The role of Geminin in Neural Crest Stem Cells, a functional approach through bioinformatics analyses.

Nikolaos I. Kanellakis
School of Medicine
University of Patras

Supervisor: Dr Stavros Taraviras
Associate Professor
I would like to dedicate this thesis to my loving parents and brother.
Acknowledgements

Voici mon secret. Il est très simple: on ne voit bien qu’avec le cœur.
L’essentiel est invisible pour les yeux.

Antoine de Saint Exupery

First of all, I would like to thank Dr. Stavros Taraviras and Dr. Zoi Lygerou, as they introduced me to what is the science of biology, and taught me how to work in the lab. It’s been a wonderfull journey working with them for the past two years. Furthermore, I would like to thank Dr. Vasilis Megalooikonomou from my undergraduate course at CEID, I am really happy and proud to have such a good teacher. Athanasios Tsakalidis is another professor from CEID I would like to thank.

Special thanks to my supervisor Dr. Alexandra Patmanidi who was always there for me. Of course i should mention the members of the Stem Cell and Cell Cycle Laboratories as they welcomed me as a part of their team.

I would not have completed this work without the valuable help and the support of my family and friends Nikos and Maria, who are now waiting for the arrival of their first baby boy. Eleni :) who has always been on my side and her support during hard times, Marietta ;) as she was always trying to cheer me up, Giannis who learned me such great things, Ioanna for her humour, Tasos for the things we did together (some good some stupid) and his girlfriend Dimitra and Ionas who was the first to correct this text. I would like to mention the supportiveness of Thomas, Giota and Chris from CEID, Amalia and Christos who never stopped encouraging me, George who will always be young and genious. In 2006, I came accross a very special and
valuable friend of mine Alexis, it will always be a pleasure discussing and exchanging points of view with him. As well i would like to thank Eve for helping me with the text and the presentation of my master thesis. Last but not least my thanks to T.A. for motivating and inspiring me to go further. Special thanks to ma cherie Celine who helped and supported me to complete this work even thought she is a little bit further than i wished her to be.

Nick K. 2014
Abstract

Geminin (Gmn1) is a nuclear protein that regulates the licensing of DNA replication, as well as cell fate decisions. To this end, it is becoming increasingly evident that Geminin regulates transcriptional programs associated with lineage specification. Neural crest cells (NCCs), which originate at the dorsalmost region of the neural tube, are a transient, embryonic cell population characterized by unusual migratory ability and developmental plasticity unique to vertebrates that gives rise to a diverse cell lineage including melanocytes, craniofacial cartilage and bone, smooth muscle, peripheral and enteric neurons and glia. The Neural crest cell population shows a tremendous potential to differentiate and contribute to almost every organ in the adult body. In fact they are so important that they have been called the forth germ layer. Neural Crest Compartment can be divided into four main functional cell lineages, which include the cranial neural crest, trunk neural crest, cardiac neural crest, vagal and sacral neural crest. The aim of this study is to identify the gene network directed by geminin in order to coordinate the neural crest stem cell derivation towards mature cells.

In order to examine the role of Geminin on neural crest cells, we have used conditional knockout mice to ablate Geminin early in the developing embryonic neural crest compartment. Specifically, we have used mice that express the Cre recombinase under the control of the Wnt1 gene regulatory elements (Wnt1:iCre) and the loxP system to delete Geminin from neural crest cells. In order to dissect the molecular basis of the phenotype, we have used cDNA microarray analyses, which revealed the deregulation of genes involved in cellular lineage
specification transcription programs and epigenetic regulation. Gene expression data mining in combination with in silico comparative analyses support further our findings and suggest a key role for Geminin in neural crest development at the transcriptional and epigenetic levels. For the unpaired T-test we set p value threshold at 0.05. We have a scientific interest only in genes that have a fold change greater than the absolute value of 1.5.

Our results show that deletion of Geminin causes in embryos severe morphological and craniofacial malformations. Embryos lacking Geminin at neural crest cells die a few days before birth. Geminin acts as a key regulatory element in neural crest stem cells during embryogenesis. Geminin is associated with pathways that regulate neural crest cell self-renew and differentiation. In other words, geminin is essential for the proper development of all the tissues that derive from the neural crest stem cells population.
Contents

1 Introduction

1.1 Stem Cell

1.1.1 Embryonic Stem Cells (ESCs)

1.1.2 Adult Stem Cells (ASCs)

1.2 Neural Crest Cells

1.2.1 Neural Crest Induction

1.2.2 Cranial Neural Crest Cells

1.2.3 Trunk Neural Crest Cells

1.2.4 Vagal and Sacral Neural Crest Cells

1.2.5 Cardiac Neural Crest Cells

1.3 Microarray

1.3.1 DNA Microarrays

1.3.1.1 Gene Expression Profiling

1.3.1.2 Chromatin immunoprecipitation on Chip

1.3.1.3 SNP detection

1.3.1.4 Comparative genomic hybridization

1.3.1.5 Alternative splicing detection

1.3.1.6 Fusion genes microarrays and Tiling arrays

1.3.1.7 Tiling arrays

List of Figures x

Nomenclature xiv
CONTENTS

1.3.2 cDNA microarrays ........................................ 16
   1.3.2.1 Work principle .................................. 16
   1.3.2.2 Microarray synthesis techniques ................. 17
1.3.3 Affymetrix Microarray Chips .......................... 19
1.3.4 Illumina BeadChip Arrays ............................... 21
1.3.5 Agilent Microarrays .................................... 21
1.3.6 Microarray Experimental Process ..................... 21
   1.3.6.1 Collection of the Microarray Data .............. 21
   1.3.6.2 Robust Multi-array Average - RMA ............... 22
   1.3.6.3 Background Noise Correction ................... 22
   1.3.6.4 Normalization ................................... 23
   1.3.6.5 Probe Summarization .............................. 23
   1.3.6.6 Hybridization Quality Control .................. 24
   1.3.6.7 Statistical Analysis .............................. 25
   1.3.6.8 Fold Change .................................... 26
   1.3.6.9 Meta-analysis of the Microarray Data ........... 27
1.4 Next Generation Sequencing, NGS ........................ 28
   1.4.1 RNA Sequencing, RNA-Seq .......................... 28
1.5 Geminin is Involved in Cell Cycle Regulation and Cell Fate Decision 29
   1.5.1 The structure of Geminin ............................ 29
   1.5.2 Geminin Regulates the Cell Cycle ................... 29
   1.5.3 The role of Geminin in differentiation ............. 30
1.6 Mouse as a model organism in biomedical research .... 31
   1.6.1 The advantages of Mouse as a model organism .... 31
   1.6.2 Jackson Laboratory ................................. 32

2 Aims and Objectives ........................................ 34

3 Materials and Methods .................................... 35
   3.1 Experimental Animals ................................. 35
      3.1.1 Cre Recombination System ......................... 36
   3.2 Mice Breedings ....................................... 38
   3.3 Mouse Embryo Dissection ............................. 39
CONTENTS

3.4 Genotyping ................................................. 39
  3.4.1 DNA Extraction ................................. 39
  3.4.2 Polymerase Chain Reaction ....................... 40
    3.4.2.1 Geminin^{Fl/Wt} PCR .................... 41
    3.4.2.2 Geminin^{Ko/Wt} PCR ................... 43
    3.4.2.3 Cre Recombinase^{+/-} PCR ............... 44
    3.4.2.4 Rosa26^{STOPYFP} PCR .................. 46
  3.4.3 Electrophoresis ..................................... 48
3.5 Affymetrix GeneChip cDNA Microarrays Hybridization .......... 49
  3.5.1 Synthesize first-strand cDNA ................. 49
  3.5.2 Synthesize second-strand cDNA ............... 50
  3.5.3 Synthesize cRNA by In Vitro Transcription .... 51
  3.5.4 cRNA Purification ............................. 51
  3.5.5 Assess cRNA yield and size distribution .... 53
  3.5.6 Synthesize 2nd-cycle cDNA .................. 54
  3.5.7 Hydrolyze using RNase H .................... 55
  3.5.8 Purify 2nd-cycle cDNA ....................... 55
  3.5.9 Assess cDNA yield and size distribution .... 57
  3.5.10 Fragment and label the single-stranded cDNA .... 57
  3.5.11 Labeling of Fragmented Single-Stranded DNA .... 59
  3.5.12 Hybridization .................................. 60
  3.5.13 Wash, Stain and Scan .......................... 61
3.6 Microarray Analysis .................................... 61
3.7 Solutions ................................................. 63

4 Results ................................................. 66
  4.1 Animal Models ...................................... 66
    4.1.1 Geminin Knock Out PCR ................... 66
    4.1.2 Geminin floxed PCR ....................... 68
    4.1.3 Wnt1 Cre PCR ................................ 68
  4.2 Total RNA Isolation ................................ 69
  4.3 MicroArray Chip preparation ...................... 69
  4.4 Differentially expressed gene profile ............ 70
CONTENTS

4.5 Transcription Factors ............................................. 77
  4.5.1 The expression of the Hox-family genes is induced .... 77
  4.5.2 Members of the Alx-family are down regulated .... 78
  4.5.3 Gata family members are down regulated .... 79
  4.5.4 Fox gene family members change their expression in the
       absence of Geminin ........................................ 80
  4.5.5 Ascl1 is upregulated ..................................... 81
4.6 Chemokines Regulation ......................................... 81
4.7 Boxplot of our data ........................................ 82
4.8 Unsupervised Hierarchical Clustering ....................... 83
4.9 Principal Component Analysis (PCA) ....................... 85
4.10 Gene expression Meta-Analysis .............................. 86
4.11 Gene Ontology Analysis .................................... 89
4.12 Phenotype Analysis ......................................... 92
4.13 Pathway Analysis ............................................ 92
  4.13.1 Geminin seems to be involved into the regulation of the
       epithelial to mesenchymal transition .................. 92
  4.14 Geminin affects Notch signaling pathway ............. 94
  4.15 The deletion of Geminin affects cardiogenesis ....... 94

5 Discussion ......................................................... 96
  5.1 Neural Crest Stem Cells and Geminin .................... 96
  5.2 Similar Studies ............................................... 98
    5.2.1 The role of Geminin in hematopoietic stem cells .... 98
    5.2.2 The role of Tgfbr2 in neural crest stem cells ...... 100
  5.3 Conclusion ...................................................... 101

References .......................................................... 102
List of Figures

1.1 Proliferation and differentiation are the main properties of stem cells. The balance between proliferation and differentiation is extremely significant for the development of a multicellular organism. 2

1.2 Stem cell potency. Towards differentiation stem cells lose differentiation abilities and obtain tissue specific characteristics. . . . . . 4

1.3 Stem Cell lineages. Differentiation, a process strictly directed by certain transcription factors by which a non specialized cell becomes a functionally specialized cell type . . . . . . . . . . . . . . 4

1.4 Dedifferentiation, transdifferentiation, reprogramming are key processes for regenerative medicine field. . . . . . . . . . . . . . . 5

1.5 Neural crest stem cells are a transient, multipotent and migratory cell population. Neural Crest stem cell derivatives contribute to almost all tissues of the developed organism. Adapted from Knecht, 2002 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

1.6 Neural Crest formation. Neural crest is induced in the region of the neural plate border. After neural tube closure, neural crest delaminates from the region between the dorsal neural tube and overlying ectoderm and migrates out towards the periphery. Adapted from Wikipedia . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

1.7 In the early developmental stages of the neural tube, segmentation of the neuroepithelium occurs. This segmentation turns into a series of neuromeres. Each segment is called a rhombomere. Every rhombomere develops its own set of ganglia and nerves. Adapted from Dr. Brian E. Staveley, Memorial University of Newfoundland 11
1.8 Neural Crest stem cell derivatives. Different neural crest populations along the rostrocaudal axis Tatjana Sauka-Spengler and Marianne Bronner, Cell, 2010 ........................................... 13
1.9 Hybridization of the target to the probe. Fluorescently labeled target sequences bind to a probe sequence. Adapted from Affymetrix manual .......................................................... 17
1.10 Affymetrix GeneChips used for gene expression profiling, the match at the bottom left is to show to size of the chip. Adapted from Affymetrix website ......................................................... 19
1.11 Mus Musculus (House Mouse), is a small mammal of the order Rodentia. Laboratory mouse, which is one of the most important model organisms in biology and medicine. ......................... 32
3.1 Geminin wild type allele has 7 exons ........................................ 36
3.2 Geminin floxed allele has 2 loxP loci ........................................ 36
3.3 Geminin ko allele is not functional ........................................... 36
3.4 The Cre Recombination system .............................................. 37
3.5 Rosa26 locus ................................................................. 38
3.6 A strip of eight pcr tubes ................................................... 41
3.7 Geminin wild type allele primer starting sites ......................... 41
3.8 Geminin floxed allele primer starting sites ............................. 42
3.9 Geminin knock out allele primer starting sites ....................... 43
3.10 The PCR thermal cycler used for the PCR reactions ............... 48
3.11 The figure shows an example of typical results on fragmented samples. ................................................................. 59
3.12 T-test used in case the groups have equal variance and unequal sample size ................................................................. 62
4.1 A Geminin knock out pcr (electrophoresis gel). A 1000bp ladder is used in order to determine the length of the DNA products. A positive and negative control is included. We tagged all the samples according to the Geminin allele that they bear .............................. 67
4.2 Geminin floxed pcr. As a positive control we used a Geminin Floxed/Floxed sample. Negative control is a Geminin Wild Type / Wild Type sample. We have tagged all the samples according to their genotype. 68

4.3 Wnt1 Cre pcr. Cre recombinase is under the control of Wnt1 regulatory elements in order to ablate Geminin specifically in neural crest stem cells. All samples have been tagged according to their genotype. 69

4.4 A volcano plot indicating the genes of interest throughout the total number of genes that have been hybridized. In our study we have 865 genes of interest, indicated with the blue dots. 72

4.5 The table presents the properties of the unpaired t-test and the resulting genes of interest, according to the different values for p-value and fold change. As the criteria that we will set are getting more strict there are fewer genes of interest. 73

4.6 A heatmap presenting the differential gene expression profile of genes of interest. Heatmap was used to visualize easily and compare the expression of many genes across our conditions - groups. The fold change of a gene is represented by the color and the color’s intensity. To indicate the minimum gene expression we use far red, while maximum gene expression is indicated by the far green. 74

4.7 Venn diagramm presenting the upregulated, downregulated and the total number of genes 75

4.8 Top 10 of down and up regulated genes 76

4.9 Summarization of the characteristics of the genes of interest. 76

4.10 Heatmap of the differential expression of the hox transcription factors. All of the hox transcription factors are up-regulated. 78

4.11 Heatmap presenting the differential expression of alx gene family 79

4.12 Transcription factors in sympathetic neuron development 80

4.13 Fox gene family regulation 81

4.14 Heatmap presenting the regulation of the chemokines present in our study. Cxcl12 and Cxcr4 signaling is responsible for the mesencephalic neural crest stem cells migration. 82
4.15 A boxplot is often used in bioinformatics to easily present the data. Here we present the box plot of our data. We have two separate groups of embryos Experimental and Wild Type. The vertical axis is the expression value of the genes. Through a boxplot it is easy to identify outliers. ......................... 83

4.16 Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. In our case there are two groups of embryos, Experimental and Wild Type. For each sample there is a heatmap representing the expression value of the genes of interest. 84

4.17 Principal Component Analysis graph. The red dots represent the experiment (Geminin knock out) mice while the blue dots represent the wild type mice ......................... 85

4.18 The top associated network functions that are affected by the deletion of geminin. The functions that vital for the homeostasis and the proper development of the embryo ..................... 87

4.19 Diseases and disorders caused by the deletion of geminin in neural crest stem cells. ................................................. 88

4.20 Physiological System Development and Function .................... 88

4.21 Top canonical pathways over represented in our data. The pathways affected indicate the key role of geminin in neural crest cells. 89

4.22 The figure presents the top 5 cellular components that are altered by the deletion of Geminin. ................................. 90

4.23 Gene ontology molecular function describes the elemental activities of a gene product. ............................................. 91

4.24 Gene Ontology Biological Process refers to all operations of molecular events. Here we present the top 5 biological processes that are affected by the deletion of Geminin in neural crest stem cells. 91

4.25 The phenotype analysis was perfomed with WebGestalt, a web based bioinformatics tools developed by Vanderbilt University. . . 92

4.26 Heatmap of statistically significant differentially expressed genes, which are involved into the regulation of EMT. .................... 93

4.27 Heatmap of statistically significant differentially expressed genes, which are involved into the notch pathway. .................... 94
LIST OF FIGURES

4.28 Heatmap of statistically significant differentially expressed genes, which are involved into cardiogenesis. ........................................... 95

5.1 Heatmap of differential expression of some of the most important transcription factors. ................................................................. 97

5.2 Venn diagram presenting statistically significant up and down regulated genes throughout the total number of genes that was hybridized to the microarray chip. .................................................. 99

5.3 There are 98 common genes between the Vav and Wnt1 microarray studies, 14 of them are transcription factors. Here we present the transcription factors with their fold change (FC) ................. 100

5.4 There are 54 common genes between the our study and Iwata et al. study. Here we present some of genes associated with craniofacial development with their fold change (FC) ......................... 101
Chapter 1

Introduction

1.1 Stem Cell

Stem Cells differ from all other cell types because they have unique properties and characteristics that reflect mainly into their ability to self-renew and differentiate. The main stem cell categories are pluripotent and multipotent. Pluripotent stem cells are derived from preimplantation embryo and they have the ability to differentiate into all three germ layers. On the other hand, somatic stem cells, either embryonic or adult, are multipotent: they give rise to specialized cells of a specific tissue in order to replenish any damaged pool and maintain homeostasis. [Shipony et al., 2014] Embryonic stem cells, are isolated from the inner cell mass of blastocysts, whereas adult stem cells, are found in various tissues [Cao et al., 2014]

Stem Cells share the next two fundamental properties:

- ability to divide and self-renew while remain undifferentiated,
- the ability to differentiate into multiple specialized cell types
Figure 1.1: Proliferation and differentiation are the main properties of stem cells. The balance between proliferation and differentiation is extremely significant for the development of a multicellular organism.

