의학석사 학위논문

Establishment of an Experimental Protocol for Neural Crest Cell Lineage Tracing in Mouse Embryos

쥐 배아에서 신경능선세포의
계통 추적을 위한 실험 프로토콜 확립

2020 년 8 월

서울대학교 대학원
의과학과 의과학전공
이 자 원

A thesis of the Master's degree

쥐 배아에서 신경능선세포의 계통 추적을 위한 실험 프로토콜 확립

Establishment of an Experimental Protocol for Neural Crest Cell Lineage Tracing in Mouse Embryos

August, 2020

The Department of Biomedical Sciences

Seoul National University

College of Medicine

Jawon Lee

Establishment of an Experimental Protocol for Neural Crest Cell Lineage Tracing in Mouse Embryos

지도교수 최 무 림

이 논문을 의학석사 학위논문으로 제출함
2020 년 5 월

서울대학교 대학원
의과학과 의과학전공
이 자 원

이자원의 석사 학위논문을 인준함
2020 년 7 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Establishment of an Experimental Protocol for Neural Crest Cell Lineage Tracing in Mouse Embryos

by
Jawon Lee

A thesis submitted to the Department of Biomedical Sciences
in partial fulfillment of the requirements
for the Degree of Master of Science in Medicine
at Seoul National University College of Medicine

July, 2020

Approval by Thesis Committee:

Professor _____Chairman

Professor _____Vice chairman

Professor _____

ABSTRACT

Establishment of an experimental protocol for neural crest cell lineage tracing in mouse embryos

Jawon Lee

Department of Biomedical Sciences

College of Medicine

Seoul National University

Introduction: Neural crest cells (NCCs) are multipotent cells, which arise in the ectoderm during embryogenesis, migrate through the embryo and give rise to a wide range of tissues including smooth muscle cells, melanocytes, Schwann cells and neurons. Accumulated evidence over the past decades suggests that intricate molecular regulation underlies NCC development. Any errors in this process result in congenital defects in humans, including cleft lips and palates, inherited forms of melanoma, DiGeorge/Velo-cardio-facial syndrome (DGS/VCFS), persistent truncus arteriosus and patent ductus arteriosus. Despite accumulating studies, accurate genetic picture underlying the complex nature of NCC migration and differentiation remains elusive. I explored NCC development using lineage-traceable cell clones in developing mouse embryos and established experimental protocols for this purpose.

Methods: NCC lineage was traced using CellTag lentivirus in *Wnt1-cre;R26R-tomato* mouse embryos. *In vitro* culture was established to study trunk NCC development of embryonic day (E) 9.5 mouse embryos for a 24-hour period and *in utero* injection was established to study NCC development of E12.5 to E18.5 mouse embryo brain for six days. Single-cell sequencing (scRNA-seq) of the E10.5 embryonic torso and the E18.5 embryonic brain was performed and sequencing results were analyzed.

Results: A method to inject CellTag lentivirus, a lineage tracing system, into E9.5 mouse embryos was designed. A culture chamber method for 24-hour *in vitro* culture was established, and tdTomato fluorescence was observed in the embryo after the culture. To model DGS in mouse, the *LgDel* (Large Deletion) embryos were recovered using the cryopreservation method, and the wildtype and *LgDel* mice were propagated. As a result of analyzing scRNA-seq data from E10.5 embryos, cardiac muscle and vascular development clusters were uncovered. In contrast, single-cell analysis of E18.5 mouse embryo brain resulted in glia cell and neuron differentiation clusters.

Discussion: I aimed to trace NCC lineage to resolve the complex molecular mechanism of NCC development in mouse embryos. For this purpose, a lineage tracing system called CellTag was used. However, in order to obtain reads mapped to the CellTag region, an additional PCR step using CellTag specific primers would be necessary. In conclusion, I established experimental protocols for lineage tracing with scRNA-seq, with minor modifications. They can be used to gain insight into the complex relationships between developing cells and to discover novel genes involved in NCC development.

Keywords: Neural crest cell, Mouse embryo development, Single-cell sequencing, DiGeorge syndrome, *In vitro* culture, *In utero* transplantation

Student Number: 2018-28442

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

BAM	Binary Alignment Map
BMP	Bone morphogenetic protein
CIGAR	Concise Idiosyncratic Gapped Alignment Report
CFU	Colony-forming units
DGS/VCFS	DiGeorge/Velo-cardio-facial syndrome
DGS	DiGeorge syndrome
DGE	Digital gene expression
DMEM	Dulbecco's Modified Eagle Medium
DPBS	Dulbecco's phosphate-buffered saline
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced Green Fluorescent Protein
E	Embryonic day
FBS	Fetal Bovine Serum
Fgf8	Fibroblast growth factor 8
GTF	Gene transfer format
GFP	Green fluorescent protein
HEK 293	Human embryonic kidney 293 cell
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IUT	*In utero* transplantation
IVF	*In-vitro* fertilization
LB	Luria-Bertani
MOI	Multiplicity of infection
$NaHCO_3$	Sodium bicarbonate

NCC	Neural crest cell
NGS	Next-Generation Sequencing
nCount	Number of RNA counts
nFeature	Number of feature RNA
PDA	Patent ductus arteriosus
PBS	Phosphate-buffered saline
PBS-BSA	Phosphate-buffered saline-Bovine serum albumin
PCR	Polymerase chain reaction
PTA	Persistent truncus arteriosus
PA	Pharyngeal arch
Poly-A	Polyadenylation
PEI	Polyethylenimine
PCA	Principal component analysis
qPCR	Quantitative polymerase chain reaction
QC	Quality control
RCF	Relative centrifugal force
RFP	Red fluorescent protein
RPMI	Roswell Park Memorial Institute Medium
SSC	Saline-sodium citrate
scRNA-seq	Single-cell RNA sequencing
SOC	Super Optimal broth with Catabolite repression
STAMP	Single-cell transcriptomes attached to microparticles
Tbx1	T-box containing transcription factor
Tg	Transgenic
TGF-β	Transforming growth factor beta
UMAP	Uniform Manifold Approximation and Projection

UMI	Unique molecular identifier
UTR	Untranslated region
WGS	Whole Genome Sequencing

INTRODUCTION

Neural crest cell development in early embryos

During embryogenesis, neural crest cells (NCCs) originate from the dorsal side of the embryo, migrate to the ventral side and give rise to multiple organs[1] (Figure 1). In the following section, I will shortly summarize the current knowledge on mouse embryogenesis.

In the beginning of mouse embryogenesis, the zygote undergoes rapid cell cycles with no significant overall growth, producing a cluster of cells the same size as the original zygote. Cleavage ends with the formation of the blastula. Then, gastrulation, a phase in embryonic development during which the single layered blastula develops into a three-layered gastrula, takes place. By the end of gastrulation, the embryo has begun differentiation to establish distinct cell lineages and has set up the basic axes of the body such as dorsal-ventral and anterior-posterior. In triploblastic organisms the gastrula is three-layered. These three germ layers are known as ectoderm, mesoderm, and endoderm. Each layer gives rise to specific tissues and organs in the developing embryo. The formation of the neural crest takes place during the process of neurulation. Neural crest cell migration involves an initial epithelial-mesenchymal transition to delaminate from the ectoderm layer. Then these neural crest cells, depending upon the rostrocaudal level within the embryo, migrate throughout the embryo with distinct morphological patterns.

Neural crest cells differentiate into critical components of facial bones, muscles, thyroid and thymus (cranial NCCs), vasculature, septa of the heart (cardiac NCCs), skin, enteric neurons and kidneys (trunk NCCs). NCC development requires a

sophisticated differentiation process through the interplays between the neural crest and the surrounding cells along the anterior-posterior axis of the embryo.

Manipulation of genes or tissues involved in NCC development in murine and avian experimental systems has enabled better understanding of the cellular and molecular mechanisms of the developmental process. For example, the abolition of Tbx1, a T-box containing transcription factor, in mice causes a variety of organ defects, strikingly resembling human DiGeorge syndrome(DGS)[2]. The deletion of Fgf8, a downstream factor of Tbx1, in mice likewise results in a phenotype similar to DGS[3]. Bone morphogenetic proteins (BMPs) are signaling proteins in the TGF-β superfamily. Perturbation of the BMP pathway causes cardiovascular and craniofacial defects due to abnormal NCC development[4,5]. Although a number of candidate genes have been linked to these processes, the causal relationships between the current genetic knowledge and the implicated human diseases remain largely unclear.

To overcome our limited understanding of NCC development, investigating the tissue- and stage-specific expression patterns of entire genes, as well as their isoform-specific expression patterns, will be crucial. It is also likely that most of the human embryos with the above-mentioned genetic abnormalities will develop extremely severe defects leading to preliminary pregnancy termination, whereas existing human patients may carry a milder phenotype that has not been fully addressed by the existing list of candidate genes. This plausible scenario concerning the discrepancy between mouse models and human genetics further suggests that there are undiscovered genetic factors that are involved in modulating the NCC development process. Therefore, studying the early development of NCC is essential.

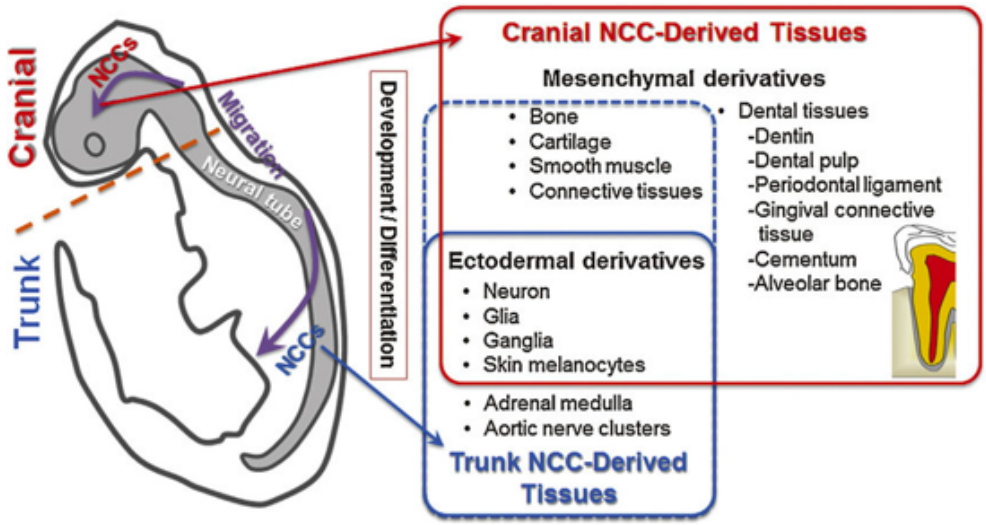

Figure 1. Contributions of different NCC populations to tissues and organs

Migrating NCC and the congenital malformations when NCC development is perturbed (adapted from K. Niibe et al., 2016).

The diseases of defective NCC development

Perturbation of NCC development will lead to a diverse spectrum of human diseases. At the point of writing, Online Mendelian Inheritance in Man (OMIM) lists 188 human diseases directly associated with NCC development, ranging from rare congenital defects, such as DiGeorge/Velo-cardio-facial syndrome (DGS/VCFS), persistent truncus arteriosus (PTA) and patent ductus arteriosus (PDA) to relatively common diseases, such as cleft lips and palates and inherited forms of melanoma.

One of the abnormalities associated with the neural crest is DiGeorge syndrome (DGS). DGS which is also known as 22q11.2 deletion syndrome is a disorder caused by the deletion of a small segment of chromosome 22. It is a congenital disorder in which there are developmental abnormalities of the third and fourth pharyngeal pouches, including thymic hypoplasia, hypoparathyroidism, and midline cardiac defects[6]. NCCs play a major role in the development of the pharyngeal arches, and defects in these cells are likely responsible for the syndrome[7]. It is typically due to the deletion of 30 to 40 genes in the middle of chromosome 22 at a location known as 22q11.2[8]. About 90% of cases are linked to this deletion, which most often occurs as a random event during the formation of reproductive cells or in early fetal development, while 10% are inherited from a person's parents[9]. The inheritance of DGS is considered autosomal dominant because a deletion in one copy of chromosome 22 in each cell is sufficient to cause the condition.

