Fakultät Angewandte Informatik
(Faculty Applied Computer Science)
Master-Studiengang

Thema (deutsch/englisch)

Identification of critical candidate genes during Melanoma development and progression based on RNA sequencing data.
Identifizierung kritischer Kandidatengene während der Entwicklung und Progression des Melanoms durch RNA-Sequenzierungsdaten.

Masterarbeit zur Erlangungdes akademischen Grades:

Master of Science
an der
Technischen Hochschule Deggendorf

vorgelegt von

Name,Vorname: El Belbesy, Mohammed

Matrikelnummer: 00801592

Ort, Datum: Deggendorf, 23.12.2022

Erstprüfer/in:

Prof. Dr. Melanie Kappelmann-Fenzl

Deggendorf Institute of Technology
Fakultät Angewandte Informatik
Faculty of Applied Computer Science

12/23/2022

Identification of critical candidate genes during Melanoma development and progression based on RNA sequencing data.

Identifizierung kritischer Kandidatengene während der Entwicklung und Progression des Melanoms durch RNA-Sequenzierungsdaten.

A thesis submitted for the Master's degree to the Faculty of Applied Computer Science at Deggendorf Institute of Technology

Under the supervision of

Prof. Dr. Melanie Kappelmann-Fenzl (PhD)
Deggendorf Institute of Technology
Faculty of Applied Computer Science

University of Erlangen (FAU)
Institute of Biochemistry

Dr.rer.nat Stefan M Fischer (PhD)
Faculty of Applied Computer Science
Deggendorf Institute of Technology

By

Mohammed Hassan El Belbesy
Matriculation number: 00801592

I hereby declare that I wrote this thesis myself with the help of no more than the mentioned literature and auxiliary means.

Deggendorf, 23.12.2022

……..

Mohammed El Belbesy

Acknowledgment

First and foremost, I want to express my profound gratitude to Dear Prof. Dr. Melanie Kappelmann-Fenzl, who supervised me during my thesis project. She gave me a unique thesis topic, and I appreciate all of her assistance, motivation, and vast expertise and knowledge. She has been a constant source of advice and patience, and I can't thank her enough for that.

I also would like to thank Dr.rer.nat Stefan Fischer for helping me through the progress of this thesis and agreeing to be there for me. I am so grateful to him for the skills and knowledge he taught me during the thesis.

Last but not least, I am so grateful to my family, my father, mother, wife, and brothers. I would like to thank them for their enduring support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

NHEMs	Normal Human Epidermal Melanocytes
RGP	Radial Growth Phase
VGP	Vertical Growth Phase
MET	Metastasis
PCA	Principal Component Analysis
ORA	Over Representation Analysis
GSEA	Gene Set Enrichment Analysis
GO	Gene Ontology
MSigDB	Molecular Signature Database
NCBI	The National Center for Biotechnology Information
NGS	Next-generation sequencing
RNA-Seq	RNA sequencing

Abstract

Melanoma is a lethal form of skin cancer. Over 90% of cancer patients die from metastatic disease, which happens when the cancer cells migrate from their primary sites to distant organs. The incidence of melanoma has rapidly grown recently, and the survival rate of melanoma patients is still poor.

Melanoma develops when melanocytes in the skin change, divide uncontrollably, and expand radially. After this radial development, vertical growth may occur, which leads to an invasion through the basement membrane into the dermis below and results in metastasis.

In this study, we aimed to identify key regulators and candidate genes responsible for the transition to the individual stages of melanoma development. Therefore, we determined differentially expressed genes (DEGs) comparing NHEMs and cells derived from different developmental steps of melanoma (RGP, VGP, and MET). DEGs were identified and analysed using the Deseq2 v1.34.0 package from Bioconductor. Moreover, Over Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) were performed to assign biological function to the determined stage dependent deregulated genes.

In total, we demonstrated 2308 up-regulated genes in RGP melanoma cells, 1903 up-regulated genes in the VGP melanoma cells, and 1793 up-regulated genes in MET. The upregulated DEGs, such as JAG1, JAG2, Col5A2, LUM, and POSTN were significantly enriched in the angiogenesis process in the MET cell line.

Aim of Work

This study aims to identify the candidate genes during melanoma development by performing a step-by-step analysis approach comparing RNA-Seq datasets of Normal human epidermal melanocytes (NHEMs), radial growth phase (RGP), vertical growth phase (VGP) and metastatic melanoma cells (MET). In-depth bioinformatic analyses will provide information on which candidate genes are involved in or even responsible for the transitions from NHEMs to RGP, RGP to VGP and VGP to MET melanoma cells within melanoma development. Further, functional analysis and clustering of the identified DEGs will give insights which biological processes lead to the changes in cell behavior associated to each developmental step of melanoma.

Identification of these candidate genes involved in the different developmental steps and their functional relevance will help to better understand the molecular mechanisms of melanoma development leading to a malignant phenotype and to develop early diagnostic tools in a melanoma- stage dependent manner.

Chapter 1: Introduction

1.1 Overview of Cancer

There are billions of cells in the human body that grow and divide as needed throughout one´s life. Cells typically die when they become abnormal or old; this process is called “apoptosis” (*InformedHealth.Org*, 2006). One of the reasons for developing cancer is when something goes wrong with this process, and the cells continue to divide, preventing the old or abnormal cells from apoptosis. As these cells proliferate uncontrolled, they may suffocate healthy cells. As a result, cancer may form, either malignant or benign (Cooper, 2000).

1.1.1 Cancer: Benign vs. Malignant

A tumor is an abnormal mass of cells in the body, also known as a neoplasm. When tumors develop a pathological character, they are also called cancer. There are two types of cancer: benign and malignant (Patel, 2020). Benign cancers do not invade or spread to the surrounding tissues as malignant ones. This migratory and invasive potential of malignant cancers to different body parts is called metastasis. Benign cancer typically do not recur after removal, whereas malignant ones can (The National Cancer Institute, 2021). However, benign cancers can become risky if they pressure vital tissues like blood, arteries, or nerves. Therefore, depending on the situation, they may or may not need treatment (Stuart, 2021). For instance, breathing difficulty may result from a massive benign lung cancer pressing the trachea. Consequently, immediate surgical removal would be necessary (Patel, 2020).

1.1.2 Types of Cancers

The term "cancer" refers to malignant tumors which can migrate and invade other tissues in the body. Cancers are classified based on the type of cell from which they develop into different classes carcinomas, sarcomas, and leukemias or lymphomas (Cooper, 2000). Cancers could be also classified according to the tissue of origin in two main groups. Cancers of the blood cells, such as leukemia, lymphoma, and multiple myeloma, are hematologic (blood) cancers. Cancers of other body organs or tissues are considered solid cancers (American Cancer Society, 2022). Breast, prostate, lung, and colorectal cancers are the most widespread solid cancers, responsible for more than half of all cancer cases (Cooper, 2000; Ferlay et al., 2021).

1.1.3 The development of Cancer

Cancer can arise from a variety of different factors. According to scientists, the combination of multiple factors results in cancer. The influencing factors may be a person's constitutional traits, environmental influences, or genetics(Stanford Medicine, n.d.). Radiation and chemical carcinogens are considered initiating agents damaging DNA and generating mutations. Some initiating agents contributing to

human cancers include solar ultraviolet radiation (the primary reason for skin cancer) and carcinogenic chemicals in tobacco smoke (Cooper, 2000).

Some human cancers are influenced by hormones, especially estrogens, which act as tumor promoters. For instance, estrogen stimulates the uterine endometrium's cells to proliferate, and excessive estrogen exposure raises a woman's risk of developing endometrial cancer (Cooper, 2000). Several viruses also can cause cancer in humans. For example, liver and cervical carcinoma, which account for 10 to 20% of cancer incidence globally, are two prevalent human diseases by viruses. Examples of DNA viruses that can lead to cancer are the human papillomavirus, hepatitis B virus, and Epstein-Barr virus. An example of RNA viruses that influence human cancer is hepatitis C (J. B. Liao, 2006).

One of the main reasons for developing cancer is a genetic disorder. The DNA carries the genes in the cell's nucleus (Weinberg, 1996). A gene is the coding region of the DNA that codes for various RNA transcripts, which may have regulatory functions or code for proteins. Proteins are a sequence of amino acids and fulfill almost all the body's structural, regulatory, and signal-transducing functions (Kappelmann-Fenzl, 2021). Thus, mutations in a gene can disturb the behavior or function of a cell by altering a protein´s function, amount or activity (Weinberg, 1996).

Three main categories of genes are crucial in initiating cancer: proto-oncogenes, tumor suppressor genes, and DNA repair genes(Weinberg, 1996). Proto-oncogenes play a crucial role in cell division and growth. Mutations in these genes lead to the expression of cancer promoting oncogenes (Heidi Chial, 2008). Conversely, tumor suppressor genes regulate cell development and proliferation to prevent from cancer progression. Specific tumor suppressor gene mutations can cause excessive cell proliferation. Moreover, mutations in DNA repair genes frequently cause chromosome changes like duplications and deletions of specific chromosome segments. Cells that accumulate such mutations have a high potential to develop into cancer cells (The National Cancer Institute, 2021; Weinberg, 1996).

1.2 Human skin

The human skin is the body's largest organ, with many different functions. For example, the skin protects from ultraviolet light and the invasion of pathogens. Additionally, it restricts water loss and controls body temperature through blood flow and sweat evaporation (Igarashi et al., 2007). The skin is made up of three main layers; The epidermis, an outer layer of the stratified epithelium; the dermis, an interior layer that is less cellular; and the hypodermis, a subcutaneous layer mainly made of adipose tissue (Yanez et al., 2017). The basal lamina, also known as the basement membrane, separates the two skin layers, epidermis and dermis, primarily serving as a dynamic interface and a diffusion barrier (Breitkreutz et al., 2013).

The epidermis is mainly composed of keratinocytes formed by the division of cells in the basal layer. In addition to keratinocytes, The epidermis also contains several different cell types, including Merkel

cells, Langerhans cells, which have immunological activities, and melanocytes, which give the keratinocytes pigment (Burns et al., 2004). From the outside to the interior, the epidermis is further separated into the stratum corneum (horny layer), stratum granulosum (granular layer), stratum spinosum (prickle cell layer), and stratum basale (basal layer, also called the stratum germinativum). The Malpighian layer is made up of the stratum basale and stratum spinosum (Ng & Lau, 2015).

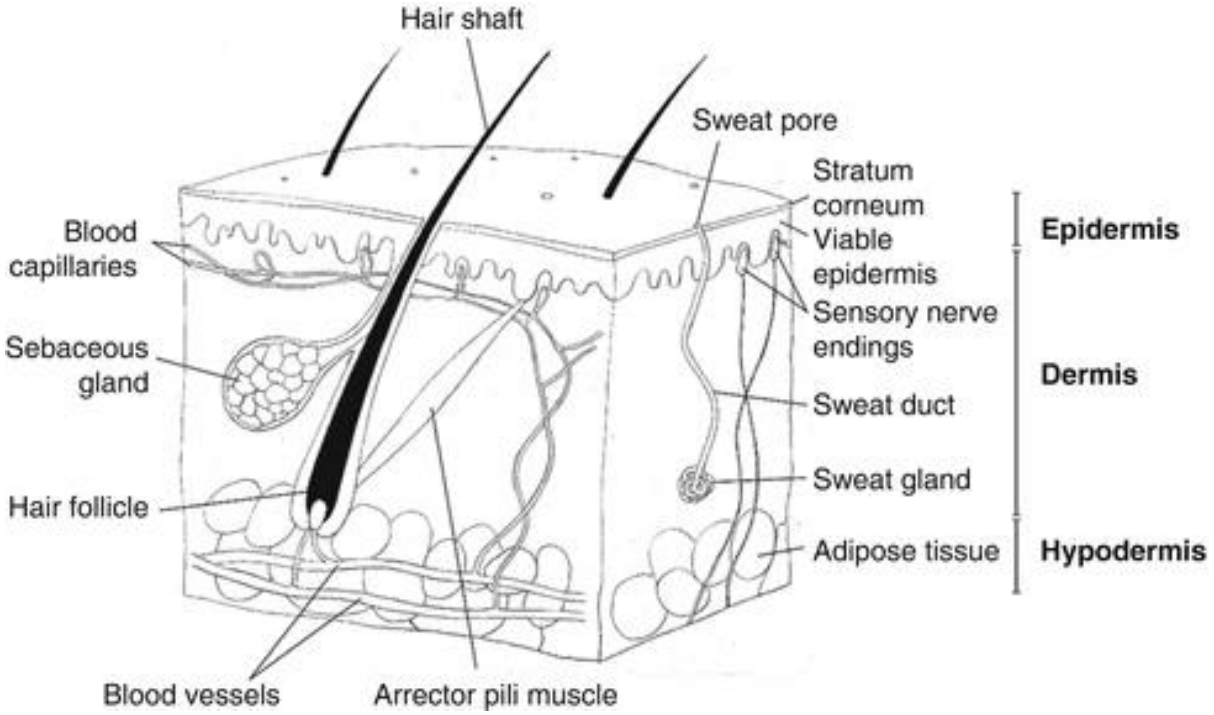


Figure 1: A diagram shows the composition of human skin. *The stratum corneum and the viable epidermis make up the epidermis* (Ng & Lau, 2015).

1.2.1 Melanocytes

In the human body, melanocytes are a diverse set of cells arising from embryonic cells known as neural crest cells (NCC). They are responsible for melanin synthesis (Plonka et al., 2009). Melanocytes are found not only in the epidermis but also in hair, iris, inner ear, nervous system, and heart (Cichorek et al., 2013). Melanocytes maintain melanin, one of the skin's primary light-absorbing pigments. Melanosomes are specialized organelles found in melanocytes, which become active and produce melanin when the skin is exposed to sunlight (Igarashi et al., 2007).