In 2006 Takahashi and Yamanaka a group from Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan managed to reprogram differentiated cells to an embryonic-like state [Takahashi and Yamanaka, 2006]. This new type of stem cell, was named induced pluripotent stem cells (iPSCs). Induced pluripotent stem cells, not only show the same morphology and growth properties as embryonic stem cells but in addition express ES cell marker genes. Induced pluripotent stem cell (iPSC) technology gave us a new hope and perspectives on how to cure various human diseases. There are many reasons that make stem cells valuable, the most important among them is their unique renewal ability. [Shoudong Ye1 and Ying, 2014]

Unlike other cell types, stem cells divide and self-renew during an organism’s life span. In vitro embryonic stem cells are capable to proliferate and remain undifferentiated for a year or even more. [Endele et al., 2014] [Fan et al., 2004] Proliferation is not dependent on differentiation [Andng et al., 2008] Stem cells that proliferate can either undergo symmetric or asymmetric cell divisions. Symmetric cell division contributes to the maintenance of stem cell pool while in the other hand asymmetric cell division gives rise to committed progenitors. Stem cells have special and accurate mechanisms that control their cell cycle phase [Tsai and McKay, 2002].

Potency is the ability of stem cells to differentiate into specialized cells (neurons, glia, erythrocytes, epithelial cells, hepatocytes, mast cells). Stem cells do not have any structures that are tissue specific. [Wu et al., 2013] Through the complex progress of differentiation a stem cell will give rise to a fully differenti-
Differentiation is not a single stage process, the cell has to go through various stages until it becomes fully specialized. The different stem cell states are:

- **Totipotency**, the ability of a cell to produce any cell type through differentiation. Totipotent cells are the zygotes. These kind of cells have the greatest differentiation potential. After the fertilization the zygote gives rise to totipotent stem cells.

- **Pluripotency**, the ability of an embryonic stem cell to differentiate to any of the three germ layers, endoderm (lungs, interior stomach lining), mesoderm (bone, muscles) and ectoderm (nervous system, epidermal tissues). Induced Pluripotent Stem Cells belong to this state of potency as they can differentiate to any of the germ layers.

- **Multipotency**, the ability of progenitor cells to differentiate into multiple, but limited cell types. Progenitor cells are found in most human tissues. For example Hematopoietic progenitors give rise to all blood cells.

- **Oligopotency**, the ability of progenitor cells to differentiate into a few cell types, like myeloid stem cells.

- **Unipotency**, the ability of a unique progenitor to differentiate to a unique cell type, for example the osteoblasts generate only bone tissue.

[Ishiuchi and Torres-Padilla, 2013]
Figure 1.2: Stem cell potency. Towards differentiation stem cells lose differentiation abilities and obtain tissue specific characteristics.

Figure 1.3: Stem Cell lineages. Differentiation, a process strictly directed by certain transcription factors by which a non specialized cell becomes a functionally specialized cell type.

The ability of stem cells to self-renew indefinitely in combination with our
ability to induce their fate opens up new horizons of their use in medicine. In multicellular organisms stem cells serve as a repairing system for damaged tissues. [Aurora and Olson, 2014] Regenerative medicine deals with the process of engineering or regenerating human cells, tissues or organs to restore or establish normal function. The term regenerative medicine was used for the first time in 1992 by Kaiser [Kaiser, 1992]. Recent findings on stem cells and their plasticity are promising and aim to inaugurate a new era in organ and tissue transplantation technology. There are three main regenerative processes:

- **Dedifferentiation**,  
- **Transdifferentiation**,  
- **Reprogramming**

![Diagram showing processes of totipotent, pluripotent, multipotent, and unipotent cells with arrows indicating dedifferentiation and reprogramming](image)

Figure 1.4: Dedifferentiation, transdifferentiation, reprogramming are key processes for regenerative medicine field.

Although a number of stem cell therapies have already been developed, it is clear that regenerative medicine is not yet ready to be used in clinical practice. It is anticipated that in the future the use of adult stem cells will be able to cure certain diseases of the central nervous system, kidney and cardiac failure as well as genetic diseases or cancer. [Hirschi et al., 2013]
1.1.1 Embryonic Stem Cells (ESCs)

Embryonic stem cells (ESCs) are able to give rise to any of the three germ layers (endoderm, mesoderm, ectoderm), as they are pluripotent. They are derived from the inner cell mass of the blastocyst [Kim, 2014]. Besides the ability of pluripotency embryonic stem cells can replicate indefinitely in cell culture [Kleinsmith and G. Barry Pierce, 1964]. ESCs are a usefull tool for biomedical research, tissue engineering and regenerative medicine. [Kobayashi et al., 2014] [Adama et al., 2014] In addition human embryonic stem cells are being used as models for various genetic disorders like cystic fibrosis. Scientists can generate induced pluripotent stem cell lines that have different genetic mutations. The use of ESCs in biomedical research rises a few ethical issues.

1.1.2 Adult Stem Cells (ASCs)

Adult stem cells are found in fully developed multicellular organisms. They are undifferentiated somatic cells, which can proliferate indefinitely and differentiate into committed progenitors, and these -in turn- to all the cell types of the tissue that they originate from. [Yulin et al., 2012] Under normal conditions they can not differentiate to all cell types. Their role is to maintain the population or repair the cells, tissues or organs in need. Some tissues or organs have more stem cells than others as they are responding really well to regenerative pressure, (bone marrow, liver) while other tissues show a poor response (neurons, heart). [Matsumoto et al., 2011] Studies claim that when adult stem cells are exposed to specific stimuli they show a remarkable plasticity in their developmental potentialities. Stem cells have been identified in many more tissues that once scientists expected. [Gonzalez-Perez, 2014]

There are various types of adult stem cells that have been indentified in many mammalian organs and tissues.

- **Hematopoietic** stem cells, give rise to all blood cells, including red blood cells, B lymphocytes, T lymphocytes, natural killer cells, neutrophils, basophils, eosinophils, monocytes, platelets and macrophages [Farrella et al., 2014]
• **Mesenchymal** stem cells, give rise to different types of mature cells like, bone, cartilage and fat cells. The connective tissue is also derived from mesenchymal stem cells.

• **Neural** stem cells, give rise to neurons (nerve cells) and non neuronal cells like astrocytes and oligodendrocytes.

• **Epithelial** stem cells give rise to various types of mature cells like, absorptive cells, goblet cells, paneth cells, and enteroendocrine cells

• **Skin** stem cells are found in the basal layer of the epidermis and at the base of hair follicles.
1.2 Neural Crest Cells

The Neural Crest Cells are a transient, multipotent, migratory cell population, which contributes to cellular diversity as it has unique features [Stuhlmiller and Garca-Castro, 2012] The Neural crest cell population shows a tremendous potential to differentiate and contribute to almost every organ in the adult body. In fact they are so important that they have been called the forth germ layer. [Hall, 2000] Neural crest cells are derived from the neural epithelium, at the dorsalmost region of the neural tube, and can only be found in vertebrates [Gilbert, 2000]. They give rise to multiple differentiated cell lines. The plasticity of neural crest stem cells is so remarkable that they are being used as a model for cell fate decision, migration and differentiation. [Theveneau and Mayor, 2012] The fate of the cells that originate from neural crest cells depends partially on where they will migrate to and finally settle.

Neural crest cells are capable to generate a prodigious number of differentiated cell types. [Bittencourt et al., 2013]

- **Peripheral Nervous System** (PNS), neurons, sensory ganglia, sympathetic and parasympathetic ganglia neuroglial and Schwann cells.

- **Epidermal pigment cells**

- **Connective tissue**, dermis, smooth muscle, lachrymal, thymus, thyroid, smooth muscle in arteries of aortic arch origin

- **Facial cartilage and bone**

- **Adrenal medulla**
Figure 1.5: Neural crest stem cells are a transient, multipotent and migratory cell population. Neural Crest stem cell derivatives contribute to almost all tissues of the developed organism. Adapted from Knecht, 2002

Neural crest cells can be divided into four main regions. [Gilbert, 2000]

i **Cranial** neural crest cells

ii **Trunk** neural crest cells

iii **Vagal and Sacral** neural crest cells

iv **Cardiac** neural crest cells

### 1.2.1 Neural Crest Induction

Neural crest formation is a significant and complex developmental process. Neural Crest induction begins early in the developing embryo during gastrulation. Multiple signaling pathways and transcription factors are involved in neural crest induction and formation. [Stuhlmiller and Garca-Castro, 2012]
Figure 1.6: Neural Crest formation. Neural crest is induced in the region of the neural plate border. After neural tube closure, neural crest delaminates from the region between the dorsal neural tube and overlying ectoderm and migrates out towards the periphery. Adapted from Wikipedia

Neural crest formation occurs in three phases. In the first place during gastrulation, neural crest cells are induced in the future neural plate border. Then, during neurulation Wnt, Notch and BMP signaling maintain the neural crest cell population. Members of the Msx, Pax and Zic families are activated during the first phase of Neural Crest while Snail2, FoxD3, Sox9 and Sox10 appear to be definitive neural crest markers. [Gilbert, 2000] Neural crest cells induction, delamination, migration and differentiation is directed by specific signaling pathways and transcription factors. Epithelial - mesenchymal transition (EMT) is a series of molecular events that occur just after the neuroepithelium is formed when the neural crest cells change from a pseudoepithelial to a mesenchymal state. [Groves and LaBonne, 2014]

Defects in neural crest formation are associated with a variety of human dis-
eases or malformation known as neurocristopathies, melanomas, neuroblastomas, Hirschsprung and Waardenburg syndrome and cleft palate. The signaling pathways that induce neural crest generation appear to be well conserved among different species, despite the fact that timing varies.

1.2.2 Cranial Neural Crest Cells

Cranial neural crest cells arise in the region between non-neural and neural ectoderm and migrate towards the developing craniofacial region. The majority of skeletal and connective tissues in the face is derived from cranial neural crest cells. There are seven rhombomeres from which cranial neural crest cells migrate to colonize the craniofacial regions. In the vertebrate embryo, a rhombomere is a transiently divided segment of the developing neural tube. [Glover, 2001]

Figure 1.7: In the early developmental stages of the neural tube, segmentation of the neuroepithelium occurs. This segmentation turns into a series of neuromeres. Each segment is called a rhombomere. Every rhombomere develops its own set of ganglia and nerves. Adapted from Dr. Brian E. Staveley, Memorial University of Newfoundland
Cranial neural crest cells that migrate dorsolaterally will generate the craniofacial mesenchyme and will finally differentiate into craniofacial structures like cartilage, bone, cranial neurons, glia, and connective tissues. Thymic cells and the bones of middle ear and jaw have cranial neural crest origination. [Steventon et al., 2014]

1.2.3 Trunk Neural Crest Cells

Trunk neural crest cells are involved in two main migratory pathways. The first pathway consists of the cells that migrate dorsolaterally and give rise to melanocytes. The neural crest cells that migrate ventrolaterally take the second differentiation pathway and give rise to the dorsal root ganglia. Some of those cells continue to migrate more ventrally and generate the sympathetic ganglia and the nerve clusters surrounding the aorta. [Miquerol and Kelly, 2013]

1.2.4 Vagal and Sacral Neural Crest Cells

Vagal and Sacral neural crest cells generate the parasympathetic ganglia of the enteric nervous system. If vagal and sacral neural crest cells do not migrate towards the colon, enteric ganglia will not be generated. Vagal neural crest cells contribute most to the induction of the enteric precursors along the entire length of the gut. [Sasselli et al., 2012]

1.2.5 Cardiac Neural Crest Cells

Cardiac neural crest cells produce the entire musculoconnective tissue wall of the large arteries as they arise from the heart. Cardiac neural crest cells expressing specific transcripton factors (Sox4, Sox11) migrate and differentiate to form the outer tract of the heart. [Paul et al., 2013]
1.3 Microarray

Microarrays are used to assay large amounts of biological data using high-throughput screening in a miniaturized, multiplexed and parallel processing detection method. The microarray chip is a 2D array on a solid substrate, usually a glass slide or silicon thin-film cell. [Ham et al., 2007]

The types of microarrays that are being used in the laboratories are:

- DNA microarrays, such as cDNA microarrays, oligonucleotide microarrays, BAC microarrays and SNP microarrays

- Chemical compound microarrays
Nowadays the technology of DNA microarrays has become not only the most widely used, but also the most sophisticated while the use of protein, peptide and carbohydrate microarrays is tending to expand. Affymetrix is a leading company in the development of DNA microarrays.[Hong et al., 2014]

1.3.1 DNA Microarrays

DNA microarrays can be used for a variety of research purposes like:

1.3.1.1 Gene Expression Profiling

DNA microarrays can be used to simultaneously monitor the expression levels of thousands of genes in a specific cell population and study parameters, such as the effects of developmental stages, treatments or diseases on gene expression. DNA microarray gene expression profiling has been used to identify the response to a medical treatment on gene expression level.[Schena et al., 1995]

1.3.1.2 Chromatin immunoprecipitation on Chip

Chromatin immunoprecipitation can be used to isolate the DNA sequences where specific proteins bind. Subsequently the DNA fragments can be hybridized to a
microarray chip. This allows the identification of particular sites on a genome where proteins bind. Identifying where on a specific genome RNA Polymerase II binds will help us to study the transcription landscape.

1.3.1.3 SNP detection

Microarray technology can be used to identify Single Nucleotide Polymorphisms (SNPs) either within or among populations. The applications of SNP microarrays include genotyping, measuring predisposition to a disease and identifying drug-candidates or somatic mutations in cancers.

1.3.1.4 Comparative genomic hybridization

Comparative genomic hybridization is being used to measure the changes of the genome content (DNA copy number) in different cell types or closely related organisms. It is usually being used to profile cancer cells.

1.3.1.5 Alternative splicing detection

An exon junction array is used to assay the expression of alternative spliced forms of a gene. In other words they are used to detect the exons of a gene and the different splice isoforms. The probes that are being used target the expected or predicted exons.

1.3.1.6 Fusion genes microarrays and Tiling arrays

Fusion genes microarrays can also be used to detect fusion transcripts, for example from cancer specimens. The design of special probes allows us to measure combined measurements of chimeric transcript junctions.

1.3.1.7 Tiling arrays

Finally genome tiling arrays are being used to detect expression of transcripts or alternatively spliced forms which may not have been previously known or predicted. The probes that are being used are overlapping and represent the genomic area of interest.
1.3.2 cDNA microarrays

A DNA microarray is a collection of microscopic DNA spots attached to a solid surface. DNA microarrays are used to measure the expression levels of large numbers of genes simultaneously. Each DNA spot contains picomoles of a specific DNA sequence, known as probes (also known as reporters or oligos). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA (also called anti-sense RNA) sample (called target) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. The process of measuring gene expression via cDNA is called expression analysis or expression profiling. [Maskos and Southern, 1992]

1.3.2.1 Work principle

Microarray technology is based on the property of hybridization and the Central Dogma of molecular biology. Hybridization between two DNA strands is the core principle of cDNA microarrays. DNA strands with complementary nucleic acid sequences pair with each other by forming hydrogen bonds between the complementary nucleotide base pairs as shown in figure 1.10. As a result a double stranded DNA fragment with a high number of complementary base pairs has a tighter non-covalent bonding between the two strands. Fluorescently labeled target sequences that bind to a probe sequence generate a signal. Total strength of the signal, from a spot, depends upon the amount of target sample binding to the probes present on that spot. Microarrays use relative quantitation in which the intensity of a feature is compared to the intensity of the same feature under a different condition, and the identity of the feature is known by its position.
The central biology dogma describes the flow of genetic information. DNA information is copied into mRNA, through a process called transcription, and proteins are synthesized using the information in mRNA as a template, through the process of translation. The use of cDNA microarrays allows scientists to measure the mRNA levels. The central biology dogma implies that mRNA levels are associated with gene expression levels.

DNA microarrays are being used to measure changes in gene expression levels, detect single nucleotide polymorphisms (SNPs) and genotype. Microarray chips differ in fabrication, workings, accuracy, efficiency, and cost.

### 1.3.2.2 Microarray synthesis techniques

There are two major types of cDNA microarray synthesis techniques, spotted and oligonucleotide microarrays.

In **spotted microarrays**, the probes are oligonucleotides spotted onto glass. Oligonucleotides are cDNA or small fragments of PCR products that correspond to specific mRNAs. The first step is the construction of the probes and then the deposition on the array surface. The most common approach uses a robotic arm to control an array of fine pins, which are deposit each probe at the desired
locations on the surface of the array. [Goldmann and Gonzalez, 2000]

Finally the "grid" of probes, which has just been created, represents the nucleic acid profiles of the prepared probes and is ready to receive complementary cDNA or cRNA "targets" derived from experimental or clinical samples. This technique is used by research scientists around the world to produce "in-house" printed microarrays from their own labs. It is easy to customize these arrays for each experiment, as each scientist is free to choose not only the probes that he will use but also the printing locations on the arrays. [Lausted, 2004]

In addition this technique provides a relatively low-cost microarray that may be customized for each study, and avoid the costs of purchasing often more expensive commercial arrays that may represent vast numbers of genes that are not of interest to the investigator.

In oligonucleotide microarrays, the probes are short sequences designed to match parts of the sequence of known or predicted open reading frames. Despite the fact that oligonucleotide microarrays are also used in "spoted microarrays" the term in most cases refers to the manufacturing technique.