DiGeorge syndrome patients suffer from various brain-related diseases such as schizophrenia (~30%) and autism (~25%). However, the cause of brain-related diseases in DiGeorge syndrome patients still remains unknown. One of the main questions addressed in this study is whether the brain is affected by neural crest cells.

Using single-cell sequencing to understand NCC development

Many genomic approaches are available to interrogate developmental processes. However, many of these approaches require tremendous amounts of cells and tissues. Moreover, these approaches only yield correlative data on gene activity, and generating a dynamic view of genomic regulation remains a great challenge due to limited temporal and regulatory resolution afforded by these approaches.

Single-cell RNA sequencing (scRNA-seq) has several roles during cell development such as displaying complex and unique cell populations and detecting regulatory relationships among the genes and the trajectories of distinct cell lineages[10–13].

In a Drop-seq experiment, prior to the division of two liquid currents into individual droplets, a modified microfluidic tool merges the currents. The first current obtains cells while the other current holds barcoded primer beads suspended in a lysis buffer. The two currents then form a droplet containing the cell and the bead. In seconds, the cell lysis allows mRNAs to be free, and RNA hybridization to the primers on the microparticle's outer layer starts. After applying a reagent, a disruption in the inner area of the cell lysis occurs and leads to burst the droplets, the beads are captured and rinsed afterward. Finally, reverse transcription of mRNAs with template switching generates STAMPs, and these STAMPs work as a template to present a PCR handle downstream of the synthesized cDNA[14,15]. To analyze the result, the first read indicates the cell barcode and unique molecular identifiers (UMI), and the second read in each pair contains the sequence from the cDNA. In order to define the transcript's gene of origin, the sequence is aligned to the genome. A Drop-seq library on a high-throughput sequencer produces over millions of paired-

end reads. The reads are aligned to a reference genome to determine the gene-of-origin of the cDNA, and the cell barcodes arrange the reads. Finally, each UMI in an individual cell is calculated for each gene[15,16] (Figure 2).

The basic unit of single-cell sequencing analysis is a single cell, but there is a considerable amount of information missing from just looking at a single cell because the limits of sensitivity do not allow it to obtain information about all RNA molecules in the cell. Therefore, the key to single-cell sequencing analysis is to classify information obtained from multiple cells and collect it among similar cells, then group the groups of similar cells into a single unit. This may lead to the question of whether or not it is different from the existing bulk RNA sequencing. But the important difference is that bulk RNA sequencing is to analyze the mean values of all cells in a specimen, while single-cell sequencing will outline each cell within the specimen, determine how many different cell types exist, and then create and analyze the mean values for each cell type. Thus, it is possible to obtain additional information about cell-type specific expression and a sensitivity comparable to traditional RNA sequencing for each cell group. I leverage single-cell sequencing analysis to decipher the dynamic molecular regulation of NCC development.

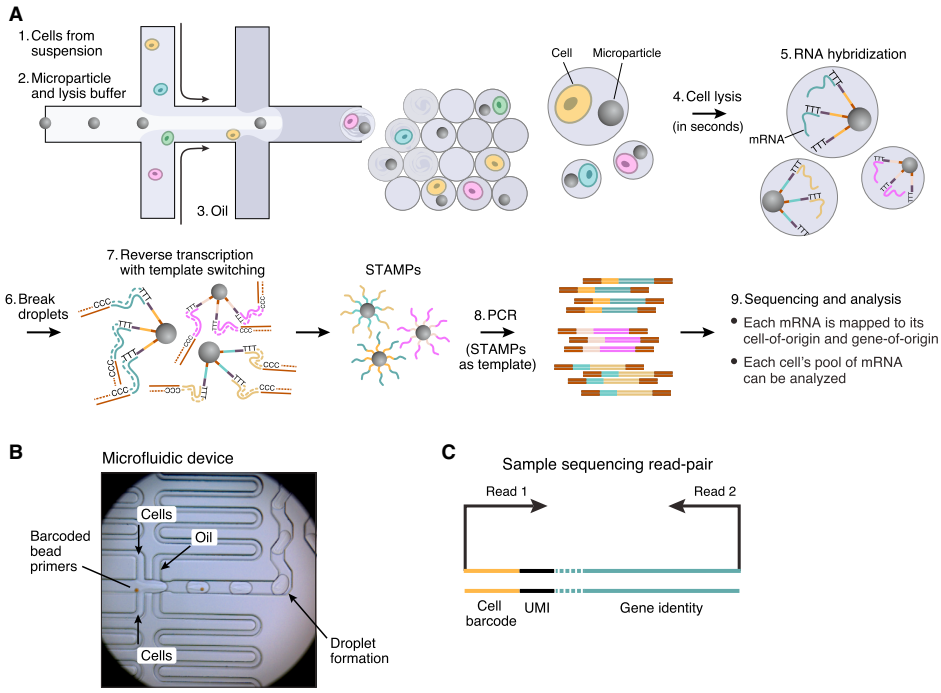

Figure 2. Workflow of Drop-Seq.

(A) Schematic of single-cell mRNA-seq library preparation with Drop-seq. (B) In Drop-seq, the microfluidic device is operated. Beads suspended in a lysis agent flow into the device from the central channel combining from the top and bottom. (C) Molecular elements of a Drop-seq sequencing library (adapted from E. Z. Macosko et al., 2015).

Principles of lineage tracing

Lineage tracing is the identification of all progeny of a single cell. It provides powerful means of understanding tissue development, homeostasis, and diseases[17]. Single-cell technologies are practically deployed in order to interpret the heterogeneity occurring during lineage conversion. Lineage relationships, though, are hard to be tracked since the single-cell sequencing dissociates the cells in a single cell unit. Thus, trajectory reconstruction is distracted.

To overcome this problem, a lineage tracing method named CellTag system was reported[18]. It is a marking method using lentivirus to check each cell with heritable barcode combinations[18,19]. As a result of CellTagging, transduced cells indicate distinct combinations, and it occurs heritable marks allowing the tracking of clonally related cells. Moreover, CellTags are easily captured within each single-cell transcriptome, allowing tracking of clonal recode as time passes, in parallel with cell identity (Figure 3).

In the *Wnt1-cre;R26R-tomato* mouse model NCCs fluoresce red, while CellTag containing cells fluoresce green because the CellTag lentivirus vector contains GFP, an 8-bp CellTag and a poly-A sequence. If the cell has both red and green fluorescence after differentiation, it implicates that the cell originated from a neural crest cell.

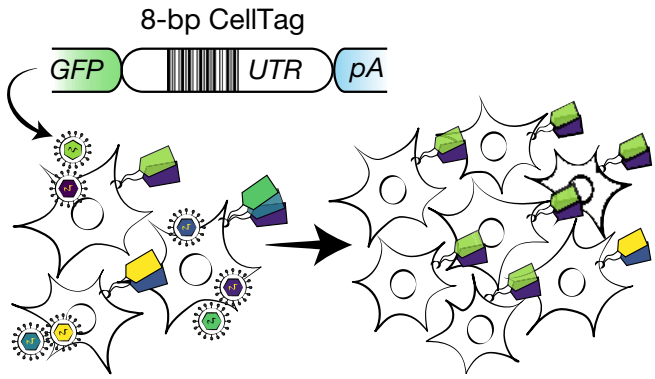

Figure 3. Concept of CellTag systems

A lentiviral construct includes an SV40 polyadenylation signal and an 8-bp random CellTag barcode in the 3′ untranslated region (UTR) of GFP (adapted from Biddy et al., 2018).

MATERIALS AND METHODS

CellTag library system setup

Amplification of CellTag library

CellTag library DNA was obtained from Addgene[18]. The library was delivered as suspended DNA in a microcentrifuge tube. Since the delivered library volume was only 10 µL, the library had to be amplified. The CellTag library DNA from bacterial transformation (Stbl3 Competent Cells, Thermo Fisher Scientific) was grown more than 16 hours in culture media. It was then followed by the extraction of the plasmid DNA using maxi-prep, producing lentivirus at a multiplicity of infection (MOI) of 3.

Bacterial transformation by liquid culture was performed as previously described[18,19]. Briefly, after thawing Stbl3 competent cells on ice, 100 µL of cells were mixed with 10-50 ng of CellTag library in a 1.5 mL microcentrifuge tube. Subsequently, the transformation mixture was placed on ice for 30 minutes. Then the cells were heat shocked for 60 seconds at 42 °C and immediately put back on ice for 1 minute. After adding SOC medium to bring the final volume to 1000 µL, the solution was incubated while shaking (~250 rpm) for 1 hour at 37 °C.

Checking the CellTag library complexity and growing bacteria were performed simultaneously. Firstly, to maintain CellTag library complexity, the number of colony-forming units (CFU) was calculated. After taking 20 µL of the recovery, serial dilutions from 1:10 to 1:1000 were prepared and plated onto LB agar plates containing ampicillin. Following overnight incubation at 37 °C the number of colonies on the plates were counted to calculate the number of CFUs. In this case, I

required approximately 100 to 200 CFUs per unique CellTag in the library to satisfy the complexity criteria. Secondly, the remaining recovery was added to 500 mL of LB + Ampicillin and grown overnight while shaking (~250 rpm) at 37 °C. The cells from the liquid culture were harvested and the library was purified using the DNA-maxi Endotoxin-Free kit (INtRONbio, Korea). The usage of an Endotoxin-free kit for large-scale plasmid DNA isolation assures sufficient harvest of high-quality DNA.

Assessment of CellTag library complexity

After amplification of the CellTag library, the complexity was assessed via sequencing. Whole genome sequencing (WGS) was performed at Theragen Etex (Suwon, Korea) to verify that the CellTag library was evenly amplified.

FASTQ files were aligned to the reference genome using Bwa Mem (v.0.7.16a). After sorting and indexing the BAM files by coordinates, the CellTag sequencing depth was used to verify the quality of sequencing. Subsequently, the reads containing a CellTag motif were extracted from the BAM files and recognized by the following tag sequence in the plasmid library vector: GAATTC NNNNNNNNACC. Extraction of CellTag reads was achieved by combining the samtools view command line tool and the grep command line utility searching for a regular expression containing the tag sequence. Finally, the total number of CellTag reads and the number of unique CellTag reads were counted and compared.

Lentivirus

Lentivirus production

Lentiviral particles were produced by transfecting HEK 293T cells with the pSMAL-CellTag construct, along with packaging constructs PAX2 and MD2. CellTag DNA was introduced into a host cell by transfection with polyethylenimine (PEI), a stable cationic polymer. After creation of the lentivirus, the lentiviral stocks were concentrated and titrated using qPCR lentivirus titration kit. Below, the details of these experiments are described.

Seed adherent cells for transfection. After preparing the adherent cells at 80% in the plate, the growth medium was aspirated from the cells and they were washed with PBS. The PBS was aspirated and 2 mL trypsin EDTA solution was added. Succeeding an incubation for 2 minutes, the cells were inspected using a microscope to verify they were rounded up, a shape which indicates an appropriate detachment of the cells from the surface of the culture dish. Subsequently, the detached cells were suspended in 8 mL fresh complete medium (DMEM/5% FBS/penicillin) by pipetting the solution up and down thoroughly. The cell suspension was then transferred to a 15 mL sterile centrifuge tube and the cells were spun at 200 × g at room temperature for 5 min. The supernatant was discarded, and the remaining cell pellet was resuspended in 1 mL fresh complete medium. A hemocytometer was used to count the cells. Next, 0.5 x 10^6 cells were added into each well of a 6-well plate and suspended in 2 mL of fresh complete medium. They were incubated for one day until adherent.

