) and their derivatives, such as adipocytes (Figure 5H), osteocytes (Figure 5I) and chondrocytes (Figure 5J). Therefore, the neuronal differentiation defects observed in the S-HSCR-iPSCs were restricted to the NC lineage, whereas the CNS lineage was not affected.

Identification of disease-associated pathways and mutations for S-HSCR

To gain mechanistic insights into the disease-related mechanisms underlying S-HSCR, the transcriptomes of the fibroblast cells, iPSCs and their NC derivatives (ENCCs) from the patient and control groups were obtained through RNA sequencing. A principal component analysis (PCA) of these profiles showed a good separation between the different types of cells. Control and patient fibroblast cells and iPSCs showed similar expression signatures and clustered together, whereas their ENCC derivatives exhibited more distinctive gene expression profiles (Supplementary Figure 7). A subsequent gene ontology (GO) analysis of the control and S-HSCR-iPSC-derived ENCCs showed a significant enrichment of GO terms associated with cell junction assembly/migration and neuronal differentiation (Figure 6A).

It is challenging to identify causative mutations for a rare and complex disease. Therefore, we established a new strategy to identify novel functional mutation(s) in S-HSCR by using our established iPSC platform, with a particular focus on HSCR#3. We used a two-tiered approach to define how genetic variation(s) functionally contribute to disease phenotypes (Figure 6B). We first identified candidate genes by comparing the transcriptomes of the control and the diseased cells. Gene differentially expressed at the different developmental stages (iPSC versus NC

derivatives) of a given cell line (longitudinal expression study) were first identified and were then used for the cross sectional comparison between the patient and the control groups (Supplementary methods). 735 genes were DE in the longitudinal study of the control derived cell line and 562 from the patient-derived cell line. When patient and control DE genes were compared, 299 genes were differentially expressed in the patient cells but not in the control cells during ENCC differentiation (Supplementary Table 4), thus suggesting that these changes may alter the differentiation path in the patient (Supplementary methods). Noteworthy, when we used a pool of 20 other iPSC lines (from our data and a public database) as controls, a total of 179 genes were found to be differentially expressed in HSCR#3-iPSC. This result might indicate that HSCR#3-iPSCs carry some genetic alterations that also lead to dysregulation at the very first step of differentiation. After shortlisting the potential candidate genes, to further narrow down the list of potentially causative gene(s), we then combined the transcriptome data with the whole-exome sequencing data obtained from sequencing the blood DNA sample of the patient. From the whole-exome sequencing study of this patient, we identified genetic variations in 267 genes predicted to be damaging (sftp://odds.cgs.hku.hk). These genes were appended to the network analysis of the differentially expressed according to our cross sectional analyses. GeneMANIA identified genes (network nodes) present in the association networks created from the cross sectional genes (Figure 6C). Among those genes, *VCL* was notable. This gene is a nexus among multiple differentially expressed genes in the networks. In addition, according to the GO enrichment of the networked genes, *VCL* interacts with multiple genes involved in “axon guidance”.

Functional dissection of the Vinculin M209L in HSCR pathogenesis

VCL is an essential protein required for focal adhesion (FA) assembly through binding to its partner proteins, including talin[19], actin[20] and paxillin[21]. Mutations interrupting FA assembly or affecting the size and position of FAs may interfere various cellular processes such as cellular movement or differentiation. In our patient, VCL contained a M209L substitution in the head domain of the protein. SWISS-MODEL modeling suggested that the M209L substitution might alter the conformation of the loop structure in the folded protein (Figure 7A). To determine whether VCL M209L affects the FAs dynamics and contributes to disease phenotypes, we corrected the *VCL* mutation in HSCR#3-iPSCs by using a CRISPR/Cas9-mediated targeting approach (Figure 7B and Supplementary Figure 8). To this end, FA assembly was examined in the diseased and disease corrected ENCCs based on the expression pattern of VCL and its interacting partners (Phalloidin (actin) and Paxillin). In addition, the mean size of FA (FA size) in ENCCs was measured based on the expression of Paxillin (a FA marker). Consistently, we found that the FA size of ENCCs derived from HSCR#3-iPSCs significantly smaller than that of the control (IMR90) ENCCs. Correction completely rescued this defect (Figure 7C). This phenomenon was recapitulated by overexpression of a VCL M209L mutant in HeLa cells, thus suggesting that the decreased FA size was the result of a mutation in *VCL* that led to the M209L substitution (Figure 7D). A subsequent immunoprecipitation (IP) assay further revealed that the *M209L* substitution decreased VCL-actin binding (Figure 7E). Intriguingly, correction of the *VCL* mutation alone in HSCR#3-iPSCs efficiently restored the differentiation (Figure 7F) and migration (Figure 7G) capacities of ENCCs. In particular, neurons derived from corrected ENCCs possessed extended axons, which were highly similar to those of the wild-type cells (Figure 7F).

Discussion:

In this study, we demonstrated how human pluripotent stem cells (hPSCs) can be used to model HSCR disease phenotypes and to correlate the genetic lesions and disease severity in the dish. More importantly, we also identified a new experimental approach by combining the genetic data obtained directly from patient blood DNA and the transcription profile of the disease-relevant cells, to identify mutation(s) functionally contributing to disease phenotypes for a complex disease.

A recent meta-analysis comparing the transcriptional patterns of various hPSC lines has revealed that transcriptional differences between genetically unrelated hPSC lines are primarily driven by genetic background rather than cellular origin or reprogramming method[22], and these differences may affect cellular behaviors. Therefore, we performed a comprehensive study on eight healthy and diseased iPSC lines and used various approaches to illustrate how iPSCs can be used to model HSCR disease. First, we defined the disease phenotypes in the culture dish with iPSC lines carrying CD mutations in *RET*, the best-characterized HSCR gene. These two *RET* isogenic mutant lines (IMR90$^{RET+/-}$ and IMR90$^{RET-/-}$) showed comparable RET levels, but RET/GDNF signaling was partially or totally compromised. Comparisons between these two isogenic mutant lines and their parental line lined the dosage effect of RET/GDNF signaling with cellular phenotypes, thus reflecting disease severity. More importantly, we also showed that ENCC functions were restored by correction of a single *RET CD* mutation (RET G731del) in a patient-specific iPSC line (HSCR#6) derived from a TCA-HSCR patient. This finding further suggested that patient-specific iPSCs are a reliable system for mechanistic studies of the genetic etiology underlying HSCR disease. In particular, HSCR-associated genetic lesions probably

affect the early events of NCC development, which can be monitored on the basis of the differentiation and migration capacities of iPSC-derived ENCCs *in vitro*.

The genetic etiology of S-HSCR probably involves rare or common variants, acting alone or in combination; thus, a disease model carrying the same genetic background of the patient would have advantages for functionally interconnecting mutation(s) with disease phenotypes. We generated patient-specific iPSCs from 2 S-HSCR patients carrying the most common *RET NCD* variant. These S-HSCR-iPSC lines allowed us to assess the combined effects of all HSCR-relevant variants present in a given patient. First, we linked the cellular phenotypes in the dish with the disease phenotypes of the patients. CNS progenitors and cranial NCCs were derived from the control and S-HSCR iPSCs by using three highly comparable differentiation protocols differing only in cell densities and RA treatment. The differentiation capacities toward CNS neurons and non-neuronal NC lineages were examined. Our data clearly indicated that the differentiation defects were restricted to certain lineages and recapitulated the patient phenotypes. We further pinpointed the disease-relevant mutations by integrating the genetic data obtained from the whole-exome sequencing of patient blood DNA and the expression profiles of ENCCs derived from patient iPSCs. Combining these two datasets successfully increased the statistical power and allowed us to identify the disease-relevant mutation(s). As shown by our analysis, the whole-exome sequencing data identified more than two hundred mutations from the patient genomic DNA. After integrating the transcriptional data obtained from iPSC-derived ENCCs, a single *VCL* (M209L) mutation was found to be the most relevant.

VCL has been implicated in several developmental processes, including cardiac myocyte and NCC development[23, 24]. *Vcl* null mutants die at embryonic day 10, owing to failure of rostral neural tubule closure. The cranial and spinal nerve

development are also compromised[25]. In our functional study, we demonstrated that the VCL M209L substitution led to decreased FA size in the patient ENCCs, and such changes may directly interfere with the cellular behavior of ENCCs and their responses to the environment, thereby causing defects in differentiation and migration. Consistently with these results, a gene ontology analysis also suggested that migration and axon guidance pathways are affected in these diseased cells. More importantly, a subsequent rescue experiment clearly demonstrated that the VCL M209L substitution functionally contributes to the disease phenotypes.

An advantage of our experimental approach was its high power for identification of the most relevant genetic changes from the disease-relevant cells. In addition, integrating the genetic data obtained directly from patient blood DNA with the transcriptomic data of the disease-relevant cells further eliminated the effect of systemic inherent variability due to different genetic background or mutations artificially introduced during the culturing process. The use of highly comparable differentiation protocols to test the differentiation plasticity of the control and diseased cells toward various lineages also minimized the error associated with *in vitro* differentiation systems. Importantly, the experimental paradigm established here is generally applicable for mechanistic studies of various complex diseases.

In summary, we identified a novel mutation in *VCL* that is a possible contributor to HSCR. In addition, we also established a new strategy by combining patient-specific iPSCs, the CRISPR/Cas9 system and whole-exome sequencing data to identify mutations functionally contributing to a complex disease.

Acknowledgement:

Authors would like to thank Mr. Nick Chi Ho LIN (the University of Hong Kong), Dr. Marco Kong (the University of Hong Kong) and Dr. Chad Cowan (Harvard Stem Cell Institute) for their bioinformatics support and technical advice on Calcium imaging and CRISPR/Cas9 platform, respectively. This work was supported by a HMRF grant (Project no.: 01121476) to E.S.W.N., and research grants from the Research Grants Council of Hong Kong Special Administrative Region, China Hong Kong (HKU17119514, T12C-714/14-R). Confocal Imaging and RNA sequencing were performed using equipment maintained by Li Ka Shing Faculty of Medicine Faculty Core Facility and Center for Genomic Sciences, the University of Hong Kong, respectively.

References:

1. Parisi MA. Hirschsprung Disease Overview. Seattle: University of Washington, Seattle, 2002 [updated 2015 Oct 1].
2. Durbec PL, Larsson-Blomberg LB, Schuchardt A, et al. Common origin and developmental dependence on c-ret of subsets of enteric and sympathetic neuroblasts. Development 1996;122:349-58.
3. Schuchardt A, D'Agati V, Larsson-Blomberg L, et al. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. Nature 1994;367:380-3.
4. Taraviras S, Marcos-Gutierrez CV, Durbec P, et al. Signalling by the RET receptor tyrosine kinase and its role in the development of the mammalian enteric nervous system. Development 1999;126:2785-97.
5. Kjaer S, Hanrahan S, Totty N, et al. Mammal-restricted elements predispose human RET to folding impairment by HSCR mutations. Nat Struct Mol Biol 2010;17:726-31.
6. Uesaka T, Nagashimada M, Yonemura S, et al. Diminished Ret expression compromises neuronal survival in the colon and causes intestinal aganglionosis in mice. J Clin Invest 2008;118:1890-8.
7. Emison ES, Garcia-Barcelo M, Grice EA, et al. Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. Am J Hum Genet 2010;87:60-74.
8. **Mali P**, **Yang L**, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. Science 2013;339:823-6.
9. Lee G, Chambers SM, Tomishima MJ, et al. Derivation of neural crest cells from human pluripotent stem cells. Nat Protoc 2010;5:688-701.