Oligonucleotide arrays are produced by printing short oligonucleotide sequences designed to represent a single gene or family of gene splice-variants by synthesizing this sequence directly onto the array surface instead of depositing intact sequences. This is the main difference between the two manufacturing techniques. Sequences may be longer (60-mer probes such as the Agilent design) or shorter (25-mer probes produced by Affymetrix) depending on the desired purpose. The longer probes are used when we wish to increase the specificity of the target genes on the other hand shorter probes may be spotted in higher density across the array and as a result they are cheaper to manufacture. [Nuwaysir et al., 2002]

The technique that Affymetrix is using to produce the oligonucleotide arrays is based on photolithographic synthesis on a silica substrate where light and light-sensitive masking agents are used to "build” a sequence one nucleotide at a time across the entire array. Each applicable probe is selectively "unmasked” prior to bathing the array in a solution of a single nucleotide, then a masking reaction takes place and the next set of probes are unmasked in preparation for a different nucleotide exposure. After many repetitions, the sequences of every probe become fully constructed. More recently, Maskless Array Synthesis from
NimbleGen Systems has combined flexibility with large numbers of probes. [Pease et al., 1994]

1.3.3 Affymetrix Microarray Chips

Affymetrix, Inc. is an American company, based in California that manufactures DNA microarrays. Affymetrix which is the world leading company in microarray chip manufacture is specialized in oligonucleotide arrays. These arrays are used for gene expression profiling. Affymetrix microarray chip are called GeneChips and are used to determine which genes exist in a sample by detecting specific pieces of mRNA. A single chip can be used to do thousands of gene profiling experiments in parallel. In other words with a single chip we can measure the regulation of many genes in parallel. GeneChips can be used only once.

![Affymetrix GeneChips](Image)

Figure 1.10: Affymetrix GeneChips used for gene expression profiling. the match at the bottom left is to show to size of the chip. Adapted from Affymetrix website

Affymetrix is using oligonucleotides, that are directly synthesized onto a glass wafer by a combination of semiconductor-based photolithography and solid phase chemical synthesis technologies. Each array contains up to 900,000 different oligos and each oligo is present in millions of copies. The oligonucleotides that Affymetrix is using have a length of 25 bases. Since we know the location of every oligonucleotide probe on the chip the hybridization patterns and signal intensities can be interpreted in terms of gene identity and relative expression.
levels.

Each gene is represented on the array by a series of different oligonucleotide probes. This means that we have multiple oligonucleotides for the same gene. Each probe pair is consisted of a perfect match oligonucleotide (PM) and a mismatch oligonucleotide (MM).

The perfect match probe, which is used to measure the expression level of the gene, has a sequence exactly complimentary to the particular gene. The mismatch probe differs from the perfect match probe by a single base substitution at the center of the oligonucleotide, disturbing the binding of the target gene transcript. This is helpfull in order to determine the background and nonspecific hybridization that contributes to the signal measured for the perfect match oligo. For the analysis of the gene expression there are two kinds of algorithms, those who take into consideration the mismatched oligos and those who do not. For example MAS5 is an algorithm which subtracts the hybridization intensities of the mismatch probes from those of the perfect match probes to determine the absolute intensity value for each probe set. RMA16 is a newer algorithm, which uses only the perfect match oligos.

The probes are chosen based on current information from Genbank and other nucleotide repositories. The sequences on the expression arrays are believed to recognize unique regions of the 3 end of the gene.

For the present study we have used the Affymetrix GeneChip MoGene 1.0ST, which is designed for the mouse genome. The algorithm used for analysis of gene expression is the RMA16. The GeneChip also includes hybridization and poly-A controls in order to check the quality of the experiment. The GeneChip MoGene 1.0 ST array targets 28,853 genes with 770,317 probes. In other words there are approximately 27 probes for each gene.

Each eukaryotic Affymetrix GeneChip probe array contains probe sets for several genes that are absent in eukaryotic samples (lys, phe, thr, and dap). These poly-A RNA controls are synthesized in vitro and are premixed at staggered concentrations. The controls are then amplified and labeled together with the total RNA samples. Examining the hybridization intensities of these controls on GeneChip arrays helps to monitor the labeling process independently from the quality of the starting RNA samples. [AffymetrixInc, 2010]
1.3.4 Illumina BeadChip Arrays

Illumina, also based in California, USA, has developed a novel bead array technology which represents a fundamentally different approach to high-density arrays. The BeadArray technology is based on silica beads. Each bead is covered with hundreds of thousands of copies of a specific oligonucleotide that act as the capture sequences in one of Illumina’s assays. Each bead chip contains 6, 8 or 12 arrays. These arrays are separated among them, so that they can be used for different samples. Illumina beadchips are available for human or mouse genome. [IlluminaInc, 2005]

1.3.5 Agilent Microarrays

Agilent Microarray Chip is another platform for gene expression profiling. Agilent’s technology is based on 60 base long oligonucleotides. Agilent has manufactured microarray chips for a variety of organisms including human and mouse. Both catalog and custom microarrays are manufactured using a proprietary non-contact industrial inkjet printing process, in which oligo monomers are deposited uniformly onto specially-prepared glass slides. [AgilentInc, 2008]

1.3.6 Microarray Experimental Process

Based on the data, a scientist then determines whether the hypothesis was supported or refused. Science is subject to change when new evidence is found.

Microarray experiments are high throughput, this means that a single experiment may give a huge amount of data. Thus the most trivial and time consuming part of a microarray experiment is the analysis of the data. The researcher will have to compare the gene expression levels between the different groups and interpret the results. There is special software, designed to help with this data analysis. [Churchill, 2002]

1.3.6.1 Collection of the Microarray Data

The raw data of a microarray experiment are the images of the scanned microarray chips. We should clarify that by raw data we mean the intensity values of the
pixels of the images of the microarray chips. Most microarray manufacturers suggest a software for the data analysis. Affymetrix is developing Affymetrix GeneChip Operating Software (GCOS), which takes as input the raw data and offers a variety of tools for the analysis of the data. The file which is used to store the raw data is .cel file. Probe summarization algorithms perform the following three key tasks:

- Background correction
- Normalization
- Probe Summarization

The algorithm that we have used to perform these tasks is RMA (Robust Multi-array Analysis).

1.3.6.2 Robust Multi-array Average - RMA

RMA is an algorithm used for the normalization, background correction and summarization of the probe-level intensity measurements from Affymetrix Gene Chips. RMA takes into account only the perfect match (PM) probes which are first background corrected, then normalized and finally summarized. The result of the RMA is a set of expression values. RMA does not consider mismatches and thus reduces the noise. [Irizarry et al., 2003]

1.3.6.3 Background Noise Correction

Background noise correction is one of the most essential tasks for microarray analysis. In microarray experiments the intensity value of the fluorescence of a probe is a combination of the fluorescence intensity value of the probe by itself and the fluorescence intensity value of the background. Background noise correction is a significant step, as it removes the noise from non-biological signals. Background noise is the result of non-specific binding between the probes and the targets, as well as effective washing of the microarray chip during the hybridization process.

Background noise may cause a lot of problems when the intensity of the biological signal is close the intensity of the non-biological signal i.e. noise. As a
consequence small changes to the measured signal may cause significant variance at the resulting gene expression levels.

The most common way to remove background noise is to set a threshold for the upper and lower value of the fluorescence intensity. Values that are off the limits are rejected and marked as non biological signal - noise. For this reason Affymetrix has designed microarray chips, which have about 17,000 probes specialized to detect the levels of the background noise.

RMA performs a background correction on a per chip basis. This implemented correction is non linear. The background correction is based on the distribution of PM values amongst probes on an Affymetrix array. As we already have mentioned PM values are a mixture of a background signal, caused by optical noise and non-specific binding, plus a signal, which is what we are trying to detect. The background is estimated as expectation of the signal (S) conditioned on observed PM values (O), using a Fast Fourier Transformation.[AgilentInc, 2010]

1.3.6.4 Normalization

There are many sources of systematic variation in microarray experiments which affect the measured gene expression levels. Normalization is the term used to describe the process of removing such variation. The aim of normalization is to adjust for effects which arise from variation in the microarray chip. In other words normalization is essential so that multiple chips can be compared to each other and analyzed together. RMA is using quantile normalization. In statistics, quantile normalization is a technique for making two distributions identical in statistical properties. Thus after the normalization the distributions of gene expression levels are identical across the chips.[Lim et al., 2007]

1.3.6.5 Probe Summarization

After the probe level intensity values have been background corrected and normalized they need to be summarized into expression values. Affymetrix Genechips are designed in such a way that each gene is matched to 11-20 such probes evenly distributed throughout the chip. These probes make up a Probe Set on the chip. Summarization is the process where we calculate the intensity value of the flou-
rescence of a probeset based on the probes. The summarization procedure that RMA uses is motivated by the assumption that observed log-transformed PM values follow a linear additive model containing a probe affinity effect, a gene specific effect (the expression level) and an error term. For RMA, the probe affinity effects are assumed to sum to zero, and the gene effect (expression level) is estimated using median polishing. Median polishing is a robust model fitting technique, that protects against outlier probes.

1.3.6.6 Hybridization Quality Control

Hybridization quality control is a necessary task for the analysis of microarray data. There are various tests for the quality control.

The 3'/5' ratio is a measure of the efficiency of the cDNA synthesis reaction. All Affymetrix arrays contain probes for the regions corresponding to the 3, middle and 5-end of housekeeping genes such as GAPDH and β-Actin. The ratio of the signal intensity for 3 probesets to that from 5 probesets provides a measure of the number of cDNA synthesis reactions that went to completion i.e. the full-length cDNA is synthesized. We expect that the 3'/5' ratio will be close to one. If not the cDNA synthesis is not considered to be completed successfully. A ratio greater than 3 indicates that either the starting RNA was degraded or that there was a problem with the cDNA synthesis reaction.

Hybridization control plots are another quality control for the microarray samples. Pre-mixed hybridization control transcripts in known staggered concentrations are added to the hybridization mix. Hybridization controls are composed of a mixture of biotin-labelled transcripts of bioB, bioC, bioD, and are prepared in staggered concentrations. These controls allow you to monitor the hybridization and washing process. The signal intensity of these controls should increase with the concentrations. Deviations from the expected intensity profile of these controls indicates a potential problem with the hybridization or washing process.

Principal Component Analysis (PCA) is another procedure used for sample quality control. The PCA outcome are the principal components, which are vectors that capture the most variance in the data. Each sample is plotted according to its values for the first three principal components. We expect that the
samples within an experimental condition should be more similar to each other than to those from different conditions. [Jolliffe, 2002]

1.3.6.7 Statistical Analysis

After the data preprocessing (background noise removal, normalization, summarization) we have the expression values for all the available genes. We are interested only in genes that are statistically significant differentially expressed between the experimental conditions. The most used statistical tests are the t-test and the ANOVA. According to the experimental design the researcher will have to choose the appropriate statistical test. When there are two groups of samples the statistical test that we use is the t-test. In any other case, when we have to deal with more than two groups we prefer ANOVA. [O’Mahony and Michael, 1986]

A t-test is any statistical hypothesis test in which the test statistic follows a Student’s t distribution if the null hypothesis is supported. It can be used to determine if two sets of data are significantly different from each other. Based on the dependencies of the samples, we may distinguish two types of t-tests. If the samples are independent we use the unpaired t-test while in case the samples are dependent we choose the paired t-test. [William et al., 1992]

Analysis of variance (ANOVA) is a collection of statistical models used to analyze the differences between group means and their associated procedures (such as "variation" among and between groups). ANOVAs are useful in comparing (testing) three or more means (groups or variables) for statistical significance. In its simplest form, ANOVA provides a statistical test of whether or not the means of several groups are equal, and therefore generalizes the t-test to more than two groups. [Montgomery and Douglas, 2001] [Rosenbaum and Paul, 2002] [Henry, 1959]

In statistical significance testing, the p-value is the probability of obtaining a test statistic result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The p-value is the probability of obtaining the observed sample results (or a more extreme result) when the null hypothesis is actually true. If this p-value is very small, usually less than or
equal to a threshold value previously chosen called the significance level, usually 0.05 it suggests that the observed data is inconsistent with the assumption that the null hypothesis is true, and thus that hypothesis must be rejected and the other hypothesis accepted as true. The smaller the p-value, the larger the significance because it tells the investigator that the hypothesis under consideration may not adequately explain the observation. In our case a gene is statistically significant differentially expressed if its’ p-value is less than 0.05. Once you have set a threshold significance level (usually 0.05), every result leads to a conclusion of either ”statistically significant” or not ”statistically significant”. Some statisticians feel very strongly that the only acceptable conclusion is significant or ‘not significant’, and oppose use of adjectives or asterisks to describe values levels of statistical significance.

<table>
<thead>
<tr>
<th>P Value</th>
<th>Working</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.001</td>
<td>Extremely Significant</td>
<td>***</td>
</tr>
<tr>
<td>0.001 to 0.01</td>
<td>Very Significant</td>
<td>**</td>
</tr>
<tr>
<td>0.01 to 0.05</td>
<td>Significant</td>
<td>*</td>
</tr>
<tr>
<td>&gt;0.05</td>
<td>Not significant</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 1.1: An informal interpretation of a p-value, based on a significance level of about 5%. If p-value is very small, usually less than or equal to a threshold value previously chosen called the significance level it suggests that the observed data is inconsistent with the assumption that the null hypothesis is true, and thus that hypothesis must be rejected and the other hypothesis accepted as true.

1.3.6.8 Fold Change

Fold change is a measure describing how much a quantity changes going from an initial to a final value. Fold change is calculated simply as the ratio of the final value to the initial value. This means that if the initial value is A and final value is B, the fold change is B/A. A benefit of expressing a change as the ratio between an initial value and a final value is that the change itself is emphasized rather than the absolute values. This property makes the fold change suitable for statistical tests that need to normalize data to eliminate systematic error.
Fold change is often used in analysis of gene expression data in microarray and RNA-Seq experiments. [Tusher et al., 2001]

In our experiment we have a fold change for every gene. We will compare the experimental group to the control group. Thus a positive fold change for a gene means a higher expression level in the experimental group and respectively a negative fold change a lower expression level in the experimental group. The threshold that we have set for the fold change is 1.5.

1.3.6.9 Meta-analysis of the Microarray Data

Having identified all the genes with a p-value less that 0.05 and an absolute fold change greater than 1.5 we are ready to continue with the meta-analysis of the data. Gene Ontology Analysis is a process for the classification of the genes. An ontology is a representation of something we know about. The Gene Ontology project provides an ontology of defined terms representing gene product properties. The ontology covers three domains.

- **Biological Process**, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units
- **Molecular Function**, the elemental activities of a gene product at the molecular level, such as binding or catalysis;
- **Cellular Component**, the parts of a cell or its extracellular environment

**Ingenuity Pathway Analysis (IPA)** is a software application that enables researchers to analyze and understand the complex biological and chemical systems at the core of life science research. IPA performs the following tasks:

- build pathways for targets, biomarkers, diseases and biological functions
- design visually enhanced pathways
- compare and contrast across list, pathways, biomarkers and analyses

[Quiagen, 2013]
1.4 Next Generation Sequencing, NGS

Next-generation sequencing applies to genome sequencing, transcriptome profiling (RNA-Seq), DNA-protein interactions (ChIP-sequencing), and epigenome characterization. The high demand for low-cost sequencing has driven the development of high-throughput sequencing (or next-generation sequencing) technologies that parallelize the sequencing process, producing thousands or millions of sequences concurrently. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing. In ultra-high-throughput sequencing as many as 500,000 sequencing-by-synthesis operations may be run in parallel.[de Magalhes et al., 2010]. Next Generation Sequencing technologies are driving biomedical research. Cancer and genetic syndromes, due to next generation sequencing, met a new era in diagnostics and understanding. [Mardis, 2008]

1.4.1 RNA Sequencing, RNA-Seq

RNA-seq (RNA Sequencing), also called "Whole Transcriptome Shotgun Sequencing", is a technology that uses the capabilities of next-generation sequencing to reveal a snapshot of RNA presence and quantity from a genome at a given moment in time. RNA-Seq can look at different populations of RNA to include total RNA, small RNA, such as miRNA, tRNA, and ribosomal profiling. In addition RNA-Seq is being used to determine exon and intron boundaries. Clinical and more specifically cancer research had a major breakthrough thanks to the development and evolution of RNA-Seq. Microarrays technology has limited coverage, as they rely on a good knowledge of an organism’s genome. In other words with microarrays technologies we can target only the identification of known common alleles. On the other hand RNA-Seq does not rely on known alleles as it can be used for de novo transcriptome assembly.
1.5 Geminin is Involved in Cell Cycle Regulation and Cell Fate Decision

1.5.1 The structure of Geminin

Geminin is a coiled-coil nuclear protein constructed of about 209 amino acids (in homo sapiens), with a molecular weight of approximately 25 kDa. Geminin is present in most eukaryotes and is highly conserved across species. Geminin was originally identified as an inhibitor of DNA replication via its central coiled coil and C-terminus. Nowadays numerous functions have been elucidated including roles in metazoan cell cycle, regulation, cell lineage commitment, and neural differentiation. [Kroll et al., 1998]

A bipartite nuclear localization signal (NLS) at the N-terminus of Xenopus geminin controls the nuclear localization of the protein. This signal is conserved in many non-mammalian vertebrates. Instead in mammalian species the nuclear localization is controlled by an alternative N-terminal motif. Geminin contains an atypical leucine-zipper coiled-coil domain consisted by five repetitions of seven aminoacids. It has been shown that geminin regulates cell cycle by binding to Cdt1 and SWI/SNF complex.

1.5.2 Geminin Regulates the Cell Cycle

In each cell cycle DNA replication is initiated from a large number of origins. Throughout the genome there are many more potential origins that are not in use by the cell at a specific time. At the final stages of mitosis and the primary stages of the G1 phase, DNA replication initiation depends on specific factors such as the origin recognition complex (ORC), Cdc6, Cdt1 and MCM proteins. These molecules form a complex that is assembled dynamically onto chromatin. It is necessary that in each cell cycle DNA will replicate only once. To achieve this cell has developed special mechanisms to prevent DNA re-replication. Geminin is a molecule that prohibits DNA re-replication by binding to the licencing factor Cdt1, which is responsible for loading the MCM proteins onto the origins of replication
The expression of geminin begins at the S phase during DNA replication. Geminin expression levels remain high until the end of M phase. Cdt1, which is a target of Geminin and Geminin are separately regulated. Geminin binds to Cdt1 and prohibits its binding to the origins of replication. As a result MCMs molecules can not bind to the origins of replication for a second time within the same cell cycle. At the end of mitosis the APC complex degrades geminin. This means that Cdt1 is free to bind to the origins of replication. [Lygerou and Nishitami, 2004] [Caillat and Perrakis, 2012]

1.5.3 The role of Geminin in differentiation

A balance of cell proliferation and differentiation is essential for the development and homeostasis of an organism. Geminin has been identified as a protein that induces cell differentiation towards neural fates. The overexpression of bone morphogenetic proteins (BMPs) promotes epidermal differentiation and blocks neural differentiation in ectoderm of both vertebrate and invertebrate organisms. Kroll et al have shown that in Xenopus embryos overexpression of Geminin in gastrula ectoderm suppresses BMP4 expression and converts the prospective epidermis into neural tissue. [Kroll et al., 1998]

In vertebrates during development organogenesis requires the strict control of cell proliferation and differentiation. TIPT2 factor expressed in mouse embryos and adult mice interacts with geminin. TIPT2 also intercats with polycomb members including Scmh1, Mph2, Ring1B and the general transcritpiton factor TBP.