Transfect cells. For the preparation of the mixture for plasmid transfection 3 µg DNA, 2.25 µg viral packaging (PAX2) and 0.75 µg viral envelope (MD2) constructs were used. Firstly, PEI (1 µg/ µL) was added to the diluted DNA and

mixed immediately by vortexing or pipetting. The volume of PEI used was based on a 3:1 ratio of PEI (μg) to total DNA (μg)[20]. Accordingly, the total volumes of DNA, PAX2 and MD2 were 6 μg and the PEI volume used was 18 μg, respectively. The mixture was incubated for 15 minutes at room temperature and added to the cells. Due to the potential toxicity of PEI if applied to cells for longer periods of times, after 4~6 hours a media change was performed. The 6-well dishes were then placed in the 37 °C incubator, 5% CO2 for incubation.

Concentration and titration. After 24 hours, the lentivirus was concentrated using Lenti-X concentrator (Clonetech) and the lentivirus-containing supernatant was harvested. The clarified supernatant was transferred to a sterile container and one volume of Lenti-X Concentrator was combined and gently mixed with three volumes of clarified supernatant. Subsequently, the mixture was incubated at 4°C for overnight. To extract the lentivirus, samples were centrifuged at 1,500 x g for 45 minutes at 4°C and the supernatant carefully removed, resulting in an off-white pellet. Next, the pellet was gently resuspended in 1/10 to 1/100th of the original volume of complete DMEM. Lastly, the virus titration was verified using qPCR Lentivirus Titration kit (ABM).

Lentivirus infection efficiency in cell line

Before infecting primary cells with the produced lentivirus, a preliminary experiment was performed in a HEK cell line. After infecting the cell line with lentivirus, the DNA was extracted. To estimate the efficiency of lentivirus infection, amplification sequencing was performed at Theragen Etex (Suwon, Korea). DNA was amplified using CellTag primers (CellTag forward: ATCTCAAATCCC TCGGAAGC, CellTag reverse: TGTTCTGCTGGTAGTGGTCG). The following

PCR conditions were applied: initial denaturation at 94 ºC for 5 min, 35 cycles of 94 ºC 30 sec, 55 ºC 30 sec and 72 ºC 20 sec, and final extension of 72 ºC 20 sec. CellTag alleles yielded a 235 bp product.

For analyzing the lentivirus infection efficiency, the method described in section "Assessment of CellTag Library Complexity" was applied. FASTQ files were aligned to the reference genome using Bwa Mem and reads containing a CellTag motif were extracted from the BAM file. Next, the number of unique CellTag reads and the number of overlapped CellTag reads were counted and compared.

Mouse embryo experiments

Injection preparation

Firstly, a glass capillary (WPI, US) with puller was plugged into the microcapillary pipette (Sigma aldrich, US). The microcapillary pipettes and a 10 mL syringe were connected and a mix of CellTag library DNA and fast green stain was sucked in with the syringe. The green stain mixture was used to confirm the amount and direction of the DNA injection (Figure 5C).

Culture chamber in E9.5 embryo for 24-hour culture

The collagen gel is an essential material needed in order to set up the culture chamber. To prepare the reagent sufficient for six chambers, the following reagents were used: (1) 2 mL 10 x complete DMEM (powder, Gibco), (2) 0.2 mL HEPES (pH 7.4, WelGene) (3) 0.2 mL NaHCO3 (Biosesang, Korea), (4) 1.6 mL Distilled water, (5) 16 mL I-PC Atelocollagen (Koken, Toshima Ward, Tokyo, Japan). Firstly, reagents

(1) through (4) were mixed while cooling in an ice water bath and then added to the collagen solution. Since the collagen solution is viscous, one has to pipette several times to remove the collagen attached to the pipette surface, while being careful not to form bubbles as it is difficult to remove them later. Subsequently, 1.5 mL of the above mixed reagents were added to two-chamber culture slide dishes. Lastly, the solution in the culture slide dishes was warmed at 37°C in a CO_2 incubator for 1 hour to allow gel formation.

Separated E9.5 mouse embryos were transferred individually onto the bottom layer of a collagen gel in two-chamber culture slide dishes (SPL, Korea). The bottom layer was prepared previously from an acid collagen solution (Koken, Toshima Ward, Japan) according to the manufacturer's specified protocol as described above. Embryos were then covered with an approximately 2 mm thick overlay of the same collagen gel matrix as used in the bottom layer, followed by an overlay of 100% rat serum (LORKLAND). Subsequently, these were topped with a mineral oil layer to prevent evaporation (Figure 5B). The above method was first described by Kawakami et al[21].

Mouse embryo dissociation methods

E10.5 and E18.5 embryo dissociation methods are similar. For E10.5 embryo dissociation, the torso of the mouse was gently separated from the mouse embryo and transferred into a 15 mL conical tube. Subsequently, 2 mL of trypsin was added to the tube, and the mix was incubated at 37°C in an incubator. In order to assure equal dissociation during the incubation period, the tissue was mixed using a pipette every 3 minutes. This step was repeated twice, while assuring that the tissue was not left in trypsin for more than 15 minutes. In order to stop trypsin treatment, 5 mL

complete DMEM was added. Next, the cells were centrifuged at 1000 RCF for 10 min at room temperature. The supernatant was aspirated, and the pellet was suspended in 1 mL DPBS. Lastly, the cells were counted using a hemocytometer and trypan blue and diluted to a final concentration of 100 cells/μL PBS-BSA. In comparison to E10.5 embryo dissociation, for E18.5 embryo dissociation the embryo brain was separated carefully, and 10 mL of trypsin was added to the tube. In order to stop trypsin treatment, 20 mL complete RPMI was added.

Mouse models

Mouse strain construction and genotyping

All the mouse experiments were performed under the standard protocols approved by IACUC (SNU-200104-1).

To specifically label NCCs in developing mouse embryos, *Wnt1-Cre*/+ mice were crossed with mice carrying the R26R-Tomato reporter allele. Both strains were obtained from the Jackson laboratory as live animals with a C57BL/6J genetic background (*Wnt1-Cre* strain: #003829, *R26R-Tomato*: #007914). I subsequently mated the two strains to produce males carrying *Wnt1-Cre/+*;*R26R-Tomato*/+ or *Wnt1-Cre/+*;*R26R-Tomato*/*R26R-Tomato* genotypes and females carrying *R26R-Tomato*/*R26R-Tomato*.

Frozen embryos of the *LgDel* (B6.129S7-Del(16Es2el-Sept5)3Bld/Cnrm) strain were purchased from Infrafrontier and processed following the cryopreservation method presented by KRIBB (Korea Research Institute of Bioscience & biotechnology). *LgDel* mice were developed from a C57BL/6J genetic background at the KRIBB institute. This *LgDel* mouse model of chromosome 22q11

deletion syndrome (22q11DS) carries a deletion on chromosome 16 spanning from *Es2el* to *Sept5*.

Mice were genotyped using DNA extracted from toe clips. Each mouse's genomic DNA was amplified using a PCR protocol consisting of an initial denaturation at 94 ºC for 5 min, 35 cycles of 94 ºC for 30 sec each, 55 ºC for 30 sec and 72 ºC for 20 sec, followed by a final extension at 72 ºC for 20 sec. The primers used for PCR amplifications are displayed in Table 1.

Primer name	Primer sequence
Wnt1 Tg Forward	5'-TAA GAG GCC TAT AAG AGG CGG-3'
Wnt1 Tg Reverse	5'-ATC AGT CTC CAC TGA AGC-3'
Cre Tg Forward	5'-GCG GTC TGG CAG TAA AAA CTA TC-3'
Cre Tg Reverse	5'-GTG AAA CAG CAT TGC TGT CAC TT-3'
Tomato Tg Forward	5'-GGC ATT AAA GCA GCG TAT CC-3'
Tomato Tg Reverse	5'-CTG TTC CTG TAC GGC ATG G-3'
LgDel Forward	5'-ACC TGT GGC CCT GGG ACT-3'
LgDel Reverse 1	5'-CCA GAC TGC CTT GGG AAA AG-3'
LgDel Reverse 2	5'-TCG GAA TTA GTC TGT CAC CTA GC-3'

Table 1. List of primer sequences

Determining the DGS mouse model exact mutation region

To construct the DiGeorge syndrome mouse model, frozen embryos were obtained from Infrafrontier. Frozen embryos were cryopreserved and IVF (*In-vitro* fertilized) at KRIBB (Korea Research Institute of Bioscience & Biotechnology). Mating a founder mouse (*LgDel* mouse) with a wildtype mouse resulted in an F1 generation of 15 littermates. Because the available information about the purchased frozen embryos only contained which genes were deleted, but not the exact position, the genotypes of the F1 generation could not be confirmed directly. To verify the genotype of *LgDel* mice, I sent extracted DNA to perform whole genome sequencing (WGS) to Theragen Etex (Suwon, Korea). The obtained sequence was covered 1x.

The resulting whole genome sequencing FASTQ files were provided by Theragen Etex for download. Subsequently, they were aligned to the reference genome mm10 (UCSC, 2009) using Bwa Mem (v.0.7.16a)[22,23]. After sorting and indexing the BAM file by coordinates, the sums of coverage depths for every 1kb were retrieved using Linux awk command line tools and samtools depth for each sample, respectively.

Single-cell analysis

Addition of GFP-CellTag transgene as new chromosomes to the mm10 genome

The pipeline was applied using a custom reference genome which was generated by adding the sequences corresponding to the GFP-CellTag transgene as new chromosomes to the mm10 genome. The significant UTRs in the GFP-CellTag transgene constructs had an ability to detect transgene expression. To achieve this, at first, a custom gene transfer format (GTF) file with our transgenes was produced by indexing of the GTF and FASTA files.

Drop-seq experiment

Cells were dissociated using trypsin, washed in DPBS containing 0.01% BSA and diluted to 100 cells/μL. Drop-seq was performed as previously described by McCarroll's lab [15]. In other words, cells and beads were diluted to 1×10^5 cells/mL and 1.2×10^5 beads/mL for one sample. After loading cells and beads, the droplet microfluidics ran for 11 minutes. For assessment of droplet quality and bead doublets, 10 μL from the droplet layer were input into a hemocytometer. Furthermore, it was confirmed, that all the droplets were of similar size. Emulsions were collected and broken down using 1 mL of Perfluorooctanol (Sigma) for 15 mL of emulsion, followed by washing in 6x saline-sodium citrate (SSC) buffer to recover the beads. Reverse transcription was then performed using the Maxima H Minus Reverse Transcriptase kit (EP0752, Life Tech). This step generates cDNA strands on the RNA hybridized to the bead primers. After treatment with 2,000 U/mL of Exonuclease I (New England Biolabs), aliquots of 2,000 beads were amplified by

PCR for 15 cycles, using Kapa HiFi Hotstart Readymix (Kapa Biosystems). The PCR products were sent to Theragen. They performed tagmentation of cDNA using Nextera XT and sequencing of the samples. Libraries were sequenced on an Illumina HiSeq 2500, with Drop-seq primers.

Drop-seq computational analysis

To analyze the Drop-seq data, several preprocessing steps were applied, that enabled selection of high-quality sequenced cells, which were then used for UMAP clustering analysis and identification of marker genes.

DropSeq FASTQ files consist of read pairs. The first read of each pair contains the cell barcode, UMI (unique molecular identifiers) and a PolyT region. UMIs are the molecular tags that are used to detect and quantify unique mRNA transcripts. Complementary, the second read contains the CellTag sequences.