10. Fattahi F, Steinbeck JA, Kriks S, et al. Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. Nature 2016.

11. Lee G, Kim H, Elkabetz Y, et al. Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. Nat Biotechnol 2007;25:1468-75.

12. Zeltner N, Lafaille FG, Fattahi F, et al. Feeder-free derivation of neural crest progenitor cells from human pluripotent stem cells. J Vis Exp 2014.

13. Chambers SM, Fasano CA, Papapetrou EP, et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol 2009;27:275-80.

14. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010;38:W214-20.

15. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protocols 2008;4:44-57.

16. Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195-201.

17. Okita K, Matsumura Y, Sato Y, et al. A more efficient method to generate integration-free human iPS cells. Nat Methods 2011;8:409-12.

18. Mica Y, Lee G, Chambers SM, et al. Modeling neural crest induction, melanocyte specification, and disease-related pigmentation defects in hESCs and patient-specific iPSCs. Cell Rep 2013;3:1140-52.

19. Burridge K, Mangeat P. An interaction between vinculin and talin. Nature 1984;308:744-6.

20. Jockusch BM, Isenberg G. Interaction of alpha-actinin and vinculin with actin: opposite effects on filament network formation. Proc Natl Acad Sci U S A 1981;78:3005-9.

21. **Turner CE**, **Glenney JR**, Jr., Burridge K. Paxillin: a new vinculin-binding protein present in focal adhesions. J Cell Biol 1990;111:1059-68.

22. Choi J, Lee S, Mallard W, et al. A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. Nat Biotechnol 2015;33:1173-81.

23. Xu X, Francis R, Wei CJ, et al. Connexin 43-mediated modulation of polarized cell movement and the directional migration of cardiac neural crest cells. Development 2006;133:3629-39.

24. Zemljic-Harpf AE, Miller JC, Henderson SA, et al. Cardiac-myocyte-specific excision of the vinculin gene disrupts cellular junctions, causing sudden death or dilated cardiomyopathy. Mol Cell Biol 2007;27:7522-37.

25. Marg S, Winkler U, Sestu M, et al. The vinculin-DeltaIn20/21 mouse: characteristics of a constitutive, actin-binding deficient splice variant of vinculin. PLoS One 2010;5:e11530.

Author names in bold designate shared co-first authorship

Figure legends:

Figure 1. Generation of disease-relevant cells for studying genetic lesions associated with HSCR disease. (A) Schematic for the generation of diseased and healthy ENCCs from iPSCs. Diseased iPSC (HSCR-iPSC) lines were generated by reprogramming the patient fibroblast cells or by introducing a disease-associated mutation into the control iPSCs through a genome-editing approach. The genetically matched control iPSC lines were the disease-corrected iPSC and the parental cells of the mutant lines. ENCCs were derived from the diseased and healthy iPSC lines. The cellular behavior and expression profiles of the diseased ENCCs were compared with their genetically matched controls. The table summarizes the key experiments performed in this study. **(B)** Overview of healthy and diseased fibroblasts, iPSC lines and their derivatives used for functional analyses and RNA-seq.

Figure 2. Generation of diseased iPSC lines. (A) Generation of isogenic iPSC lines carrying homozygous or heterozygous *RET* mutations. **(B)** IMR90$^{RET+/-}$ line: a heterozygous mutation was introduced into exon 12 of the *RET* gene. **(C)** IMR90$^{RET-/-}$ carries a truncation deletion and a 2 amino-acid deletion mutation in allele 1 (A1) and allele 2 (A2), respectively. The DNA and amino acid sequences of the two isogenic RET mutant iPSC lines (IMR90$^{RET+/-}$ and IMR90$^{RET-/-}$) are shown. **(D)** DNA and amino acid sequences of the *RET* gene in an iPSC (HSCR#6) line derived from a TCA-HSCR patient carrying the *RET G731del* mutation in exon 12, affecting the tyrosine kinase domain of RET. **(E)** Western blot analysis of protein lysates obtained from HeLa cells overexpressing wild-type (WT) or mutant *RET*. The red arrow indicates the full-length RET. **(F)** Targeting strategy for correcting the *RET G731del*

mutation in the HSCR#6-iPSC line. **(G)** Correction of the *RET G731del* mutation was confirmed by Sanger sequencing.

Figure 3. Derivation of ENCCs from iPSC lines. **(A)** Schematic of the generation of ENCCs and their derivatives. **(B)** Stepwise differentiation protocol for the generation of vagal and cranial NCs in the presence or absence of retinoic acid (RA), respectively. **(C)** Transcriptional profiles of vagal and cranial HNK-1^{+}/$p75^{NTR+}$ NCCs were obtained by RNA sequencing. The Venn diagram shows differentially expressed genes (DEGs) between vagal and cranial NCCs. The expression levels of various NCC markers and *HOX* genes in these two NCC populations are shown in the bar charts. **(D)** Flow cytometry analysis of the co-expression of CD49d and RET in the HNK-1^{+}/$p75^{NTR+}$ NCCs. **(E)** Immunocytochemistry was used to monitor the expression of SOX10. **(F)** *In vitro* differentiation assays indicated the differentiation plasticity of ENCCs, as monitored by the expression levels of markers of various neuronal subtypes. TUJ1: beta-tubulin; PGP9.5: protein gene product 9.5; HU: HU neuronal protein; PHOX2B: paired-type homeobox protein; TH: tyrosine hydroxylase; VIP: vasoactive-intestinal polypeptide; ChAT: choline acetyltransferase. **(G)** Flow cytometry analyses to quantify HNK-1^{+}/p75NTR^{+} ENCCs derived from control and diseased iPSC lines. The mean values from 6 independent assays are shown.

Figure 4. RET-deficient ENCCs exhibit defects in migration and differentiation. **(A)** Flow cytometry analyses were used to measure RET-expressing ENCCs derived from control (IMR90), two isogenic *RET* mutant (IMR$90^{RET+/-}$ and IMR$90^{RET+/-}$), HSCR#6 and *RET*-corrected HSCR#6 iPSC lines. The percentages of RET^{+} ENCCs

are shown in the bar chart (mean ± SEM from 4-6 independent experiments). **(B & C)** The neuronal differentiation of control and disease ENCCs after GDNF treatment (100 μg/ml) was monitored on the basis of the expression of TUJ1. Scale bars = 50 μm. Quantitative data are shown in the bar chart. **(D)** Scratch assays were performed to measure the migratory ability of iPSC-derived ENCCs. The bar charts show wound closure measured as the percentage of scar width over time (18 h). **(E)** *In vitro* differentiation analysis of ENCCs derived from two S-HSCR (HSCR#3 and HSCR#5) iPSC lines. Their differentiation plasticity was measured on the basis of the expression of the pan-neuronal markers TUJ1, PGP9.5 (protein gene product 9.5), and HU and the ENS-related neuronal markers VIP (vasoactive intestine peptide) and ChAT (choline acetyltransferase). Quantitative data are shown in the bar charts. **(F)** Scratch assays on ENCCs derived from control and S-HSCR iPSCs. The bar chart shows wound closure, measured as the percentage of scar width over time (18 h). The data shown in the bar charts are from at least three independent assays.

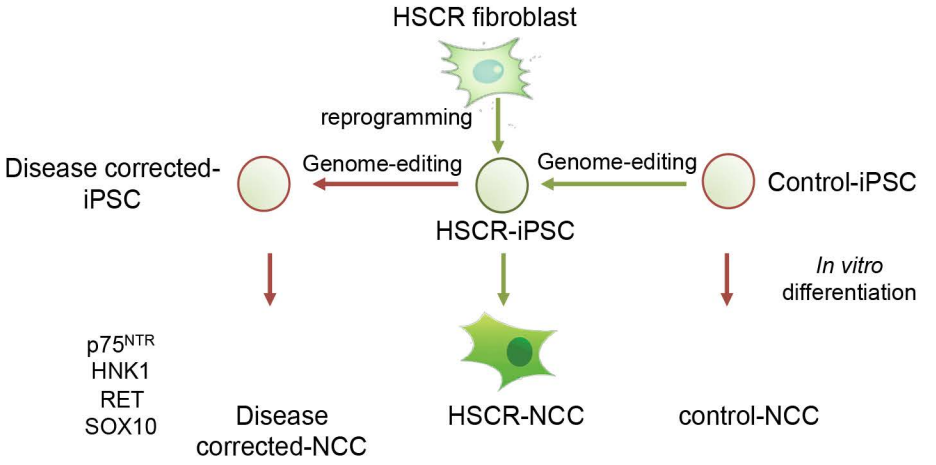
Figure 5. Differentiation defects of patient cells are lineage specific. (A) Schematic of the derivation of CNS progenitors and derivatives. **(B)** Comparable CNS progenitors ($PAX6^{+}$) were obtained from control and diseased iPSC lines. **(C)** CNS progenitors derived from control and diseased iPSCs showed similar capacities to form various subtypes of neurons. (n=3) **(D)** Cranial NCCs were generated from control and diseased iPSC lines and subsequently differentiated into non-neuronal NC lineages. **(E)** All S-HSCR-iPSC lines showed comparable capacities to form cranial NCCs and co-expressed SOX10. **(F)** A comparable amount of NCMPs was obtained from control and diseased lines. The NCMPs did not show obvious differences in generating adipocytes, osteocytes and chondrocytes.

Figure 6. RNA sequencing data identify disease-associated pathways. (A) GO analyses of genes differentially expressed in S-HSCR ENCCs. **(B)** Schematic of the integrated analysis of RNAseq data and whole-exome sequencing data obtained from HSCR-iPSC-derived ENCCs and patient blood DNA, respectively. The degree of overlap of the RNA and exome sequencing indicates a causal gene/pathway and disease relationship. **(C)** GeneMANIA shows genes (network nodes) present in the association networks created from the longitudinal and cross sectional analyses by combining the genetic and transcription data from HSCR#3.

Figure 7. *VCL M209L* mutation causes delayed migration and defective formation of neuronal axons. (A) Schematic showing that the M209L substitution resides in the head domain of VCL in proximity to the talin binding region. SWISS-MODEL indicates that the *M209L* substitution alters the protein conformation of VCL. **(B)** Sanger sequencing confirmed that the HSCR#3-iPSCs and corrected cells are carrying mutant and wild-type *VCL*, respectively. **(C)** Immunocytochemistry analysis of VCL, Phalloidin, and Paxillin in control, HSCR#3 and *VCL*-corrected HSCR#3-iPSC derived ENCCs. Upper panel: The co-localization of VCL with Phalloidin at the focal adhesion (FA) is highlighted and magnified as shown in the insets. Lower panel: FAs were identified based on the expression of Paxillin and the areas marked with a dashed line were measured using ImageJ and considered as the FA size. The bar chart shows the average FA size in control, HSCR#3, and *VCL*-corrected ENCCs. **(D)** Immunocytochemistry analysis of phalloidin and paxillin in HeLa cells overexpressing wild-type or mutant *VCL*. Upper panel shows the colocalization of VCL with Phalloidin at the FAs. Lower panel indicates that the size

of FA marked by Paxillin was smaller in the mutant VCL-overexpressing HeLa cells. The bar chart shows the average FA size in the wild-type or mutant VCL-overexpressing HeLa cells. **(E)** Western blot (WB) analysis confirmed comparable expression levels of wild-type or mutant VCL in transfected cells. Co-immunoprecipitation (co-IP) indicated decreased actin binding of mutant VCL. **(F)** An *in vitro* differentiation assay revealed that correction of the *VCL* mutation restores the differentiation capacity of ENCCs. Corrected cells possess more axon outgrowth, marked by arrowheads. The percentages of cells expressing various neuronal markers derived from HSCR#3 and corrected iPSCs are shown in the bar chart. **(G)** Scratch assays on HSCR#3 and *VCL*-corrected iPSC-derived ENCCs. The bar chart shows wound closure measured as the percentage of scar width over time (18 h). The data shown in the bar charts are from at least three independent assays.