Sox2 is a transcription factor associated with the developmental plasticity of stem cells and the regulation of chromatin state. Mouse embryos lacking geminin can not overpass the stage of eight cells and fail to form an inner cell mass. [Kroll, 2007] [de Renty et al., 2014]
1.6 Mouse as a model organism in biomedical research

Mus Musculus or the house mouse is a small mammal of the order Rodentia. It characteristically has a pointed snout, small rounded ears, and a long almost hairless tail. It is the most widely used mammalian model in biomedical and genetic research. The sequence of the human and the mouse genome revealed that there is a great degree of conservation between those two mammals, as they share about 99% of their genome ([Waterston et al., 2002]), even though that their evolutionary lineages diverged about 96 million years ago.

Other organisms like yeast or flies are an excellent model to study developmental processes and the cell cycle. In case that we wish to examine the complex mammalian physiological systems mice are one of the best models to use as they can naturally develop a wide variety of diseases including cancer, hypertension, diabetes, and osteoporosis. Scientists have achieved to induce to mice diseases that normally do not strike them like Alzheimer or even alcoholism, by manipulating their genome and enviroment. [Gonipeta et al., 2013] [Zhou et al., 2014]

1.6.1 The advantages of Mouse as a model organism

The use of Mus Musculus as a model organism offers a lot of advantages, among them their close resemblance to humans anatomically, physiologically and genetically. In addition, their genome can easily be manipulated and analyzed. The cost for maintenance of a colony of mice is low, while they are small animals that have a short generation time and an accelerated life span. The duration of the pregnancy in the mouse is 21 days. The pups are able to reproduce 5 weeks after their birth. One mouse year equals to 30 human years. [Greaves et al., 2014]
Figure 1.11: Mus Musculus (House Mouse), is a small mammal of the order Rodentia. Laboratory mouse, which is one of the most important model organisms in biology and medicine.

The complete sequence of the C57BL/6J mouse genome significantly enhanced our abilities not only to understand and explain the various phenotypes but also to manipulate and genetically modify it. Nowadays there are thousands of mice strains and genetically engineered mutants available. Chemical mutagenesis is a process where a mouse is exposed to DNA damaging chemicals, in order to produce new models of genetic diseases. The ability to generate transgenic mice where a new gene has been inserted into their genome gives scientists a great tool to study the function of specific genes. Genes can not only be knocked-in but also knocked-out with homologous recombination. State of the art methods and techniques like in vitro fertilization, blastocysts injection and cryopreservation are an extremely useful tool for the reproduction of valuable mouse strains. It is by far the animal most commonly genetically altered for scientific research.

1.6.2 Jackson Laboratory

Jackson Laboratory is a biomedical research institute which has played a crucial role in the development of the mouse as a model organism. Worldwide the laboratory is the most important source for more than 7,000 strains of genetically modified mice. The Mouse Genome Informatics database is developed and maintained by the bioinformatics department of the laboratory. We should mention that one of the most famous mouse strain C57BL/6J also known as Black 6 was developed by the Jackson laboratory in the 1920s. The Laboratory’s mission is to
1. Introduction

discover precise genomic solutions for disease and empower the global biomedical community in the shared quest to improve human health. [Li et al., 2014]
Chapter 2
Aims and Objectives

Neural crest stem cells are a transient migratory cell population that gives rise to a variety of cell lineages like, melanocytes, craniofacial cartilage and bone, smooth muscle, peripheral as well as enteric neurons. Self-renew and differentiation of neural crest stem cells is under severe control during the process of embryogenesis. The control of these two main processes is achieved through the directed coordination of proliferation, transcription of specific genes and the regulation of chromatin structure.

A recent study in our lab [Phd Thesis Stathopoulou, 2012 ] has shown that geminin acts as a key regulatory element in neural crest stem cells during embryogenesis. Geminin is associated with pathways that regulate neural crest cell self-renew and differentiation.

The aim of this study is to identify the gene network conducted by geminin during neural crest stem generation using whole transcriptome analysis. Gene expression profile was analysed using cDNA microarray technology. Thus we compared the gene expression profile of genetically modified mouse embryos, lacking geminin from neural crest stem cells, with wild type mouse embryos. Bioinformatics tools were used to identify processes regulated by geminin.
Chapter 3

Materials and Methods

3.1 Experimental Animals

All experimental animals that were used in our study had a mixed genetic background. Genetic background refers to the genetic make-up (strain) of each mouse (all its alleles at all loci) except the mutated gene of interest that we study. On the one hand it is clear that phenotype variability is depended on the strain of the mouse but on the other hand a mixed background mouse can provide the quickest preview of possible strain-dependent phenotypes.

For our research purposes we have generated four transgenic mice lines which differ in the genomic loci of Geminin. The lineages of mice that we generated are the following:

- Mice homozygous for the wild type (WT) allele of Geminin, Geminin\textsuperscript{wt/wt}.
- Mice homozygous for the floxed allele of Geminin, Geminin\textsuperscript{fl/fl}. The floxed allele of Geminin has two loxP sites, one before the third exon and the other after the forth exon.
- Mice heterozygous for the gene of Geminin, these mice have both the Wild Type and the floxed allele, Geminin\textsuperscript{fl/wt}.
- Mice heterozygous for the gene of Geminin, these mice have both the Floxed and the Knock Out allele, Geminin\textsuperscript{fl/ko}. The Knock Out allele lacks the third and the forth exon; thus it is not functional.
3. Materials and Methods

3.1.1 Cre Recombination System

Cre recombinase is an enzyme which is able to recognise the LoxP sites and remove the genomic region between them. In other words, Cre recombinase in combination with the geminin floxed alleles can be used to generate tissue specific conditional knock out mice. This means that if we wish to ablate Geminin on a specific tissue we need to cross mice expressing the floxed Geminin allele with mice expressing Cre Recombinase under the control of the appropriate regulatory elements on the tissue of interest.
3. Materials and Methods

In order to achieve our goal, i.e. delete geminin from the neural crest stem cells during embryogenesis, we used mice that express Cre Recombinase under the control of the wnt1 gene. Wnt1 is expressed specifically in neural crest stem cells at the early stages of embryogenesis at day E8.0-8.5. ([Danielian et al., 1998]) We only used mice heterozygous for Cre Recombinase because homozygous mice die. In order to investigate the role of geminin at E10.5 we generated two groups of mice, experimental and wild type. These groups differ at the expression of geminin in neural crest stem cells. Wild type mice can express geminin in neural crest stem cells but, in contrast, experimental mice lack geminin from these cells. Wnt1cre mice were an offer of Dr. Vasilios Pachnis (National Institute for Medical Research, NIMR-MRC, United Kingdom).

In order to mark Neural Crest Stem Cells with a reporter protein we used mice that have the Rosa26\(^{Stop}\)YFP genetic locus ([Srinivas et al., 2001]). These mice were an offer of Dr. Vasilios Pachnis (National Institute for Medical Research, NIMR-MRC, United Kingdom). These R26-stop-YFP mice have a loxP-flanked STOP sequence followed by the Yellow Fluorescent Protein gene (YFP) inserted into the Rosa26 locus. When bred to mice expressing Cre recombinase, the STOP sequence is deleted and YFP expression is observed in the cre-expressing tissue. Expression of YFP is blocked by an upstream loxP-flanked STOP sequence, be-
cause it stops the transcription of the gene.

![Figure 3.5: Rosa26 locus](image)

When the enzyme of cre recombinase is active, it deletes the STOP sequence in the tissue of interest, in our case neural crest stem cells, and YFP expression is observed. Transgenic mice that express Cre Recombinase, under the control of Wnt1 regulatory elements, will also express the reporter protein YFP in the neural crest cells during embryogenesis. We should note that mice that are homozygous for the R26-stop-EYFP mutant allele are viable, fertile, normal in size and do not display any gross physical or behavioral abnormalities.

### 3.2 Mice Breedings

With the intention of generating wild type and experimentional mice we crossed geminin\(^{\text{ko/wt}}\);Wnt1Cre\(^{+/\text{-}}\) with geminin\(^{\text{fl/fl}}\) mice. The next table contains the genotypes for all possible descendants and the group that they belong to.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>geminin(^{\text{ko/fl}});Wnt1Cre(^{+/\text{-}})</td>
<td>Experimental</td>
</tr>
<tr>
<td>geminin(^{\text{ko/fl}});Wnt1Cre(^{-/-})</td>
<td>Wild Type</td>
</tr>
<tr>
<td>geminin(^{\text{wt/fl}});Wnt1Cre(^{+/\text{-}})</td>
<td>Wild Type</td>
</tr>
<tr>
<td>geminin(^{\text{wt/fl}});Wnt1Cre(^{-/-})</td>
<td>Wild Type</td>
</tr>
</tbody>
</table>

Table 3.1: All possible genotypes for the geminin gene and the group that they belong to.

All research protocols about the animal handling have been approved by the Ethics Committee of the National Institute for Medical Research (NIMR), London, United Kingdom and the University of Patras, Greece.
3. Materials and Methods

3.3 Mouse Embryo Dissection

When we setup a breeding we place a male and a female mouse in the same cage. The same day is considered to be the day of conception. The next day, which is the first day of the pregnancy is E0.5 development stage. In our study we have dissected embryos at E10.5. The process for the dissection is:

- Sacrifice the female mouse using cervical dislocation.
- Spray the mouse with 70% Ethanol.
- Lift the skin over the belly and make a small incision.
- Enlarge the incision and pull the string of embryos out and transfer to a Petri dish on ice filled with ice-cold PBS.
- Dissect the embryos out of the uterus under a stereomicroscope in a Petri dish with fresh pre-cooled (4°C) PBS, make sure to remove all the membranes and transfer the cleaned embryo to an empty Petri dish on ice.

3.4 Genotyping

3.4.1 DNA Extraction

Genomic DNA extraction from animal tissue is essential for genotyping. For that reason we use a sample from the murine tail. The protocol that has been used is the following:

- Each tissue sample is placed in an eppendorf which contains a solution of 500ul tail buffer and 6 ul of K proteinase (10mg/ml Sigma P2308) to be solubilised. The samples are incubated for at least 16 hours in the waterbath at 55°C.
- After the incubation add to each sample 500ul of Phenol Chloroform Isoamyl Alcohol Solution (PCI, Sigma P2069) and vortex for 10 minutes.
3. Materials and Methods

- Centrifuge the samples at 13,000 rpm for 30 minutes at 25°C. Then transfer the liquid supernatant to a new eppendorf.

- Add 500ul of Chloroform (Sigma C2432) and vortex for ten minutes.

- Centrifuge the samples at 13,000 rpm for 15 minutes at 25°C. Then transfer the liquid supernatant to a new eppendorf.

- Add 500ul of isopropanol and shake the eppendorfs.

- Incubate the samples for 30 minutes to one hour at −80°C.

- Centrifuge the samples at 13,000 rpm for 30 minutes at 4°C and remove the supernatant.

- Wash the precipitate with 500ul of EtOH 70%.

- Centrifuge the samples at 13,000 rpm for 30 minutes at 4°C and remove the supernatant.

- Dry the precipitate and then add 100ul of sterile ddH₂O. Incubate the samples for one hour at 55°C to optimise the stripping.

- DNA samples should be stored at −20°C.

3.4.2 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was used to identify the alleles of geminin. PCR is a biochemical method used in molecular biology for the amplification, thousands to millions of times, of a specific sequence of DNA. PCR was developed by Kary Mullis and is based on the complementarity of the DNA strands. DNA polymerase can add a nucleotide only to a pre-existing 3’-OH group. Primers are short DNA sequences that help DNA polymerase start synthetizing the new strand.
3.4.2.1 Geminin\(^{Fl/Wt}\) PCR

In order to identify floxed and wild type alleles of geminin we used the following pair of primers:

- 5’- TTTGGACGCATGGGACAGACC -3’ and
- 5’- GTCCAGGCTGCAGTGTCTCAC -3’

This set of primers recognises the area that is flanked by the LoxP locus which are after the 4\(^{th}\) exon of geminin. The product of the wild type allele is 549 base pairs long while the product of the floxed allele is 597. The next table describes the quantities of each reagent that we need for the Geminin\(^{Fl/Wt}\) PCR.
3. Materials and Methods

Figure 3.8: Geminin floxed allele primer starting sites

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Initial Concentration</th>
<th>Final Concentration</th>
<th>Final Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>-</td>
<td>300-500 ng</td>
<td>-</td>
</tr>
<tr>
<td>10x PCR Buffer (- MgCl₂)</td>
<td>10x</td>
<td>1x</td>
<td>5 ul</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>50 mM</td>
<td>1.5 mM</td>
<td>1.5 ul</td>
</tr>
<tr>
<td>dNTPS</td>
<td>10 mM</td>
<td>0.2 mM</td>
<td>1 ul</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>100 pmoles/ul</td>
<td>1 pmoles/ul</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>100 pmoles/ul</td>
<td>1 pmoles/ul</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>Taq Polymerase</td>
<td>5 units/ul</td>
<td>0.05 units</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>H₂O</td>
<td>-</td>
<td>-</td>
<td>Up to 50 ul</td>
</tr>
</tbody>
</table>

Table 3.2: Geminin\textsuperscript{Fl/Wt} PCR reagents

The program for the PCR conditions that we used is:

- 94 °C for 5 minutes,
- For 34 cycles,
  - 94 °C for 45 seconds,
  - 67 °C for 30 seconds,
  - 72 °C for 60 seconds,
- 72 °C for 10 minutes,
- 4 °C for ever.
3.4.2.2 Geminin\textsuperscript{Ko/Wt} PCR

In order to identify the knock out allele of geminin we used the following pair of primers:

- 5’-GGAATATTTAGAGTTCTGAAATGAGATG -3’ which recognises a 15 bases sequence before the third exon and
- 5’- CCAACTCAGTCACTGCTCTGTT -3’ which recognises a 15 bases sequence after the forth exon of geminin

When Cre Recombinase is expressed it removes the area between the two LoxP sites. In our case, this area consists of the third and the forth exon. The sequence that remains has a length of only 249 base pairs. Of course these primers can also identify the same spots in the wild type allele. In that case the sequence of interest is 1900 base pairs long and for that reason usually pcr does not give a product (too long for conventional pcr).

![Figure 3.9: Geminin knock out allele primer starting sites](image)

The next table describes the quantities of each reagent that we need for the Geminin\textsuperscript{Ko/Wt} PCR.
3. Materials and Methods

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Initial Concentration</th>
<th>Final Concentration</th>
<th>Final Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>-</td>
<td>300 ng</td>
<td>-</td>
</tr>
<tr>
<td>10x PCR Buffer (-MgCl₂)</td>
<td>10x</td>
<td>1x</td>
<td>5 ul</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>50 mM</td>
<td>1.5 mM</td>
<td>1.5 ul</td>
</tr>
<tr>
<td>dNTPs</td>
<td>10 mM</td>
<td>0.2 mM</td>
<td>1 ul</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>100 pmoles/ul</td>
<td>1 pmoles/ul</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>100 pmoles/ul</td>
<td>1 pmoles/ul</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>Taq Polymerase</td>
<td>5 units/ul</td>
<td>0.05 units</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>H₂O</td>
<td>-</td>
<td>-</td>
<td>Up to 50 ul</td>
</tr>
</tbody>
</table>

Table 3.3: Geminin<sup>Ko/Wt</sup> PCR reagents

The program for the PCR conditions that we used is:

- 94°C for 5 minutes,
- For 34 cycles,
  - ✓ 94°C for 45 seconds,
  - ✓ 62.7°C for 45 seconds,
  - ✓ 72°C for 60 seconds,
- 72°C for 10 minutes,
- 4°C for ever.

### 3.4.2.3 Cre Recombinase<sup>+/−</sup> PCR

Cre recombinase PCR is used to check whether our mice have or not cre recombinase gene. The set of primers that we have used amplifies a DNA sequence of 411 bases, part of the gene of interest. The set of primers is:

- Forward Primer: 5’- AGGTGTAGAGAAGGCACTCAGC -3’
- Reverse Primer: 5’- CTAATCGCCATCTTCCAGCAGG -3’
3. Materials and Methods

As an internal control we used the microsomal epoxide hydrolase (MEH) gene, which is ubiquitously expressed. The presence of the MEH gene is independent of the allele of geminin or the presence of cre recombinase. The primers that we used for MEH amplify a sequence of 341 bases. The set of primers is:

- Forward Primer 5’- AAGTGAGTTTGCATGGGCAGC -3’
- Reverse Primer 5’- CCCTTTAGCCCCCTCCCTCTG -3’

The next table describes the quantities of each reagent used for the cre recombinase PCR.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Initial Concentration</th>
<th>Final Concentration</th>
<th>Final Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>-</td>
<td>300 - 500 ng</td>
<td>-</td>
</tr>
<tr>
<td>10x PCR Buffer (-MgCl₂)</td>
<td>10x</td>
<td>1x</td>
<td>5 ul</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>50 mM</td>
<td>2 mM</td>
<td>2 ul</td>
</tr>
<tr>
<td>dNTPs</td>
<td>10 mM</td>
<td>0.4 mM</td>
<td>2 ul</td>
</tr>
<tr>
<td>Forward Primer Cre</td>
<td>100 pmoles/ul</td>
<td>0.4 pmoles/ul</td>
<td>0.2 ul</td>
</tr>
<tr>
<td>Reverse Primer Cre</td>
<td>100 pmoles/ul</td>
<td>0.4 pmoles/ul</td>
<td>0.2 ul</td>
</tr>
<tr>
<td>Forward Primer MEH</td>
<td>100 pmoles/ul</td>
<td>0.4 pmoles/ul</td>
<td>0.2 ul</td>
</tr>
<tr>
<td>Reverse Primer MEH</td>
<td>100 pmoles/ul</td>
<td>0.4 pmoles/ul</td>
<td>0.2 ul</td>
</tr>
<tr>
<td>Taq Polymerase</td>
<td>5 units/ul</td>
<td>0.05 units</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>H₂O</td>
<td>-</td>
<td>-</td>
<td>Up to 50 ul</td>
</tr>
</tbody>
</table>

Table 3.4: Cre Recombinase PCR reagents

The conditions used for the pcr are:

- 94 °C for 4 minutes,
- For 30 cycles,
3. Materials and Methods

✓ 94°C for 30 seconds,
✓ 60°C for 30 seconds,
✓ 72°C for 30 seconds,

• 72°C for 5 minutes,
• 4°C for ever.