The primary role of melanocytes is to produce melanin. Tyrosinase, Tyrosinase-Related Protein-1 (TYRP1), and Tyrosinase-Related Protein 2/Dopachrome Tautomerase are the three enzymes that work together to create the pigments inside melanosomes. Typically, two types of pigments are produced: the orange/yellow pigment pheomelanin, which has weak photoprotective characteristics, and the brown/black pigment eumelanin, which exhibits photoprotective properties (Cui et al., 2007).

In the human skin, pigments play a crucial role in protection from the damaging effects of ultraviolet radiation (UVR). Melanin diminishes ultraviolet radiation induced cellular DNA damage and genomic

instability by absorbing and scattering UV radiation (Cui et al., 2007). The main mediator of this reaction is the tumor-suppressor p53 activation in keratinocytes, which, in response to ultraviolet radiation, induces the transcription of proopiomelanocortin (POMC), the precursor to hormones with pro-pigmenting properties like α-melanocyte stimulating hormone (αMSH) and adrenocorticotropic hormone (ACTH) (Cui et al., 2007).

Melanocytes develop from embryonic neural crest cells (melanoblasts). Melanoblasts will then move and develop and differentiate into melanocytes. After that, mature melanocytes are transported to keratinocytes, and eventually, cell death. Melanocyte embryonic development allows us to learn more about skin diseases like melanoma (Cichorek et al., 2013).

1.3 Skin caner

Skin cancer is one of the most common types of cancer among humans (Dildar et al., 2021). The prevalence of skin cancer is increasing due to two major risk factor groups: biological and non-biological (Carr et al., 2020).

1.3.1 Non-Biological risk factors

The skin protects against non-biological risk factors like environmental stressors such as noise pollution, night-time artificial lighting, and exposure to air pollutants (Flies et al., 2019). As a result, these pollutants raise the risk of cutaneous disorders, including skin cancer (Parrado et al., 2019).

Nonetheless, frequent ultraviolet exposure from sunlight is the main factor for the onset and spread of skin cancer globally. Various molecular processes, such as the activation of the p53 pathways, increased DNA damage, inflammatory reactions, genetic alterations, oxidative stress, immunosuppression, and induction of the apoptotic pathway, significantly alter cell physiology to cause cell death cycle arrest. In addition, reactive oxygen species (ROS), which interact with lipid and protein molecules to produce intermediates that combine with DNA to form adducts and damage DNA, are created during exposure to ultraviolet radiation (Khan et al., 2021).

1.3.2 Biological risk factors

Biological risk factors of skin cancer can influence protein synthesis, which may have harmful effects on the proliferation of skin cells and ultimately leads to melanoma or non-melanoma skin cancer (S. Wu et al., 2016). Melanoma and non-melanoma skin cancers usually have imbalanced signaling pathways related to the regulation of gene expression. For example, the dysregulation of the PTCH1 gene mutation causes unregulated skin cell proliferation and the emergence of many Basal Cell Carcinomas. Similar to this, in men, CDKN2A gene mutations are the most frequently found cause, whereas, in women, MDM2 gene alterations are predisposed to melanoma development at an earlier age (Khan et al., 2021).

Deviations from the original function of skin cells lead to activation of oncogenes or suppression of the tumor suppressor genes. This process is associated with some factors such as growth factor independence, resistance to growth inhibitors, continuous angiogenesis, the unlimited ability for replication, metastasis, and tissue invasion (Khan et al., 2021).

In a molecular context, G protein-coupled melanocortin-1 receptors (MCIRs) in the membrane of melanocytes are important elements of melanocyte physiology and are well identified in the major melanoma risk factor: UV-induced tanning pathways. In response to the external signal of melanocyte-stimulating hormone, MCIR controls melanocyte growth (a-MSH). Any UV-induced damage to melanocytes sets off a series of molecular events, including P53 stabilization and transcriptional activation of pro-opiomelanocortin, which develop into a number of signaling molecules and activate the synthesis of melanocyte-stimulating hormone (a-MSH) (Khan et al., 2021).

Through MCIR, a-MSH production starts another cycle of melanocyte proliferation. Cyclic adenosine monophosphate (cAMP) levels rise as a result of MCIR activation, and CREB regulated transcript activator (CRTC) controls the expression of the transcriptional activation factor microphthalmia (MITF). The second important factor that significantly raises the risk of melanoma is MCIR polymorphism. Patients with MCIR variations show reduced pheomelanin UV protection and have a lower burden of UV signature mutations (Khan et al., 2021).

1.3.3 Types of skin cancer

Skin cancer is typically divided into two main types: melanoma and non-melanoma skin cancer (Silpa & Chidvila, 2013). Nearly 95% of skin cancers are non-melanoma skin cancers, which are brought on by genetic and environmental causes. The majority of non-melanoma skin cancers are Squamous Cell Carcinomas and Basal Cell Carcinomas, which together account for 99% of all non-melanoma skin cancer (Khan et al., 2021). Basal Cell Carcinoma often arises from the basal layer of the epidermis. It is the most prevalent type of skin cancer, which commonly occurs in the head and neck, trunk, and extremities (Silpa & Chidvila, 2013). Squamous Cell Carcinoma is the second most prevalent cancer in the United States, frequently found in black and Asian Indians. It usually occurs in the sun-exposed regions of the head and neck (Silpa & Chidvila, 2013).

A dangerous and fatal form of skin cancer is melanoma. According to the American Cancer Society records, melanoma skin cancer accounts for the high mortality rate among skin cancers (Dildar et al., 2021).

1.4 Melanoma

Melanoma is a deadly form of skin cancer that makes up 4-5% of all cancer cases, and metastatic melanoma is responsible for 80% of skin cancer fatalities. Over 90% of cancer patients die from

metastatic disease, which is the spread of cancer cells from their primary sites to distant organs (Qiu et al., 2015).

Patients with metastatic melanoma have a high mortality rate because most tumors are resistant to radiotherapy and chemotherapy, even though early-stage melanomas that have not spread to the lymph nodes can be removed with little risk of recurrence (Uong & Zon, 2010). Melanoma occurrence has increased dramatically in recent years, and the survival rate of patients with melanoma remains low (J. Chen et al., 2019).

Melanoma starts when melanocytes undergo a transformation, divide abnormally and grow radially in the skin. This radial development may subsequently be followed by vertical growth, which could result in an invasion through the basement membrane into the dermis underneath and subsequent metastasis (Uong & Zon, 2010). Melanoma manifests as an atypical plaque during the radial growth phase; cells may invade the dermis but do not produce nodules. However, the lesion expands vertically during the vertical growth phase, transforming into a metastatic tumor (Elder et al., 1984).

Many genes have previously been discovered and are anticipated to be potential targets for melanoma metastasis prevention: PGC1A, which is known to encode PGC1α, a metabolic transcriptional coactivator that prevents oxidative stress and inhibits the metastasis of melanoma (C. Luo et al., 2016), A5 integrin which is upregulated in melanoma metastasis by BRIC5-encoded surviving in an Akt-dependent manner (McKenzie et al., 2013) and KISS1 overexpression decreasing the metastatic potential of melanoma cells, to name just a few examples (J.-H. Lee et al., 1996).

1.4.1 Melanoma Therapy

The primary treatment for localized melanoma is surgical excision of the cancerous cells and any normal tissues surrounding them. Patients with lesions larger than 0.8 mm in diameter or smaller but ulcerated undergo a biopsy from the sentinel lymph node (C. Lee et al., 2013). Unfortunately, melanoma tends to penetrate and metastasize, and as a result, the prognosis worsens the more profound the disease goes under the skin. Surgery cannot entirely eradicate metastatic melanoma, and metastatic cells are highly resistant to various therapeutical option. The median survival time for patients with metastatic melanoma typically falls between 6 and 10 months (Bertolotto, 2013).

Alkylating drugs like dacarbazine (Deticene), and fotemustine (Muphoran) temozolomide (Temodal), are used in traditional chemotherapy because they cause cytotoxic effects by preventing cell proliferation. However, only 10% of objective responses are promoted by these chemotherapeutic medications, and overall survival does not increase (Tsao et al., 2004).

Precision medicine, also known as personalized medicine, is a medical technique that divides individuals with the same diagnosis into distinct groups and then designs interventions, therapies, or

other medical decisions, particularly for each patient, depending on that patient's expected response or risk of disease (Moscow et al., 2018).

Personalized therapy is now being developed due to the discovery of the molecular changes in melanoma and the discovery of intracellular signaling pathways and interactions between the tumor immune microenvironment (Bertolotto, 2013; Palumbo et al., 2016). Currently, two categories of cutaneous melanomas are mutually distinctive: those with an activating BRAF mutation (mostly BRAF V600E), which accounts for 40–50% of all melanoma patients, and those with mutations other than BRAF (Davies et al., 2002). The discovery of the RAS-RAF-MEK-ERK (MAP kinase) signaling pathway and its targeting has been significant progress for the latest treatments for stage III and IV melanoma (Funck-Brentano et al., 2021).

To address the molecular flaws that are present in melanoma, numerous targeted therapies have been developed (Rebecca et al., 2012) a list of potential targets and therapy is shown in table 1.

Table 1: Established and potential targets for melanoma treatment and their activity level. *Therapeutic approaches seek to diminish the activation of the signaling pathway, its transduction, or its ultimate transcriptional effects (Helgadottir, H., Rocha Trocoli Drakensjö, I., & Girnita, A., 2018).*

Type of target	Examples of drugs/agents	Comment
Established effective targets on plasma membrane		
c-kit	Imatinib (Glivec, Imatinib)	In mucosal melanoma
IGF-1	Linsitinib	Phase I i association with Erlotinib
Epidermal growth factor	Gefitinib (Iressa), Erlotinib (Tarceva)	Approved for lung cancer, studied on melanoma, both cutaneous and uveal
Potential effective targets on plasma membrane		
GNAQ/GNA11	PKC inhibitor AEB071 (sotrastaurin)	In uveal melanoma
Established effective targets within signaling transduction pathways		
BRAF	Vemurafenib, dabrafenid, encorafenib	In skin, melanoma binds to and inhibits activated BRAF
MEK	Trametinib, cobimetinib, binimetinib	Often associated with BRAF inhibitors to overcome acquired resistance
Potential effective targets within signaling transduction pathways		
NRAS	Farnesyltransferaze inhibitors (R115777)	Most frequently mutated at hotspots in exon 1 (codon 12) and exon 2 (codon 61), which results in the prolongation of its active GTP-bound state
PI3K	Pictilisib	In melanomas with *PTEN* aberrations
ALK	Crizotinib	In uveal and spitzoid melanoma

CDK4/6	Abemaciclib, palbociclib	In melanomas with *CDKN2A* aberrations
Established effective nuclear targets		
None described so far		
Potential effective nuclear targets		
MITF	CH5552074	Inhibition of cell growth by reducing the expression level of MITF protein
TERT	*In vitro* test with Telomerase inhibitor IX	Acral and cutaneous melanoma
BAP1	*In vitro* with ubiquitin vinyl sulfone (Ub-VS)	In uveal melanoma
Histone deacetylases	Entinostat	In association with pembrolizumab in melanoma
Established effective immune targets		
CTLA-4	Ipilimumab	T-cell activator and blocks B7-1 and B7-2 T-cell co-stimulatory pathways
PD-1	Pembrolizumab, nivolumab	Binds to PD-1 and as such activates T-cell-mediated immune responses
PDL-1	Atezolizumab	
IDO	Epacadostat	IDO
Potential effective immune targets		
SD-101		*Via* the toll-like receptor 9
OX40		Co-stimulatory molecule that can be expressed by activated immune cells
CD137		Member of the TNFR super family
GITR		Glucocorticoid induced TNF receptor

Despite the significant improvement in treatment over the last couple of years, metastatic melanoma remains a significant clinical challenge. Consequently, there is a need for further research on disease etiology and pathogenesis, leading to the identification and validation of drug targets and biomarkers. Such efforts will hopefully improve preventive measures, early diagnosis, and personalized treatments (Moscow et al., 2018).

1.5 Molecular diagnostics in clinical oncology

Molecular diagnostics relies on the detection of individual molecules. Oncohematologists first identified the potential of molecular genetic techniques since specific chromosomal translocations may greatly benefit the diagnosis of various leukemias and lymphomas (Fey & Wainscoat, 1988). Clinical DNA testing has significantly advanced with the development of PCR (polymerase chain reaction) (Sokolenko & Imyanitov, 2018). PCR is a laboratory method to amplify specific DNA segments for

various laboratory and clinical applications (Ghannam & Varacallo, 2018). There are two routes where molecular tests have become a part of standard patient management in oncology (Figure 2). First, identifying subjects with hereditary cancers in a daily clinical practice and secondly molecular diagnostics choosing the best therapy based on the molecular biomarkers of the tumor tissues (Sokolenko & Imyanitov, 2018).