A



1. Generation of iPSC lines using fibroblasts from a TCA and two S-HSCR patients by reprogramming
2. Introduction of *RET* CD mutations into the IMR90-iPSC using CRISPR/Cas9 genome editing platform
3. Characterization of NCCs derived from healthy and diseased iPSC lines
4. Global expression profiling on fibroblasts, iPSCs and ENCCs by RNAseq
5. CD mutation analysis by whole-exome sequencing on genomic DNA from blood of a S-HSCR patient
6. Identification and characterization of mutations functionally contributing to ENCC defects
7. Correction of disease-associated mutations

B

	Diseased cells		Healthy cells
Fibroblast	HSCR#3-Fib		
	HSCR#5-Fib		
	HSCR#6-Fib		
iPSCs	HSCR#3-iPSC		IMR90-iPSC
	HSCR#5-iPSC	IMR90$^{RET-/-}$-iPSC	HSCR#3-iPSC-VCL Corrected
	HSCR#6-iPSC	IMR90$^{RET+/-}$-iPSC	HSCR#6-iPSC-RET Corrected
ENCC	HSCR#3-ENCC		IMR90-ENCC
	HSCR#5-ENCC	IMR90$^{RET-/-}$-ENCC	HSCR#3-ENCC-VCL Corrected
	HSCR#6-ENCC	IMR90$^{RET+/-}$-ENCC	HSCR#6-ENCC-RET Corrected

Figure 1

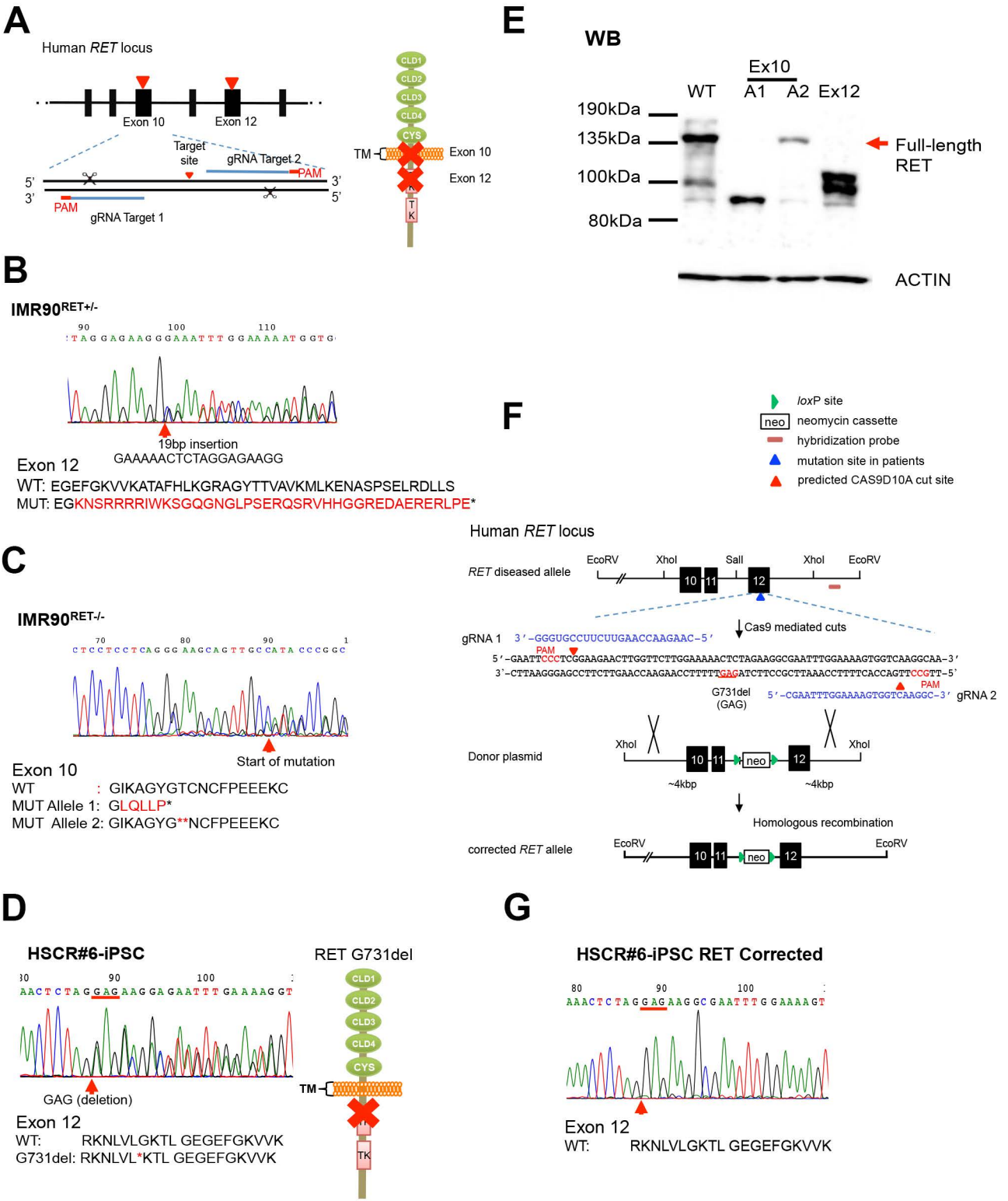

A
Human RET locus
Exon 10
Exon 12
Target site
gRNA Target 2
PAM
5'
3'
PAM
gRNA Target 1
CLD1
CLD2
CLD3
CLD4
CYS
TM
Exon 10
Exon 12
T
K
B
IMR90RET+/-
90
100
110
19bp insertion
GAAAAACTCTAGGAGAAGG
Exon 12
WT: EGEFGKVVKATAFHLKGRAGYTTVAVKMLKENASPSELRDLLS
MUT: EGKNSRRRRIWKSGQGNGLPSERQSRVHHGGREDAERERLPE*
C
IMR90RET-/-
70
80
90
Start of mutation
Exon 10
WT : GIKAGYGTCNCFPEEEKC
MUT Allele 1: GLQLLP*
MUT Allele 2: GIKAGYG**NCFPEEEKC
D
HSCR#6-iPSC
RET G731del
90
100
GAG (deletion)
Exon 12
WT: RKNLVLGKTL GEGEFGKVVK
G731del: RKNLVL*KTL GEGEFGKVVK
CLD1
CLD2
CLD3
CLD4
CYS
TM
TK
E
WB
Ex10
WT
A1
A2
Ex12
190kDa
135kDa
100kDa
80kDa
Full-length RET
ACTIN
F
loxP site
neo
neomycin cassette
hybridization probe
mutation site in patients
predicted CAS9D10A cut site
Human RET locus
RET diseased allele
EcoRV
XhoI
SalI
XhoI
EcoRV
10
11
12
gRNA 1
3'-GGGUGCCUUCUUGAACCAAGAAC-5'
Cas9 mediated cuts
PAM
5'-GAATTCCCTCGGAAGAACTTGGTTCTTGGAAAAACTCTAGAAGGCGAATTTGGAAAAGTGGTCAAGGCAA-3'
3'-CTTAAGGGAGCCTTCTTGAACCAAGAACCTTTTTGAGATCTTCCGCTTAAACCTTTTCACCAGTTCCGTT-5'
PAM
G731del
(GAG)
5'-CGAATTTGGAAAAGTGGTCAAGGC-3'
gRNA 2
Donor plasmid
XhoI
neo
XhoI
~4kbp
~4kbp
Homologous recombination
corrected RET allele
EcoRV
neo
EcoRV
G
HSCR#6-iPSC RET Corrected
80
90
100
Exon 12
WT: RKNLVLGKTL GEGEFGKVVK