3.4.2.4 Rosa26STOPYFP PCR

In order to genotype the experimental mice for the Rosa26STOPYFP gene we used a set of 3 primers, which are:

• Forward: 5’- GCTCTGAGTTGTTATCAGTAAGG -3’
• Reverse 1: 5’- GCGAAGAGTTTGTCCTCAACC -3’
• Reverse 2: 5’- GGAGCGGGAGAAATGGATAGT ’3’

Forward and Reverse 1 primers are recognising the recombinant allele and amplify a sequence of 350 bp, while Forward and Reverse 2 primers are used to recognise the wild type allele and amplify a sequence of 500 bp.

The next table describes the quantities of each reagent used for the Rosa26STOPYFP PCR.
3. Materials and Methods

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Initial Concentration</th>
<th>Final Concentration</th>
<th>Final Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>-</td>
<td>500 ng</td>
<td>-</td>
</tr>
<tr>
<td>10x PCR Buffer (-MgCl₂)</td>
<td>10x</td>
<td>1x</td>
<td>1.5 ul</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>50 mM</td>
<td>2 mM</td>
<td>1 ul</td>
</tr>
<tr>
<td>dNTPs</td>
<td>10 mM</td>
<td>0.4 mM</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>Forward Primer Cre</td>
<td>100 pmoles/ul</td>
<td>0.4 pmoles/ul</td>
<td>2 ul</td>
</tr>
<tr>
<td>Reverse Primer 1 Cre</td>
<td>100 pmoles/ul</td>
<td>0.4 pmoles/ul</td>
<td>1 ul</td>
</tr>
<tr>
<td>Reverse Primer 2 MEH</td>
<td>100 pmoles/ul</td>
<td>0.4 pmoles/ul</td>
<td>1 ul</td>
</tr>
<tr>
<td>Taq Polymerase</td>
<td>5 units/ul</td>
<td>0.05 units</td>
<td>0.5 ul</td>
</tr>
<tr>
<td>H₂O</td>
<td>-</td>
<td>-</td>
<td>Up to 25 ul</td>
</tr>
</tbody>
</table>

Table 3.5: Rosa²⁶<sup>STOPYFP</sup> PCR reagents

The conditions used for the per are:

- 94°C for 4 minutes,

- For 35 cycles,
  - 94°C for 30 seconds,
  - 58°C for 40 seconds,
  - 72°C for 45 seconds,

- 72°C for 5 minutes,

- 4°C for ever.
3.4.3 Electrophoresis

Gel Electrophoresis and specifically in our case agarose gel electrophoresis is a method used in molecular biology, to separate a mixed population of DNA fragments in a matrix of agarose. DNA sequences are separated by length. Usually the separated DNA bands are viewed under UV light. In order to make the gel, we use a TBE solution to dilute the agarose. Higher agarose percentage enhances resolution of lower bands (shorter DNA sequences). Conversely a lower agarose percentage enhances resolution and separation of higher bands (longer DNA sequences). In our study we have used 2% agarose gels for the electrophoreses. This means that we add 2 g of agarose to 100 ml of TBE. Then the solution is boiled in a microwave oven, so that agarose dilutes and finally we add 10 ul of Ethidium Bromide (EtBr 10 mg/ml). Ethidium Bromide binds on the DNA molecules and fluorescences under UV light.
3. Materials and Methods

3.5 Affymetrix GeneChip cDNA Microarrays Hybridization

Ambion WT Expression Kit enables you to prepare RNA samples for Affymetrix whole transcriptome microarray analysis. The kit generates cDNA from total RNA. The WT Expression Kit uses a reverse transcription priming method that specifically primes non-ribosomal RNA from the samples, including both poly(A) and non-poly(A) mRNA. The primer sequences used provide complete and unbiased coverage of the whole transcriptome. Reverse transcription polymerase chain reaction (RT-PCR) is a technique commonly used in molecular biology to detect RNA expression levels. RT-PCR is used to clone expressed genes by reverse transcribing the RNA of interest into its DNA complement through the use of reverse transcriptase. Subsequently, the newly synthesized cDNA is amplified using traditional PCR.

3.5.1 Synthesize first-strand cDNA

- At room temperature, prepare the First-Strand Master Mix in a nuclease-free tube. Include 5% excess volume to correct for pipetting losses. Combine the components in the sequence shown in the table below. Mix thoroughly by gently vortexing. Centrifuge briefly (5 sec) to collect the mix at the bottom of the tube.

<table>
<thead>
<tr>
<th>First-Strand Master</th>
<th>Mix component Volume for one reaction (uL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Strand Buffer Mix</td>
<td>4</td>
</tr>
<tr>
<td>First-Strand Enzyme Mix</td>
<td>1</td>
</tr>
<tr>
<td>Total Volume</td>
<td>5</td>
</tr>
</tbody>
</table>

- Transfer 5 L of the First-Strand Master Mix to the supplied PCR tubes.

- Add 5 uL RNA to each tube or well containing the First-Strand Master Mix for a final reaction volume of 10 uL. If necessary, use nuclease-free water to bring the RNA to 5 uL. Mix thoroughly by gently vortexing. Centrifuge briefly to collect the reaction at the bottom of the tube/plate.
3. Materials and Methods

- Incubate for 1 hr at 25 C, then for 1 hr at 42 C, then for at least 2 min at 4 C in a thermal cycler.

- Immediately after the incubation, centrifuge briefly (5 sec) to collect the first-strand cDNA at the bottom of the tube/plate. Place the sample on ice for 2 min to cool the plastic tubes.

3.5.2 Synthesize second-strand cDNA

Single-stranded cDNA will be converted to double-stranded cDNA, which acts as a template for transcription. The reaction uses DNA polymerase and RNase H to simultaneously degrade the RNA and synthesize the second-strand cDNA.

- On ice, prepare the Second-Strand Master Mix in a nuclease-free tube. Combine the components in the sequence shown in the table below. Include 10% excess volume to correct pipetting losses. Mix thoroughly by gently vortexing. Centrifuge briefly (5 sec) to collect the mix at the bottom of the tube and proceed immediately to the next step.

<table>
<thead>
<tr>
<th>Second-Strand Master Mix component</th>
<th>Volume for one reaction (uL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclease-free Water</td>
<td>32.5</td>
</tr>
<tr>
<td>Second-Strand Buffer Mix</td>
<td>12.5</td>
</tr>
<tr>
<td>Second-Strand Enzyme Mix</td>
<td>5</td>
</tr>
<tr>
<td>Total Volume</td>
<td>50</td>
</tr>
</tbody>
</table>

- Transfer 50 uL of the Second-Strand Master Mix to each (10 uL) first-strand synthesis cDNA sample. Mix thoroughly by gently vortexing or flicking the tube 3 or 4 times. Centrifuge briefly to collect the reaction at the bottom of the tube/plate and proceed immediately to the next step.

- Incubate for 1 hr at 16 C, then for 10 min at 65 C, then for at least 2 min at 4 C in a thermal cycler.
3. Materials and Methods

- Immediately after the incubation, centrifuge briefly (5 sec) to collect the double-stranded cDNA at the bottom of the tube/plate. Place the sample on ice to cool the plastic tube.

3.5.3 Synthesize cRNA by In Vitro Transcription

In this procedure, antisense cRNA is synthesized and amplified by in vitro transcription (IVT) of the second-strand cDNA template using T7 RNA polymerase.

- At room temperature, prepare an IVT Master Mix in a nuclease-free tube. Prepare master mix for all the second-strand cDNA (60-L) samples. Combine the components in the sequence shown in the table below. Include 5% excess volume to correct pipetting losses. Mix thoroughly by gently vortexing. Centrifuge briefly (5 sec) to collect the mix at the bottom of the tube, then proceed immediately to the next step.

<table>
<thead>
<tr>
<th>IVT Master Mix component</th>
<th>Volume for one reaction (uL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IVT Buffer Mix</td>
<td>24</td>
</tr>
<tr>
<td>IVT Enzyme Mix</td>
<td>6</td>
</tr>
<tr>
<td>Total Volume</td>
<td>30</td>
</tr>
</tbody>
</table>

- Transfer 30 L of the IVT Master Mix to each 60-L Second-Strand cDNA sample. Mix thoroughly by gently vortexing, then centrifuge briefly to collect the reaction at the bottom of the tube/plate.

- Incubate the IVT reaction for 16 hr at 40 C, then overnight at 4 C in a thermal cycler. After the incubation, centrifuge briefly (5 sec) to collect the reaction at the bottom of the tube/plate. You may either place the samples on ice or immediately freeze them at 20 C for overnight storage.

3.5.4 cRNA Purification

In this procedure, enzymes, salts, inorganic phosphates, and unincorporated nucleotides are removed to improve the stability of the cRNA.
3. Materials and Methods

- At room temperature, immediately before use, prepare the cRNA Binding Mix in a nuclease-free tube for all the samples in the experiment.

<table>
<thead>
<tr>
<th>cRNA Binding Mix component</th>
<th>Volume for one reaction (uL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleic Acid Binding Beads</td>
<td>10</td>
</tr>
<tr>
<td>Nucleic Acid Binding Buffer Concentrate 50</td>
<td>50</td>
</tr>
</tbody>
</table>

- Add 60 uL of cRNA Binding Mix to each sample. Pipette up/down 3 times to mix.

- Transfer each sample to a well of a U-Bottom Plate.

- Add 60 uL of 100% isopropanol to each sample. Pipette up/down 3 times to mix.

- Gently shake for 2 min to thoroughly mix. The cRNA in the sample binds to the Nucleic Acid Binding Beads during this incubation.

- Move the plate to a magnetic stand to capture the magnetic beads. When capture is complete (after 5 min), the mixture is transparent, and the Nucleic Acid Binding Beads form pellets against the magnets in the magnetic stand. The exact capture time depends on the magnetic stand that you use, and the amount of cRNA generated by in vitro transcription.

- Carefully aspirate and discard the supernatant without disturbing the magnetic beads, then remove the plate from the magnetic stand.

- Add 100 uL of Nucleic Acid Wash Solution to each sample, then shake at moderate speed for 1 min.

- Move the plate to a magnetic stand and capture the Nucleic Acid Binding Beads as in the previous step.

- Carefully aspirate and discard the supernatant without disturbing the Nucleic Acid Binding Beads, then remove the plate from the magnetic stand.
3. Materials and Methods

- Move the plate to a shaker and shake the plate vigorously for 1 min to evaporate residual ethanol from the beads. Dry the solution until no liquid is visible, yet the pellet appears shiny. Additional time may be required. Do not over-dry the beads.

- Add to each sample 40 uL of preheated (55 to 58 C) Elution Solution to elute the purified cRNA from the Nucleic Acid Binding Beads. Incubate without shaking for 2 min.

- Vigorously shake the plate for 3 min then check to make sure that the Nucleic Acid Binding Beads are fully dispersed. If they are not, continue shaking until the beads are dispersed and/or pipette up/down 3 times. Pellets can be disrupted by manual pipetting using a single-channel P-200 or equivalent.

- Move the plate to a magnetic stand to capture the Nucleic Acid Binding Beads.

- Transfer the supernatant, which contains the eluted cRNA, to a nuclease-free multiwell plate.

- You may place the reaction on ice, or freeze the samples at 20 C for overnight storage.

3.5.5 Assess cRNA yield and size distribution

The cRNA yield depends on the amount and quality of poly(A) RNA in the input total RNA. The proportion of poly(A) RNA in total RNA is affected by factors such as the health of the organism and the organ from which it is isolated, cRNA yield from equal amounts of total RNA may vary considerably. Determine the concentration of a cRNA solution by measuring its absorbance at 260 nm. Applied Biosystems recommends evaluating the absorbance of 1.5 uL of cRNA sample using a NanoDrop Spectrophotometer.
3. Materials and Methods

3.5.6 Synthesize 2nd-cycle cDNA

In this procedure, sense-strand cDNA is synthesized by the reverse transcription of cRNA using random primers. The sense-strand cDNA contains dUTP at a fixed ratio relative to dTTP. For the 2nd-cycle cDNA synthesis is required 10 ug of cRNA.

- On ice, prepare 455 ng/uL cRNA. This is equal to 10 g cRNA in a volume of 22 uL. If necessary, use nuclease-free water to bring the cRNA sample to 22 uL. If the cRNA is too diluted, then concentrate the cRNA by vacuum centrifugation.

- On ice, using PCR tubes or plate, combine
  - 22 uL of cRNA (10 ug) and
  - 2 uL of Random Primers

Mix thoroughly by gently vortexing. Centrifuge briefly (5 sec) to collect the reaction at the bottom of the tube/plate. Place on ice.

- Incubate 5 min at 70 C, then 5 min at 25 C, and finally 2 min at 4 C in a thermal cycler. After the incubation, centrifuge briefly (5 sec) to collect the 2nd-Cycle cDNA at the bottom of the tube/plate.

- At room temperature, prepare the 2nd-Cycle Master Mix in a nuclease-free tube. Include 10% excess volume to correct for pipetting losses. Combine the components in the sequence shown in the table below.

<table>
<thead>
<tr>
<th>2nd-Cycle Master Mix component</th>
<th>Volume for one reaction (uL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd-Cycle Buffer Mix</td>
<td>8</td>
</tr>
<tr>
<td>2nd-Cycle Enzyme Mix</td>
<td>8</td>
</tr>
<tr>
<td>Total Volume</td>
<td>16</td>
</tr>
</tbody>
</table>

Mix thoroughly by gently vortexing. Centrifuge briefly (5 sec) to collect the mix at the bottom of the tube/plate.
3. Materials and Methods

- Transfer 16 uL of 2nd-Cycle Master Mix to each (24-uL) cRNA/Random Primer sample. Mix thoroughly by gently vortexing. Centrifuge briefly to collect the reaction at the bottom of the plate/tube.

- Incubate for 10 min at 25 C, then 90 min at 42 C, then 10 min at 70 C, then for at least 2 min at 4 C in a thermal cycler.

- Immediately after the incubation, centrifuge briefly (5 sec) to collect the cDNA at the bottom of the tube/plate. Place the samples on ice.

3.5.7 Hydrolyze using RNase H

RNase H degrades the cRNA template leaving single-stranded cDNA.

- On ice, add 2 uL of RNase H to the 2nd-Cycle cDNA from above. Mix by pipetting up/down 3 times to ensure that all of the enzyme is dispensed from the pipet tip.

- Mix thoroughly by gently vortexing. Centrifuge briefly (5 sec) to collect the reaction at the bottom of the tube/plate.

- Incubate for 45 min at 37 C, then 5 min at 95 C, then for at least 2 min at 4 C in a thermal cycler.

- After the incubation, centrifuge briefly (5 sec) to collect the RNase H Hydrolyzed 2nd Cycle cDNA at the bottom of the tube/plate. Place the samples on ice. You may store the samples at -20 C overnight.

3.5.8 Purify 2nd-cycle cDNA

After synthesis, the second-strand cDNA is purified to remove enzymes, salts, and unincorporated dNTPs. This step prepares the cDNA for fragmentation and labeling.

- At room temperature, prepare the cDNA Binding Mix in a nuclease-free tube for all the samples in the experiment. Prepare the amount needed
for all samples in the experiment plus 10% excess volume to compensate pipetting losses.

<table>
<thead>
<tr>
<th>cDNA Binding Mix component</th>
<th>Volume for one reaction (uL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleic Acid Binding Beads</td>
<td>10</td>
</tr>
<tr>
<td>Nucleic Acid Binding Buffer Concentrate</td>
<td>50</td>
</tr>
</tbody>
</table>

- Add 18 uL of nuclease-free water to each sample for a final volume of 60 uL.
- Add 60 uL of cDNA Binding Mix to each sample. Pipette up/down 3 times to mix.
- Transfer each sample to a well of a U-bottom plate.
- Add 120 uL of 100% ethanol to each sample. Pipette up/down 3 times to mix.
- Gently shake for 2 min to thoroughly mix. The cDNA in the sample binds to the Nucleic Acid Binding Beads during this incubation.
- Move the plate to a magnetic stand to capture the magnetic beads. When the capture is complete (after 5 min), the mixture is transparent, and the Nucleic Acid Binding Beads form pellets against the magnets in the magnetic stand. The exact capture time depends on the magnetic stand that you use.
- Carefully aspirate and discard the supernatant without disturbing the magnetic beads, then remove the plate from the magnetic stand.
- Add 100 uL of Nucleic Acid Wash Solution to each sample, then shake the samples at moderate speed for 1 min.
- Move the plate to a magnetic stand to capture the Nucleic Acid Binding Beads.
3. Materials and Methods

- Carefully aspirate and discard the supernatant without disturbing the Nucleic Acid Binding Beads, then remove the plate from the magnetic stand.

- Move the plate to a shaker, then shake the plate vigorously for 1 min to evaporate residual ethanol from the beads. Dry the solution until no liquid is visible, but the pellet appears shiny.