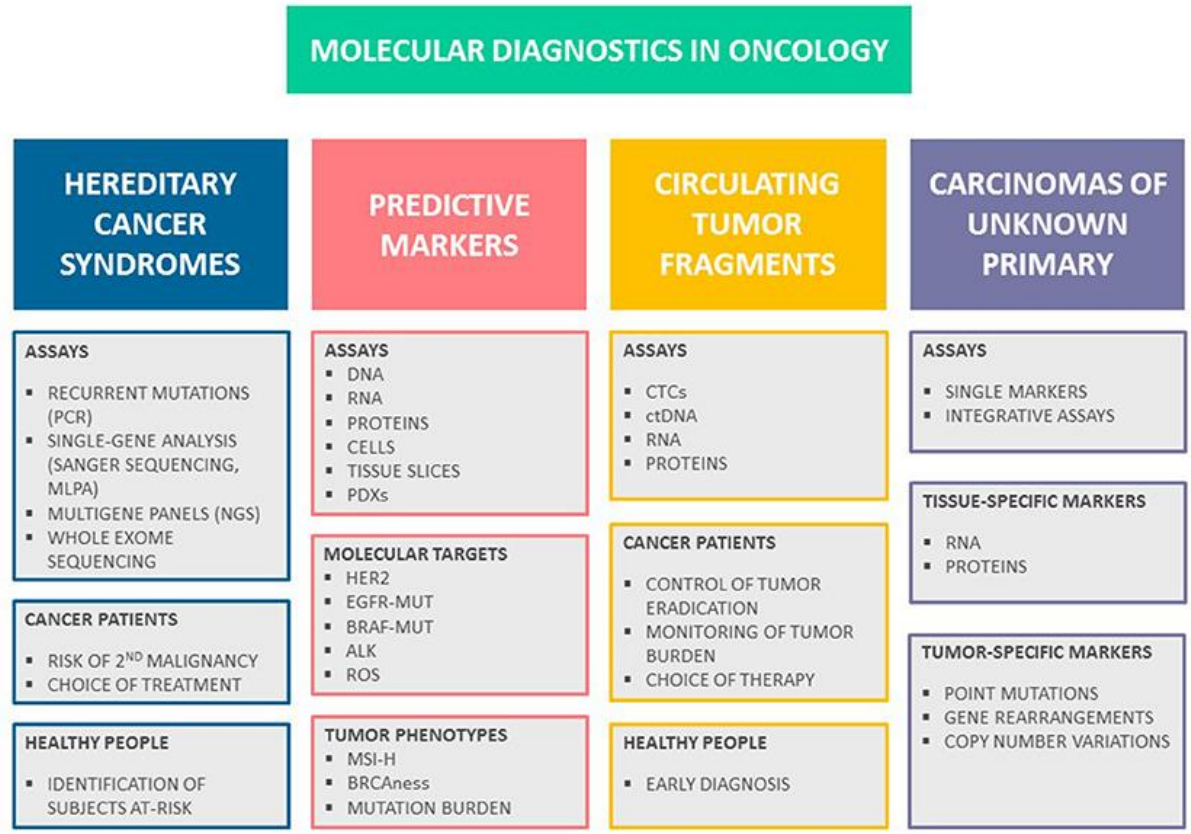


Figure 2: Molecular diagnostics in oncology. *There are several significant routes in cancer medicine that employ molecular-based assays. Testing for hereditary cancer syndromes is routinely used to identify at-risk persons and personalize treatment. Several predictive tests involve either the analysis of individual drug targets or the identification of specific tumor phenotypes, which aid the choice of anticancer drugs. Recent mutation testing and RNA analysis developments offer novel tools for diagnosing cancers of the unknown primary site* (Sokolenko & Imyanitov, 2018).

The fight against cancer continues to be a major concern despite the significant achievements and efforts made in research so far. A key goal of cancer research is to understand the complexity of cancer by using advanced technologies and computing techniques (Garland, 2017).

1.5.1 High-throughput sequencing

In 1953 Watson and Crick, the double-helix, described the shape of the nucleic acid. This enables scientists to understand at the molecular level that gene interactions are the core of life (Watson & Crick, 1953). Sanger sequencing refers to the first-generation sequencing technology. Sanger invented the chain termination method first in 1977 (Sanger et al., 1977). Then Maxam and Gilbert created the chemical degradation approach (Maxam & Gilbert, 1977). Since its debut, the DNA microarray has

significantly facilitated development in a variety of sectors (Russo et al., 2003). Several next-generation sequencing systems have been developed since the first high-throughput sequencing platform debuted in 2005 (Margulies et al., 2005) (Figure 2). Third-generation sequencing, enables the long-read sequencing of molecules (Schadt et al., 2011).

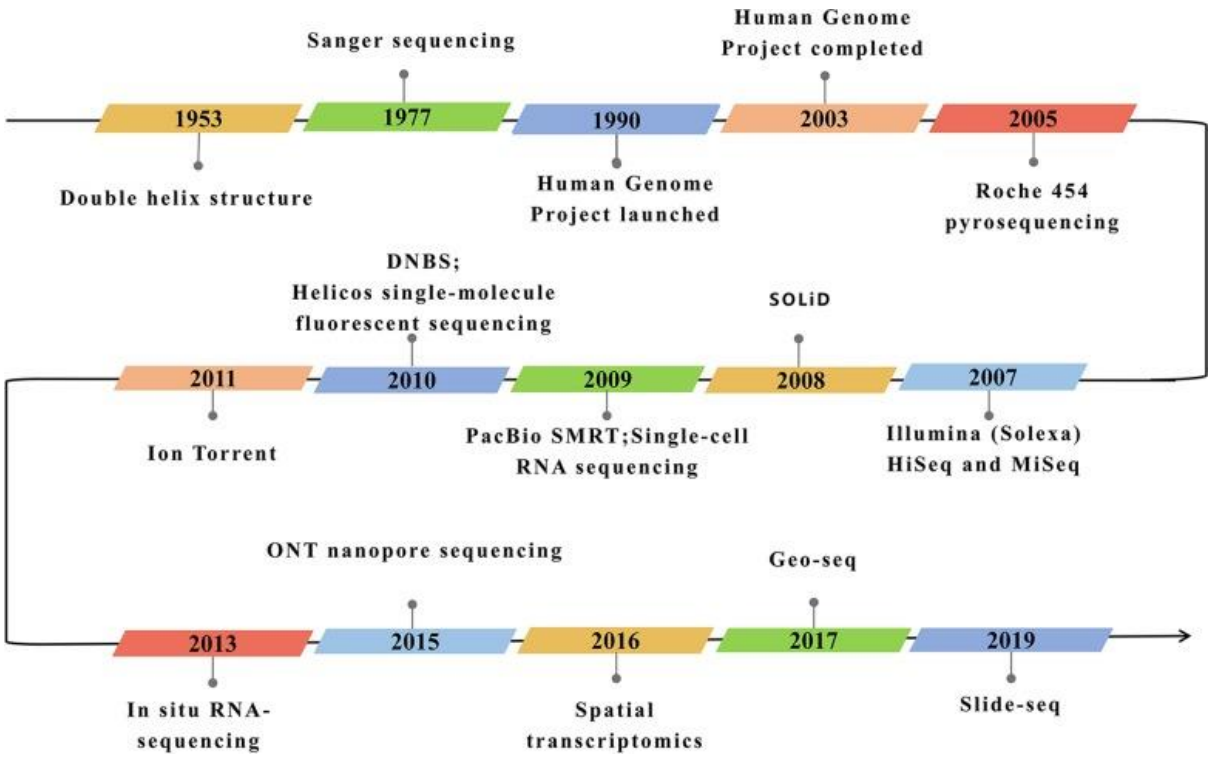


Figure 3: The development timeline of RNA sequencing technologies (Hong et al., 2020). *In 1977, Sanger developed the chain termination technique, regarded as the first generation of sequencing. Maxam and Gilbert then developed the chemical degradation strategy. Since the introduction of the first high-throughput sequencing platform in 2005, numerous next-generation sequencing systems have been created. Third-generation sequencing is an alternative to NGS that allows for long-read sequencing of individual RNA molecules.*

High-throughput sequencing techniques are nowadays routinely used, for example, to detect gene expression levels, variants, gene fusions, and copy number variations. of the human genome and transcriptome and, thus, a promising tool for predicting prospective melanoma treatment targets (Mery et al., 2019). The plenty of the sequence data derived from the diverse organisms' genomes has provoked a significant increase in research and the innovation of computer-based analytical tools (Rehm, 2001).

Omics technologies are high-throughput biochemical assays simultaneously measuring similar-type molecules from a biological sample. For instance, genomics profiles DNA, transcriptomics quantifies RNA transcripts, proteomics quantifies proteins, and metabolomics quantifies metabolites (Conesa & Beck, 2019). Access to omics data has greatly advanced thanks to advancements in high-throughput DNA sequencing, computation, and algorithms. NGS technologies enable the opportunity to develop the ideal cancer drug for the ideal patient by

accurately analyzing the pan-genomic profiles of tumors at the genomic and transcriptomic levels (Buescher & Driggers, 2016).

1.5.2 Whole genome, whole exome and RNA-Sequencing

A human´s entire genome can be sequenced by whole-genome sequencing, whereas whole-exome sequencing can be used to sequence the coding regions of the genome. This method seeks to evaluate the entire panel of genes related to cancer. Another strategy involves sequencing specific regions of chosen genes, concentrating on cancer gene "hotspot" regions with recurrent mutations. The common objective is still to be able to perform almost any kind of analysis in order to find potential therapeutic targets at the genomic and transcriptomic levels that can be used to categorize tumors and forecast outcomes (Ulahannan et al., 2013). RNA sequencing makes use of high-throughput sequencing techniques to reveal information about a cell's transcriptome. RNA sequencing offers significantly higher coverage and better resolution of the dynamic nature of the transcriptome when compared to earlier Sanger sequencing- and microarray-based approaches (Kukurba & Montgomery, 2015).

The most common application of RNA-Seq is to analyze differences in gene expression (DGE). The typical procedure starts with RNA extraction in the lab, then moves on to mRNA enrichment and ribosomal RNA depletion, cDNA synthesis, and the creation of an adaptor-ligated sequencing library. The library is then sequenced on a high-throughput machine with a read depth of 10–30 million reads per sample (usually Illumina) (Stark et al., 2019). Then, the computational process of analyzing RNA sequencing data includes the quality control of the raw data, read alignment and transcript assembly, measurement of expression, and analysis of differential expression (Hong et al., 2020).

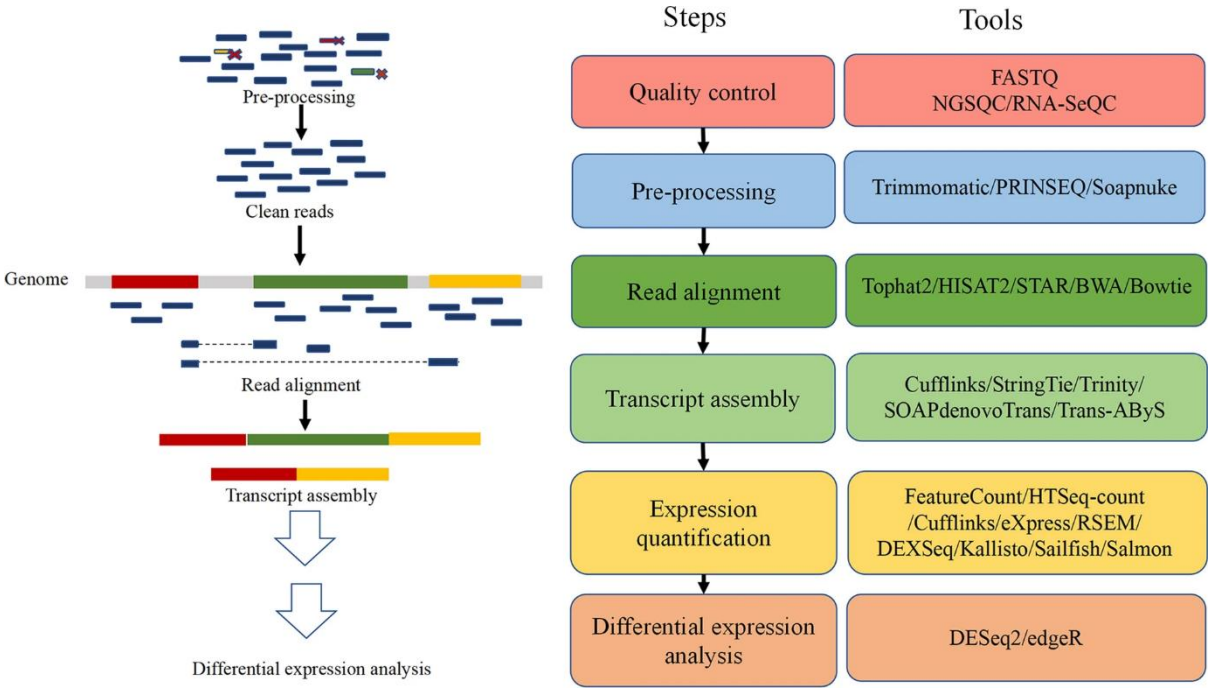


Figure 4: Bioinformatics tools commonly used in RNA-Seq data analysis. *RNA sequencing provides essential information for the study and therapy of cancer. It will be extensively utilized for research on numerous different types of cancer with the onset of the era of precision medicine. Single cell sequencing and RNA sequencing give biological data on tumor cells, study factors influencing intratumor expression heterogeneity, and pinpoint the molecular causes of the emergence of various oncological illnesses* (Hong et al., 2020).

Chapter 2: Materials and Methods

This chapter describes the methodology and material that were used for RNA-Seq data preprocessing and in depth bioinformatical analysis. Moreover, functional analysis techniques and tools are illustrated in depth. Figure 5 shows the workflow of the performed NGS data analysis (Figure 5).