Figure 2

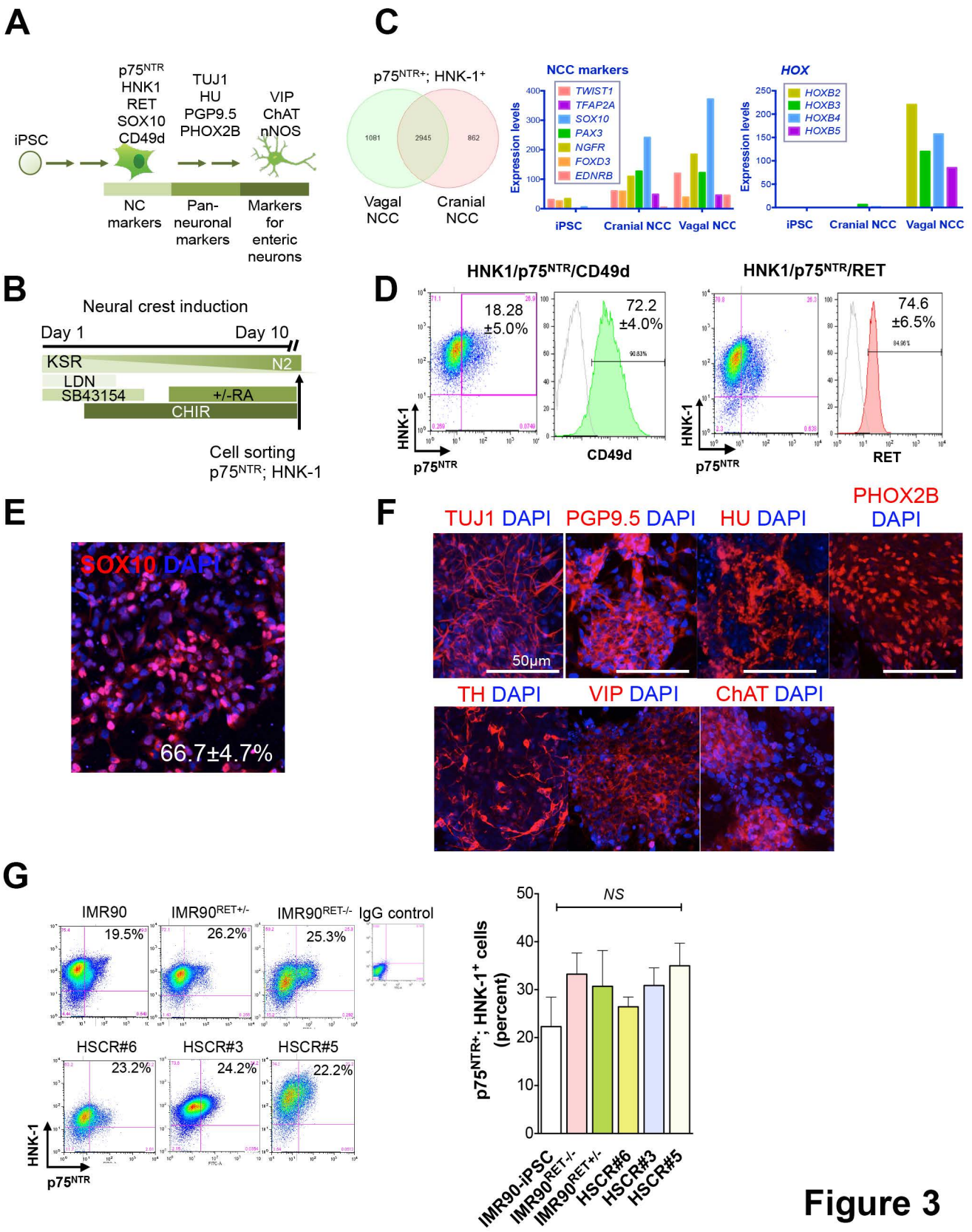

A
iPSC
p75NTR HNK1 RET SOX10 CD49d
TUJ1 HU PGP9.5 PHOX2B
VIP ChAT nNOS
NC markers
Pan-neuronal markers
Markers for enteric neurons
B
Neural crest induction
Day 1
Day 10
KSR
N2
LDN
SB43154
+/-RA
CHIR
Cell sorting
p75NTR; HNK-1
C
p75NTR+; HNK-1+
1081
2945
862
Vagal NCC
Cranial NCC
NCC markers
TWIST1
TFAP2A
SOX10
PAX3
NGFR
FOXD3
EDNRB
Expression levels
iPSC
Cranial NCC
Vagal NCC
HOX
HOXB2
HOXB3
HOXB4
HOXB5
D
HNK1/p75NTR/CD49d
18.28 ±5.0%
72.2 ±4.0%
HNK-1
p75NTR
CD49d
HNK1/p75NTR/RET
74.6 ±6.5%
RET
E
SOX10 DAPI
66.7±4.7%
F
TUJ1 DAPI
PGP9.5 DAPI
HU DAPI
PHOX2B DAPI
50μm
TH DAPI
VIP DAPI
ChAT DAPI
G
IMR90
IMR90RET+/-
IMR90RET-/-
IgG control
19.5%
26.2%
25.3%
HSCR#6
HSCR#3
HSCR#5
23.2%
24.2%
22.2%
HNK-1
p75NTR
NS
p75NTR+; HNK-1+ cells (percent)
IMR90-iPSC
IMR90RET-/-
IMR90RET+/-
HSCR#6
HSCR#3
HSCR#5

Figure 3

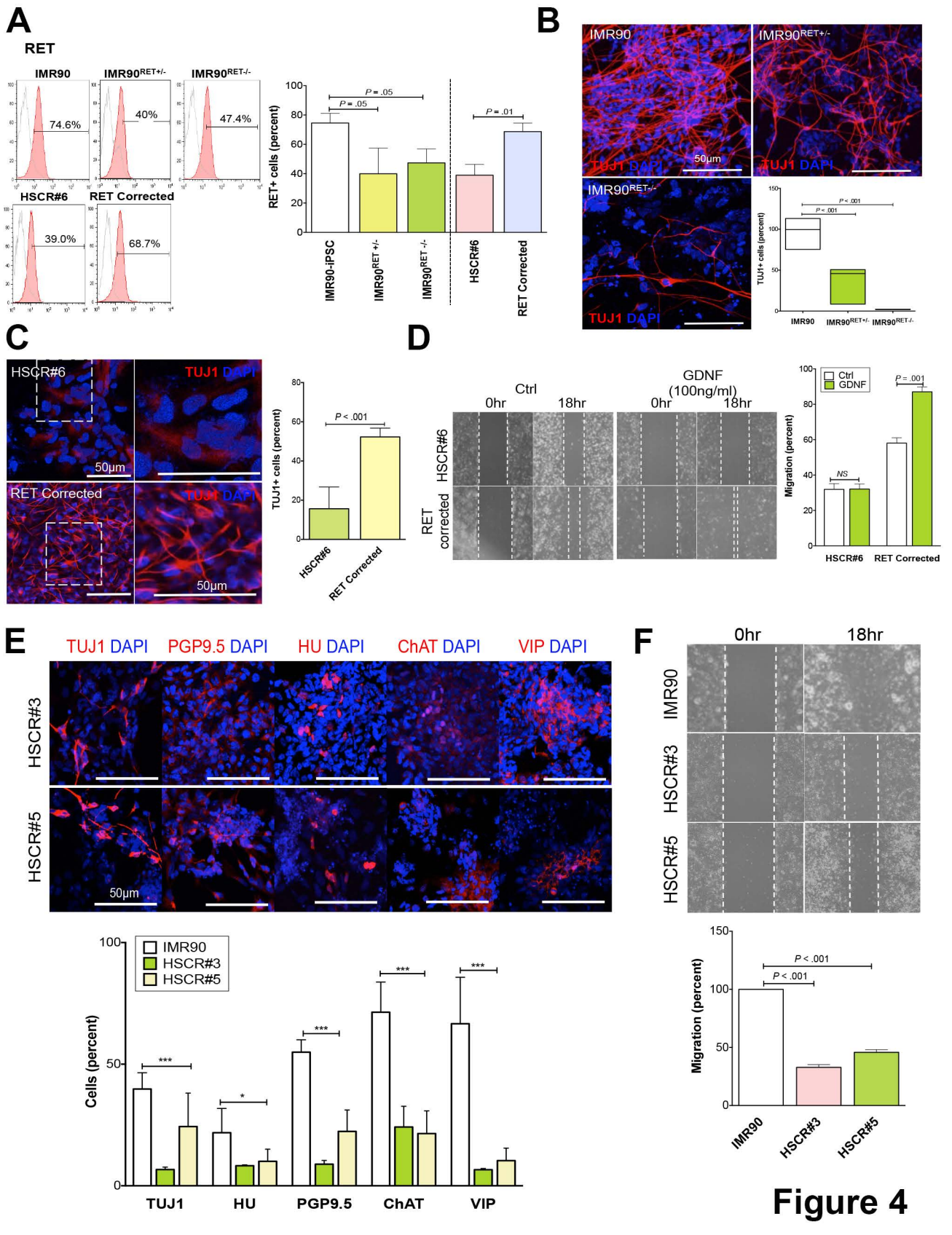

A
RET
IMR90
IMR90RET+/-
IMR90RET-/-
74.6%
40%
47.4%
HSCR#6
RET Corrected
39.0%
68.7%
RET+ cells (percent)
P = .05
P = .05
P = .01
IMR90-iPSC
IMR90RET +/-
IMR90RET -/-
HSCR#6
RET Corrected
B
IMR90
IMR90RET+/-
IMR90RET-/-
TUJ1 DAPI
50µm
TUJ1+ cells (percent)
P < .001
P < .001
IMR90
IMR90RET+/-
IMR90RET-/-
C
HSCR#6
TUJ1 DAPI
50µm
RET Corrected
TUJ1 DAPI
50µm
TUJ1+ cells (percent)
P < .001
HSCR#6
RET Corrected
D
Ctrl
GDNF (100ng/ml)
0hr
18hr
0hr
18hr
HSCR#6
RET corrected
Migration (percent)
Ctrl
GDNF
P = .001
NS
HSCR#6
RET Corrected
E
TUJ1 DAPI
PGP9.5 DAPI
HU DAPI
ChAT DAPI
VIP DAPI
HSCR#3
HSCR#5
50µm
Cells (percent)
IMR90
HSCR#3
HSCR#5
***
*
***
***
***
TUJ1
HU
PGP9.5
ChAT
VIP
F
0hr
18hr
IMR90
HSCR#3
HSCR#5
Migration (percent)
P < .001
P < .001
IMR90
HSCR#3
HSCR#5

Figure 4

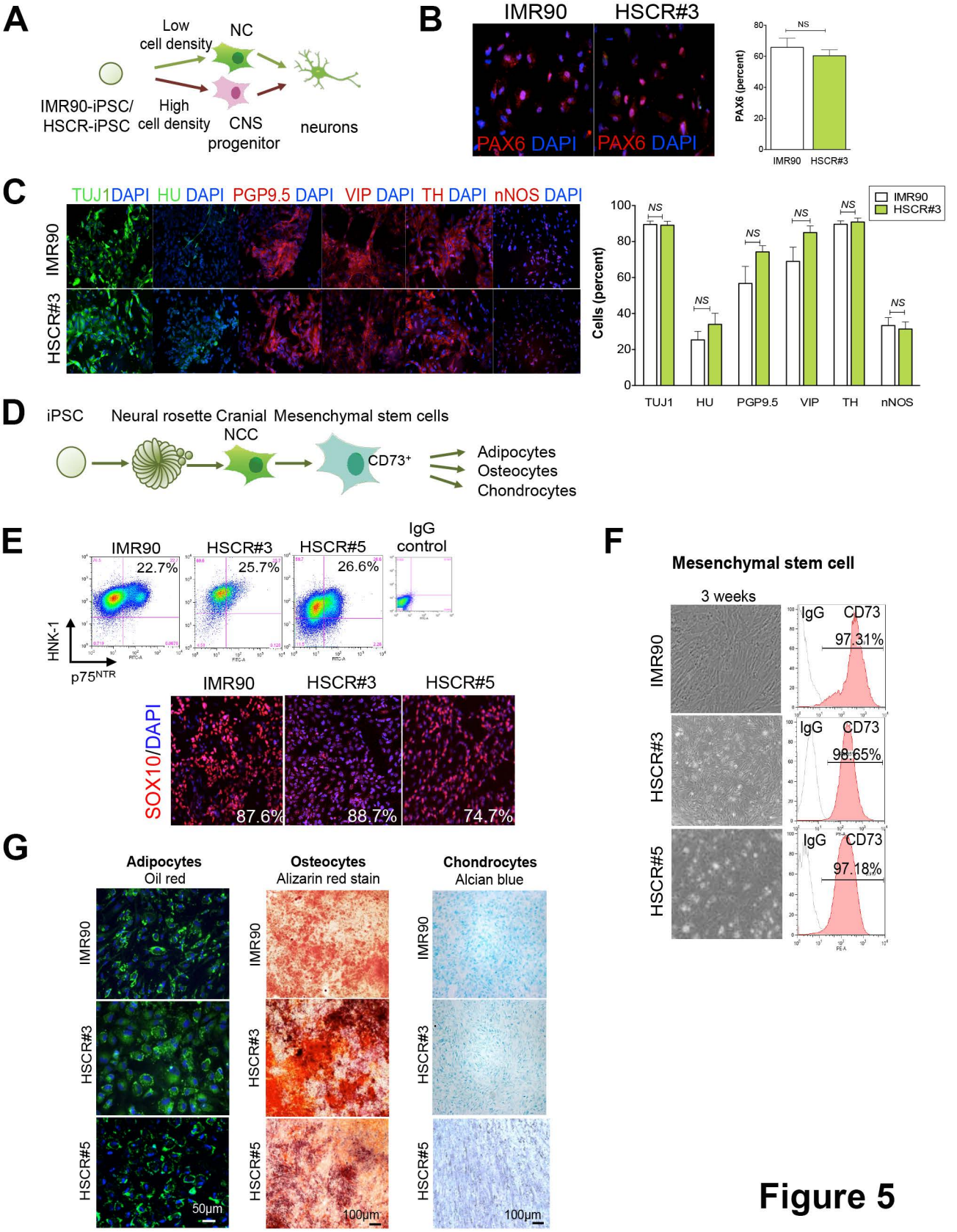

A
Low cell density
NC
IMR90-iPSC/ HSCR-iPSC
High cell density
CNS progenitor
neurons
B
IMR90
HSCR#3
PAX6 DAPI
PAX6 DAPI
NS
PAX6 (percent)
IMR90
HSCR#3
C
TUJ1DAPI HU DAPI PGP9.5 DAPI VIP DAPI TH DAPI nNOS DAPI
IMR90
HSCR#3
Cells (percent)
IMR90
HSCR#3
NS
TUJ1
HU
PGP9.5
VIP
TH
nNOS
D
iPSC
Neural rosette Cranial NCC
Mesenchymal stem cells
CD73+
Adipocytes
Osteocytes
Chondrocytes
E
IMR90
HSCR#3
HSCR#5
IgG control
22.7%
25.7%
26.6%
HNK-1
p75NTR
SOX10/DAPI
87.6%
88.7%
74.7%
F
Mesenchymal stem cell
3 weeks
IMR90
HSCR#3
HSCR#5
IgG
CD73
97.31%
98.65%
97.18%
G
Adipocytes
Oil red
Osteocytes
Alizarin red stain
Chondrocytes
Alcian blue
IMR90
HSCR#3
HSCR#5
50μm
100μm
100μm