- Elute the purified cDNA from the Nucleic Acid Binding Beads by adding 30 uL of preheated (55 to 58°C) Elution Solution to each sample. Incubate for 2 min at room temp. without shaking.

- Vigorously shake the plate for 3 min. Make sure that the Nucleic Acid Binding Beads are fully dispersed. If they are not, continue shaking until the beads are dispersed and/or pipette up-and-down 3 times.

- Move the plate to a magnetic stand to capture the Nucleic Acid Binding Beads.

- Transfer the supernatant, which contains the eluted cDNA, to a nuclease-free multiwell plate.

- Place the reaction on ice or you may freeze the samples at 20°C for overnight storage.

3.5.9 Assess cDNA yield and size distribution

For most tissue types, the recommended 10 ug of input cRNA should yield ≥6 ug of cDNA. Determine the concentration of a cDNA solution by measuring its absorbance at 260 nm. We recommend evaluating the absorbance of 1.5 uL of cDNA sample using a NanoDrop Spectrophotometer.

3.5.10 Fragment and label the single-stranded cDNA

- Set up fragmentation reaction in 0.2 mL strip tubes
3. Materials and Methods

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume/Amount in 1 Rxn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-Stranded DNA</td>
<td>5.5 ug</td>
</tr>
<tr>
<td>RNase-free Water</td>
<td>up to 31.2 uL</td>
</tr>
<tr>
<td>Total Volume</td>
<td>31.2 uL</td>
</tr>
</tbody>
</table>

- Prepare the Fragmentation Master Mix

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume/Amount in 1 Rxn</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNase-free Water</td>
<td>10 uL</td>
</tr>
<tr>
<td>10X cDNA Fragmentation Buffer</td>
<td>4.8 uL</td>
</tr>
<tr>
<td>UDG, 10 U/uL</td>
<td>1.0 uL</td>
</tr>
<tr>
<td>APE 1, 1,000 U/uL</td>
<td>1.0 uL</td>
</tr>
<tr>
<td>Total Volume</td>
<td>16.8 uL</td>
</tr>
</tbody>
</table>

- Add 16.8 uL of the above Fragmentation Master Mix to the samples prepared in first 1. Flick or gently vortex the tubes and spin down.

- Incubate the reactions in a thermal cycler at:
  - 37°C for 60 minutes
  - 93°C for 2 minutes
  - 4°C for at least 2 minutes

- Flick-mix, spin down the tubes, and transfer 45 uL of the sample to a new 0.2 mL strip tube. The remainder of the sample can be used for size analysis using a Bioanalyzer. The range in peak size of the fragmented samples should be approximately 40 to 70 nt.
3. Materials and Methods

3.5.11 Labeling of Fragmented Single-Stranded DNA

- Prepare the labeling reactions. A master mix using the 5X TdT Buffer, TdT and DNA Labeling reagent can be prepared just before aliquoting 15 uL into the 0.2 mL strip tubes containing the 45 uL of Fragmented Single-Stranded DNA.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume in 1 Rxn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragmented Single-Stranded DNA (from Procedure G)</td>
<td>45.0 uL</td>
</tr>
<tr>
<td>5X TdT Buffer</td>
<td>12.0 uL</td>
</tr>
<tr>
<td>TdT</td>
<td>2.0 uL</td>
</tr>
<tr>
<td>DNA Labeling Reagent, 5 mM</td>
<td>1.0 uL</td>
</tr>
<tr>
<td>Total Volume</td>
<td>16.8 uL</td>
</tr>
</tbody>
</table>

- After adding the labeling reagents to the fragmented DNA samples, flick-mix and spin them down.
3. Materials and Methods

- Incubate the reactions in a thermal cycler at:
  - 37°C for 60 minutes
  - 70°C for 10 minutes and
  - 4°C for at least 2 minutes

3.5.12 Hybridization

Three heating blocks are required: one at 65°C, one at 99°C, and the third one at 45°C

- Prepare the Hybridization Cocktail in a 1.5 mL, use RNase-free microfuge tube.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume for One 169 Format Array</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragmented and Labeled DNA Target</td>
<td>17.0 uL</td>
</tr>
<tr>
<td>Control Oligonucleotide B2 (3 nM)</td>
<td>1.7 uL</td>
</tr>
<tr>
<td>20X Eukaryotic Hybridization</td>
<td>5.0 uL</td>
</tr>
<tr>
<td>2X Hybridization Mix</td>
<td>50.0 uL</td>
</tr>
<tr>
<td>DMSO</td>
<td>7.0 uL</td>
</tr>
<tr>
<td>Nuclease-free Water</td>
<td>up to 100.0 uL</td>
</tr>
<tr>
<td>Total Volume</td>
<td>100.0 uL</td>
</tr>
</tbody>
</table>

- Flick or gently vortex the tubes and spin down.

- Heat the Hybridization Cocktail at 99°C for 5 minutes. Cool to 45°C for 5 minutes, and centrifuge at maximum speed for 1 minute.

- Equilibrate the GeneChip ST Array to room temperature immediately before use. Label the array with the name of the sample that will be hybridized.

- Inject the appropriate amount of the specific sample into the array through one of the septa.
### Array Format

<table>
<thead>
<tr>
<th>Array Format</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>49 (Standard)</td>
<td>200.0 uL</td>
</tr>
<tr>
<td>64</td>
<td>200.0 uL</td>
</tr>
<tr>
<td>169</td>
<td>80.0 uL</td>
</tr>
</tbody>
</table>

- Place array in 45 C hybridization oven, at 60 rpm, and incubate for 17 hours 1 hour.

#### 3.5.13 Wash, Stain and Scan

- After 17 hours of hybridization remove the array from the hybridization oven. Vent the array by inserting a clean pipette tip into one of the septa and extract the hybridization cocktail with a pipettor through the remaining septum. The used hybridization cocktail can be rehybed on another array if necessary. Store the cocktail at -20 C. Refill the probe array completely with the appropriate volume of Wash Buffer A.

- After the washing, continue with the staining process. Use the staining materials provided by Affymetrix, specialized for the targeted GeneChip.

- Finally scan the GeneChips in order to get the image of the fluorescence intensities.

#### 3.6 Microarray Analysis

With the aim to study the differential gene expression profile we compared the gene expression of mouse embryos lacking geminin from neural crest stem cells with wild type embryos. In other words we have generated two separate groups of mouse embryos based on their genotype. After embryo dissection at E10.5, we used Fluorescence Activated Flow Cytometry (FACS) to isolate neural crest cells from the whole embryo. The genotype of the experimental group of mouse embryos, is Geminin$^{kn/fl}$;Wnt1Cre$^{+/-}$; Rosa26$^{+/+}$ while the genotype of mouse embryos at the control group is Geminin$^{wt/fl}$;Wnt1Cre$^{+/-}$; Rosa26$^{+/+}$. For our
purposes we collected 4 wild type embryos and 5 experimental. Subsequently we isolated the RNA from the neural crest stem cells that were separated through FACS. As a result we ended with 4 wild type RNAs and 5 experimental RNAs that were used to hybridize to different Affymetrix GeneChips.

The unpaired T-test was used to determine the statistical significance of the gene expression profile, according to the MAQC Project. A t-test is any statistical hypothesis test where the test statistic follows a Student’s t distribution if the null hypothesis is supported. A test statistic is a single measure of some attribute of a sample (i.e. a statistic) used in statistical hypothesis testing. Student’s T-test can be used to determine if two sets of data are significantly different from each other, and is most commonly applied when the test statistic follows a normal distribution if the values of the scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistic follows a Student’s t distribution. The independent samples t-test is used when two separate sets of independent and identically distributed samples are obtained, one from each of the two populations being compared. In order to select the T-test with the optimal fitness to our data we checked whether our groups have equal variances.

\[ t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]

Figure 3.12: T-test used in case the groups have equal variance and unequal sample size

MicroArray Quality Control (MAQC) project is helping improve the microarray and next-generation sequencing technologies and foster their proper applications in discovery, development and review of FDA regulated products. MAQC project aims to:

- provide quality control (QC) tools to the microarray community so that to avoid procedural failures
3. Results

- develop guidelines for microarray data analysis
- establish QC metrics and thresholds
- evaluate the advantages and disadvantages of various data analysis methods

3.7 Solutions

**PBS 10X**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume</td>
<td>500 ml</td>
</tr>
<tr>
<td>NaCl</td>
<td>40 gr</td>
</tr>
<tr>
<td>KCl</td>
<td>1 gr</td>
</tr>
<tr>
<td>Na$_2$HPO$_4$</td>
<td>7.2 gr</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>1 gr</td>
</tr>
</tbody>
</table>

Table 3.6: PBS 10x solution, PH regulation is not necessary

**PBS 1x**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume</td>
<td>100 ml</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>90 ml</td>
</tr>
<tr>
<td>PBS 10x</td>
<td>10 ml</td>
</tr>
</tbody>
</table>

Table 3.7: PBS 1x solution, PH should be regulated at 7.2 to 7.4 using HCL

**TBE 10x**

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume</td>
<td>500 ml</td>
</tr>
<tr>
<td>Tris - base</td>
<td>54 g</td>
</tr>
<tr>
<td>Boric Acid</td>
<td>27.6 g</td>
</tr>
<tr>
<td>EDTA 0.5 M</td>
<td>20 ml</td>
</tr>
<tr>
<td>pH 8.0</td>
<td></td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>Up to 500 ml</td>
</tr>
</tbody>
</table>

Table 3.8: TBE 10x solution
3. Results

**TBE 0.5x**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume:</td>
<td>500 ml</td>
</tr>
<tr>
<td>TBE 10x</td>
<td>25 ml</td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>475 ml</td>
</tr>
</tbody>
</table>

Table 3.9: TBE 0.5x solution, used for electrophoresis

**EtOH 70%**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EtOH 100%</td>
<td>70 ml</td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>30 ml</td>
</tr>
</tbody>
</table>

Table 3.10: EtOH 70% used for cleaning

**Tail Buffer**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris HCl</td>
<td>100mM</td>
</tr>
<tr>
<td>EDTA</td>
<td>5 mM</td>
</tr>
<tr>
<td>SDS</td>
<td>0.2%</td>
</tr>
<tr>
<td>NaCl</td>
<td>200 mM</td>
</tr>
</tbody>
</table>

Table 3.11: Tail Buffer used for tail tissue dissolution

**1M Tris HCl pH 8.5**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume</td>
<td>500 ml</td>
</tr>
<tr>
<td>Tris</td>
<td>60.5 g</td>
</tr>
<tr>
<td>HCL</td>
<td>pH regulation to 8.5</td>
</tr>
</tbody>
</table>

Table 3.12: Tris HCl

**0.5 M EDTA pH 8**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume</td>
<td>500 ml</td>
</tr>
<tr>
<td>EDTA</td>
<td>93 g</td>
</tr>
</tbody>
</table>

Table 3.13: EDTA
## 3. Results

**20% SDS**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume</td>
<td>500 ml</td>
</tr>
<tr>
<td>SDS</td>
<td>93 g</td>
</tr>
</tbody>
</table>

Table 3.14: SDS

**5M NaCl**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Volume</td>
<td>500 ml</td>
</tr>
<tr>
<td>NaCl</td>
<td>100 g</td>
</tr>
</tbody>
</table>

Table 3.15: NaCl
Chapter 4

Results

4.1 Animal Models

Polymerase Chain Reaction (PCR) genotyping the mice and the embryos that have been used for this experiment. Here we present the results of the PCRs that we did with the aim to genotype our mice. All possible genotypes for geminin that we could get from our breedings are:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{geminin}^{\text{ko/Il}};\text{Wnt1Cre}^{+/\text{-}}$</td>
<td>Experimental</td>
</tr>
<tr>
<td>$\text{geminin}^{\text{ko/Il}};\text{Wnt1Cre}^{\text{-/-}}$</td>
<td>Wild Type</td>
</tr>
<tr>
<td>$\text{geminin}^{\text{wt/Il}};\text{Wnt1Cre}^{+/\text{-}}$</td>
<td>Wild Type</td>
</tr>
<tr>
<td>$\text{geminin}^{\text{wt/Il}};\text{Wnt1Cre}^{\text{-/-}}$</td>
<td>Wild Type</td>
</tr>
</tbody>
</table>

Table 4.1: All possible genotypes for the geminin gene and the group that they belong to

The genotype of the mice has been confirmed by the following PCRs.

4.1.1 Geminin Knock Out PCR

Geminin knock out (KO) PCR was performed with the aim to find out which mice carry a geminin knock out allele. The expected size of Geminin KO amplicon is
249 base pairs. The conditions of the KO PCR have already been described in Materials and Methods.

The PCR products were loaded on an agarose gel, for the electrophoresis. The first slot of the gel is loaded with a 1000bp ladder that we use as an indicator in order to determine the length of the DNA products. In the next slots we load the samples. In addition a slot is reserved to load a positive control (a sample that we a priori know that has the desired allele), a negative control (a sample we a priori know does not have the allele of interest) and finally a non template control. Non template control is a sample that has only the pcr mastermix and no DNA is added. Non template control is used with the intention to check whether our PCR was contaminated. In figure 4.1 we present a photo of the electrophoresis gel. We observe the gel under UV light as ethidium bromide, binded to DNA, will glow. The band that corresponds to the knock out allele is expected at 249 base pairs.

Figure 4.1: A Geminin knock out pcr (electrophoresis gel). A 1000bp ladder is used in order to determine the length of the DNA products. A positive and negative control is included. We tagged all the samples according to the Geminin allele that they bear.
4.1.2 Geminin floxed PCR

Geminin floxed PCR determines the presence of a floxed geminin allele. In the mice used for our experiments the expected size of Geminin floxed amplicon is 597 bases while the allele of Geminin wild type is 549 base pairs. The conditions of Geminin floxed PCR are shown in Material and Methods. In figure 4.2 there is an electrophoresis gel from a Floxed PCR. The length of the Geminin floxed allele (597 base pairs) is 52 base pairs longer than that of geminin wild type (549 base pairs). As a result the higher band is the Geminin floxed allele and the lower the Geminin wild type allele.

Figure 4.2: Geminin floxed pcr. As a positive control we used a Geminin Floxed/Floxed sample. Negative control is a Geminin Wild Type / Wild Type sample. We have tagged all the samples according to their genotype.

4.1.3 Wnt1 Cre PCR

Cre recombinase under the control of Wnt1 regulatory elements was used to ablate geminin specifically in neural crest stem cells. The PCR conditions have already been described in Materials and Methods. Here we present the gel that was used for the electrophoresis. We use a 1000bp ladder, a posive, a negative as well as a non template control.
Figure 4.3: Wnt1 Cre pcr. Cre recombinase is under the control of Wnt1 regulatory elements in order to ablate Geminin specifically in neural crest stem cells. All samples have been tagged according to their genotype.

4.2 Total RNA Isolation

Reliable results in gene expression analysis applications such as real-time RT-PCR and microarray analysis depend strongly on the quality of the RNA sample used.

RNA isolation was done using Qiagen’s RNeasy micro kit. For all the experiments we obtained high quality RNA from a limited number of cells. RNA quality affects how efficiently an RNA sample is amplified. High-quality RNA is free of contaminating proteins, DNA, phenol, ethanol, and salts.

4.3 MicroArray Chip preparation

The MicroArray Chip preparation was performed at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. The process that was
followed is described in Materials and Methods chapter. For the present study we have used the Affymetrix GeneChip MoGene 1.0ST, which is designed for the mouse genome. The algorithm used for analysis of gene expression is the RMA16. The GeneChip also includes hybridization and poly-A controls in order to check the quality of the experiment. The GeneChip MoGene 1.0 ST array targets 28,853 genes with 770,317 probes. In other words there are approximately 27 probes for each gene.

4.4 Differentially expressed gene profile

A .DAT file is the first file generated when a microarray is scanned. It is a table comprised of the fluorescence measurement and feature identity for each pixel of a scan. A microarray feature is an area on the array occupied by a population of oligonucleotide probes with the same sequence. A feature is comprised of many pixels. The data in a .DAT file is processed to generate the values found in the .CEL file. A .CEL file contains a single value for each feature. Each value in a .CEL file is a statistical "summary" of the fluorescence from a single feature. The .CEL file contains the information for viewing the images that result from scanning the hybridized arrays.

The list of differentially expressed genes contains all the genes that have a statistically significant difference between the experimental and the wild type group. This is the main list of genes that we will study and subsequently use for all analyses. It is obvious that this list contains genes that have an association with the deletion of geminin in neural crest stem cells. As already has been mentioned unpaired T-test was used in order to select only the genes with a statistically significant differential expression. The fold change of each gene is based on their fluorescence intensities on the microarray chip. Fold change indicates the difference in expression between the two groups. In our case, that we are interested in changes that occur in the experimental group, fold change indicates how many times more or less a gene is expressed in the experimental group compared to the control group.

For the unpaired T-test we set p value threshold at 0.05. This means that all genes with a p value less that 0.05 are considered to be statistically significant.
We have also set a threshold for the fold change. We have a scientific interest only in genes that have a fold change greater than the absolute value of 1.5. In other words we select all genes with a fold change greater than 1.5 or less than -1.5. Genes that have a fold change less than 1.5 or greater than -1.5 do not have a scientific interest as their change is not considered significant.

In statistics, a volcano plot is a type of scatter-plot that is used to quickly identify changes in large datasets composed of replicate data. It plots statistical significance versus fold-change on the y- and x-axes, respectively. These plots are increasingly common in omic experiments such as genomics. In our case, the volcano plot indicates the genes of interest (p-value < 0.05 and |fc| > 1.5) throughout the total number of genes that have been hybridized. The scale of y and x axes is logarithmic in order to better represent the whole population within a smaller space. The red dots represent the genes that have a p-value < 0.05 and the blue dots indicate the genes of interest (p-value < 0.05 and |fc| > 1.5). Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left- or the right-hand side. These represent values that display large magnitude fold changes (hence being left- or right- of center) as well as high statistical significance (hence being towards the top). Figure 4.4 is a volcano plot presenting the results of our analysis. The horizontal axis is the fold change (here log2 scale), as we already mentioned, we are interested in genes with a fold change greater than 1.5, the blue squares represent those genes. The vertical axis is the p-value (here log10 scale). The threshold for the p-value is 0.05. Blue and red squares are the genes overpassing that threshold. The grey squares are the genes that fail to pass both our criteria while the red squares are the genes that are statistically significant but their fold change is less than 1.5.
Figure 4.4: A volcano plot indicating the genes of interest throughout the total number of genes that have been hybridized. In our study we have 865 genes of interest, indicated with the blue dots.