Our first step of preprocessing was a conversion of the file format from FASTQ to uBAM (untagged bam) using Picard tools (v.2.9.0). Next, the resulting uBAM files were further aligned to the mm10 reference genome using STAR aligner (v.2.5.4b) resulting in aligned and tagged BAM files. The last step of preprocessing consisted of the extraction of reads per cell barcode. This step extracted the number of reads for any BAM tag in a BAM file. Applying the "knee method" described by Macosko *et al.*, 50,000 cells were selected for further analysis[15]. This method selects cells on the left of the "knee", the inflection turning point, of the cumulative fraction of reads. Following alignment, digital gene expression (DGE) matrices were generated and analyzed. This procedure and the essential equipment are additionally explained online in the Drop-seq Alignment Cookbook and Seurat-guided clustering tutorial (https://satijalab.org/seurat/ v3.0/pbmc3k_ tutorial.html, 2020-04-02).

The R package Seurat was used to cluster and visualize cells. Metadata for individual cells was assembled based on the expression data. The metadata contained information related to clone identity, number of RNA counts (nCount), number of feature RNA (nFeature) and the percentage of mitochondrial genes. After removing low-quality cells from the dataset, the data was normalized. Highly variable genes were then selected as input for dimensionality reduction using principal component analysis (PCA). Finally, these principal components acted as input to cluster the cells, visualizing these clusters using UMAP with the clustering resolution set to 0.5.

RESULTS

CellTag library system setup

Sequencing for determining the CellTag library status

The total length of the CellTag library plasmid vector was 8,432 bp. It consisted of an 8 bp variable region within the 3′ UTR of GFP in the pSMAL lentiviral construct. The specific position of the 8 bp variable region was from 2,526 to 2,533 bp. The sequencing quality of this region was checked to see whether a similar read depth as in the other regions was achieved. Overall, mean coverage depth was found to be 374,885 reads per base pair with the 8 bp barcode region showing a sufficient coverage (Figure 4A).

The theoretical number of possible tag sequences in the 8 bp variable region is 65,536 because four random bases are available (=4^8). However, the low transformation efficiency leads to the observation of only about 20,000 unique CellTags in practice, a number also previously described by Biddy et al. Of a possible 65,536 unique combinations, this experiment resulted in the detection of 21,418 unique CellTag sequences. The highest count for a unique CellTag was 407 and the lowest was 1. Furthermore, the CellTag frequency was checked by focusing on how often each CellTag was observed in a population of transduced cells. After removing the CellTags that appear only once and are likely due to sequencing and PCR errors, cells expressing more than 20 CellTags, and less than 2 CellTags were filtered out. The number of unique CellTags and the ratio of CellTags from 2 to 19 are 8,536 representing 40% of the total CellTags. (Figure 4B).

Sequencing injected CellTags in cell line

Before performing injection of the lentivirus into the primary cells, transduction of CellTag constructs to a HEK293 cell line was performed as a preliminary experiment. After injecting lentivirus in HEK293 cells, GFP fluorescence could be observed (Figure 4C). Amplification of the target using PCR and CellTag primers, enabled me to see, whether the cells contained various CellTags randomly. Sequencing of PCR products resulted in average 7x coverage. In conclusion, 61,216 unique CellTag sequences were detected. The highest count of a unique CellTag was 250,444 and the smallest was 1. The number and ratio of CellTags from 2 to 19 were 33,263 or 54% .

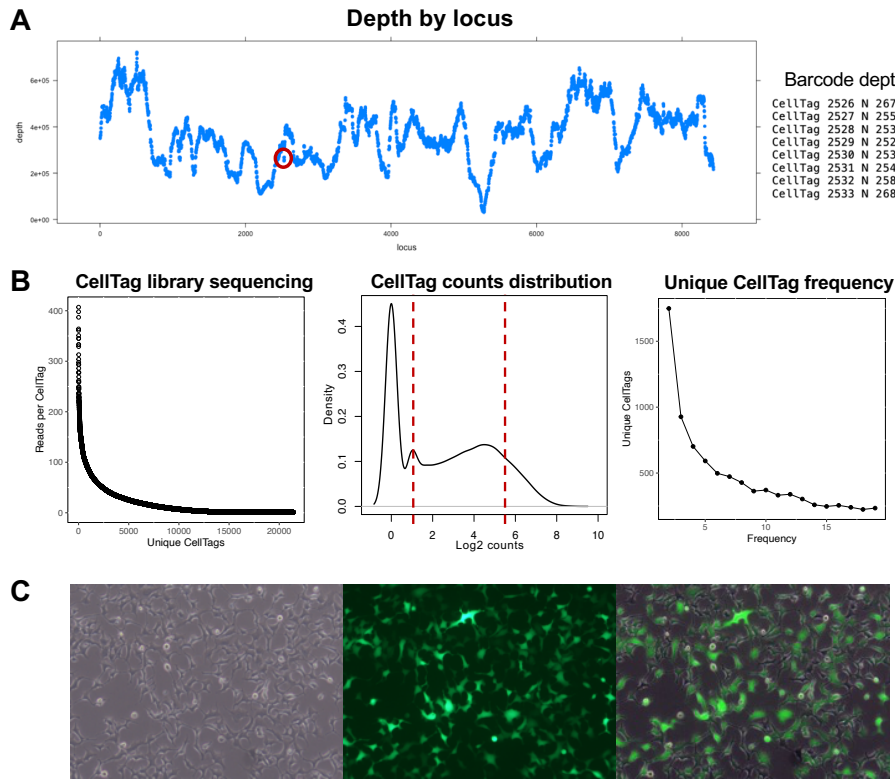

Figure 4. Evaluation of CellTag library complexity

(A) Read depth profile plot for CellTag library vector WGS. Barcode regions are evenly covered. (B) Result of CellTag library sequencing. 21,418 unique CellTags are detected (left). Unique CellTag counts distribution. Red lines are 2 CellTags to 19 CellTags (middle). Unique CellTag frequency in 2 to 19 CellTags (right). (C) Transduced lentivirus involved the CellTag library in HEK293 cell. Cells are detected bright field (left), GFP fluorescence (middle), and a merge of the two (right).

Mouse embryo culture setup

Construction of *in vitro* culture workflow

In this case, mouse is the optimal model system because it has been successfully used to model human cardiovascular and craniofacial defects and it is widely accepted that developmental processes are well conserved between mouse and human[24].

Several experimental approaches have been used to address the question of when and where cell fate decisions are made in the NCCs[25]. At E9.0, NCCs are in the migratory phase, an undifferentiated status. At E9.5, NCCs arrive at the developing pharyngeal arch (PA) structures, and differentiation processes are initiated. Finally, at E10.5, NCCs are under active differentiation in the PA structures. By examining E10.5 embryos, one can address how the differentiation progresses once the NCCs arrive in the target organs. To do so, I established an experimental method to grow E9.5 embryos *in vitro* culture for 24 hours. The whole experimental process is shown Figure 5A.

Day 1: Once the E9.5 embryos of the desired genotype were identified, the gravid uterus containing the embryos was isolated from the mother mouse. After separating all embryos, the ones containing the RFP-positive NCCs were selected under a fluorescent microscope. RFP-positive embryos represent the experimental group used to detect NCCs lineage tracing, while RFP-negative embryos were used as control group. After sorting the embryos, CellTag lentivirus was injected on the 4% agarose gel by digging a pit. Injected embryos were cultured for 24 hours in the chamber (Figure 5B, C).

Day 2: After 24 hours, the embryos were retrieved, and the torso regions were isolated (Figure 5D). The reason for only using the embryo torso is, that *Wnt1-Cre*

moves to the head in considerable amounts, not to the neural crest cells. The parts above the 1st pharyngeal arch and below the heart were eliminated and the rest of the parts was minced using a razor blade. Subsequently, the embryos were dissociated to create single-cell suspensions and the cells were prepared for the following Drop-seq experiment. More than 1.5×10^5 cells were collected for the Drop-seq experiment.

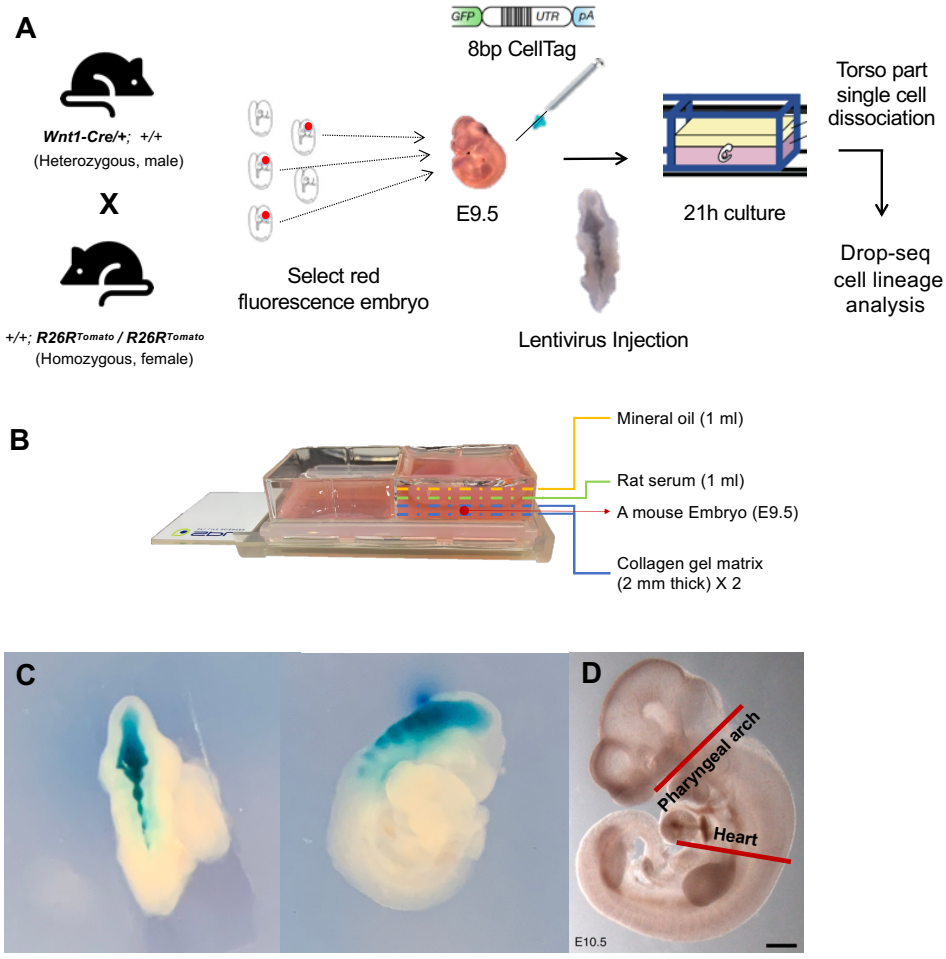

Figure 5. *In vitro* culture in E9.5 embryos

(A) Schematic of *in vitro* culture workflow. (B) Chamber culture glass slides containing mineral oil. (C) Result of injection of mixtures of lentivirus and fast green stain in neural crest at E9.5 embryo. The dorsal side of an embryo (left). The lateral side of an embryo (right). (D) Torso region of an embryo.

In vitro culture results

To observe the NCC migration, I constructed an *in vitro* embryo culture system using chamber culture glass slides (Figure 5B). Whole embryo culture was already performed by other groups[26]. This system made it possible to keep embryos developing for more than 24 hours. I compared the migration of NCCs between E9.5 and E10.5 under a fluorescent microscope for the presence of RFP. The size of the embryo increased, and the RFP-shaped light was spread widely (Figure 6). Through this, I confirmed that the neural crest cells migrated for 24 hours. The lentivirus, which was injected into the embryo, contains GFP. GFP fluorescence can be checked after 24 hours of culture.