Figure 5

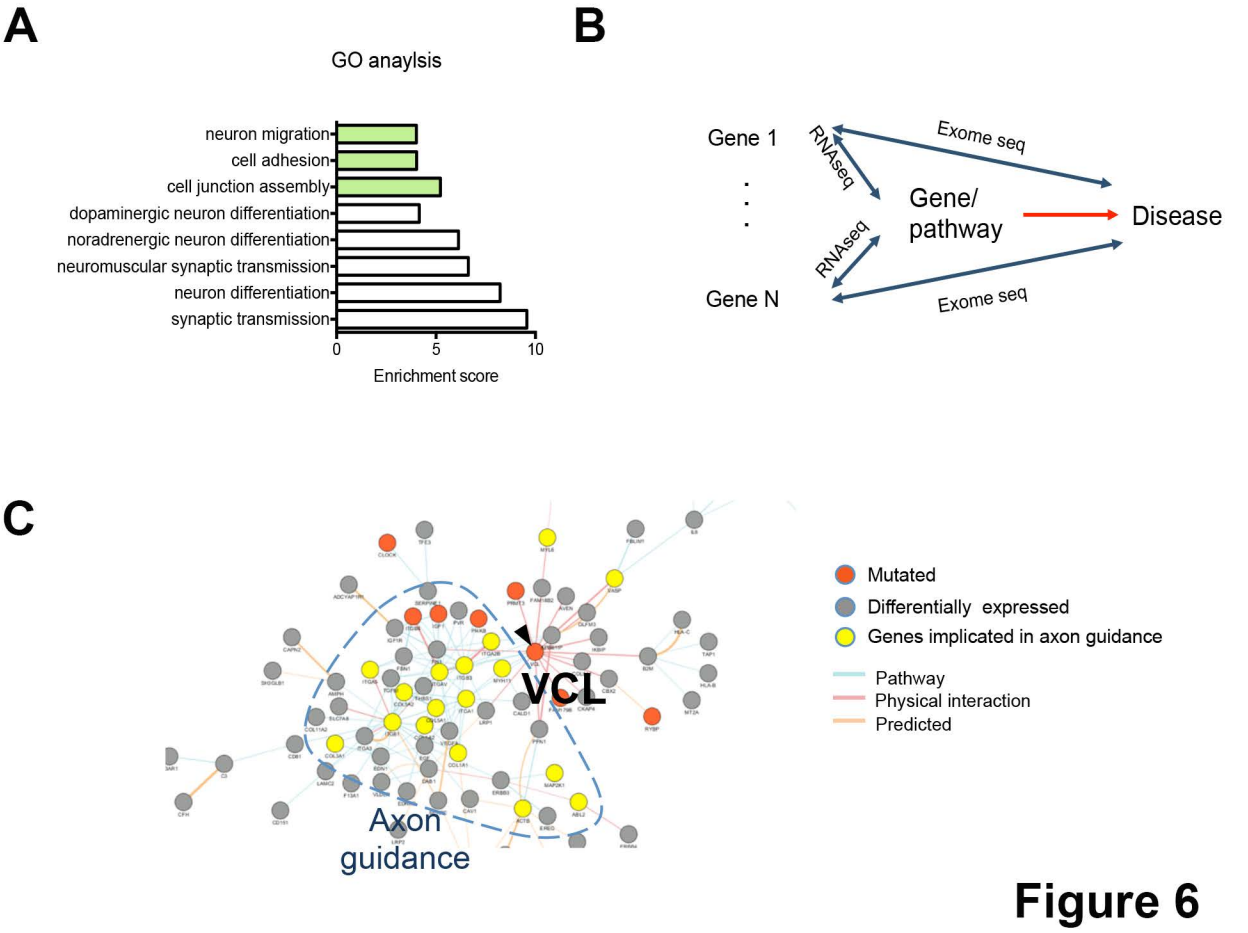

A
GO anaylsis
neuron migration
cell adhesion
cell junction assembly
dopaminergic neuron differentiation
noradrenergic neuron differentiation
neuromuscular synaptic transmission
neuron differentiation
synaptic transmission
0
5
10
Enrichment score
B
Gene 1
Gene N
RNAseq
RNAseq
Exome seq
Exome seq
Gene/
pathway
Disease
C
VCL
Axon
guidance
Mutated
Differentially expressed
Genes implicated in axon guidance
Pathway
Physical interaction
Predicted