After the statistical analysis we conclude that there are 865 genes of interest (p-value < 0.05 and |fc| > 1.5), 105 out of these genes are predicted genes so we exclude them from our further analyses. As a result we have 760 genes of interest. 284 genes out of the 760 genes of interest are down regulated (37%) and the rest 476 genes are upregulated (63%).

Figure 4.5 presents the properties of the statistical test that was used, (unpaired T-test) and the total number of genes of interest, according to the different values for the p-value and fold change. As the value of p-value decreases fewer genes were identified. We also observe that with our criteria we have 865 genes of interest.
Figure 4.5: The table presents the properties of the unpaired t-test and the resulting genes of interest, according to the different values for p-value and fold change. As the criteria that we will set are getting more strict there are fewer genes of interest.

A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors. In Biology heatmaps are typically used to represent the level of expression of many genes across a number of comparable samples. The genes that we have included are the genes of interest (p-value < 0.05 and |fc| > 1.5). The fold change of a gene is represented by the the color and the color’s intensity. To indicate the minimum gene expression we use far red, while maximum gene expression is indicated by the far green. Figure 4.6 is a heatmap presenting the level of expression of the genes of interest. In other words the horizontal axis of the heatmap presents the two groups and the vertical all the genes that overpass all the thresholds (p-value, fold change) that we have set.
Figure 4.6: A heatmap presenting the differential gene expression profile of genes of interest. Heatmap was used to visualize easily and compare the expression of many genes across our conditions - groups. The fold change of a gene is represented by the color and the color’s intensity. To indicate the minimum gene expression we use far red, while maximum gene expression is indicated by the far green.

A Venn diagram or set diagram is a diagram that shows all possible logical relations between a finite collection of sets. Venn diagrams are used to illustrate simple set relationships in probability, logic, statistics, linguistics and computer science. Figure 4.8 is a venn diagramm showing the total number of genes as well as the genes that are statistically significant upregulated and downregulated.
The list of top ten down regulated genes includes some important transcription factors for proper craniofacial development (Alx1, Alx3, Alx4). Moreover as shown in figure 4.8 in top ten down regulated genes is a Cxcl13 a chemokine associated with neural crest cells migration. Furthermore Wnt1 is included in the top ten up regulated genes.
There are 260 genes with an absolute fold change equal to or more than 2.0. These genes have a severe impact at the phenotype as they have a high expression change. Figure 4.9 summarizes the characteristics of the list of the genes of interest.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Fold Change</td>
<td>1.5</td>
</tr>
<tr>
<td>P-value</td>
<td>0.05</td>
</tr>
<tr>
<td>Total number of genes (1.5 fold change)</td>
<td>760</td>
</tr>
<tr>
<td>Down regulated genes (1.5 fold change)</td>
<td>284</td>
</tr>
<tr>
<td>Up regulated genes (1.5 fold change)</td>
<td>476</td>
</tr>
<tr>
<td>Maximum fold change</td>
<td>+3.69</td>
</tr>
<tr>
<td>Minimum fold change</td>
<td>-5.50</td>
</tr>
<tr>
<td>Total number of genes (2.0 fold change)</td>
<td>260</td>
</tr>
<tr>
<td>Total number of genes (3.0 fold change)</td>
<td>19</td>
</tr>
</tbody>
</table>

Our analysis suggests that Geminin plays a key role in neural crest stem cells development as its deletion alters the expression of important genes for neural crest development among them several transcription factors.
4.5 Transcription Factors

In molecular biology and genetics, a transcription factor is a protein that binds to specific DNA sequences, thereby controlling the flow of genetic information from DNA to messenger RNA. A defining feature of transcription factors is that they contain one or more DNA-binding domains, which attach to specific sequences of DNA adjacent to the genes that they regulate. Transcription factors operate alone or with other proteins in a complex, by promoting, or blocking the recruitment of RNA polymerase to specific genes. In our list of genes of interest there are 76 genes with transcription factor activity. These genes are interacting selectively with a specific DNA sequence in order to modulate transcription.

4.5.1 The expression of the Hox-family genes is induced

It is well-established that genes of the Hox family play an important part in regulating embryonic development and proliferation. They specify embryonic structures along the body axis and are associated with normal and malignant cell growth. In our data most of the Hox genes are upregulated. [Vieux-Rochas et al., 2013] [Minoux et al., 2013]
4.5.2 Members of the Alx-family are down regulated

Alx1, Alx3 and Alx4 are all known to be expressed in the facial mesenchyme of vertebrate embryos. In our study there are 3 members of the Alx family which are significantly down regulated. The fold change of Alx1 is -4.25. Alx1, which is also a transcription factor, is necessary for survival of the forebrain mesenchyme and may also be involved in the development of the cervix. Mutations in this gene lead to neural tube defects such as acrania and meroanencephaly. In ad-
dition morpholino knock-down of zebrafish alx1 expression causes a profound craniofacial phenotype including loss of the facial cartilages and defective ocular development. [Dee et al., 2013]

Alx3 is a nuclear protein with a homeobox DNA-binding domain that functions as a transcriptional regulator involved in cell-type differentiation and development. Alx3 is down regulated 3.02 times. Alx3 is expressed in mouse embryos from day 8 of gestation in a characteristic pattern, predominantly in neural crest-derived mesenchyme and in lateral plate mesoderm. [Berge et al., 1999]

![Alx gene family regulation](image)

Figure 4.11: Heatmap presenting the differential expression of alx gene family

Alx4 encodes a paired-like homeodomain transcription factor expressed in the mesenchyme of developing bones, limbs, hair, teeth and mammary tissue. Mutations in this gene cause a form of frontonasal dysplasia with alopecia and hypogonadism, suggesting a role for this gene in craniofacial development, mesenchymal-epithelial communication, and hair follicle development. [Beverdam et al., 2001] In our study Alx4 is downregulated 3.65 times.

### 4.5.3 Gata family members are down regulated

It has been shown that a network of crossregulatory transcription factors like Mash1, Hand2, Pahox2a and Phox2b directs the specification of sympathetic neurons from neural crest precursors. Genes of the gata family have been identified as factors in autonomic neuron differentiation.
Tsarovina et al. have characterised Gata2 and Gata3 genes as essential members of the transcription factor network which regulates sympathetic neuron development [Tsarovina et al., 2004].

### 4.5.4 Fox gene family members change their expression in the absence of Geminin

FOX (Forkhead box) proteins are a family of transcription factors that play important roles in regulating the expression of genes involved in cell growth, proliferation, differentiation, and longevity. Many FOX proteins are important to embryonic development. These proteins also have essential transcription activity by being able to bind condensed chromatin during cell differentiation processes. The defining feature of FOX proteins is the forkhead box, a sequence of 80 to 100 amino acids forming a motif that binds to DNA.
Ascl1 is upregulated

Ascl1 is a transcription regulator which is upregulated 1.58 times. Ascl1 is essential for the generation of olfactory and autonomic neurons and also plays a role in the initiation of neuronal differentiation. Ascl1 mediates transcription activation by binding to the E box (5’-CANNTG-3’). Ascl1 plays an important role for neuronal as well as catecholaminergic differentiation. Ascl1 knockout mice display severe deficits in sympathetic ganglia. In addition Ascl1 is necessary for the development of the vast majority of chromaffin cells. [Huber et al., 2002]

4.6 Chemokines Regulation

Chemokines are a family of small cytokines, or signaling proteins secreted by cells. Their name is derived from their ability to induce directed chemotaxis in nearby responsive cells, they are chemotactic cytokines. The major role of chemokines is to act as a chemoattractant to guide the migration of cells. Cells that are attracted by chemokines follow a signal of increasing chemokine concentration towards the source of the chemokine. It is known that Cxcl12 - Cxcr4 signaling is responsible for mesencephalic neural crest stem cells migration. [Rezzoug et al.,
In zebrafish loss of Cxcr4a results in aberrant cranial neural crest cells migration defects in the neurocranium, as well as cranial ganglia dysmorphogenesis. Moreover, overexpression of either Cxcl12 or Cxcr4a causes aberrant cranial neural crest cells migration and results in ectopic craniofacial cartilages. [Killian et al., 2009]

Theveneau et al. studied the interaction between neural crest cells and placodal cells, an epithelial tissue that contributes to sensory organs. They proved that neural crest cells chase placodal cells by chemotaxis, and placodal cells run when contacted by neural crest cells. The chase and run is controlled by Cxcr4 and its ligand Cxcl12. [Theveneau et al., 2013]

### 4.7 Boxplot of our data

In descriptive statistics, a boxplot (or box plot) is a convenient way of graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram. Outliers may be plotted as individual points. In figure 4.16 we present our data in a boxplot graph. There are two groups of embryos, Experimental and Wild Type (Control) and for each embryo there is a boxplot.

![Boxplot of Chemokines](image)

**Figure 4.14:** Heatmap presenting the regulation of the chemokines present in our study. Cxcl12 and Cxcr4 signaling is responsible for the mesencephalic neural crest stem cells migration.
The vertical axis is the expression values of the genes. As a result a boxplot is an easy way to show the range of the expression values of the genes.

Box and whisker plots are uniform in their use of the box: the bottom and top of the box are always the first and third quartiles, and the band inside the box is always the second quartile (the median). [Benjamini, 1988]

Figure 4.15: A boxplot is often used in bioinformatics to easily present the data. Here we present the box plot of our data. We have two separate groups of embryos Experimental and Wild Type. The vertical axis is the expression value of the genes. Through a boxplot it is easy to identify outliers.

4.8 Unsupervised Hierarchical Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields including bioinformatics.

Hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. These algorithms connect
"objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will form, which can be represented using a dendrogram.

Our samples are separated in two groups, Experimental and Wild Type (Control). Figure 4.17 presents the unsupervised hierarchical clustering that we performed with our data. For each sample there is a heatmap presenting the expression value of the genes of interest. In other words the vertical axis is the expression value of the genes of interest, the expression value is represented by the colour intensity.

Figure 4.16: Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. In our case there are two groups of embryos, Experimental and Wild Type. For each sample there is a heatmap representing the expression value of the genes of interest.
4.9 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Figure 4.18 presents the PCA graph for our data. The red dots represent the experiment (Geminin knock out) mice while the blue dots represent the wild type mice. We observe that the two groups of samples are well separated.

Figure 4.17: Principal Component Analysis graph. The red dots represent the experiment (Geminin knock out) mice while the blue dots represent the wild type mice.
Principal Component Analysis reveals the internal structure of the data in a way that best explains the variance in the data. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

4.10 Gene expression Meta-Analysis

Ingenuity Pathway Analysis (IPA) is a web-based software application that may be used to analyze, integrate, and understand data derived from gene expression, microarrays; metabolomics, proteomics, as well as RNA-Seq experiments. We used IPA to quickly identify relationships, mechanisms, functions, and pathways of relevance, in order to move beyond statistical analysis to novel biological insights.

The list of genes of interest was used as input for the IPA software. The high statistical significance of the meta-analysis results verify our hypothesis that geminin is essential for the development and the migration of neural crest stem cells.

The next figure describes the top 5 associated network functions. It is obvious that the network functions that are strongly affected by the deletion of geminin in neural crest stem cells are essential for the development of the embryo. The score given is based on the fold change of the gene and it’s position in the pathway.
Figure 4.18: The top associated network functions that are affected by the deletion of geminin. The functions that are vital for the homeostasis and the proper development of the embryo.

Embryos lacking geminin from neural crest stem cells seem to suffer from severe cardiovascular abnormalities and as a result die a few days before birth.

Ingenuity pathway analysis may also associate the statistically significant differentially expressed genes with diseases and disorders. Clearly, the top 5 diseases are totally associated with the phenotype as their statistical significance is high. P-value is based on the numbers of the genes of that are present in our analysis compared to the total number of genes involved in the specific cluster.
Figure 4.19: Diseases and disorders caused by the deletion of geminin in neural crest stem cells.

It is also worth mentioning the physiological system development functions that are affected by the deletion of geminin. A lot of developmental functions are known and their role has been well established. Figure 4.21 shows the developmental functions that are affected by the deletion of geminin.

<table>
<thead>
<tr>
<th>Diseases and Disorders</th>
<th>P-Value</th>
<th>Number of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>5.20E-20 - 4.13E-05</td>
<td>452</td>
</tr>
<tr>
<td>Gastrointestinal Disease</td>
<td>5.20E-20 - 1.79E-05</td>
<td>263</td>
</tr>
<tr>
<td>Skeletal and Muscular Disorders</td>
<td>1.76E-18 - 3.78E-05</td>
<td>94</td>
</tr>
<tr>
<td>Developmental Disorder</td>
<td>1.74E-13 - 3.78E-05</td>
<td>133</td>
</tr>
<tr>
<td>Neurological Disease</td>
<td>5.36E-11 - 3.46E-05</td>
<td>211</td>
</tr>
</tbody>
</table>

Figure 4.20: Physiological System Development and Function

<table>
<thead>
<tr>
<th>Physiological System Development and Function</th>
<th>P-Value</th>
<th>Number of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryonic Development</td>
<td>1.32E-52 - 4.38E-05</td>
<td>254</td>
</tr>
<tr>
<td>Organismal Development</td>
<td>1.32E-52 - 4.76E-05</td>
<td>311</td>
</tr>
<tr>
<td>Connective Tissue Development and Function</td>
<td>2.34E-40 - 4.38E-05</td>
<td>114</td>
</tr>
<tr>
<td>Organ Development</td>
<td>2.34E-40 - 4.38E-05</td>
<td>227</td>
</tr>
<tr>
<td>Organ Morphology</td>
<td>2.34E-40 - 4.76E-05</td>
<td>208</td>
</tr>
</tbody>
</table>

All developmental functions that are affected are vital not only for the proper development of the embryo but also for its vitality.

Ingenuity Pathway Analysis enables us to determine over represented signaling and metabolic canonical pathways. The pathways and the genes that are involved are based on prior knowledge. The affected canonical pathways and their high statistical significance prove once more the essential role of geminin in neural...
crest stem cells development and migration. Figure 4.22 presents the canonical pathways that are affected by the deletion of geminin.

<table>
<thead>
<tr>
<th>Top Canonical Pathways</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcriptional Regulatory Network in Embryonic Stem Cells</td>
<td>1.04E-08</td>
</tr>
<tr>
<td>Axonal Guidance Signaling</td>
<td>1.52E-07</td>
</tr>
<tr>
<td>Regulation of the Epithelial-Mesenchymal Transition Pathway</td>
<td>1.61E-07</td>
</tr>
<tr>
<td>Human Embryonic Stem Cell Pluripotency</td>
<td>3.35E-07</td>
</tr>
<tr>
<td>Basal Cell Carcinoma Signaling</td>
<td>9.87E-06</td>
</tr>
</tbody>
</table>

Figure 4.21: Top canonical pathways over represented in our data. The pathways affected indicate the key role of geminin in neural crest cells.

The epithelial-mesenchymal transition (EMT) is a process by which epithelial cells lose their cell polarity and cell-cell adhesion, and gain migratory and invasive properties to become mesenchymal stem cells. EMT is a major process which characterises neural crest stem cells.

Neuronal connections are formed by the extentions of axons. Axon guidance is a subfield of neural development concerning the process by which neurons send out axons to reach the correct targets. Axons often follow very precise paths in the nervous system, and how they manage to find their way so accurately is being researched.

### 4.11 Gene Ontology Analysis

Gene ontology (GO) is a major bioinformatics initiative to unify the representation of gene and gene product attributes across all species. The Gene Ontology Project provides an ontology of defined terms representing gene product properties. For the Gene Ontology analysis we used bioCompendium.
is a publicly accessible, high-throughput experimental data analysis platform. It helps in prioritizing the targets from gene expression analysis studies. The system is designed to work with large lists of genes or proteins for which it collects a wide spectrum of biological information. bioCompendium is able to compare, analyze and enrich the given gene list against background and show the results of comparison. bioCompendium provides knowledge from biological databases, cluster the proteins based on sequence similarity or domain composition, performs the pathway or GeneOntology enrichments, provides the information about interacting drugs, metabolites, ligands, other chemicals and proteins. It also provides the transcription factor binding site (TFBS) profiles, diseases associations and orthology information. [GeneOntology, 2013]

The ontology covers three domains:

- cellular component, the parts of a cell or its extracellular environment,

<table>
<thead>
<tr>
<th>Accession</th>
<th>Term</th>
<th>Adjusted P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0044421</td>
<td>extracellular region part</td>
<td>1.00E-10</td>
</tr>
<tr>
<td>GO:0043005</td>
<td>neuron projection</td>
<td>5.00E-10</td>
</tr>
<tr>
<td>GO:0045202</td>
<td>synapse</td>
<td>9.00E-09</td>
</tr>
<tr>
<td>GO:0005576</td>
<td>extracellular region</td>
<td>4.00E-08</td>
</tr>
<tr>
<td>GO:0044456</td>
<td>synapse part</td>
<td>1.00E-07</td>
</tr>
</tbody>
</table>

Figure 4.22: The figure presents the top 5 cellular components that are altered by the deletion of Geminin.

- molecular function, the elemental activities of a gene product at the molecular level, such as binding or catalysis,
Figure 4.23: Gene ontology molecular function describes the elemental activities of a gene product.

- biological process, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units.