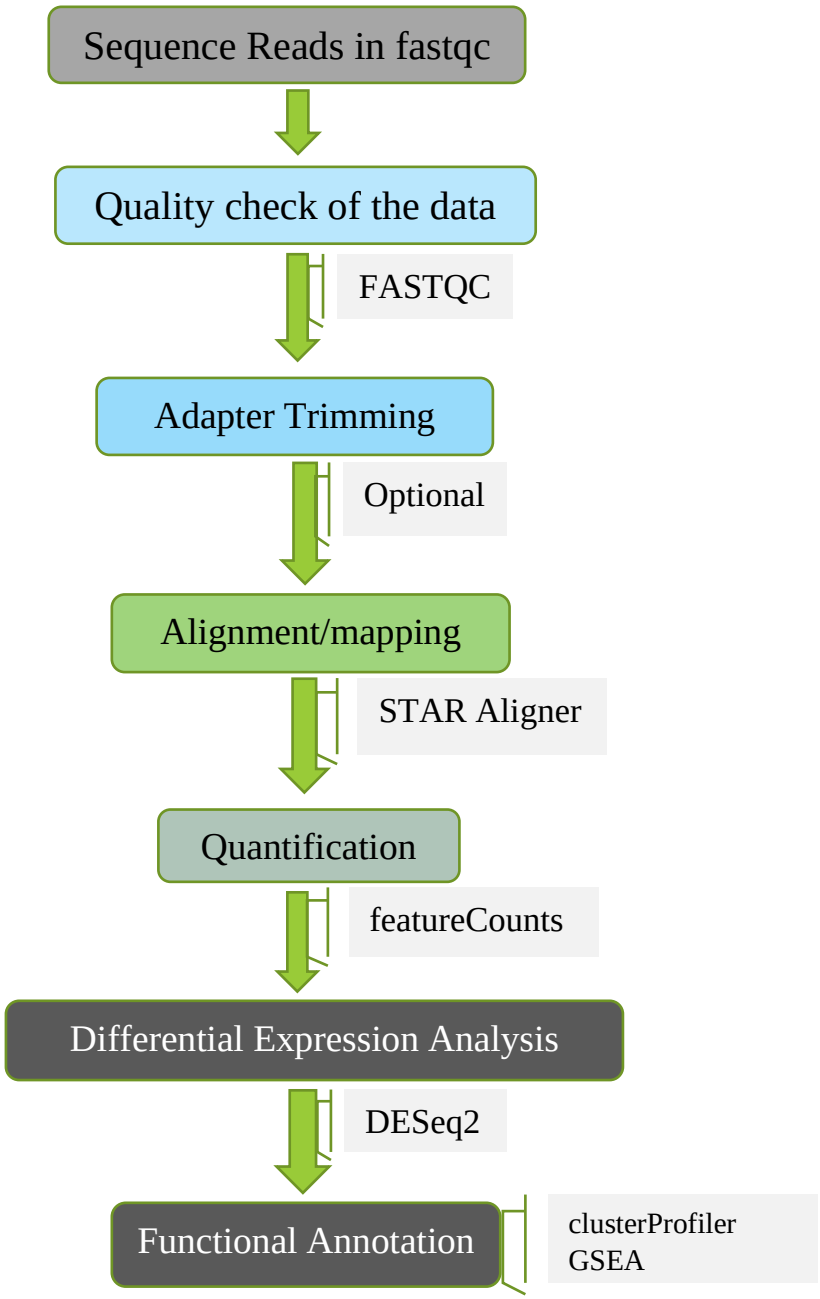


Figure 5: Pipeline illustrating the methods and the bioinformatical tools used in this thesis study. *RNA-Seq analysis was performed on normal and developmental melanoma cell lines. The analysis begins with a quality check using the fastqc tool, followed by alignment, then quantification. Next, the downstream analysis was performed by the deseq2 package from the Bioconductor in R, followed by functional annotation analysis.*

2.1 RNA-Sequencing data

Human melanoma cell lines Sbcl-2, WM3211, WM1366, WM793, WM1158, and WM9 (which were obtained from Dr. M. Herlyn, Wistar Institute, Philadelphia, USA) represent the radial growth phase RGP (Sbcl-2), vertical growth phase VGP (WM3211, WM1366, WM793), and melanoma MET (WM1158, WM9). At 37 °C and 5% CO2, the cell lines were kept in a culture medium made up of MCDB153 (Sigma-Aldrich, Steinheim, Germany) with 20% Leibovitz's L-15 (PAA Laboratories, Coelbe, Germany), 2% FCS, 1.68mM CaCl2 (Sigma), and 5 µg/ml insulin (Sigma- Aldrich, Steinheim, Germany). NHEMs represent the normal cells, were cultured in melanocyte growth media M2 at 37 °C and 5% CO2, and were obtained from neonatal foreskin (NHEMs, PromoCell, Heidelberg, Germany) (Kappelmann-Fenzl et al., 2019).

The RNA-Seq dataset used in this study is a standard Illumina data set (Paired-end) obtained from Prof. Dr. Melanie Kappelmann-Fenzl. The RNA-Seq dataset was shared through the sFTP server from the supercomputer cluster server of the Deggendorf Institute of Technology. In addition, the data is publically available in the NCBI database (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA839865).

All experimental working steps were performed at the University of Erlangen (FAU) at the Institute of Biochemistry previous to the data analysis performed in this study. The bioinformatical analyses are based on already published sequencing data of the aforementioned cell lines.

The preparation of the sequencing library involved at least two biological replicates. RNA-Seq was carried out using Illumina HiSeq2000 with the paired-end module, following Illumina's paired-end RNA-Sequencing (RNA-Seq) protocols (Illumina, Inc.)(Kappelmann-Fenzl et al., 2019).

The RNA-Seq dataset was shared through the sFTP server from the supercomputer cluster server of the Deggendorf Institute of Technology, each file in the FASTQ file format. The used data are publicly available on the NCBI database (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA839865).

2.2 Quality check of the data

The sequence and quality scores are displayed in the .FASTQ file format as a single ASCII character. Each sequence or read is represented by four lines arranged on top of one another in the text files produced by sequencing workflows (Akalin, 2020). The PHRED quality score is the probability of the log-transformed error for a base call, with high values denoting low error probabilities and vice versa (BG & Green, 1998).

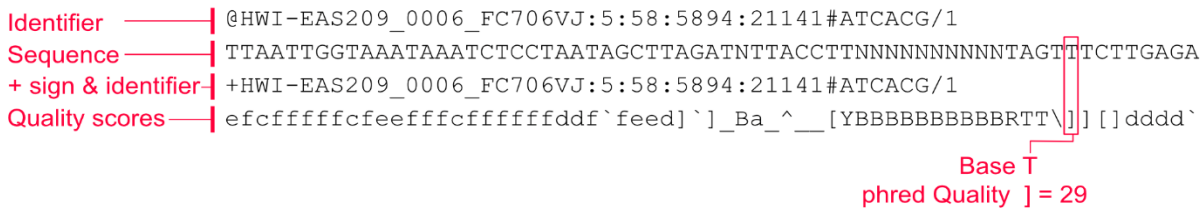


Figure 6 : FASTQ format and a short description for each line in the file format (Akalin, 2020)**.** *The first line starts with '@' character, followed by a sequence identifier and an optional description. Sequencing technology utilizes this line and usually contains specific information for the technology, like flow cell IDs, lane numbers, and information on reading pairs. The second line is the sequence of letters. The third line begins with a '+' character; it denotes the end of the sequence. The fourth line contains the quality values for the sequence in Line 2, where each letter corresponds to a quality score. These scores represent the likelihood of the base being called wrong. Qphred = −10log10 e, where e is the probability that the base is called incorrect. Since the score is in the minus log scale, the higher the score, the more unlikely that the base is called wrong* (Akalin, 2020).

In this study, the raw reads (14 FASTQ files) were quality-checked using FastQC tool (v0.11.9) from Babraham Bioinformatics institute (Andrews, 2014) to ensure high read quality. The quality reports produced by the FASTQC program assess the per base and per tile sequence quality, per sequence GC content, sequence length distribution, sequence duplication level, overrepresented sequences, and adapter content (Kappelmann-Fenzl, 2021).

2.3 Mapping

Paired-end reads were mapped to the reference human genome (GRCh38/ gencode v29.0) using the STAR Aligner tool (v2.7.10a) (Dobin et al., 2012)to determine from where the sequence reads originated in the human genome (Akalin, 2020).

There are two steps in the STAR workflow: First, generating a genome index by using the reference genome sequence (FASTA file of GRCh38/ gencode v29.0) and annotation GTF file (gencode.v29.annotation.gtf). Next, the sequencing reads were mapped to the human genome, and the STAR aligner generates a list of mapped files in BAM format.

2.4 Quantification

In order to perform differential expression (DE) analysis, it is crucial to generate a matrix containing the counts of RNA-Seq fragments for each sample. For this reason, the mapped reads were counted to generate the count table using the featureCounts tool (v2.0.0) (Y. Liao et al., 2014). featureCounts can utilize single or paired-end reads. In this study, because of the use of a paired-end sequencing approach, each read specifies a RNA fragment that is prefaced by the two reads, so featureCounts will count fragments instead of counting the reads.

2.5 Differential Expression Analysis

After count table generation the matrix was used to perform differential expression analysis using appropriate statical methods.

Differential expression analysis aims to identify whether the differences in gene expression (read counts) between defined conditions are significant (McDermaid et al., 2019a). The differential expression analysis is executed in R (v4.1.3) using R packages developed for the statistical analyses required to determine the differentially expressed genes between conditions.

Deseq2 (v1.34.0), a Bioconductor package, is a widely used method for differential expression analysis. The DESeq2 package offers statical techniques for testing for differential expression based on negative binomial generalized linear models; estimates of dispersion and logarithmic fold changes include initial distributions based on data (Love et al., 2014). Analysis steps with DESeq2 are shown in Figure 7.

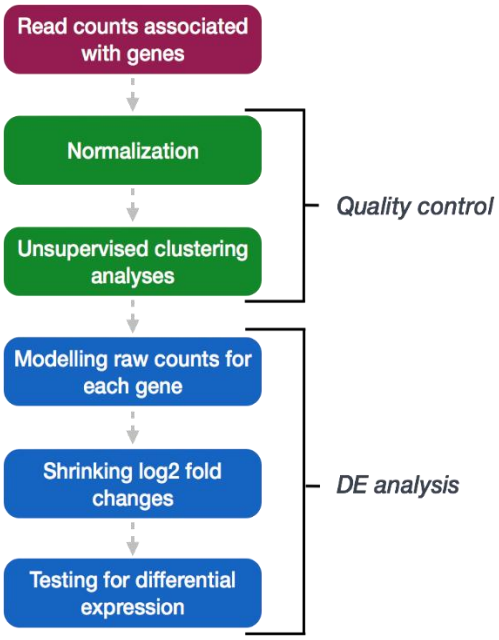


Figure 7: DESeq2 analysis workflow. *Before analyzing the count data, DESeq2 normalizes the data to account for variations in sample-to-sample RNA composition and library size. The normalized counts will then be used to create quality control (QC) charts at the gene and sample levels. Finally, the differential expression analysis is carried out using the necessary DESeq2 package algorithms (HBC, 2017)* .

In order to increase the stability and interpretability of estimates, DESeq2 employs shrinkage estimation for dispersions and fold changes which is important especially for lowly expressed genes because they have high variability and shrinkage can decrease it. This makes it possible to conduct a more quantitative study that concentrates on the degree of differential expression rather than its existence (Love et al., 2014). Furthermore, through modelling the dependence of the dispersion on the average

expression strength across all samples, the DESeq technique finds and corrects dispersion estimates that are extremely low (Anders & Huber, 2010).

2.5.1 Quality assessment of the samples

In this study, the experimental design included 14 samples representing four cell lines derived from whether NHEMs, RGP, VGP, or MET melanoma. Performing quality control (QC) checks on the count data is essential in the DESeq2 workflow to ensure the quality of the samples. To measure the quality of our experiment, we can look at how the samples relate in terms of gene expression. Visualization approaches for unsupervised clustering analysis, such as hierarchical heatmaps and principal component analysis PCA, are used to do this. These QC approaches are used to determine how comparable the biological replicates are to one another, as well as to detect outlier samples and main sources of variation in the data set.

2.5.1.1 Data visualization

In order to assess the similarity of gene expression between different samples in a dataset, hierarchical clustering with heatmaps is used. The heatmap is made by combining the gene expression correlation values for all pairwise combinations of samples in the data set, with 1 being the perfect correlation. The heatmap's colors reflect the correlation values, while the hierarchical tree shows which samples are more similar to one another. The biological replicates should be grouped together, whereas the sample conditions should be separated. Because the majority of genes should not be differentially expressed, samples should have a high correlation.

To create the heatmap, the pheatmap (Kolde & Kolde, 2018) package was used after generating the correlation values. The annotation arguments determine which metadata factors should be used as annotation bars. To select the condition column in colData1, we use the select() function from the dplyr package (v1.0.9). The heat map's output shows that the biological replicates are clustered together, but the conditions not, which means the experimental condition is the major source of variation.

2.5.1.2 Principal component analysis (PCA)

In order to continue evaluating the quality of the samples, principal component analysis PCA is used to see how the samples cluster and whether the condition of interest corresponds to the principal components explaining the variation in the data. PCA (Hotelling, 1936) is a technique for emphasizing the variation in a dataset. The first principal component, PCA1, represents the data's most significant amount of variance. PC2, the dataset's second most variation, must be perpendicular to PC1 to best describe the variance in the dataset. The number of principal components in the dataset equals the number of samples. PC1 means plotting a line through n-dimensional space to find the most significant amount of variation.

The principal component with the most variant genes has the most significant influence on the direction of that principal component. Genes are given quantitative scores based on their influence on the various PCs. The product of the influence and the normalized read counts for each gene is multiplied by all genes to get a 'per sample' PC value. The gene expression profiles of samples that cluster together are more similar than those that cluster apart, especially for the most variant genes.

The good quality is to see clusters together, and conditions separate on PC1; this is a good method to explore the data quality. This method can also be used to identify sample outliers and significant sources of variation.

2.5.2 Venn Analysis

The gplots (v3.1.3) package (Warnes et al., 2005) was used to calculate and draw Venn schemes. The venn function from gplots package is convenient for generating Venn diagrams and getting intersections between the datasets. Venn diagrams generated in this study were used to compare DEG genes and investigate the intersection of genes between the three phases from normal melanocytes to RGP, from RGP to VGP, and from VGP to MET. This step is essential to find only the specific significant genes related to each phase.

2.5.3 Functional annotation

The Gene Ontology (GO) project offers organized, controlled vocabularies and classifications for the community to annotate genes, gene products, and sequences. These resources include a variety of molecular and cellular biology domains (Consortium, 2004).

Ontologies are offered by the GO project to define characteristics of gene products in three distinct fields of molecular biology. Molecular Function (MF) refers to molecular activities like catalytic or binding functions. A biological process (BP) is an organized assembly of molecular processes that achieves a biological objective. Cellular Component (CC) defines subcellular structures and macromolecular complexes (Consortium, 2004). A statistical technique known as over-representation (or enrichment) analysis examines whether genes from pre-defined sets with biological importance, such as those belonging to a particular GO keyword, are present in a subset of data more frequently than expected (Pomyen et al., 2015).