Figure 6

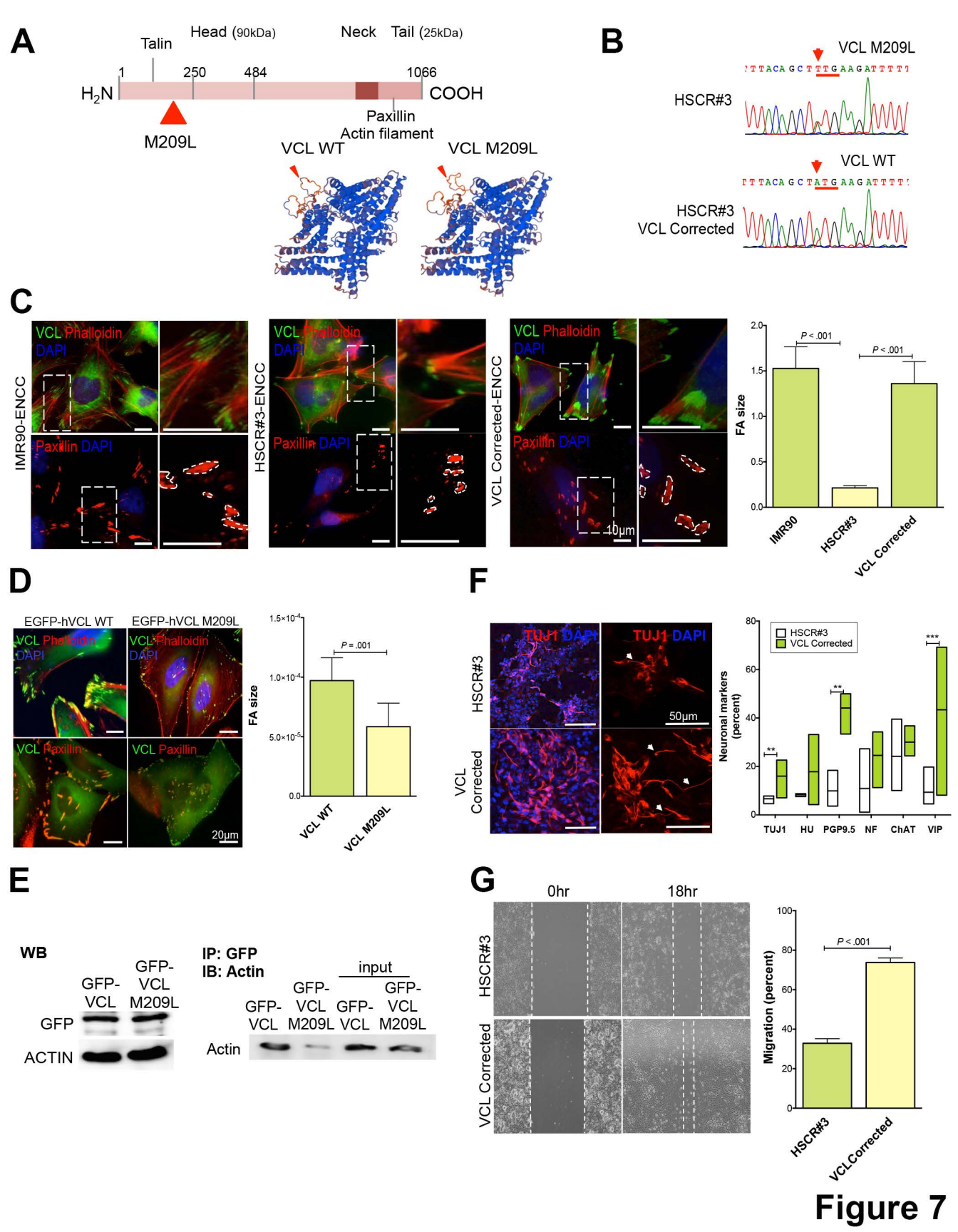
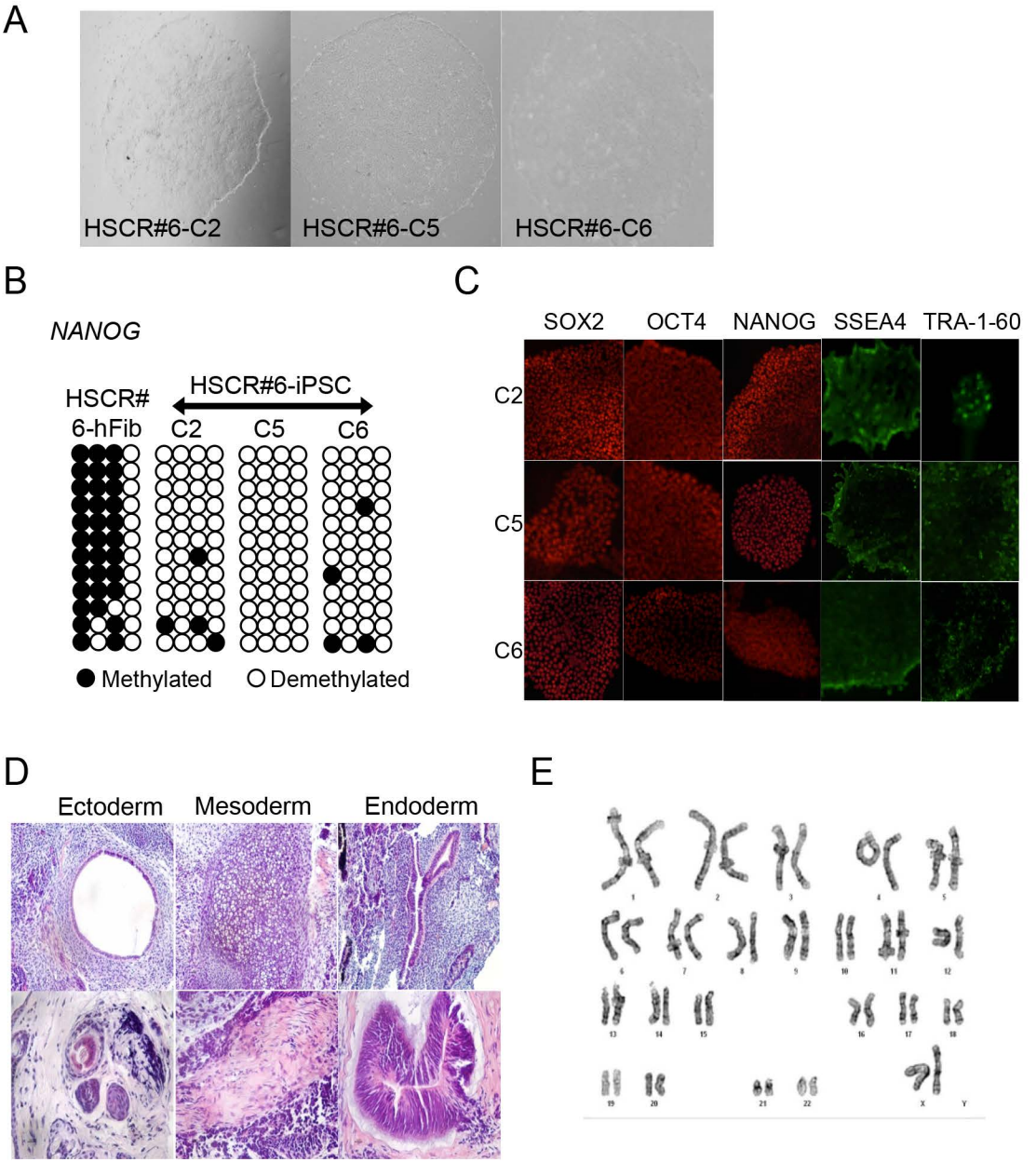
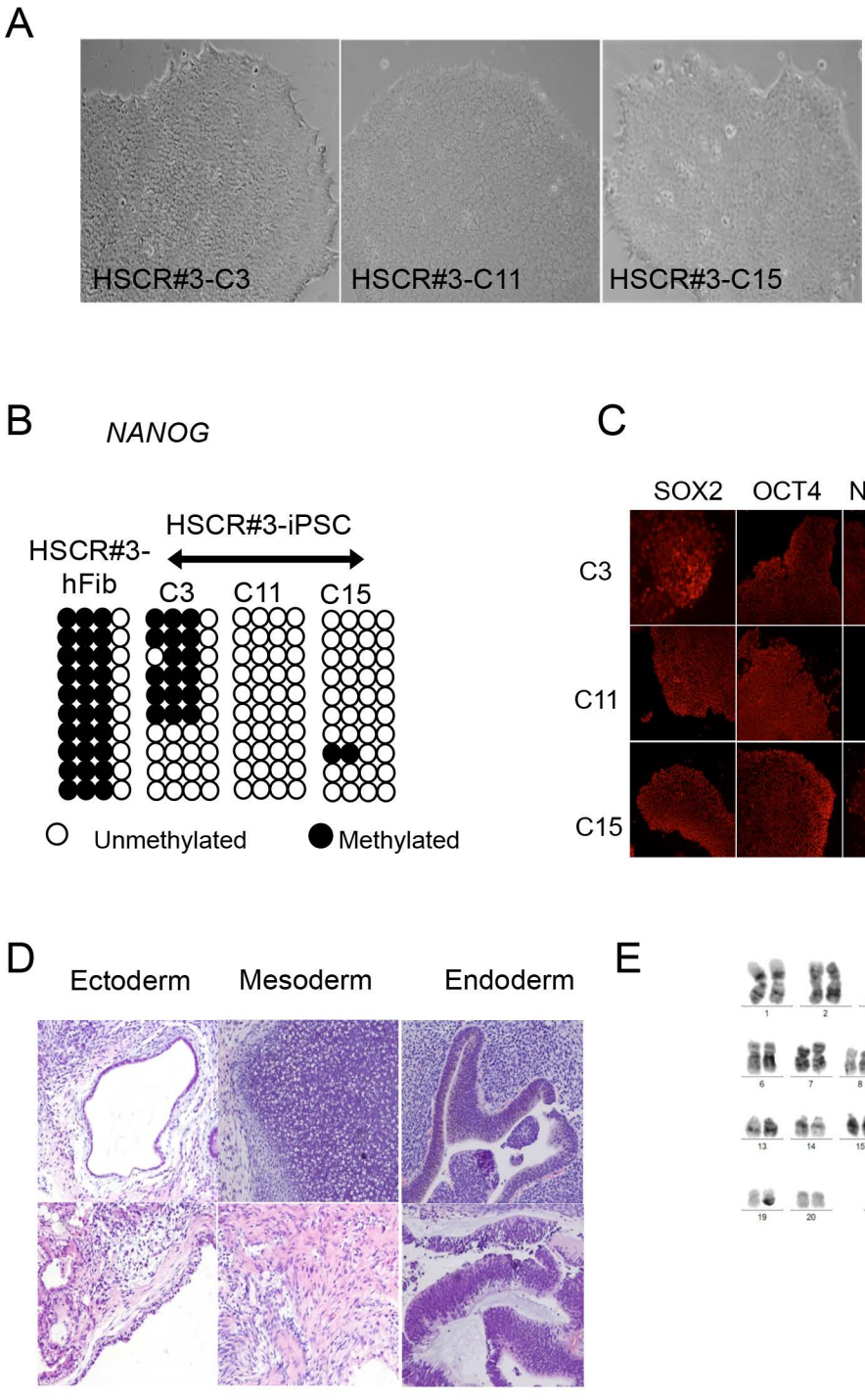

A
Talin
Head (90kDa)
Neck
Tail (25kDa)
1
250
484
1066
H2N
COOH
Paxillin
Actin filament
M209L
VCL WT
VCL M209L
B
VCL M209L
HSCR#3
VCL WT
HSCR#3
VCL Corrected
C
IMR90-ENCC
HSCR#3-ENCC
VCL Corrected-ENCC
VCL Phalloidin
DAPI
Paxillin DAPI
10μm
FA size
P < .001
P < .001
IMR90
HSCR#3
VCL Corrected
D
EGFP-hVCL WT
EGFP-hVCL M209L
VCL Phalloidin
DAPI
VCL Paxillin
20μm
FA size
P = .001
VCL WT
VCL M209L
F
HSCR#3
VCL Corrected
TUJ1 DAPI
50μm
Neuronal markers (percent)
HSCR#3
VCL Corrected
TUJ1
HU
PGP9.5
NF
ChAT
VIP
E
WB
GFP-VCL
GFP-VCL M209L
GFP
ACTIN
IP: GFP
IB: Actin
input
Actin
G
0hr
18hr
HSCR#3
VCL Corrected
Migration (percent)
P < .001
HSCR#3
VCLCorrected

Figure 7

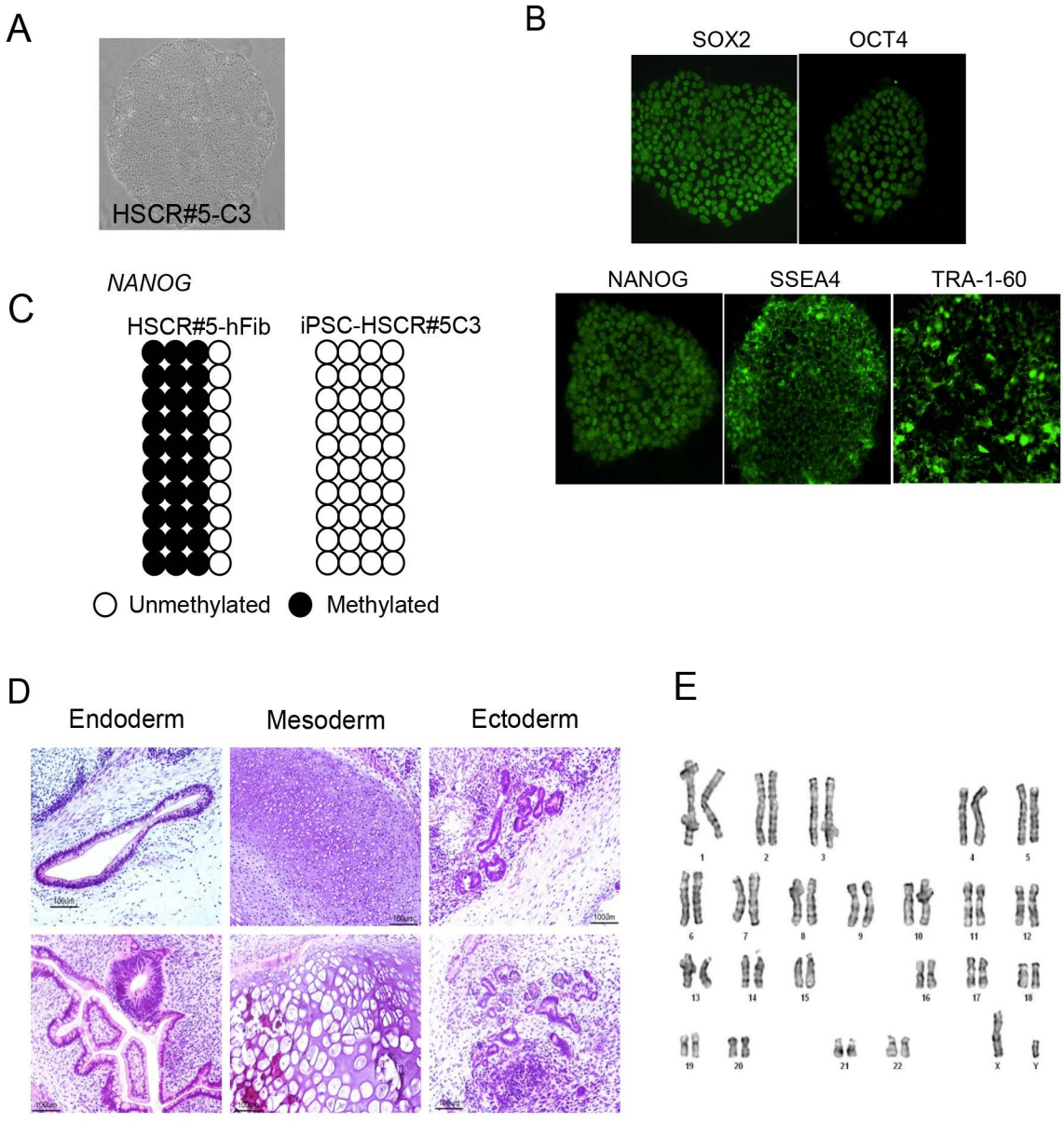


Supplementary Figure 1. Establishment and characterization of HSCR-iPSC#6. (A) Representative pictures of HSCR-iPSC clones. (B) Promoter of *NANGO* is unmethylated in all iPSC clones. (C) Immunocytochemistry analysis on the expression of stem cell markers in 3 independent HSCR-iPSC clones. (D) H & E staining of sections from a tetratoma showed that iPSCs are pluriopotent. (E) Karyotypeof iPSC.