### Molecular Function

<table>
<thead>
<tr>
<th>Accession</th>
<th>Term</th>
<th>Adjusted P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0043565</td>
<td>sequence-specific DNA binding</td>
<td>1.00E-25</td>
</tr>
<tr>
<td>GO:0003700</td>
<td>transcription factor activity</td>
<td>2.00E-21</td>
</tr>
<tr>
<td>GO:0030528</td>
<td>transcription regulator activity</td>
<td>1.00E-16</td>
</tr>
<tr>
<td>GO:0003677</td>
<td>DNA binding</td>
<td>4.00E-11</td>
</tr>
<tr>
<td>GO:0019199</td>
<td>transmembrane receptor protein kinase activity</td>
<td>6.00E-09</td>
</tr>
</tbody>
</table>

Figure 4.24: Gene Ontology Biological Process refers to all operations of molecular events. Here we present the top 5 biological processes that are affected by the deletion of Geminin in neural crest stem cells.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Term</th>
<th>Adjusted P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0048731</td>
<td>system development</td>
<td>1.00E-52</td>
</tr>
<tr>
<td>GO:0032501</td>
<td>multicellular organismal process</td>
<td>2.00E-52</td>
</tr>
<tr>
<td>GO:0032502</td>
<td>developmental process</td>
<td>1.00E-44</td>
</tr>
<tr>
<td>GO:0030182</td>
<td>neuron differentiation</td>
<td>1.00E-33</td>
</tr>
<tr>
<td>GO:0022008</td>
<td>neurogenesis</td>
<td>1.00E-33</td>
</tr>
</tbody>
</table>
4.12 Phenotype Analysis

For the Phenotype Analysis of our data we used WebGestalt. WebGestalt is a "WEB-based GEne SeT AnaLysis Toolkit". It is designed for functional genomic, proteomic and large-scale genetic studies from which large gene lists (e.g. differentially expressed gene sets,) are continuously generated. Phenotype analysis is based on data provided from the Mouse Genome Informatics database curated by the Jackson Laboratory.

The following figure presents the main results from the phenotype analysis.

<table>
<thead>
<tr>
<th>Phenotype Analysis</th>
<th>Adjusted P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalian Phenotype</td>
<td></td>
</tr>
<tr>
<td>Abnormal Craniofacial Morphology</td>
<td>1.05E-09</td>
</tr>
<tr>
<td>Abnormal Neuron Differentiation</td>
<td>1.73E-09</td>
</tr>
<tr>
<td>Abnormal Hindbrain Development</td>
<td>3.17E-10</td>
</tr>
<tr>
<td>Abnormal Axial Skeleton Morphology</td>
<td>5.40E-16</td>
</tr>
<tr>
<td>Lethality During Fetal Growth</td>
<td>2.82E-31</td>
</tr>
</tbody>
</table>

Figure 4.25: The phenotype analysis was performed with WebGestalt, a web based bioinformatics tools developed by Vanderbilt University.

4.13 Pathway Analysis

4.13.1 Geminin seems to be involved into the regulation of the epithelial to mesenchymal transition

Epithelial-to-mesenchymal transition (EMT) is a complex cellular and molecular program during which epithelial cells lose their differentiated characteristics including cell-cell adhesion, planar and apical-basal polarity and lack of motility, and acquire instead mesenchymal features, including motility and invasiveness. EMT occurs in an orchestrated fashion with one of its earliest events involving the disruption of tight junctions that connect epithelial cells. Adherens junction complexes, which contain E-cadherin and B-catenin are also disrupted along
with a reorganization of the actin cytoskeleton. EMT is observed in: embryonic implantation and gastrulation and neural crest cell motility, wound healing, tissue-regeneration, inflammation and fibrosis and the conversion of differentiated epithelial cancer cells into migratory mesenchymal cancer cells which may lead to cancer invasion, systemic cancer cell dissemination and metastasis.

Soon after induction and specification in the ectoderm, at the border of the neural plate, the neural crest stem cells population leaves its original territory through a delamination process. The delamination involves a partial or complete epithelium-to-mesenchyme transition (EMT) regulated by a complex network of transcription factors including several proto-oncogenes. [Eric Theveneau, 2012] Major inducers of EMT that have been identified include TGF-B, Notch, Wnts and growth factors like EGF, IGF, HGF, PDGF and FGF. Figure 4.26 is a heatmap of the genes in our analysis that are involved in the regulation of EMT.

| Regulation of the Epithelial-Mesenchymal Transition Pathway |
|-----------------|-----------------|-----------------|
| Fzd10           | 1.653473        | Fgfr3           | 1.817624        |
| Lox             | -2.61233        | Foxc2           | -2.43786        |
| Sna2            | -2.46362        | Gsc             | -2.42865        |
| Twist2          | -2.33067        | Fzd4            | -2.03577        |
| Wnt3            | 1.781059        | Wnt3a           | 1.601016        |
| Sna1            | -1.69412        | Fgf18           | 2.05149         |
| Fgf14           | 1.52331         | Hgf             | -2.13902        |
| Twist1          | -2.13169        | Notch1          | 1.581741        |
| Mmp2            | -1.62518        | Wnt1            | 2.694847        |
| Fzd1            | -1.54866        | Egfr            | -1.70974        |
| Wnt8b           | 1.944703        | Pdgfrb          | -2.50075        |

Figure 4.26: Heatmap of statistically significant differentially expressed genes, which are involved into the regulation of EMT.
4.14 Geminin affects Notch signaling pathway

The Notch gene family encodes transmembrane receptors that are involved in cell fate choices in vertebrates and invertebrates. Notch signaling is an evolutionarily conserved mechanism for cell-cell interaction. The Notch receptor is a single transmembrane-spanning protein with a large extracellular domain, which requires to be proteolytically processed in order to be functional. Only the cleaved fragments are presented at the cell surface where they interact with ligand, Jagged or Delta, which are transmembrane proteins expressed on the signal sending cell. Ligand interaction leads to a second cleavage of the receptor close to the plasma membrane at the extracellular side. [Yan et al., 2009] Figure 4.27 presents the statistically significant differentially expressed genes implicated in notch pathway.

<table>
<thead>
<tr>
<th>Notch Signaling Pathway</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dll1</td>
<td>2.440635</td>
</tr>
<tr>
<td>Dbx4</td>
<td>1.643405</td>
</tr>
<tr>
<td>Hes5</td>
<td>1.579881</td>
</tr>
<tr>
<td>Mfn2</td>
<td>1.986308</td>
</tr>
<tr>
<td>Lfng</td>
<td>2.005307</td>
</tr>
<tr>
<td>Dll3</td>
<td>2.426931</td>
</tr>
<tr>
<td>Notch1</td>
<td>1.581741</td>
</tr>
</tbody>
</table>

Figure 4.27: Heatmap of statistically significant differentially expressed genes, which are involved into the notch pathway.

4.15 The deletion of Geminin affects cardiogenesis

Transcription factors that initiate the cardiac fate include NKX2.5, GATA family members, T-box family members (Tbx5, Tbx20), MCMI, agamous, deficiens, serum response factor (MADS), and MEF2 family members. Extracellular signals that act upstream of these factors include BMP, FGF, and WNT families of
growth factors and WNT antagonists such as Dkk1 and Crescent.

FGF2 is required for the expression of cardiac transcription factors. BMP receptors are essential for cardiac organogenesis. In presence of BMPs, the Type-II receptor (BMPRII) phosphorylates the Type-I receptors (ALK3, ALK6), which activate signaling by intracellular effector SMADs and TAK1. SMAD1, SMAD5 and SMAD8 transduce signals from BMPs specifically, while SMAD4 is a general partner of ligand-specific SMADs. The BMP pathway is negatively regulated by Noggin. The maintenance of a cardiac program is the result of well-orchestrated interactions between all the above factors. Figure 4.28 is a heatmap of our genes of interest associated with cardiogenesis.

<table>
<thead>
<tr>
<th>Factors Promoting Cardiogenesis in Vertebrates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fzd4</td>
<td>-2.03577</td>
</tr>
<tr>
<td>Fzd10</td>
<td>1.653473</td>
</tr>
<tr>
<td>Wnt3</td>
<td>1.781059</td>
</tr>
<tr>
<td>Bmp2</td>
<td>-1.59396</td>
</tr>
<tr>
<td>Fzd1</td>
<td>-1.54866</td>
</tr>
<tr>
<td>Dkk1</td>
<td>-2.10604</td>
</tr>
<tr>
<td>Bmp6</td>
<td>1.587218</td>
</tr>
<tr>
<td>Bmp5</td>
<td>-1.60013</td>
</tr>
</tbody>
</table>

Figure 4.28: Heatmap of statistically significant differentially expressed genes, which are involved into cardiogenesis.
Chapter 5

Discussion

5.1 Neural Crest Stem Cells and Geminin

Neural crest stem cells are an embryonic cell population, vital for proper vertebrate development. They are initially localized to the dorsal neural folds. Prior to the migration neural crest cells undergo an epithelial-to-mesenchymal transition (EMT) and then migrate to their final destinations in the developing embryo. [Wu et al., 2014]

Geminin is an essential protein for the cells as it has two crucial and very important roles. Geminin regulates the cell cycle by directly binding to Cdt1 licensing factor and thus inhibiting the DNA replication. In addition Geminin interacts with transcriptional regulators of differentiation and chromatin remodelling factors. Development and homeostasis of multicellular organisms relies on an intricate balance between cell proliferation and differentiation. Geminin is a protein that directs the balanced interactions, implicated in proliferation-differentiation decisions during development [de Renty et al., 2014].

Knowing the double role of Geminin we intended to investigate and understand the role of Geminin in mouse neural crest stem cells. For this reason we generated two groups of mouse embryos of the age E10.5 dpc, wild type embryos and embryos whose geminin was ablated from neural crest stem cells. As we have shown geminin is required in neural crest stem cells for the vitality of the embryo. Embryos lacking geminin from neural crest cells showed a severe phenotype. In addition, the deletion of geminin affect the development of the enteric nervous
system. Hirschsprung’s disease is a disorder of the abdomen that occurs when part or all of the large intestine or antecedent parts of the gastrointestinal tract have no nerves and therefore cannot function. In Hirschsprung’s disease, the migration of the neural crest stem cells towards the colon is not completed. [PhD Stathopoulou, 2014]

The following heatmaps present some of the transcription factors present in our analysis.

![Heatmap of differential expression of some of the most important transcription factors.](image)

Figure 5.1: Heatmap of differential expression of some of the most important transcription factors.

Pax genes have been shown to play an important role in mammalian development during organogenesis for example Pax9 is essential for the development of the thymus. Mice lacking Pax9 gene from neural crest stem cells die shortly after birth (Balling 1998). Similarly Twist1 is a key regulator of craniofacial development. Twist1-null mouse embryos exhibit failure of cephalic neural tube closure and abnormal head development and die at E11.0.

It has been shown that foxc1 mutations lead to corneal angiogenesis, and that mice homozygous for either a global (Foxc1(-/-)) or neural crest specific (NC-Foxc1(-/-)) null mutation display excessive growth of corneal blood and lymphatic vessels. [Seo et al., 2012] There is also another study claiming that Foxc1 null embryos display bony syngnathia together with defects in maxillary and mandibular structures, as well as agenesis of the temporomandibular joint. [Inman et al.,
5. Discussion

2013] In addition, it has been demonstrated that Foxc1 and Foxc2 are essential for arterial cell specification, lymphatic vessel formation, angiogenesis and cardiac outflow tract development. [Kume, 2009] We should mention the fact that FoxD3 regulates neural crest determination in Xenopus embryos. Its overexpression in the latter, as well as in ectodermal explants, induces expression of neural crest markers. [Sasai et al., 2001] Expression of FoxD3 in the presumptive neural crest region starts at the late gastrula stage in a manner similar to that of Slug, and overlaps with that of Zic-r1.

5.2 Similar Studies

5.2.1 The role of Geminin in hematopoietic stem cells

With the intention to clarify the role of Geminin in the hematopoietic system, we used conditional knockout mice to ablate Geminin early in the developing embryonic hematopoietic compartment. We used mice that express the Cre recombinase under the control of the Vav1 gene regulatory elements (Vav1:iCre) and the loxP system to delete Geminin from hematopoietic stem and progenitor cells (HSPCs). The knockout mice are embryonically lethal, mainly due to anemia and dyserythropoiesis. In vivo and in vitro assessment of the hematopoietic hierarchy showed that the knockout mice exhibit a dramatic decrease in the numbers of committed progenitors and mature blood cells. Conversely, the hematopoietic stem cell pool was increased in the knockout mice, suggesting a defect in their differentiation potential. In order to dissect the molecular basis of this phenotype, we have used cDNA microarray analyses, which revealed the deregulation of genes involved in hematopoietic lineage specification transcription programs and epigenetic regulation. Gene expression data mining in combination with in silico comparative analyses support further our findings and suggest a key role for Geminin in hematopoiesis at the transcriptional and epigenetic levels.

Microarray analysis showed that in total 1929 genes were found to be differentially expressed in HSPCs upon Geminin ablation of which 1403 genes were upregulated and 526 genes were downregulated (p<0.05 and FC>1.5).
5. Discussion

Deletion of Geminin from the early stages in the fetal liver hemopoietic compartment in E15.5 dpc embryos results in defective blood cell lineage maturation. Whole transcriptome analyses revealed that this phenotype is supported by transcriptional deregulation of numerous genes, many of which are lineage-specific transcription factors and epigenetic regulators. To this end, we have conducted in silico comparative studies of our cDNA microarray data in order to decipher the molecular mechanisms that underline the mutant embryos phenotype. Our data suggest that during development of the embryonic hemopoietic system Geminin regulates lineage specification transcriptional programs.

There are 98 genes that are statistically significant differentially expressed in both microarray studies (Wnt1 and Vav). Gene ontology analysis revealed that there are 14 common transcription factors, as presented in the next figure.
5. Discussion

Figure 5.3: There are 98 common genes between the Vav and Wnt1 microarray studies, 14 of them are transcription factors. Here we present the transcription factors with their fold change (FC).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Wnt1 FC</th>
<th>Vav FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elk3</td>
<td>-1.52362</td>
<td>2.84814</td>
</tr>
<tr>
<td>Hoxa7</td>
<td>1.76428</td>
<td>2.429319</td>
</tr>
<tr>
<td>Hoxa6</td>
<td>2.420047</td>
<td>2.962158</td>
</tr>
<tr>
<td>Pou2f2</td>
<td>1.929805</td>
<td>1.570596</td>
</tr>
<tr>
<td>Rora</td>
<td>-1.53287</td>
<td>1.996023</td>
</tr>
<tr>
<td>Fil1</td>
<td>-2.39522</td>
<td>2.242972</td>
</tr>
<tr>
<td>Hoxa5</td>
<td>2.142666</td>
<td>2.713867</td>
</tr>
<tr>
<td>Hoxa10</td>
<td>1.955481</td>
<td>1.610184</td>
</tr>
<tr>
<td>Gata2</td>
<td>-1.84126</td>
<td>1.804829</td>
</tr>
<tr>
<td>Tshz3</td>
<td>1.55957</td>
<td>1.791321</td>
</tr>
<tr>
<td>Meccm</td>
<td>-1.80108</td>
<td>6.248843</td>
</tr>
<tr>
<td>Notch1</td>
<td>1.581741</td>
<td>3.28368</td>
</tr>
<tr>
<td>Gata3</td>
<td>-2.37158</td>
<td>2.392006</td>
</tr>
<tr>
<td>Hoxa9</td>
<td>2.028755</td>
<td>4.774076</td>
</tr>
</tbody>
</table>

5.2.2 The role of Tgfbr2 in neural crest stem cells

The encoded protein of the Tgfbr2 gene is a transmembrane protein that has a protein kinase domain, forms a heterodimeric complex with another receptor protein and binds TGF-beta. This receptor/ligand complex phosphorylates proteins, which then enter the nucleus and regulate the transcription of a subset of genes related to cell proliferation. Mutations in this gene have been associated with Marfan Syndrome, Loeys-Deitz Aortic Aneurysm Syndrome, and the development of various types of tumors. [Iwata et al., 2014]

Iwata et al. investigated the role of Tgfbr2 in neural crest stem cells. They generated Tgfbr2 conditional knock out mice using Cre Recombinase under the control of Wnt1. They showed that mice with a deletion of Tgfbr2 in cranial neural crest cells (Tgfbr2fl/fl;Wnt1-Cre mice) develop cleft palate as the result of abnormal TGFb signaling activation. They propose that Tgfbr2 mutant palatal mesenchymal cells spontaneously accumulate lipid droplets, resulting from reduced lipolysis activity. Tgfbr2 mutant palatal mesenchymal cells failed to respond to the cell proliferation stimulator sonic hedgehog, derived from the palatal
5. Discussion

We have compared the data from our study to the data of Iwata et al. and found that we have 54 common genes. The following figure presents some of the common genes that are associated with craniofacial development.

![Figure 5.4](image)

Figure 5.4: There are 54 common genes between the our study and Iwata et al. study. Here we present some of the genes associated with craniofacial development with their fold change (FC).

### 5.3 Conclusion

Geminin is a molecule that through the interaction with certain transcription factors and chromatin remodeling proteins regulates cell proliferation and cell differentiation. [Bene et al., 2004] [Luo et al., 2007]

Our results indicate that Geminin is an essential molecule involved into the formation of neural crest stem cells. We should highlight that Geminin, a protein known to direct the balance between proliferation and differentiation is required to neural crest stem cells for their proper formation, migration and transition from epithelial to mesenchymal state. The Geminin knockout mice are embryonically lethal, they die a few days before birth. In addition they suffer from severe morphological and craniofacial malformations.
References


 genesis in ectomesenchyme and lateral plate mesoderm. *Dev Biol*, 199:11–25. 79


reveals insights into the inflammatory contribution of shed endometrium. *Am J Pathol.* 31


106


REFERENCES


Tsai, R. Y. and McKay, R. D. (2002). A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells. *Genes and Development*, 16:2991–3003. 2


Zhou, Fu, Huang, Shen, Xiao, Song, and Liu (2014). Increased cystic fibrosis transmembrane conductance regulators expression and decreased epithelial sodium channel alpha subunits expression in early abortion: Findings from a mouse model and clinical cases of abortion. PloS One. 31