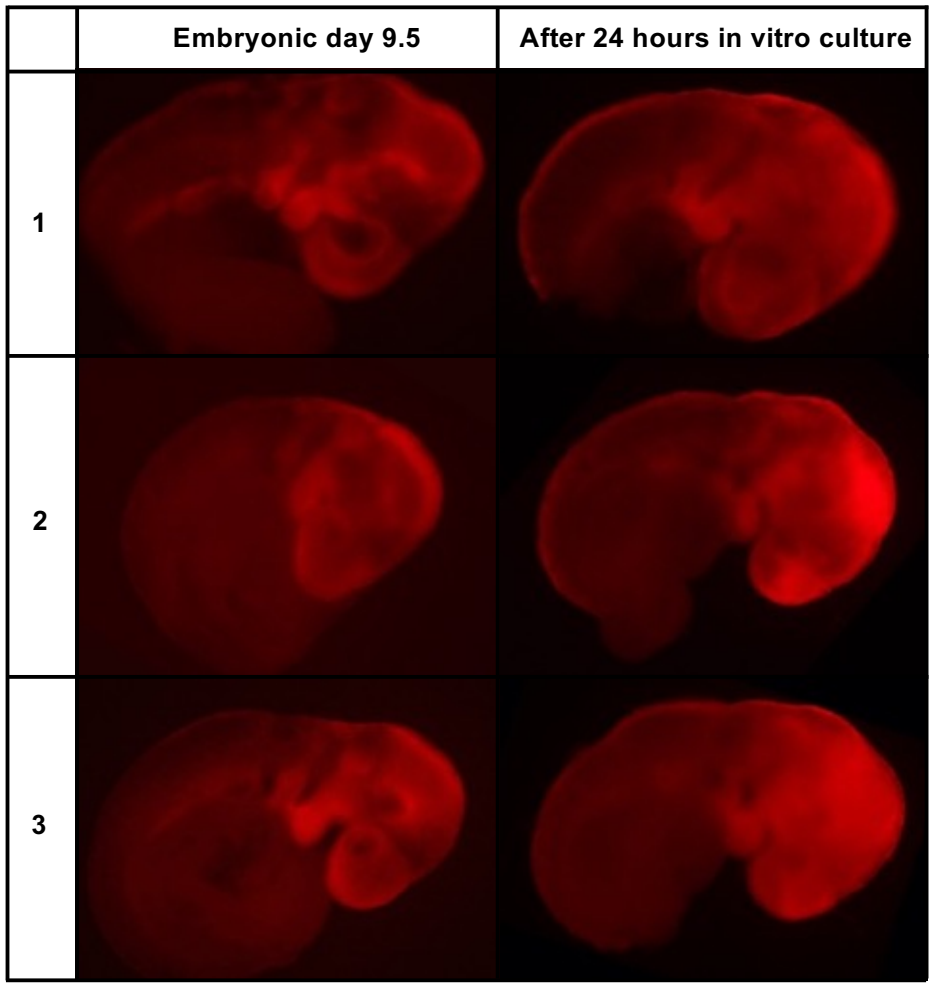

Figure 6. tdTomato fluorescence after 24 hours culture

tdTomato fluorescence was detected in E9.5 and E10.5 embryos. For 24 hours, the size of the embryo increased, and the RFP-shaped light was spread widely.

Construct *in utero* transplantation workflow

To understand NCC development in the brain, I used E12.5 embryos and raised them for 6 days until E18.5. The process of brain development is that the neural tube has formed the primary brain vesicles at E11.5: prosencephalon, mesencephalon and rhombencephalon. Between stages of E12.5 and E15.5, there is an increased expansion of neuronal precursors and cell migration forming cortical layers. Simultaneously, neurons differentiate to allow for axonal branching and synapse formation. In E18.5 embryos, the brain is going through the migration and maturation process[27].

The whole experimental process is shown in Figure 7. After identification of E12.5 embryos of the desired genotype, the pregnant mouse was anesthetized. A 1 cm incision was made into the lower abdomen and the gravid uterus was delivered through the incision. The total number of fetuses was counted by first identifying the right and left ovaries to ensure visualization of the entire uterus. *In utero* lentiviral injection into the ventricular zone of the brain was performed in the left ovaries. Right ovaries remained as a control to compare with the left injected embryos. Next, I carefully placed the uterus back into the abdominal cavity. The incision was closed in two layers with an absorbable suture. First, the fascia was closed without injuring the underlying bowel or bladder. Secondly, the skin was sutured. After 6 days, E18.5 embryos were retrieved, and the brain was separated. Subsequently, the embryos were dissociated to make single-cell suspensions and the cells were prepared for the following Drop-seq experiment.

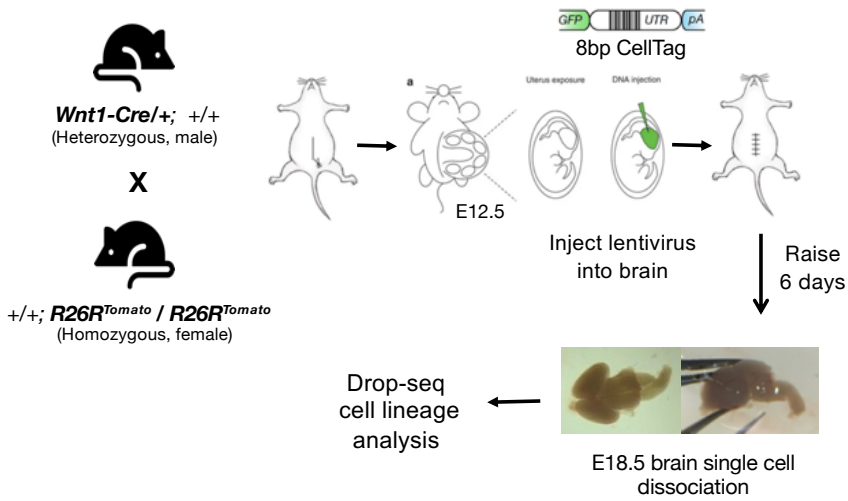

Figure 7. Processing IUT in E12.5 embryos

Schematic of *in utero* transplantation (IUT).

IUT results

The IUT experiment did not accurately confirm that the lentivirus was injected sufficiently because the embryo was surrounded by the placenta. Therefore, to ensure that the lentivirus was injected into the ventricular zone, the E12.5 mice were isolated and tested after injection (Figure 8A). Upon confirming that the lentivirus spread well into the target region, I isolated the embryo brains after six days of raising them (Figure 8B). The brain was carefully separated including the ventricle, cortex, midbrain, and brainstem (Figure 8C, D).

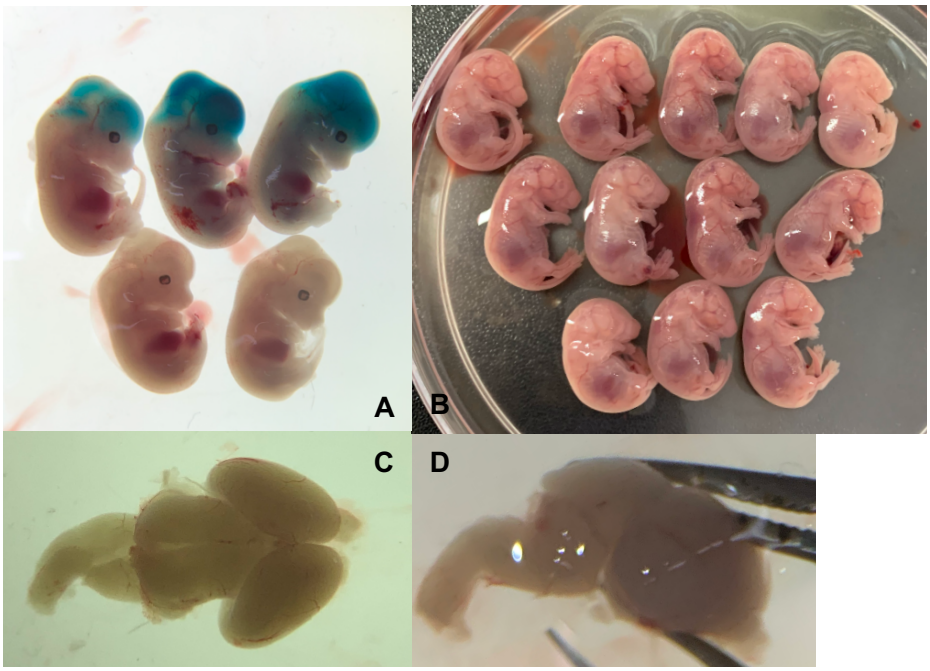

Figure 8. Results of IUT in E12.5 embryos

(A) Result of injection of mixtures of lentivirus and fast green stain in E12.5 embryo brain. (B) E12.5 embryos were raised for 6 days to until E18.5. (C) The dorsal side of an embryo brain (left). (D) The lateral side of an embryo brain (right).

Mouse models setup

Normal and DGS mouse model

To understand the intricate molecular regulation of neural crest cell development, I constructed a normal and a DGS mouse model. *Wnt1-Cre* strain and *R26R-Tomato* strain were crossed. *Tomato* strain mice can be maintained in a homozygous state (Figure 9A). *Wnt1-Cre* is a widely used NCC-specific *Cre* driver line and *R26R-Tomato* mice contain *Tomato* knocked into the ROSA26 locus, allowing activation of Tomato upon presence of Cre. In conclusion, RFP fluorescence appears in the specific regions of Wnt1 expression. The date of mating was recorded by plug formation and the pregnant females of E9.5 were dissected.

For constructing the DGS model, I inter-crossed the *LgDel/+* strain and *Wnt1-Cre/+* strain (above). After obtaining the *Wnt1-Cre;LgDel* strain, I further inter-crossed it with the *R26R-Tomato* strain (Figure 9B).

NCC development was compared by analyzing the difference in single-cell data obtained from the normal mouse model and the DGS mouse model.

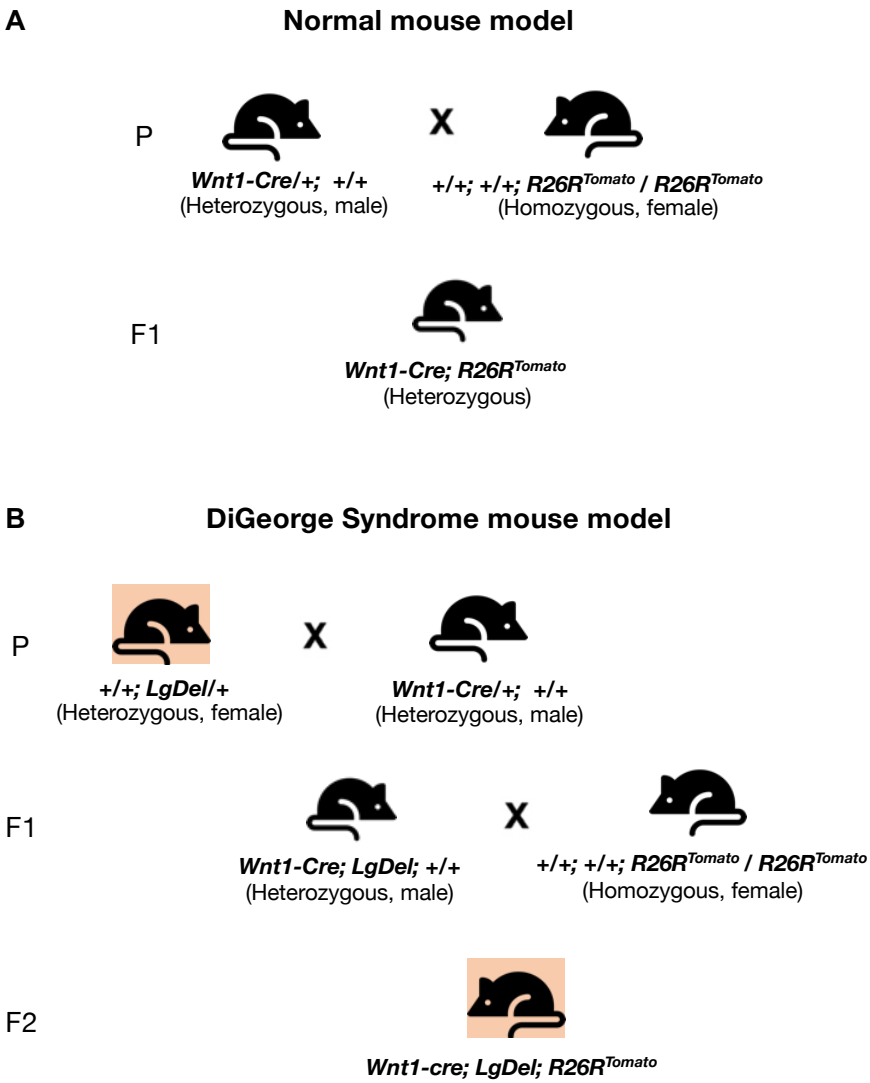

Figure 9. Mouse mating schemes

(A) Mating schemes for control mouse model. (B) DGS mouse model.