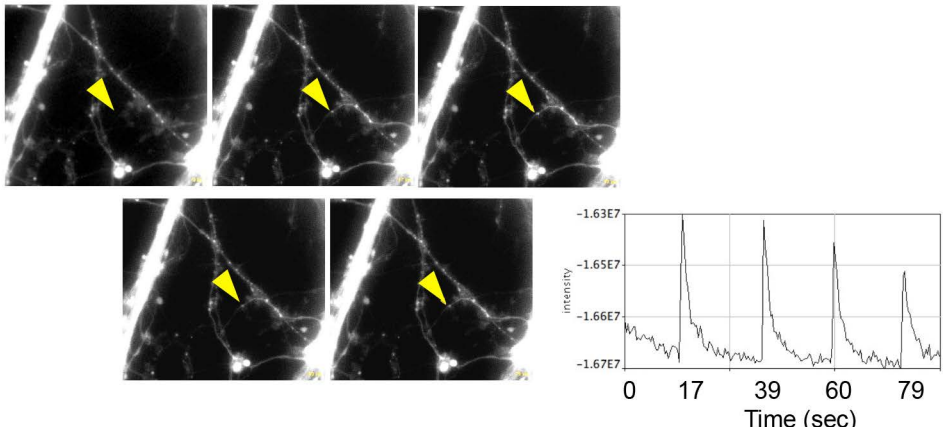
Supplementary Figure 2. Establishment and characterization of HSCR-iPSC#3. (A) Representative pictures of HSCR#3-iPSC clones. (B) Promoter of *NANGO* is unmethylated in all iPSC lines. (C) Immunocytochemistry analysis on the expression of stem cell markers in three HSCR#3-iPSC clones. (D) H & E staining of sections from a tetratoma showed that iPSCs are pluriopotent. (E) Karyotypeof iPSC.

Supplementary Figure 2



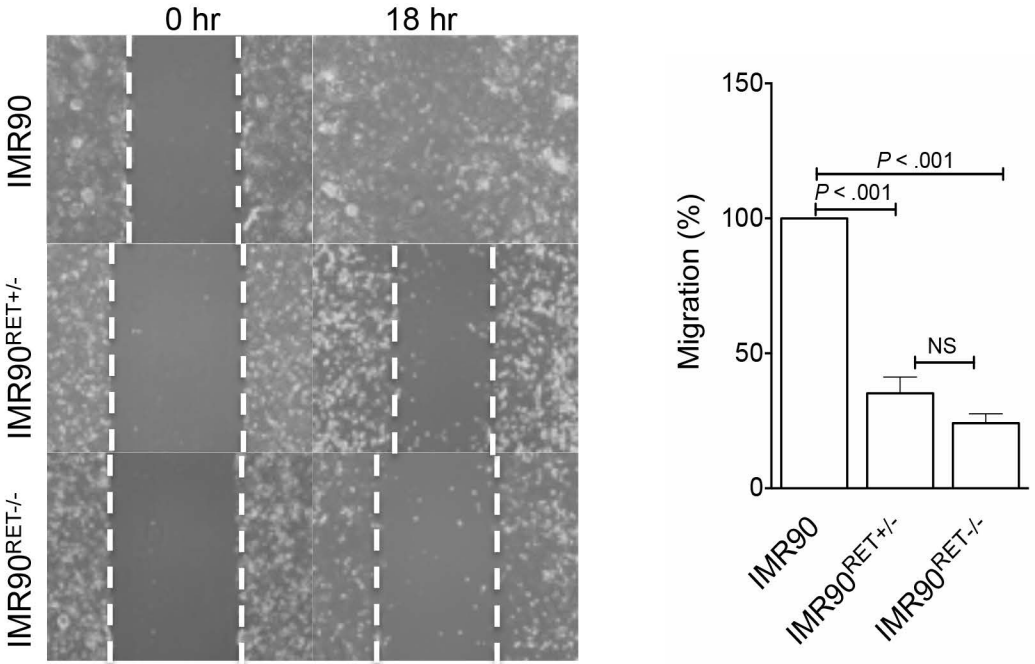
Supplementary Figure 3. Establishment and characterization of HSCR-iPSC#5. (A) Representative pictures of HSCR#5-iPSC. (B) Promoter of *NANGO* is unmethylated in iPSC. (C) Immunocytochemistry analysis on the expression of stem cell markers in one of the HSCR#5-iPSC clones. (D) H & E staining of sections from a tetratoma showed that iPSCs are pluriopotent. (E) Karyotype of iPSC.

Supplementary Figure 3



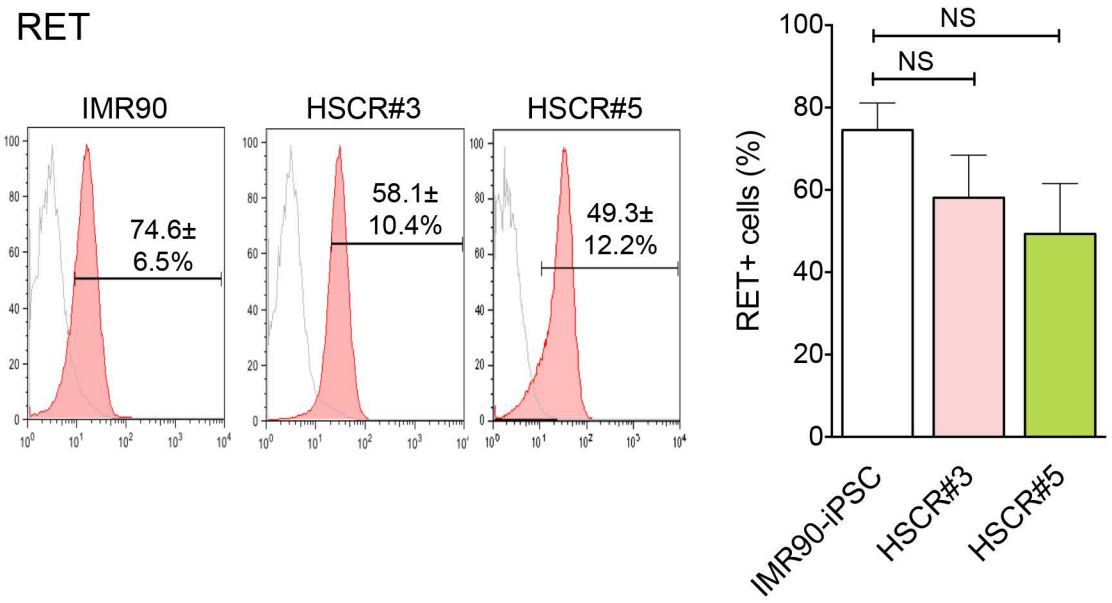
Supplementary Figure 4. Calcium transients of neurons derived from the ENCCs upon electric stimulation. Ca2+ dynamics was monitored using the Ca2+-sensitive dye Fluo-4 AM. Fluorescent signals (marked by arrows) were detected every 20 seconds.

Supplementary Figure 4



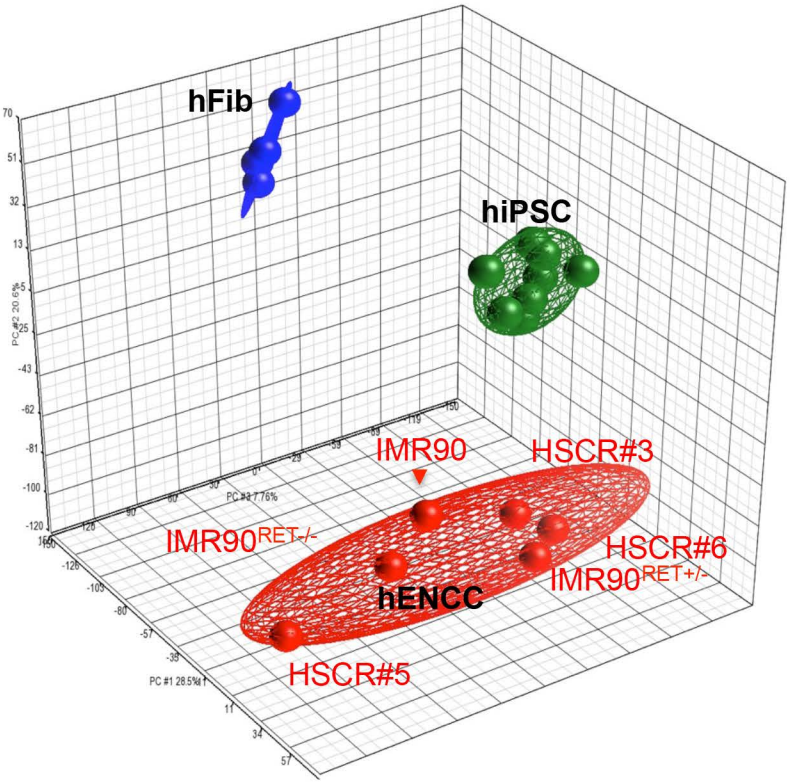
Supplementary Figure 5. Scratch assays were performed to measure the migratory ability of iPSC-derived ENCCs from the control (IMR90) and two isogenic RET mutant lines. The bar charts show wound closure measured as the percentage of scar width over time (18 h).

Supplementary Figure 5



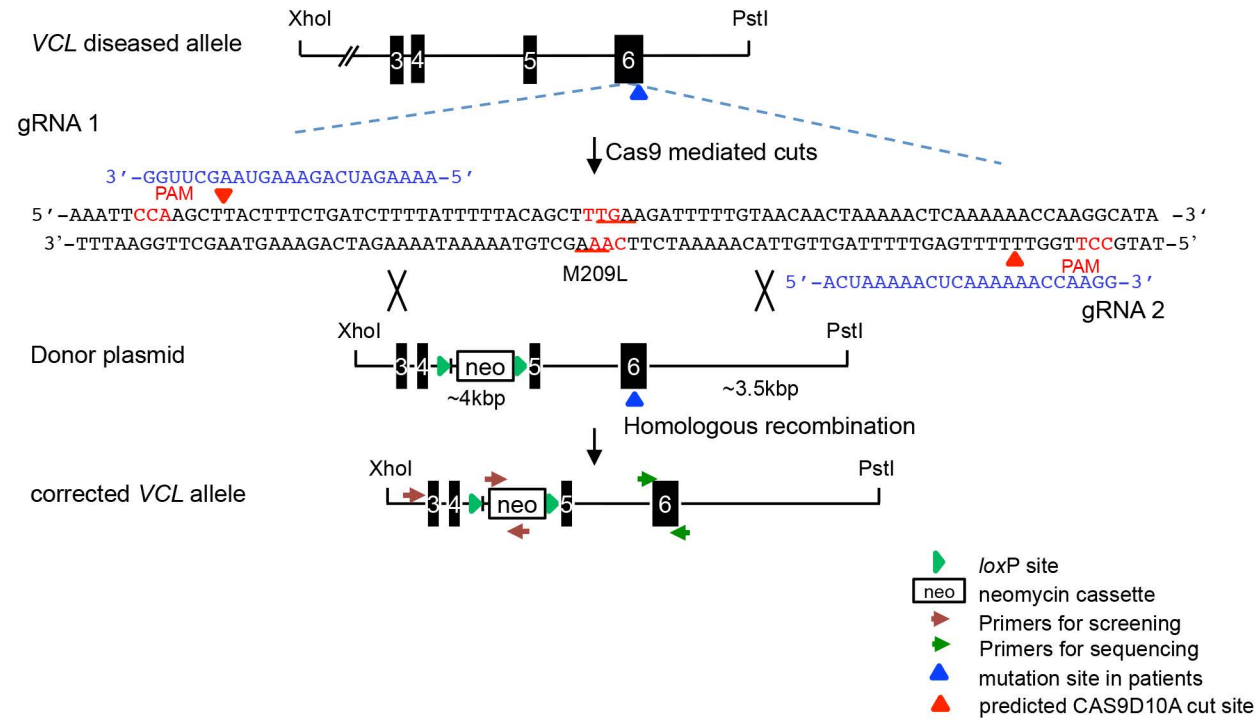
Supplementary Figure 6. Comparable RET expression levels in NCCs derived from control IMR90-iPSC and HSCR-iPSC#3 & #5. Flow cytometry analysis on the expression of RET in NCCs derived from control(IMR90) and patient iPSC (HSCR#3 & HSCR#5).