In this study, to perform the over-representation analysis clusterProfiler (T. Wu et al., 2021) v4.2.2 R package from Bioconductor, DOSE (Yu et al., 2015) v3.20.1 R package, pathview (W. Luo & Brouwer, 2013) v1.34.0 R package, AnnotationHub (Martin Morgan and Lori Shepherd, 2022) v3.2.2, ensembldb (Rainer et al., 2019)v 2.18.4, enrichplot (Guangchuang Yu, 2022) v1.14.2, ggnewscale (Elio Campitelli, 2022) v0.4.7, tidyverse(Wickham et al., 2019) v1.3.1 R packages were used. The functional analysis by the aforementioned packages requires a list of background genes and a list of significant expressed genes. For the background dataset, we used all genes tested for differential expression (all genes in our

results table). For the significant gene list, genes with p-adjusted values less than 0.05 were used. Human gene annotations were obtained using the Bioconductor org.Hs.eg.db R package v3.14.0 (Marc Carlson, 2021). The GO annotation file containing genes associated to biological processes (BP) was used for functional analysis.

Another tool for functional annotation is Gene Set Enrichment Analysis (GSEA). GSEA is an analytical technique used to evaluate gene expression data at the level of gene sets, which are collections of genes with related biological properties, chromosomal locations, or regulatory mechanisms. The gene sets have been identified based on the biological information that is currently known about biological pathways and co-expression from experimental techniques (Subramanian et al., 2005).

This method works by ranking genes according to the relationship between their expression and the classification or phenotypes and then determining whether or not members of a preset set of genes (S) are primarily found at the top or bottom of rankings in a ranked gene list (L). The approach also calculates adjusted Multiple Hypothesis Testing scores, an Enrichment Score (ES), and a Significant Level or p-value of the ES. The expression level (ES) represents the degree to which a set of genes S are over-represented at the top of a ranked gene list (L) (Figure 8) (Subramanian et al., 2005).

In this analysis, GSEA (Mootha et al., 2003; Subramanian et al., 2005) java-based stand-alone program v4.2.3 from a UC San Diego and Broad Institute joint project was used. The metric used to rank genes was (log2_ratio_of_classes).

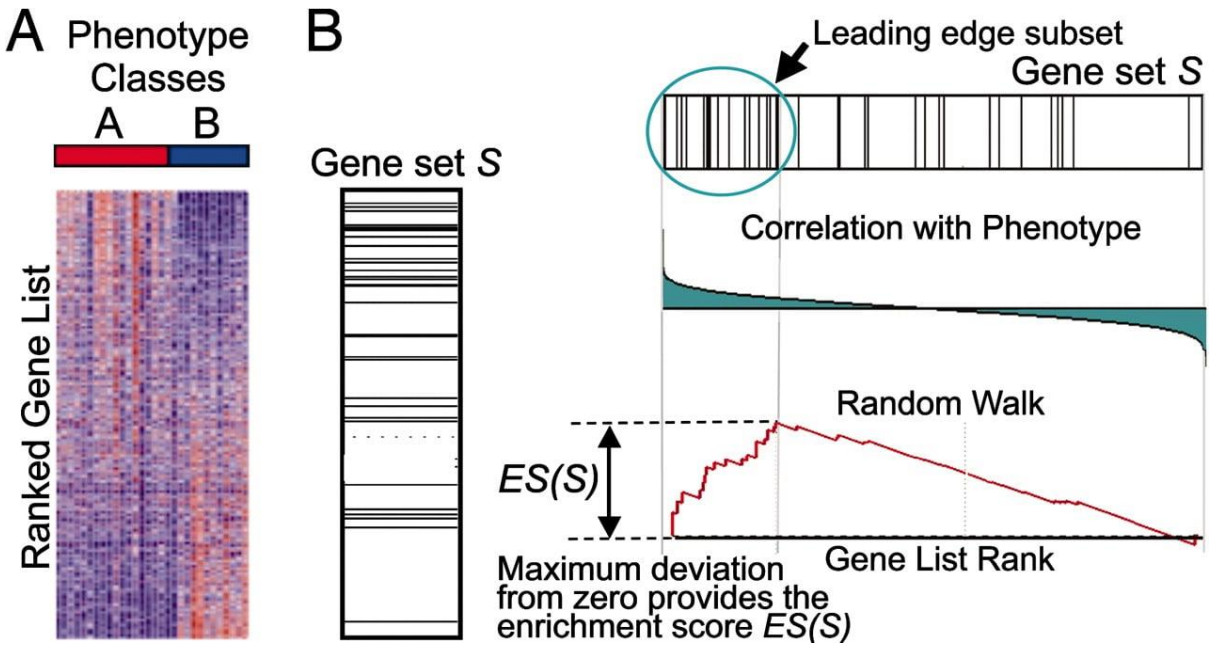


Figure 8: An overview illustrates the GSEA approach. *(A) A set of expression data sorted according to how well it correlates with phenotype, heat map, and the locations of specific genes from set S within the sorted list. The position of the maximum enrichment score (ES) and the leading-edge subset are shown in the (B) plot of the running sum for S in the data set.*

Chapter 3: Results

The chapter is divided into two parts: the results of the RNA sequencing, which include the amount, quality, and alignments of reads, and the results of the differential gene expression and similarity-based gene correlations, as well as functional annotation analysis in the second section. This includes the results from the over-representation analysis and the gene set enrichment analysis.

3.1 RNA-Seq Data Analysis

The following section contains the quality reports for the raw FASTQ RNA-Seq reads, the alignment results and the differential expression analysis.

3.1.1 RNA-Seq Quality Check

14 100 bp paired-end reads were processed by Illumina HiSeq2000 Genome Analyzer and the FASTQ files were quality checked using the FASTQC tool (v0.11.9) (Andrews, 2014) . The quality of all reads was above average. According to FASTQC-generated reports, no adapter contamination or bad quality reads could be detected. To summarize and visualize the results across all samples, the MultiQC tool was applied to the FASTQC reports (Table 2). The mean quality value across bases position is shown in Figure 9.

Type of target	Total sequences (Millions)	Sequence length (BP)	Average GC% Content
Normal Cell Lines (NHEMS)			
NHEM_76_R1	14.4	100 bp	49%
NHEM_76_R2	14.4	100 bp	48%
NHEM_77_R1	21.6	100 bp	49%
NHEM_77_R2	21.6	100 bp	48%
Radial Growth Phase (RGP)			
Sbcl2_64_R1	16.0	100 bp	49%
Sbcl2_64_R2	16.0	100 bp	48%
Sbcl2_70_R1	20.8	100 bp	45%
Sbcl2_70_R2	20.8	100 bp	45%
Vertical Growth Phase (VGP)			
WM793_67_R1	15.2	100 bp	48%

WM793_67_R2	15.2	100 bp	47%
WM793_73_R1	18.3	100 bp	47%
WM793_73_R2	18.3	100 bp	47%
WM1366_66_R1	9.2	100 bp	49%
WM1366_66_R2	9.2	100 bp	48%
WM1366_72_R1	20.7	100 bp	49%
WM1366_72_R2	20.7	100 bp	48%
WM3211_65_R1	11.5	100 bp	48%
WM3211_65_R2	11.5	100 bp	47%
WM3211_71_R1	15.5	100 bp	49%
WM3211_71_R2	15.5	100 bp	48%
Melanoma (MET)			
WM9_69_R1	24.2	100 bp	49%
WM9_69_R2	24.2	100 bp	48%
WM9_75_R1	9.0	100 bp	45%
WM9_75_R2	9.0	100 bp	45%
WM1158_68_R1	20.6	100 bp	49%
WM1158_68_R2	20.6	100 bp	48%
WM1158_74_R1	10.2	100 bp	49%
WM1158_74_R2	10.2	100 bp	48%

Table 2: General statistics illustrates the quality of the samples across all the FASTQC reports. *Total sequences in millions, sequence length, and GC% content of the FASTQ files. The average GC% content lies between 45% and 49%.*

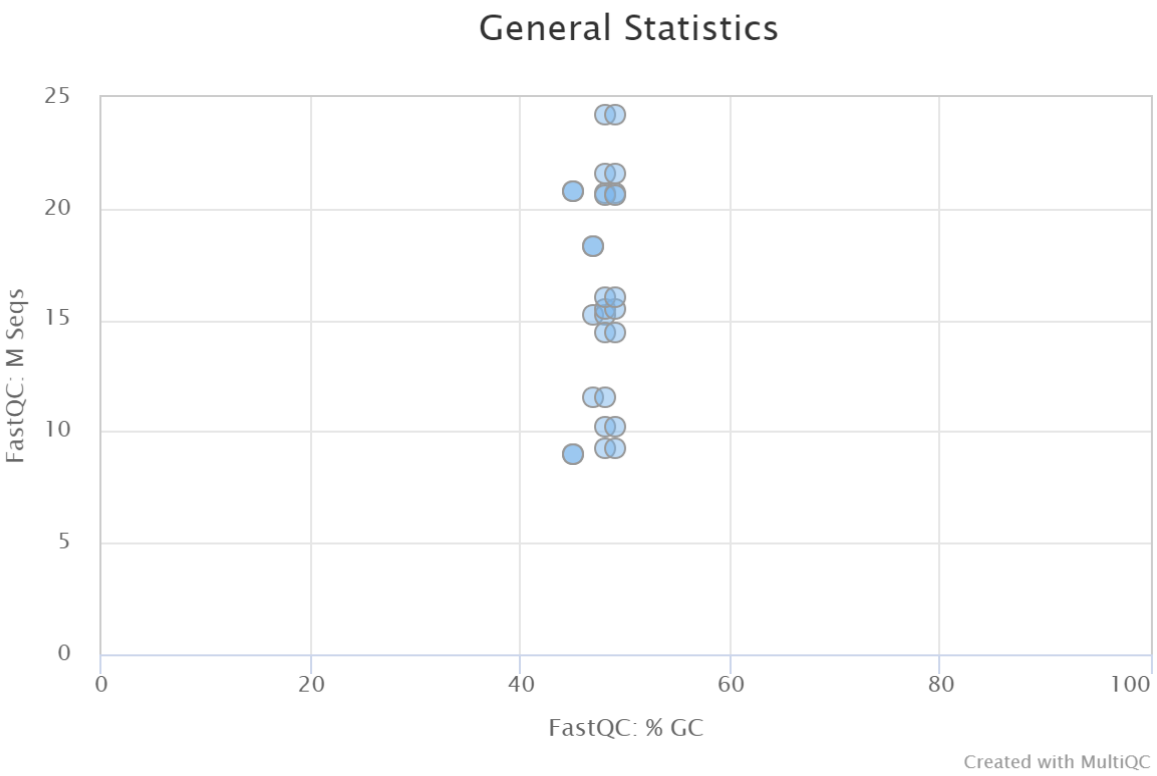


Figure 9: A scatter plot between %GC on the x-axis and total sequences in millions on the y-axis. *The average GC% content lies between 45-49%. The total sequences in millions lie between 9 million and 20 million reads.*

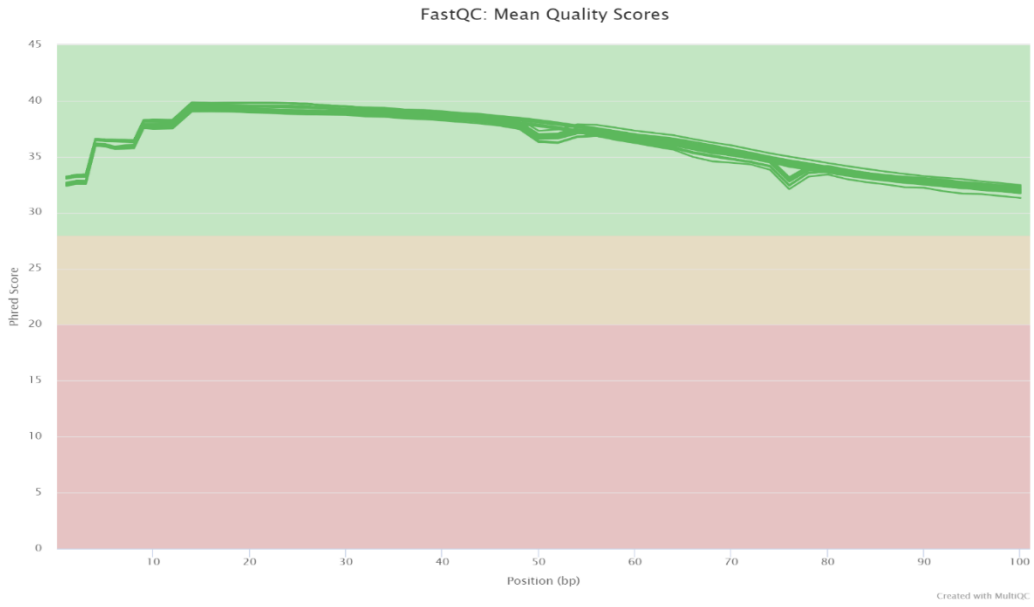


Figure 10: Sequence Quality Histogram illustrates the Phred quality score of each base pair. *The PHRED quality score is the probability of the log-transformed error for a base call, with high values denoting low error probabilities and vice versa* (BG & Green, 1998). *The figure shows that the mean quality scores lie in the green area, indicating an overall good quality.*

3.1.2 Alignment

STAR (v2.7.10a) (Dobin et al., 2012) used to map the reads to the reference genome GRCh38/ gencode v29.0.

Table 3: Representation of the Uniquely mapped reads resulting from STAR Alignment. *STAR provides information about uniquely mapped reads in million reads and % of reads.*

Type of target	Uniquely Mapped reads (Millions)	Uniquely Mapped reads (%)
Normal Cell Lines (NHEMS)		
NHEM_76	12.2	84.3%
NHEM_77	18.1	83.7%
Radial Growth Phase (RGP)		
Sbcl2_64	12.9	81.0%
Sbcl2_70	12.5	60.0%
Vertical Growth Phase (VGP)		
WM793_67	11.2	73.6%
WM793_73	11.5	62.9%
WM1366_66	6.8	74.2%
WM1366_72	14.4	69.8%
WM3211_65	8.1	69.9%
WM3211_71	10.6	68.8%
Metastasis (MET)		
WM9_69	19.2	79.7%
WM9_75	4.3	47.6%
WM1158_68	16.2	78.6%
WM1158_74	8.2	80.7%

The Alignment Scores plot of STAR illustrates considerable information about the uniquely mapped reads. The percentage of uniquely mapping, multi-mapping, and unmapped reads can be smoothly compared between samples to get an excellent overview of the quality of the samples (Figure 11).