Checking the DGS mouse model exact mutation region

LgDel mice, which model the heterozygous microdeletion of genes at human chromosome 22q11.2 associated with 22q11DS, referred to as DiGeorge/22q11.2 Deletion Syndrome, displayed cranial nerve and craniofacial dysfunction. The genetics of 22q11DS, and the close homology between human chromosome 22 and portions of mouse chromosome 16 permit fairly precise modeling of the causal CNV using targeted mutagenesis in the mouse[28,29].

The *LgDel* mouse carries a 700,000 bp deletion spanning from *Es2el* to *Sept5* corresponding to an approximate region from chr16:17,900,709-18,629,954 (mm10). The sum of the 1kb coverage depth should theoretically be 1000 since the 1x coverage has been sequenced for each sample. Regions with a coverage depth of less than 1000 represent candidate intervals for potential breakpoints. Using this approach, I was able to uncover the candidate breakpoint intervals for the beginning and end of the deleted region. The breakpoint at the beginning and the end were discovered to be located at chr16:17,905,000-17,911,00 and chr16:18,622,000-18,628,000, respectively. Comparing the coverage depth in the candidate breakpoint interval regions of 15 littermates, mouse 2, 4, 9, and 10 showed low coverage spanning the deletion region (Figure 10).

In order to find the specific breakpoint positions, the candidate intervals were extracted from the BAM files. Using the clipping coordinates specified in the CIGAR strings of soft clipped reads in the intervals, I was able to detect the exact breakpoint positions at chr16:17,907,688 and chr16:18,626,548 (Figure 10). This information was further utilized to create primers targeting the above positions and verify the genotype of the mice.

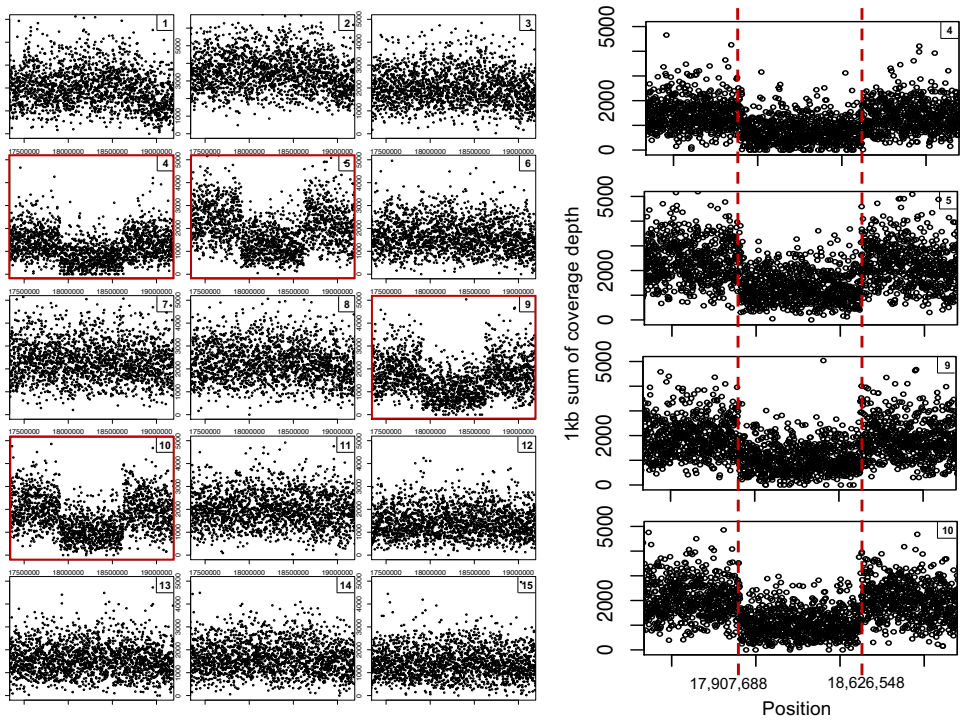

Figure 10. Identifying *LgDel* mouse using NGS

Whole-genome sequence in 15 littermates. The coverage depth is 1X. Sum of coverage depth per 1kb. X-axis represents genomic position and y-axis displays the 1kb sum of coverage depth. Mouse 2, 4, 9 and 10 have a deletion region in chr16 (left). 1kb sum of coverage depth plot of WGS of *LgDel* mouse samples. They have a deletion region in chr16: 17,907,688 ~ 18,626,548 (right).

Single-cell analysis

Drop-seq alignment and digital gene expression matrix generation

For the integrated analysis of the GFP-CellTag transgene in the mouse embryo singlecell data, the CellTag sequences were added to the mm10 genome as a new chromosome[18]. As depicted in Figure 11A, the transgene consists of a start region (2,249-2,500 bp) containing the transgene poly-A region (2,249-2,500 bp), the CellTag (2,520-2,528 bp) and the EGFP region (2,540-3,259 bp). Only the start region was used as a custom reference. Usage of Illumina sequences with a read length of 150 bp ensured sufficient coverage of the transgene chromosome. The distribution of mapped reads in the CellTag sequences is shown in Figure 11B, C.

The presence of reads mapped to the additional chromosome can be used as a confirmation that the lentivirus entered cells. Indeed, the samples without lentivirus had no mapped reads in the CellTag chromosome. However, additional experiments are needed because no reads were mapped to the CellTag region. I therefore suggest supplementing the protocol with an additional PCR step amplifying the CellTag region with CellTag specific primers after conducting PCR in the Drop-Seq experiment.

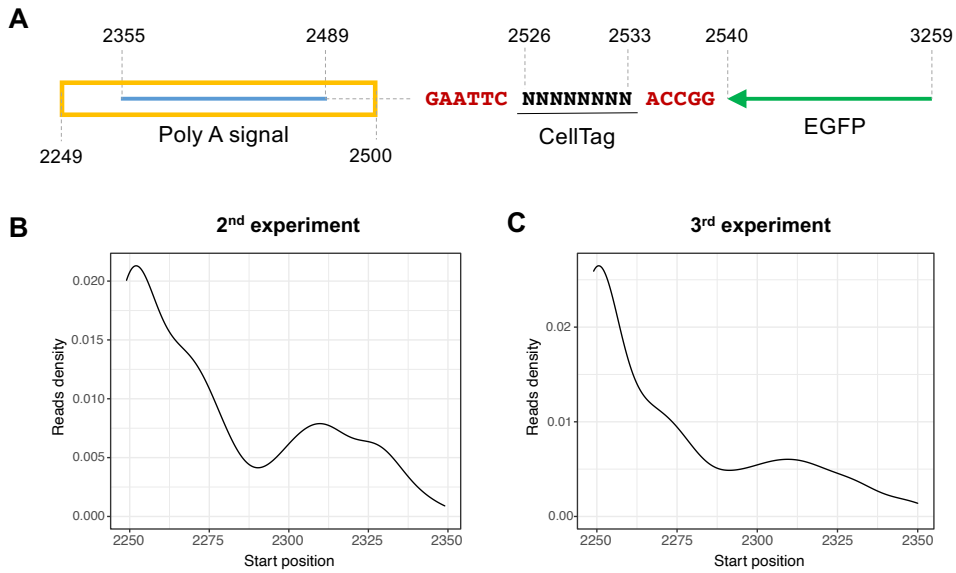

Figure 11. Aligning reads to custom reference genome

(A) A schematic of polyA, CellTag and EGFP parts of the CellTag library vector. The red indicates CellTag motif, and the yellow box indicates the mapped part of the custom reference. Reads from 2,249 bp to 2,500 bp were read as CellTag, which was designated as custom reference. (B), (C) In the second and third experiments of E9.5 embryo, the ratio of mapping CellTag reads. In both experiments, the start position which is 2,249bp had the largest number of reads, and the distribution was as above.

Visualization of scRNA-seq data

1) Mouse embryo trunk at E10.5 results

I performed single-cell experiments twice for mouse embryo trunks at E10.5 using Drop-seq and analyzed the outcome using Seurat. I performed quality control (QC) and selected cells for further analysis. The QC metrics and filtering of cells were based on two criteria. The first criterion was the number of unique genes detected in each cell. Low-quality cells or empty droplets will often contain very few genes. Cell doublets or multiplets may exhibit an aberrantly high gene count. Therefore, I excluded cells that had unique feature counts over 2,500 or less than 200. The second criterion was the percentage of reads that map to the mitochondrial genome. Low-quality or dying cells often exhibit extensive mitochondrial contamination. I filtered cells that had >7% mitochondrial counts in the first and second experiments (Figure 12A). After filtering, the number of cells used in the analysis was 417 cells in the first experiment and 1387 cells in the second experiment. Seurat offers several non-linear dimensional reduction techniques, such as UMAP (Uniform Manifold Approximation and Projection) to visualize and explore these datasets. After performing UMAP analysis, I annotated the resulting clusters using cluster biomarkers (Figure 13).

For E9.5 mouse embryos, UMAP analysis was based on Drop-Seq data from 1,804 cells. For clusters 0, 1, 2, and 3, it proved difficult to find differentially expressed features because the overall amount of expressed was too high. According to the cluster biomarkers, clusters 4, 5, and 7 are related to neurogenesis, vascular development and cardiac muscle, respectively (Figure 12B).