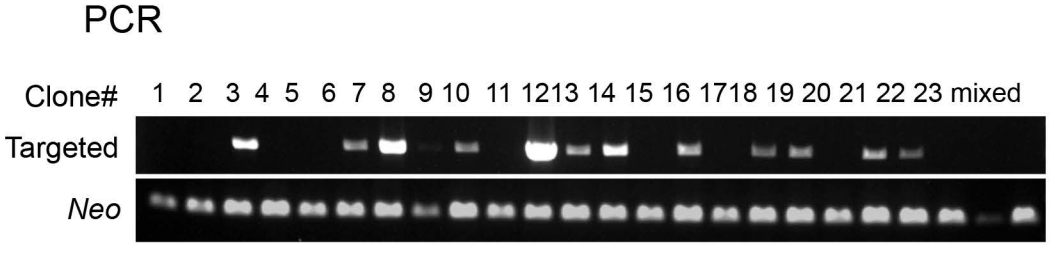
Supplementary Figure 6



Supplementary Figure 7: PCA of the transcription profiles of fibroblasts, iPSCs and ENCCs from controls and patients.

Supplementary Figure 7





Supplementary Figure 8. Correction of VCL M209L mutation in HSCR#3-iPSC. Schematic diagram of the targeting strategy to correct the M209L mutation in *VCL*. Primers designed for screening (brown arrows) and sequencing (green arrows) are shown. Representative PCR picture are shown in the lower panel. PCR product of 1.6 kb were detected in the targeted clones. The DNA sequence of VCL was then confirmed by to Sanger sequencing.

Supplementary Figure 8

hiPSC lines	Gender	Phenotype	RET Rs2435357 (C/T)	Known mutation
IMR90	F	WT	C/C	-
IMR90$^{RET+/-}$	F	-	C/C	RET E734Kfs*#40 =>/+
IMR90$^{RET-/-}$	F	-	C/C	RET G606X/G601-602del
HSCR#6	F	TCA	C/C	RET G731del/+
HSCR#3	M	S-HSCR VSD/ASD	T/T	VCL M209L
HSCR#5	M	S-HSCR	T/T	N.D.
HSCR#3 VCL Corrected	M	-	T/T	VCL WT
HSCR#6 RET Corrected	F	-	C/C	RET WT

Supplementary Table 1: List of human induced pluripotent stem cell (hiPSC) lines used in this study. F: female; M: male; WT: wild-type; S-HSCR: short HSCR; TCA: total colonic aganglionsis; VSD/ASD: ventricular septal defect/atrial septal defect; N.D.: not determined.

Supplementary Table 2: List of oligos used in this study

Target gene	Oligo sequence	Annealing Temp (°C)	Product Size (bp)	Experiment used
RET Exon 10	gRNA #1: 5'-CATAGCCAGCTTTAATCCCCCGGG-3' gRNA #2: 5'-ACCTGCAACTGCTTCCCTGAGGT-3'	-	-	Gene targeting
RET Exon 12	gRNA #1: 5'-CAAGAACCAAGTTCTTCCGAGGG-3' gRNA #2: 5'- CGAATTTGGAAAACTCCTCAAGG-3'	-	-	Gene targeting
VCL Exon 6	gRNA #1: 5'-CAAGAACCAAGTTCTTCCGTGGG-3' gRNA #2: 5'-CGAATTTGGAAAAGTGGTCAAGGC-3'	-	-	Gene targeting
RET Exon 10	Forward: 5'- AGGCTGAGTGGGCTACGTCT-3' Reverse: 5'- TGCTGTTGAGACCTCTGTGG-3'	58	208	PCR sequencing
RET Exon 12	Forward: 5'- AGCCATTCACAGTCCAGGTC-3' Reverse: 5'- CACGTAAGCAATACGGGTCA-3'	58	285	PCR sequencing
VCL Exon 6	Forward: 5'- AAAGCCCAAAACATCTAAAGTG-3' Reverse: 5'- GGGGATTTTGCTTATTTCTCC-3'	58	295	PCR sequencing
VCL	Forward: 5'- CAGTCAGACCCTTACTCAGTGC-3' Reverse: 5'- GCTAAAGCGCATGCTCCAGAC-3'	58	1588	PCR
VCL	Forward: 5'-GTTGTTACAAAAATCTTCAAAGCTGAAATGAGAACTGGCAGC-3' Reverse: 5'- GCTGCCAGTTCTCATTTCAGCTTTGAAGATTTTTGTAACAAC-3'	58	-	Mutagenesis
sihVCL	sRNA#1: 5'-GGAAGAAAUCACAGAAUCAUU-3' sRNA#2: 5'-CCAGAUGAGUAAAGGAGUAUU-3'	-	-	knockdown
VCL Exon 12	gRNA #1 5'-GAAAGATCAGAAAGTAAGCT-3 gRNA #2 5'-ACTAAAAACTCAAAAAACCAAGG- 3'			
RET	Forward: 5'-ACACCAAGGCCCTGCGGCG-3' Reverse: 5'-GGAAGGTCATCTCAGCTGAG-3'	56	681	PCR
HOXB3	Forward: 5'- CCAGTGCCACTAGCAACAGC -3' Reverse: 5'- CGTTTGCCTCGACTCTTTCATC -3'	60	128	PCR
HOXB4	Forward: 5'- GACCATTACCTCGACACCCG -3' Reverse: 5'- AACACTGCACTATCTGGCAGTCAG -3'	60	157	PCR
HOXB5	Forward: 5'- AAACTCCTTCTCGGGGCGTTAT -3' Reverse: 5'- CATCCCATTGTAATTGTAGCCGT -3'	60	139	PCR
SOX10	Forward: 5'- ACTTCGGCAACGTGGACATT -3' Reverse: 5'- CAGCCACATCAAAGGTCTCCAT -3'	72	60	PCR
SNAIL2	Forward: 5'- ATGAGGAATCTGGCTGCTGT -3' Reverse: 5'- CAGGAGAAAATGCCTTTGGA -3'	60	119	PCR
ACTIN	Forward: 5'- ACTCTTCCAGCCTTCCTTCC -3' Reverse: 5'- CGTCATACTCCTGCTTGCTG -3'	60	291	PCR

Supplementary Table 3: List of antibodies used in this study

Target Protein	Description	Company Cat no.	Working conc	Experiment used
RET	APC-conjugated anti-human RET	Neuromics FC15017	-	Flow cytometry
RET	Goat anti-RET	Neuromics GT15002	1:1000	WB
HNK-1	APC-conjugated anti-human CD57 (HNK1)	BD Pharmingen 560845	-	FACS/flow cytometry
$p75^{NTR}$	FITC-conjugated anti-human CD271 ($p75^{NTR}$)	Miltenyi Biotec 130-091-917	-	FACS/flow cytometry
Integrin α	FITC-conjugated anti-human CD49d	Biolegend 304414		FACS/flow cytometry
SOX2	Goat anti-SOX2	R&D system MAB2018	1:500	IF
SOX10	Mouse anti-SOX10	R & D system MAB2864	1:500	IF
OCT-4	Goat anti-OCT-4	Santa Cruz SC-5279	1:100	IF
NANOG	Goat anti-NANOG	R&D system AF1997	1:500	IF
SSEA-4	Mouse anti-SSEA-4	R&D system FAB1435P	1:200	IF
TRA-1-60	Mouse anti-TRA-1-60	Millipore MAB4360	1:100	IF
HNK1	Mouse monoclonal anti-HNK1	Sigma C6680	1:200	IF
$p75^{NTR}$	Rabbit anti- $p75^{NTR}$	Abcam Ab8874	1:500	IF
TUJ1	Mouse monoclonal anti-TUJ1	Convance PRB-435P	1:500	IF
HU	Mouse monoclonal anti-HU	Invitrogen A21275	1:500	IF
TH	Rabbit anti-TH	Novus NB300-109	1:200	IF
VIP	Rabbit anti-VIP	Millipore AB1581	1:200	IF
PGP9.5	Rabbit anti-PGP9.5	Abcam ab10404	1:100	IF
ChAT	Rabbit anti-ChAT	Millipore AB143	1:200	IF
nNOS	Rabbit anti-nNOS	Cell Signalling 4231	1:200	IF
Neurofilament	Mouse anti-Neurofilament	Sigma N5139	1:200	IF
PHOX2B	Goat anti-PHOX2B	Santa Cruz Sc-13224	1:100	IF
GFAP	Rabbit anti-GFAP	DAKO Z0334	1:200	IF
S100B	Rabbit anti-S100B	Sigma S2644	1:500	IF
PAX6	Rabbit anti-PAX6	DSHB	1:200	IF
SM22α	Rabbit polyclonal anti-SM22 alpha	Abcam Ab14106	1:500	IF
CALPONIN	Rabbit monoclonal anti Calponin	Abcam Ab46794	1:100	IF
PAXILLIN	Mouse monoclonal anti Paxillin line 5H11	Millipore 05-417	1:2000	IF
PHALLOIDIN	Alexa Flour 594 Phalloidin	ThemoFisher	1:100	IF

		A12381		
VINCULIN	Mouse monoclonal anti-Vincullin	Sigma A4731	1:2000	IF/IP
GFP	Rabbit polyclonal anti-GFP	Abcam AB290	1:300	IP
ACTIN	Mouse monoclonal anti-actin	Millipore MAB1501	1:10000	WB
Mouse IgG1 κ	APC-conjugated mouse IgG1 κ isotype control	BD Pharmingen 554681	-	FACS/flow cytometry
Mouse IgG/IgM	FITC-conjugated goat anti-mouse IgG/IgM	BD Pharmingen 555988	-	FACS/flow cytometry
Mouse IgM	APC-conjugated rat anti-mouse IgM	BD Pharmingen 550676	-	FACS/flow cytometry
Mouse IgG	HRP conjugated anti-mouse antibody	DAKO	1:2500	WB
Rabbit IgG	HRP conjugated anti-rabbit antibody	DAKO	1:2500	WB