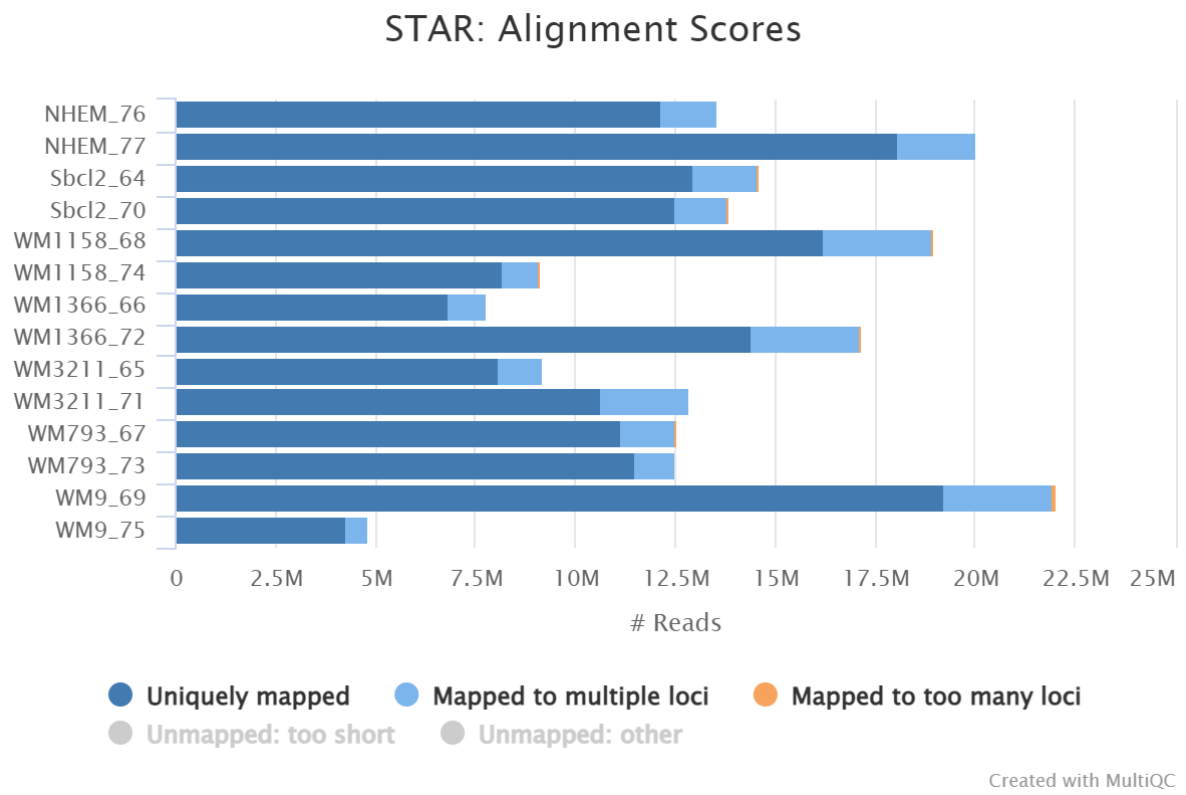


Figure 11: The Alignment Scores of Mapped reads in Millions. *The figure illustrates the total number of uniquely mapped reads, mapped reads to multiple locations, and unmapped reads.*

3.1.3 Read Counts

Calculating the number of reads per gene is done by the --quantMode GeneCounts option within the mapping command. Consequently, the STAR algorithm counts the number of reads per gene while mapping (Figure 12).

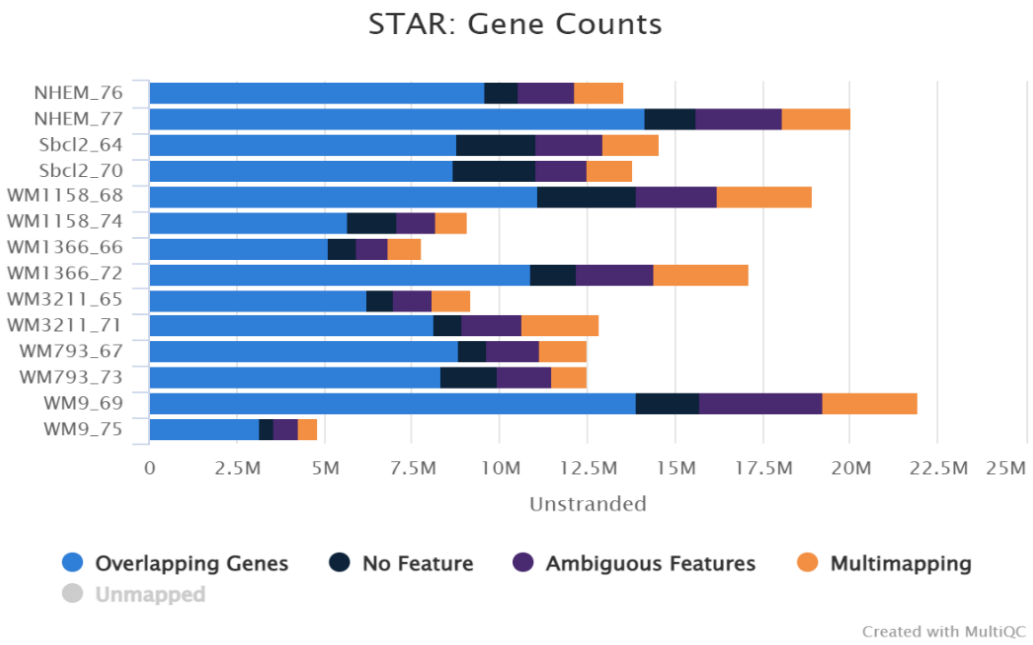


Figure 12: Gene counts for RNA-Seq samples. *The figure shows the overlapping, ambiguous features, the no feature, and the multi-mapping regions. STAR counts the number of reads per gene while mapping. A read is counted if it overlaps (1nucleotide or more) one gene.*

3.1.4 Differential Expression Analysis

In RNA-seq analysis, determining the overall degree of similarity between samples is a crucial step. First, the generated count table was imported in R. Next, the samples were analyzed for similarities in terms of their overall gene expression. Therefore, visualization approaches for unsupervised clustering analyses, such as hierarchical heatmaps and principal component analysis (PCA), was used.

These Quality Control approaches determine how comparable the biological replicates are to one another and detect outlier samples and major sources of variation in the data set.

To better visualize the clustering, first the *log* is used to transform the normalized counts before using these visualization methods. DESeq2 (v1.34.0) from Bioconductor (Love et al., 2014), utilizes a regularized log transform (rlog) of the normalized counts for sample-level quality control.

3.1.4.1 Clustering analysis

Hierarchical clustering was used to assess the similarity and gene expression between different samples in the dataset. The results were represented in a heatmap by combining the gene expression correlation values for all pairwise combinations of samples in the data set, with 1 being the perfect correlation.

The heatmap's colors reflect the correlation values, while the hierarchical tree shows which samples are more similar. This is because the biological replicates should be grouped together, whereas the sample from different conditions should be separated.

In Figure 13, samples are represented in both rows and columns. The analysis results indicate that high correlations across the board (> 0.9) imply no outlying sample(s). The samples are clustered together by condition.

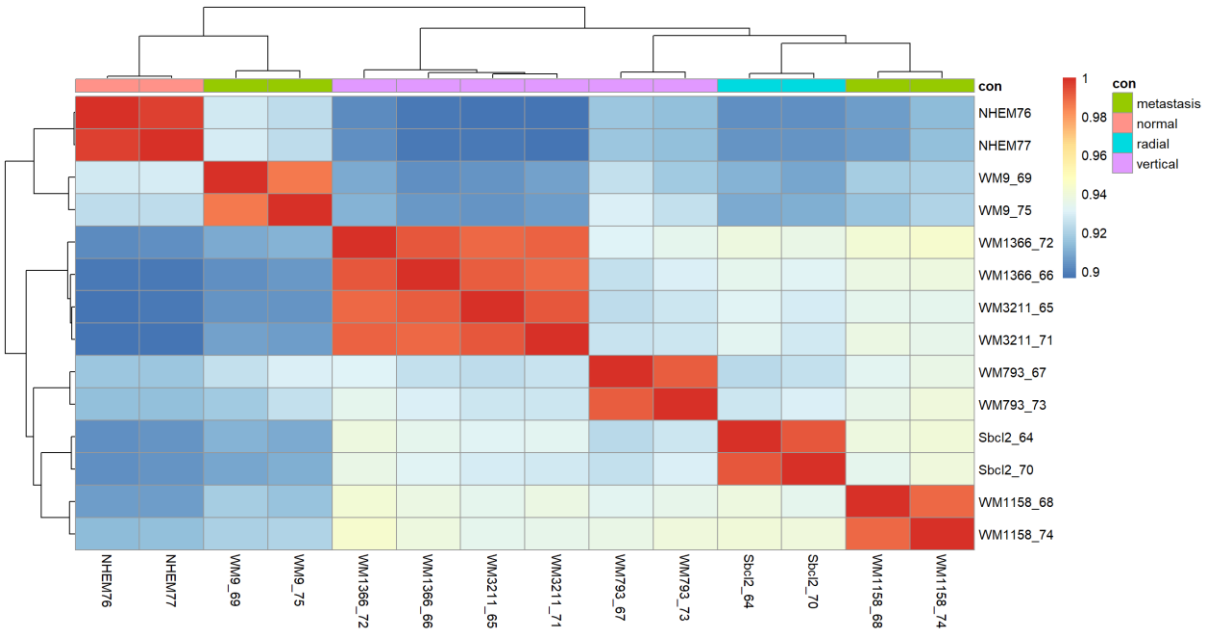


Figure 13: Heatmap depicts the correlation between samples. *The heatmap shows that samples were presented in both rows and columns, and the dendrogram at the top and left indicates how samples were clustered together. The diagonal has only red colors, which depicts the distance of the sample with itself. The legend at the top of the right side of the heatmap indicates that the red color shows short distances and high correlations among samples. The dendrogram (hierarchical cluster) shows how samples cluster together.*

PCA is also used to evaluate the quality of the samples. It is a technique for identifying the variation in a dataset and is utilized to illustrate how samples cluster and whether a condition of interest corresponds to the principal components explaining the most variation in the data. The samples (data points) are projected onto the 2D plane in a PCA to spread out in the two directions that account for most of the differences.

The first principal component, PCA1, represents the greatest amount of variance in the data in the x-axis direction. The second principal component, PC2, the dataset's second most variation, must be perpendicular to PC1 to best describe the variance in the dataset not included in PC1 in the y-axis direction. PC2 has a much smaller spread.

The Principal Component Analysis plot in Figure 14 shows that the condition of the cell lines is the major source of variation.

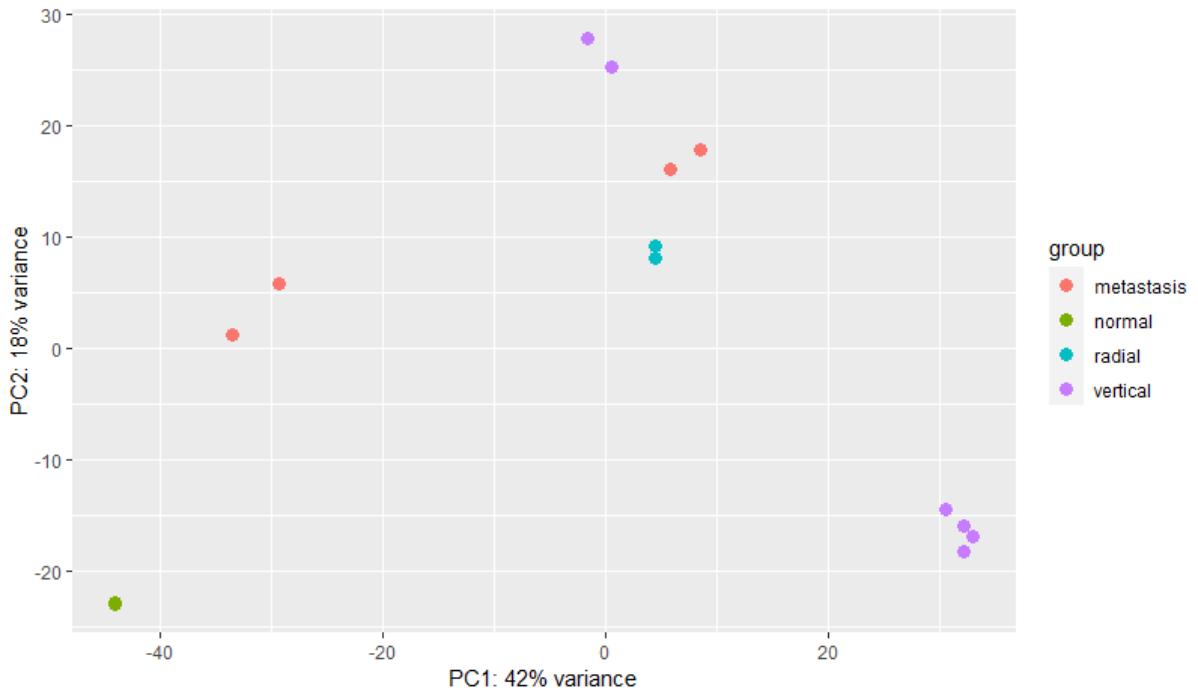


Figure 14: PCA plot. *The samples from Normal Cells (NHEMS), Radial growth phase(RGP), or Vertical growth phase cluster together. The output regarding the quality of the samples shows that the biological replicates tend to cluster together. The samples are colored by condition. The PC1 is 42 %, representing the greatest amount of variance in the data (42% of the data variance). PC2, is the dataset's second most variation (18% of the data variance), which is perpendicular to PC1 to best describe the variance in the dataset not included in PC1.*

3.1.4.2 Gene-wise dispersion estimation

Dispersion is a measure of spread or variability in the data. To determine how the data varies, it is crucial to look at the variance of gene expression with the mean of normalized counts. Variance is the square of the standard deviation, representing how far away the individual samples' expression is from the means of normalized counts. The variance is expected to increase with the gene's mean expression in RNA-Seq data. In the DESeq2 model, a dispersion metric describes a measure of variance for a given mean to assess the variability in expression (Love et al., 2014).