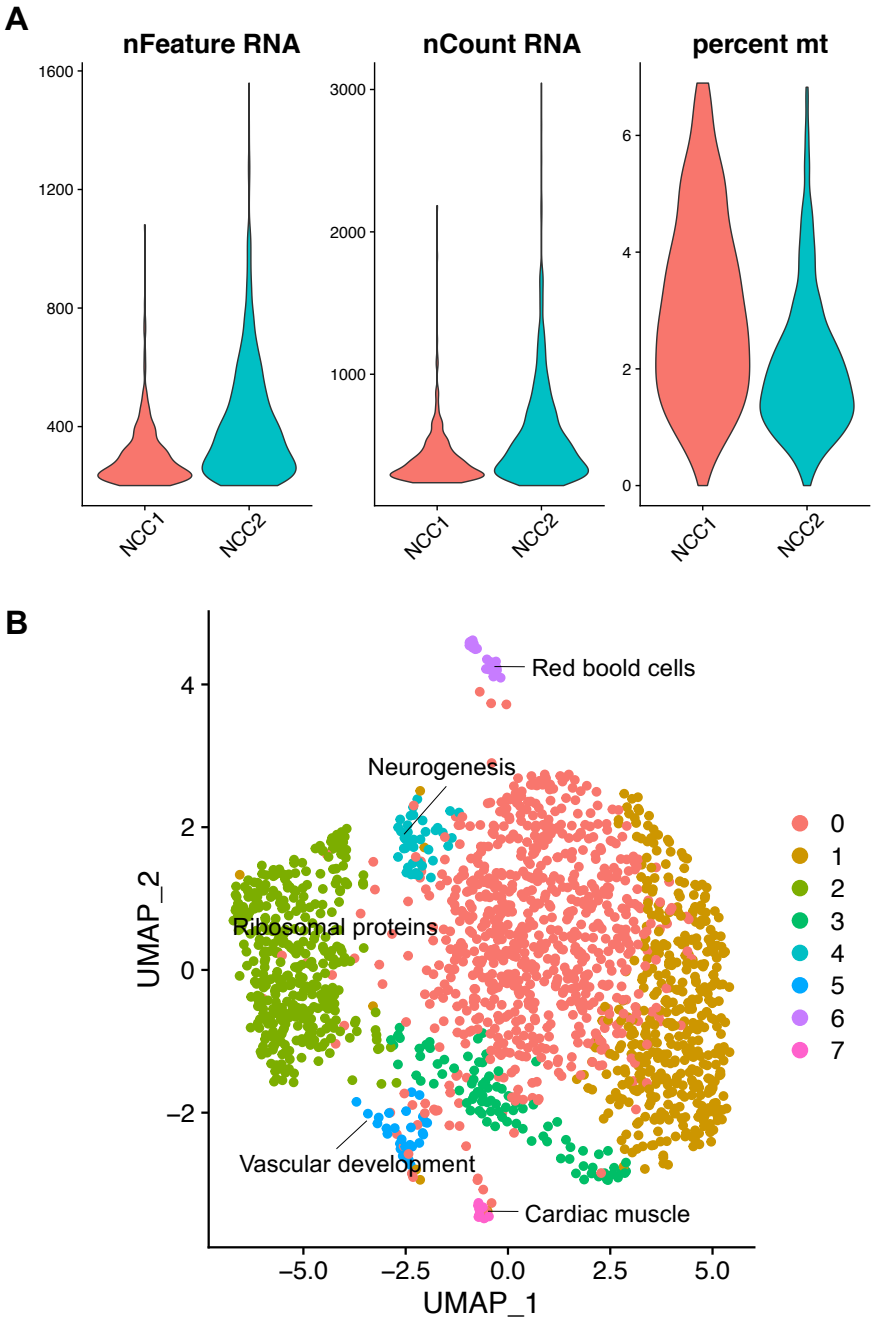
A
nFeature RNA
nCount RNA
percent mt
1600
1200
800
400
3000
2000
1000
6
4
2
0
NCC1
NCC2
B
Red boold cells
Neurogenesis
Ribosomal proteins
Vascular development
Cardiac muscle
UMAP_2
UMAP_1
4
2
0
−2
−5.0
−2.5
0.0
2.5
5.0
0
1
2
3
4
5
6
7

Figure 12. scRNA-seq clusters of E10.5 embryos

(A) QC metrics of the first and the second E10.5 embryo experiment. nFeature_RNA is the number of genes detected in each cell and nCount_RNA is the total number of molecules detected within a cell. Percent.mt is the percentage of reads that map to the mitochondrial genome. (B) Umap projection of trunk neural crest cells of E10.5 embryos (n = 9 embryos, 2,221 cells) colored by cell type prediction.

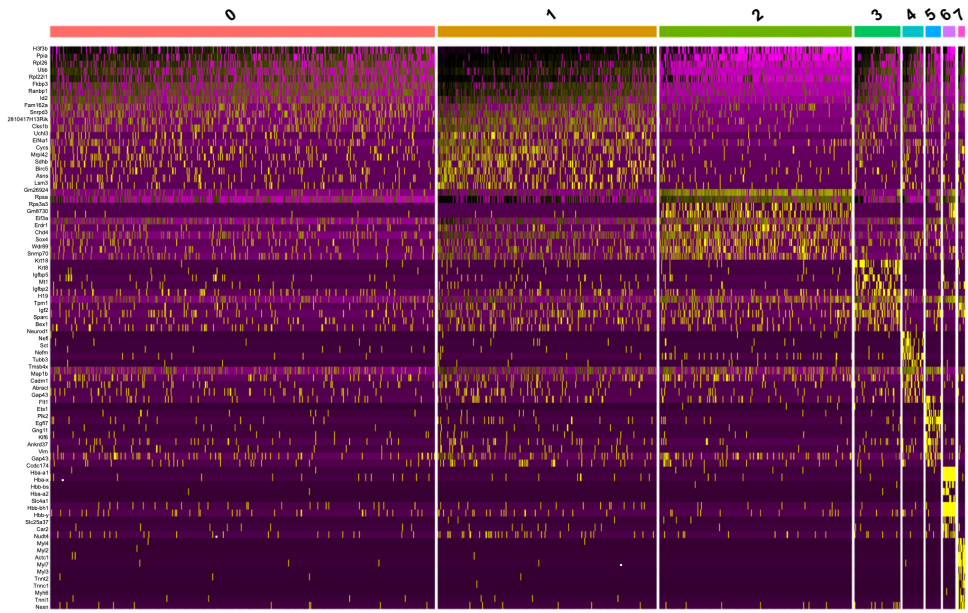

Figure 13. Finding differentially expressed genes in E10.5 embryos

The expression heatmap plotting the top 10 markers in E10.5 embryo trunk NCC. The canonical markers for each cell type, which can be used to assess the clustering.

2) Mouse embryo brain at 18.5 results

I performed Drop-seq single-cell experiments for E18.5 mouse embryo brains and analyzed the resulting data using Seurat. To ensure high-quality analysis, I filtered cells that had unique feature counts over 2,500 or less than 200 or >10% mitochondrial counts. After filtering, the number of cells used in the analysis was reduced to 1,927 cells (Figure 14A).

After performing UMAP clustering, I annotated each cluster using cluster biomarkers (Figure 15). Similar to the analysis of E10.5 embryo trunks, for E18.5 mouse embryo brains, it was difficult to find differentially expressed features for the clusters 0, 1, 2, and 3 because of the high number of expressed genes. However, according to the cluster biomarkers, clusters 4, 5 and 7 are related to neurogenesis, vascular development and cardiac muscle, respectively (Figure 14B).

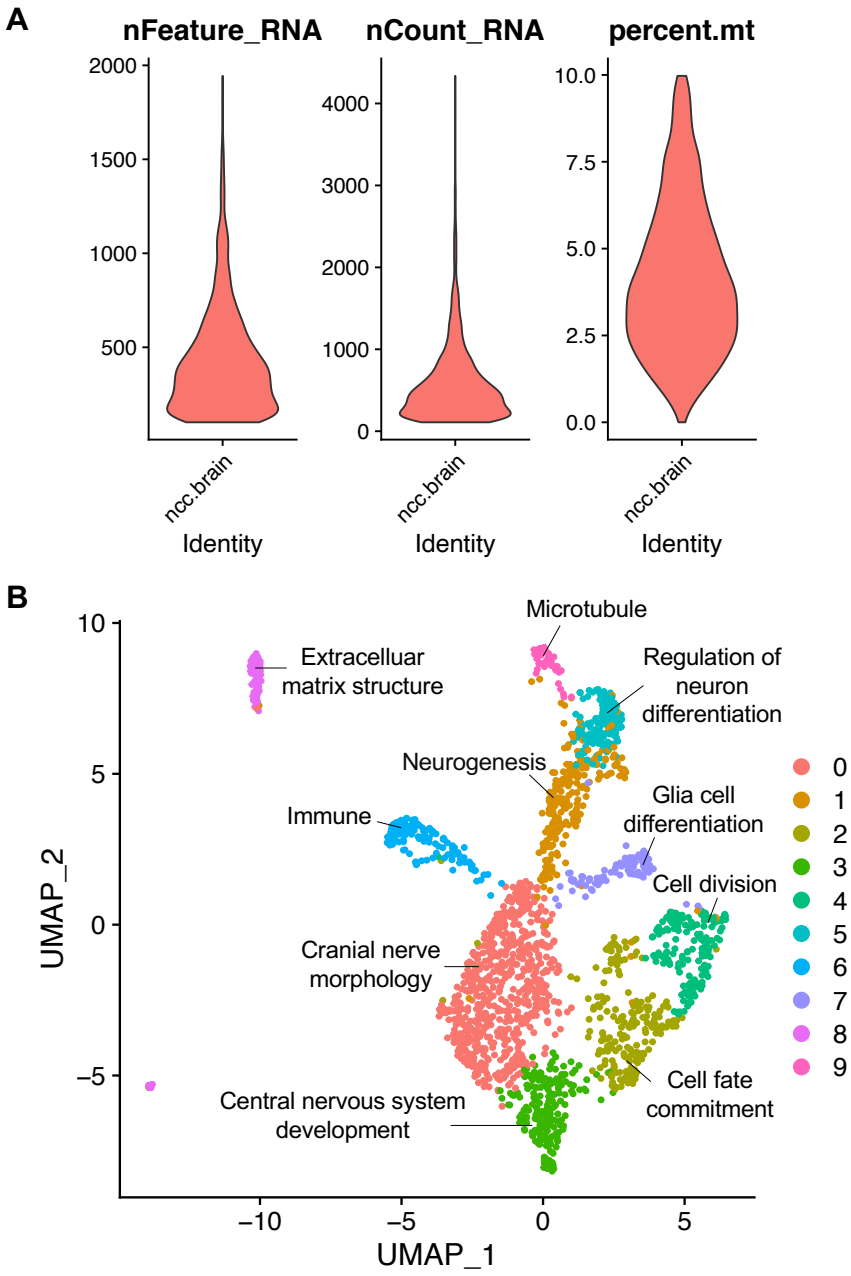
A
nFeature_RNA
nCount_RNA
percent.mt
2000
1500
1000
500
4000
3000
2000
1000
0
10.0
7.5
5.0
2.5
0.0
ncc.brain
Identity
B
Microtubule
Extracelluar matrix structure
Regulation of neuron differentiation
Neurogenesis
Glia cell differentiation
Immune
Cell division
Cranial nerve morphology
Cell fate commitment
Central nervous system development
UMAP_2
UMAP_1
10
5
0
−5
−10
0
1
2
3
4
5
6
7
8
9

Figure 14. scRNA-seq clusters of E18.5 embryo brains

(A) QC metrics of E18.5 embryo brain experiments. nFeature_RNA is the number of genes detected in each cell and nCount_RNA is the total number of molecules detected within a cell. Percent.mt is the percentage of reads that map to the mitochondrial genome. (B) Umap projection of E18.5 embryo brain (n = 2 embryos, 1,927 cells) colored by cell type prediction.

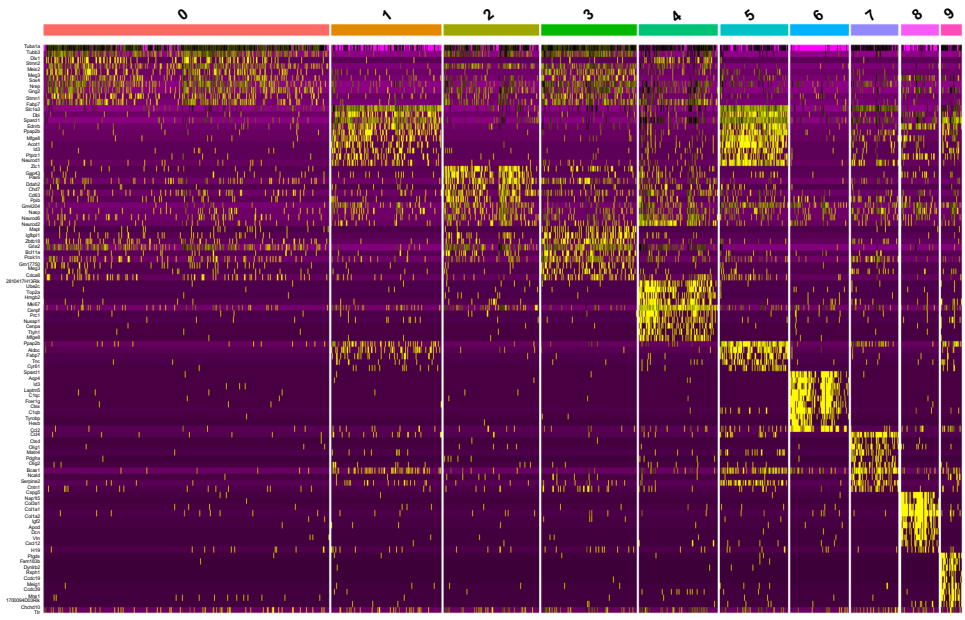

Figure 15. Finding differentially expressed genes in E18.5 embryos

The expression heatmap plotting the top 10 markers in E18.5 embryo brain. The canonical markers for each cell type, which can be used to assess the clustering.