In Figure 14, the dispersion for each gene is determined using maximum likelihood estimation to model the dispersion depending on expression level (mean counts). Gene expression variance increases as the mean decreases. Because gene-wise dispersion estimates are often misleading in RNA-Seq experiments with only a few replicates, DESeq2 uses information from all genes to determine the most likely dispersion estimates for a given mean expression value, as shown by the red line.

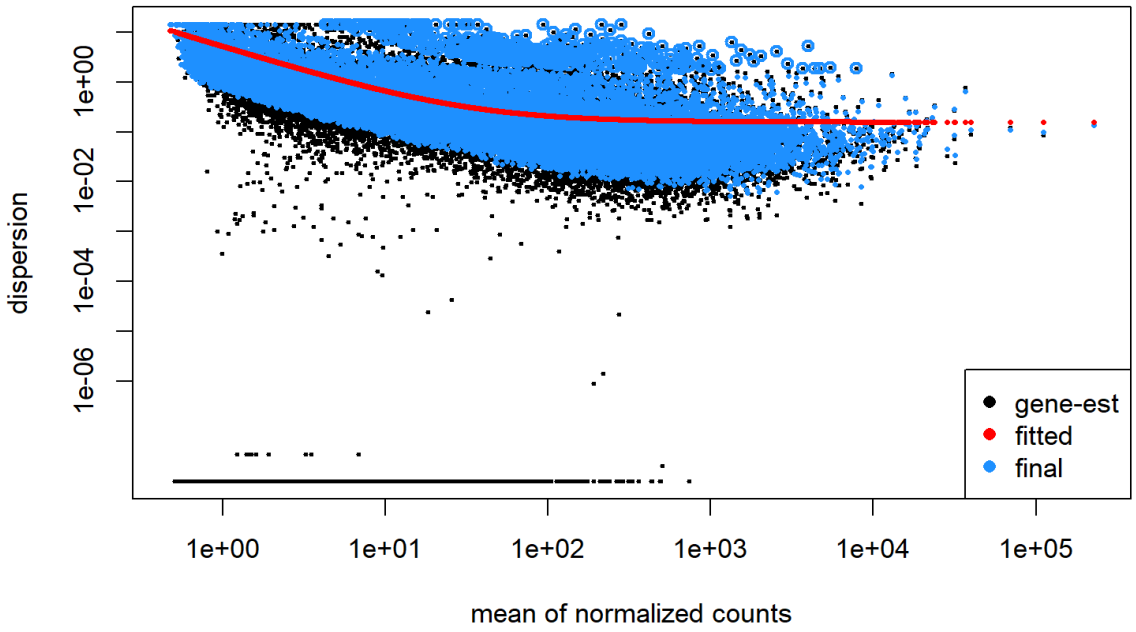


Figure 15: Dispersion vs. mean of normalized counts. *Each black dot represents a gene with associated mean and dispersion values, the shrunk genes are represented in the blue dots, and the fitted estimates are represented in the red dots. As the mean increases, the dispersion values decrease. The increase in variance, on the other hand, increases dispersion.*

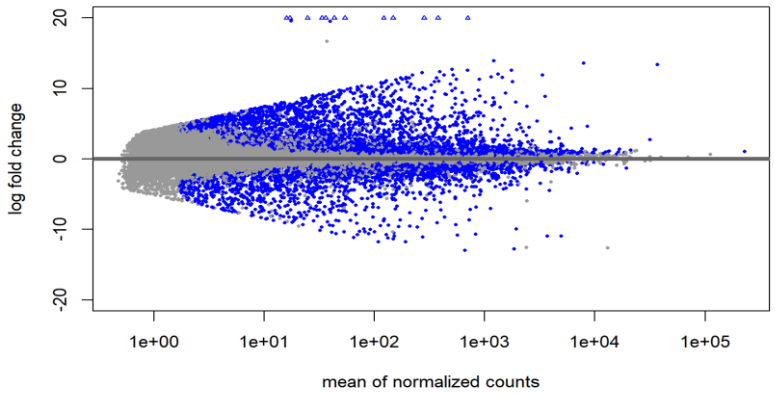
The MA plot depicts the mean of normalized counts versus log2foldchanges for all genes analyzed. Genes with similar expression levels will be presented around the horizontal line. Along the y-axis, data points with extreme values, when falling above the threshold, represent the genes with highly differential expression levels and indicate the number of genes being upregulated. On the other hand, data points that fall below the -1 threshold show high levels of downregulated genes (McDermaid et al., 2019).

Figure 15 illustrates three MA plots for each comparison (Normal to RGP, RGP to VGP, and VGP to metastasis). The MA-plot for the shrunken log2 fold changes is more beneficial, because of eliminating the noise related to log2 fold changes from low count genes (Love et al., 2014). The figure shows the impact of log fold change (LFC) shrinkage, which improves the estimated fold changes.

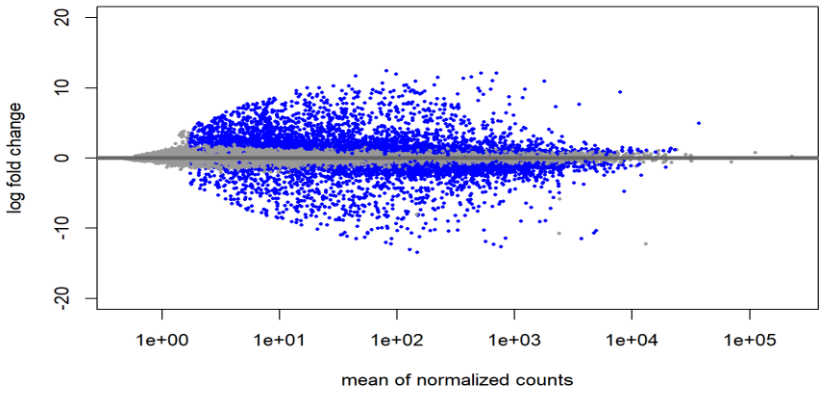
The MA-plot for the shrunken log2 fold changes is more beneficial, because of eliminating the noise related to log2 fold changes from low count genes (Love et al., 2014).

Normal (NHEMs)vs Radial growth phase (RGP)

Unshrunken results

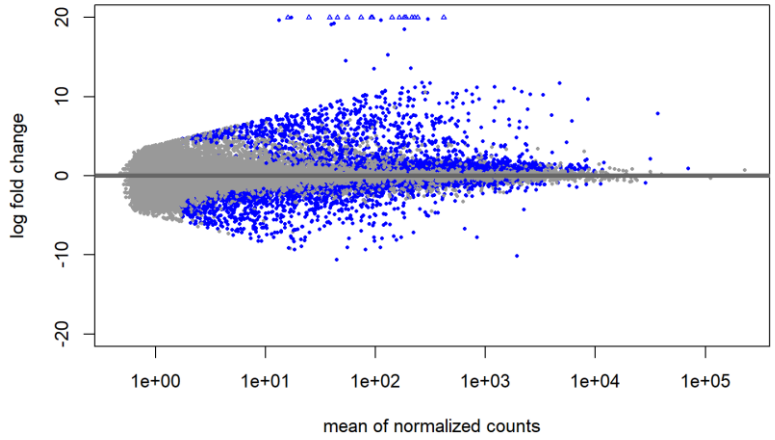


Shrunken results

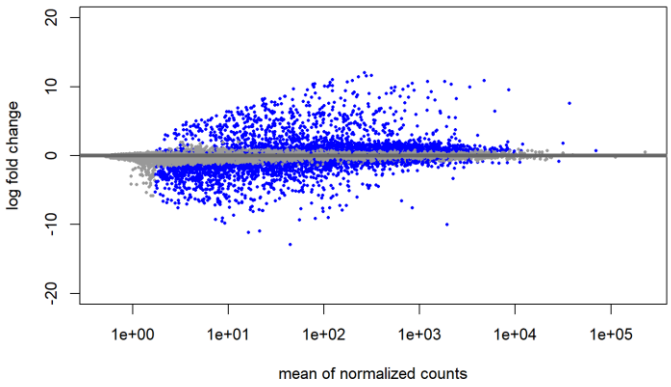


Radial growth phase (RGP) vs Vertical growth phase (VGP)

Unshrunken results

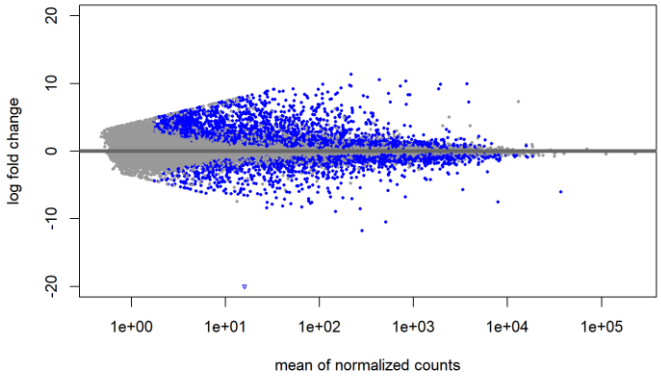


Shrunken results



Vertical growth phase (VGP) vs Metastasis (MET)

Unshrunken results



Shrunken results

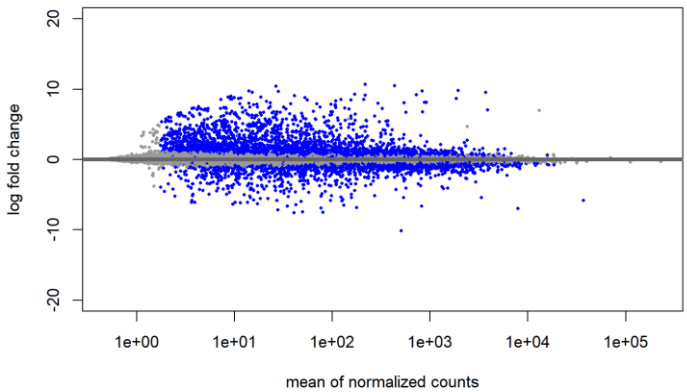


Figure 16: MA plot. *The differentially expressed genes are colored in blue, and grey dots represent the genes that are not differentially expressed. The horizontal line differentiates between the up-regulated and down-regulated genes. In the shrunken plots, the shrunken log fold changes are more precise; however, shrinking the*

log fold changes will not affect the number of differentially expressed genes returned, only the log fold change values. The log fold change values are more restricted, especially for lowly expressed genes.

3.1.4.3 Differential Expression Analysis

The goal of differential expression analysis is to determine which genes are expressed between conditions, which is crucial to figure out phenotypic variation (Costa-Silva et al., 2017).

In the Deseq2 analysis, (padj < 0.05) was applied, and the order of the genes in the table was done in ascending order of according to their p-adjusted value.

During the first developmental step from normal melanocytes to RGP melanoma cells, differential expression analysis resulted in 2308 up-regulated and 2401 down-regulated genes in the RGP melanoma cells of overall 26574 genes with nonzero total read count. Exemplarily, downregulated genes in the RGP cell line were PLXNC1, CDK2, CYTH3, and CTSL and up-regulated genes were SHROOM3, SATB2, HMGA2, TPD52L1and AFAP1L2 in RGP cell line. The details of differentially expressed genes are represented in Table 4 and Figures 17, 18, and 19 below.

Ensgene	**Symbol**	**log2FoldChange**	**Description**
ENSG00000136040	PLXNC1	-6.54869806453937	plexin C1
ENSG00000123374	CDK2	-3.07202216131381	cyclin dependent kinase 2
ENSG00000138771	SHROOM3	7.50046162719322	shroom family member 3
ENSG00000119042	SATB2	5.23804823237876	SATB homeobox 2
ENSG00000008256	CYTH3	-3.52212843667665	cytohesin 3
ENSG00000135047	CTSL	-4.79309956048035	cathepsin L
ENSG00000178878	APOLD1	-5.84545073389205	apolipoprotein L domain containing 1
ENSG00000149948	HMGA2	8.63289090070051	high mobility group AT-hook 2
ENSG00000156853	ZNF689	-3.49973846677968	zinc finger protein 689
ENSG00000108861	DUSP3	-2.26206510671385	dual specificity phosphatase 3
ENSG00000111907	TPD52L1	5.38525771556555	TPD52 like 1
ENSG00000134160	TRPM1	-7.9940207365021	transient receptor potential cation channel subfamily M member 1
ENSG00000169129	AFAP1L2	5.7008888073351	actin filament associated protein 1 like 2
ENSG00000162409	PRKAA2	7.37075793645473	protein kinase AMP-activated catalytic subunit alpha 2
ENSG00000136738	STAM	-2.26705977392634	signal transducing adaptor molecule
ENSG00000266094	RASSF5	6.99531853122955	Ras association domain family member 5

ENSG00000130396	AFDN	2.93569559994615	afadin adherens junction formation factor
ENSG00000159164	SV2A	5.63668159519788	synaptic vesicle glycoprotein 2A
ENSG00000142627	EPHA2	5.32380682151506	EPH receptor A2
ENSG00000142949	PTPRF	4.15607423391129	protein tyrosine phosphatase receptor type F

Table 4: Top 20 differentially expressed genes (NHEMs vs. RGP). *Up and down-regulated genes are identified in the table based on the log2foldchange, where negative values represent the down-regulated genes in RGP cell line, and positive values represent the up-regulated genes in RGP cell line.*