DISCUSSION

Neural crest cells (NCCs) are important in mouse development. NCCs are multipotent cells that can differentiate into multiple cells, and many studies have been conducted on what early progenitor cells become in the adult body. However, it has not been studied exactly when the NCCs are differentiated. Here I performed two experiments to study NCC development.

First, I examined the development of trunk neural crest cells of mouse embryos in an early development stage. Neural crest cells in the E9.5 torso part of the embryo were investigated and tracked to reveal what kind of cell types they differentiate into at E10.5. The results obtained by this experiment unravel important information about NCC development and can aid the study of DiGeorge syndrome, which results from incomplete development of neural crest cells.

Second, I studied the NCC development in the mouse embryo brain during late development stages. NCCs in E12.5 embryo brains were traced to track what kind of cell types they differentiated into until E18.5. DiGeorge syndrome patients suffer from various brain-related diseases, but the causes have not been revealed. Knowledge gained from this experiment can further advance the study of brain damage in DiGeorge syndrome.

The common concept of these two experiments is to study the neural crest cell development by labeling NCCs and CellTag lentivirus with RFP and GFP fluorescence, respectively. To research this, I created the *Wnt1-Cre;R26R-Tomato* mouse model, produced CellTag lentivirus, and set up an *in vitro* culture and *in utero* transplantation experiments. One of the most important factors, which can decide over success and failure of the experiment, is to ensure a high Lentivirus

concentration. Due to the lentivirus being directly injected into the primary cells, a concentration of at least 5 x 10^9 U/mL is required.

The first experimental method used in this study was the injection of lentivirus. An E9.5 mouse embryo was placed in the pit of the agarose gel with the back facing upwards, and then injected using the microcapillary. In summary, the 24-hour *in vitro* culture showed that the RFP was more widespread at E10.5 compared to E9.5. I also investigated single cells of E10.5 embryo torsos and E18.5 embryo brains to confirm which cell types exist at each stage.

Here I introduce the six troubleshootings that I experienced in establishing the protocols. First, I developed a method of creating high concentrations of lentivirus through various reagent combinations. In the preceding paper, they used X-tremeGENE as a lentivirus transfection reagent, but lipofectamine 3000 and packaging plasmids PAX2 and MD2 were used. Second, if you add a small amount of fastgreen, the effect of lentivirus will not change and the color will look green, so it is easy to inject. Third, I developed a method for injecting lentivirus into E9.5 embryo. An E9.5 mouse embryo was placed in the pit of the 4% agarose gel with the back facing upwards, and then the virus was injected using a microcapillary. Fourth, the 10x DMEM must be carefully stored. The 10x DMEM should be stored at -80 degrees because it is highly nutritious and can be easily contaminated. Fifth, in the *in utero* experiment, if a lentivirus is injected into both uteruses, the embryos experience stress. Thus, to minimize time and treatment, only one uterus side should be selected for lentivirus injection. Lastly, when creating a custom reference, if you add only the sequence corresponding to GFP-CellTag-poly(A), it causes an error. However, adding the entire sequence of the CellTag lentiviral vector will resolve the problem.

As a further improvement of the here described protocol, I suggest adding an additional PCR step using CellTag specific primers after the step of PCR in the Drop-seq experiment. This will ensure, that the captured mRNA can be mapped to the CellTag region of the GFP-CellTag transgene included in the new chromosome. Also, a further study is needed to research the differences between the results for this experiment from wild type mice and DGS model mice.

The successful completion of the project provided the first systematic genome-wide molecular interrogation of NCC development and greatly enhances our understanding of NCC development. The resulting genomic map of transcriptional activity will be critical in the comprehensive discovery of genetic players in NCC development which is essential for understanding the functional consequence of mutations found in human patients and for developing early therapeutic solutions for the human patients.

BIBLIOGRAPHY

1. Niibe, K. *et al.* The potential of enriched mesenchymal stem cells with neural crest cell phenotypes as a cell source for regenerative dentistry. *Japanese Dental Science Review* vol. 53 25–33 (2017).
2. Jerome, L. a & Papaioannou, V. E. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat. Genet.* **27**, 286–91 (2001).
3. Abu-Issa, R., Smyth, G., Smoak, I., Yamamura, K. & Meyers, E. N. Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. *Development* **129**, 4613–25 (2002).
4. Choi, M., Stottmann, R. W., Yang, Y.-P., Meyers, E. N. & Klingensmith, J. The bone morphogenetic protein antagonist noggin regulates mammalian cardiac morphogenesis. *Circ. Res.* **100**, 220–228 (2007).
5. Choi, M. & Klingensmith, J. Chordin is a modifier of tbx1 for the craniofacial malformations of 22q11 deletion syndrome phenotypes in mouse. *PLoS Genet.* **5**, e1000395 (2009).
6. Naeim, F., Nagesh Rao, P., Song, S. X. & Phan, R. T. Immunodeficiency Disorders. in *Atlas of Hematopathology* 803–812 (Elsevier, 2018). doi:10.1016/B978-0-12-809843-1.00058-9.
7. Kochilas, L. *et al.* Role of neural crest during cardiac development in a mouse model of DiGeorge syndrome. *Dev. Biol.* **251**, 157–166 (2002).
8. Genetic and Rare Diseases Information Center (GARD). 22q11.2 deletion syndrome.
9. Genetics Home Reference. July 2013. 22q11.2 deletion syndrome. https://rarediseases.info.nih.gov/diseases/10299/22q112-deletion-syndrome.
10. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
11. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science (80-.).* **347**, 1138–1142 (2015).
12. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* vol. 168 613–628 (2017).
13. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental and Molecular Medicine* vol. 50 1–14 (2018).
14. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *BioTechniques* vol. 30 892–897 (2001).
15. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
16. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).

17. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* vol. 148 33–45 (2012).
18. Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).
19. Kong, W. *et al.* CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. *Nat. Protoc.* **15**, 750–772 (2020).
20. Xiong, L. *et al.* Retromer in Osteoblasts Interacts With Protein Phosphatase 1 Regulator Subunit 14C, Terminates Parathyroid Hormone's Signaling, and Promotes Its Catabolic Response. *EBioMedicine* **9**, 45–60 (2016).
21. Kawakami, M., Umeda, M., Nakagata, N., Takeo, T. & Yamamura, K. I. Novel migrating mouse neural crest cell assay system utilizing P0-Cre/EGFP fluorescent time-lapse imaging. *BMC Dev. Biol.* **11**, 68 (2011).
22. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
23. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
24. Kee, Y., Byung, J. H., Sternberg, P. W. & Bronner-Fraser, M. Evolutionary conservation of cell migration genes: From nematode neurons to vertebrate neural crest. *Genes Dev.* **21**, 391–396 (2007).
25. Wilson, Y. M., Richards, K. L., Ford-Perriss, M. L., Panthier, J. J. & Murphy, M. Neural crest cell lineage segregation in the mouse neural tube. *Development* **131**, 6153–6162 (2004).
26. Kalaskar, V. K. & Lauderdale, J. D. Mouse embryonic development in a serum-free whole embryo culture system. *J. Vis. Exp.* 1–7 (2014) doi:10.3791/50803.
27. Cheung, T. T., Weston, M. K. & Wilson, M. J. Selection and evaluation of reference genes for analysis of mouse (Mus musculus) sex-dimorphic brain development. *PeerJ* **2017**, 1–19 (2017).
28. Lindsay, E. A. *et al.* Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature* **401**, 379–383 (1999).
29. Two, G. R. Tbx1 haploinsuf ® ciency in the DiGeorge syndrome region causes aortic arch defects in mice. **6836**, 97–101 (2001).

국문 초록

신경능선세포 (neural crest cell; NCC) 는 발생 중의 배아에서 다분화능을 가지고 이동하는 세포 집단으로, 평활근 세포, 멜라노 세포, 슈반 세포와 뉴런 등 광범위한 조직 형성에 기여한다. 지난 수십 년간 축적된 증거는 분자적 조절이 NCC 발달의 기초가 된다는 것을 시사한다. 이 과정에서 어떤 오류도 구순구개열과 구개, 유전형의 흑색종, 디죠지 증후군(DiGeorge syndrome; DGS)을 포함한 인간의 선천적인 결함을 초래한다. 발달하는 쥐 배아에서 계통 추적이 가능한 세포 클론을 이용하여 NCC 를 연구했고, 이를 위해 실험 프로토콜을 확립했다.

신경능선세포 계통은 *Wnt1-Cre;R26R-Tomato* 쥐에서 셀태그 (CellTag) 렌티바이러스를 사용하여 추적되었다. 배아일 (embryonic; E) 9.5 쥐 배아에서 몸통 NCC 발달을 연구하기 위해 24 시간 동안의 체외 배양법을 확립했고, E12.5 부터 E18.5 까지 6 일 동안 쥐 배아의 뇌에서 NCC 발달을 연구하기 위하여 자궁 내 주입법을 확립했다. E10.5 배아 몸통과 E18.5 배아 뇌의 단일 세포 염기서열(single-cell sequencing; scRNA-seq)을 수행하고 염기서열 결과를 분석하였다.

계통 추적 시스템인 CellTag 렌티바이러스를 E9.5 쥐 배아에 주입하는 방법을 설계했다. 24 시간 체외 배양에 대한 배양 방법을 확립했고, 배양 후 배아에서 tdTomato 형광이 더 뚜렷하게 관찰되었다. 생쥐의 DGS 모델링을 위해 *LgDel* 배아를 복원하여, wild type 쥐와 *LgDel* 쥐를 번식시켰다. E10.5 쥐 배아의 몸통에서 scRNA-seq 데이터를 분석한 결과 심장 근육과

혈관 발달 클러스터가 발견되었다. 대조적으로 E18.5 쥐 배아의 뇌에서 scRNA-seq 데이터를 분석한 결과 교질 세포 분화와 뉴런 분화 클러스터가 나왔다.

본인은 본 연구에서 생쥐 배아에서 NCC 발달의 복잡한 분자 메커니즘을 해결하기 위해 NCC 계통을 추적하는 것을 목표로 했다. 이를 위해 CellTag 라는 계통 추적 시스템을 이용했다. 그러나 CellTag 영역에 맵핑된 리드를 얻으려면, CellTag 특이적인 프라이머를 사용하는 추가 PCR 단계가 필요할 것이다. 결론적으로, 계통 추적을 통한 scRNA-seq 실험 프로토콜을 확립했고, 이 프로토콜을 각 연구에 맞게 약간 수정하여 사용한다면, 발달 세포 사이의 복잡한 관계를 분석하고 NCC 발달에 관련된 새로운 유전자를 발견하는 데 사용될 수 있을 것이다.

주요어: 신경능선세포, 쥐 배아 발달, 단일 세포 서열 분석, 디죠지 증후군, 체외 배양, 자궁 내 이식

학번: 2018-28442