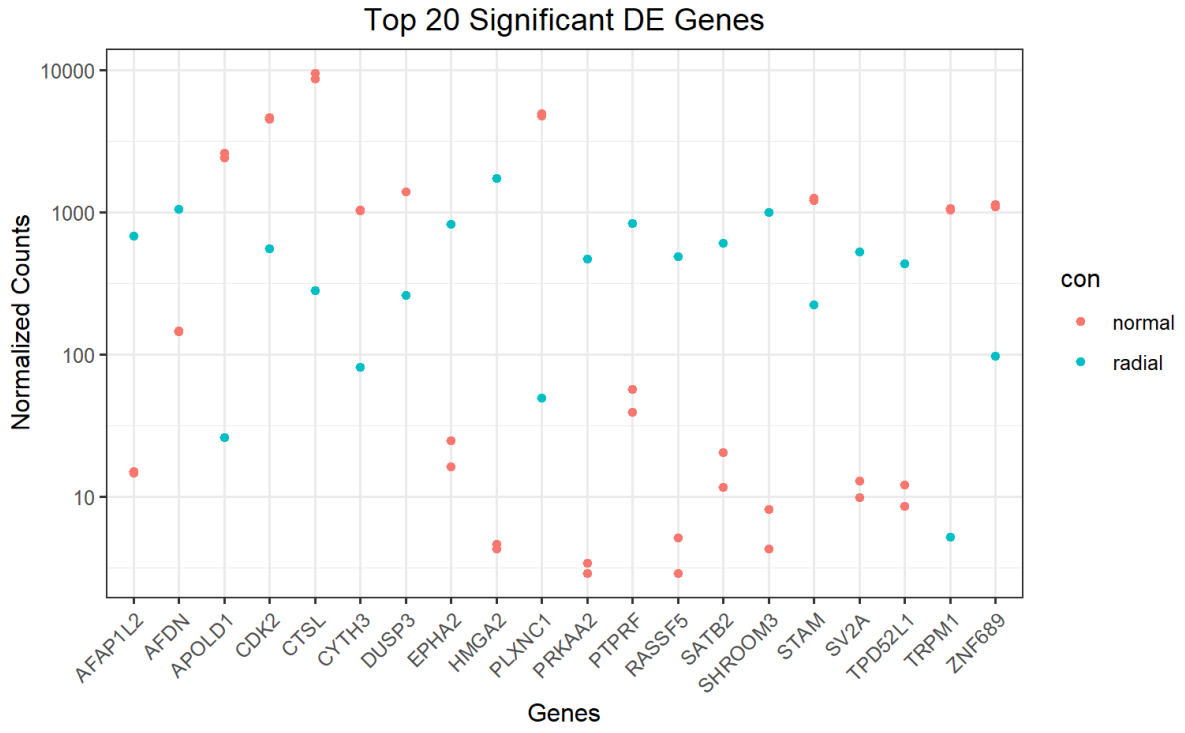


Figure 17: A scatter plot represents the top 20 significant differentially expressed genes.

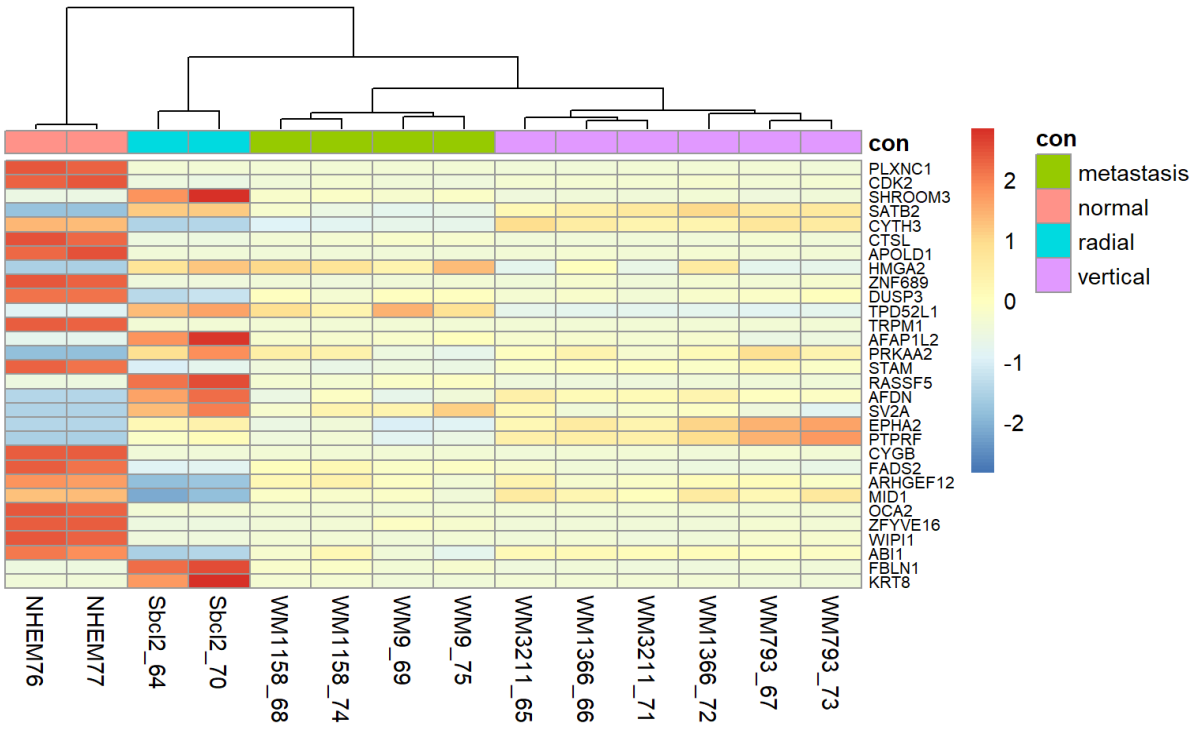


Figure 18: Heatmap depicting the top 30 differentially expressed genes. *The heatmap is a result of the developmental step from NHMEs to RGP melanoma cells. The expression level of genes in blue color indicates very low expressed genes, and red indicates very highly expressed genes.*

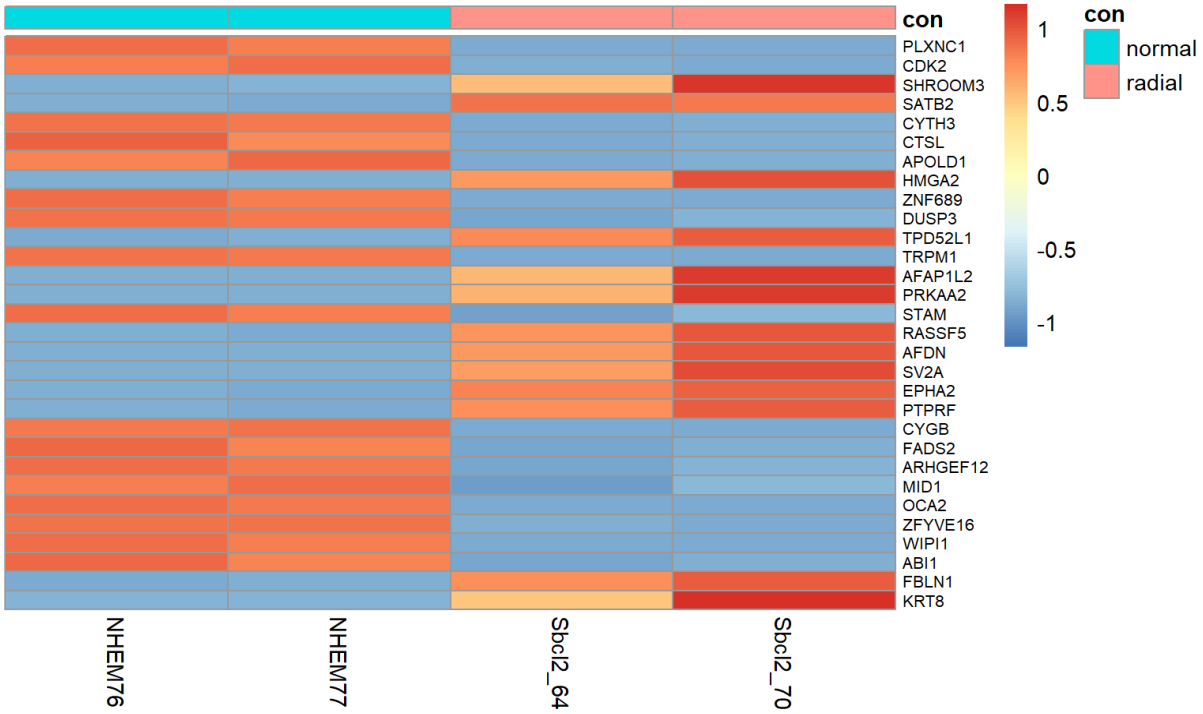


Figure 19: Heatmap of Top 30 genes (NHEMs to RGP). *The heatmap represents the development step from NHEMs to RGP melanoma.*

In (RGP vs. VGP), out of 26574 genes with nonzero total read count, there were 1903 genes up-regulated and 1308 genes down-regulated in the VGP cell line. The top differentially expressed genes are illustrated in Table 5 and Figures 20,21.

Ensgene	Symbol	log2FoldChange	Description
ENSG00000164741	DLC1	4.3299798845537	DLC1 Rho GTPase activating protein
ENSG00000008256	CYTH3	3.11244309320016	cytohesin 3
ENSG00000124575	H1-3	-10.0116410891087	H1.3 linker histone cluster member
ENSG00000138771	SHROOM3	-4.11616724590371	shroom family member 3
ENSG00000162576	MXRA8	-6.88139684911637	matrix remodeling associated 8
ENSG00000206538	VGLL3	4.67959589300863	vestigial like family member 3
ENSG00000266094	RASSF5	-4.94501134087836	Ras association domain family member 5
ENSG00000173267	SNCG	-8.93997117775874	synuclein gamma
ENSG00000105993	DNAJB6	1.19820321000951	DnaJ heat shock protein family (Hsp40) member B6
ENSG00000111907	TPD52L1	-3.7414939143596	TPD52 like 1
ENSG00000115738	ID2	-5.88786850570675	inhibitor of DNA binding 2
ENSG00000137801	THBS1	9.56423511763241	thrombospondin 1
ENSG00000101871	MID1	1.89448075637211	midline 1
ENSG00000103449	SALL1	-8.11328131113994	spalt like transcription factor 1
ENSG00000004399	PLXND1	-4.09635901323615	plexin D1
ENSG00000163359	COL6A3	-6.13203027384761	collagen type VI alpha 3 chain
ENSG00000154678	PDE1C	8.31000228654636	phosphodiesterase 1C
ENSG00000113739	STC2	9.29181527716571	stanniocalcin 2
ENSG00000173801	JUP	-6.20517523362771	junction plakoglobin
ENSG00000064042	LIMCH1	6.03377943834074	LIM and calponin homology domains 1

Table 5: Top 20 differentially expressed genes (RGP vs. VGP). *The most common downregulated genes are H1-3, SHROOM3, MXRA8, and RASSF5 and upregulated genes were DLC1, CYTH3 , HMGA2, VGLL3 and DNAJB6.*

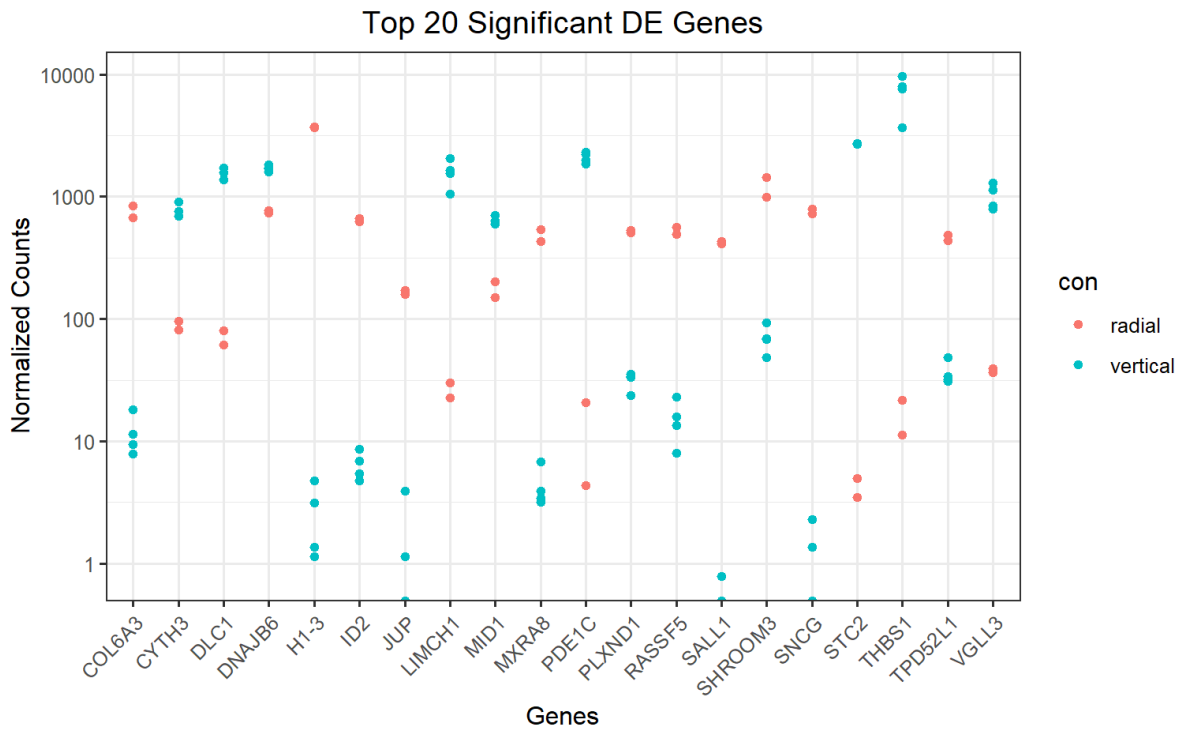


Figure 20: A scatter plot represents the top 20 significant differentially expressed genes (RGP vs. VGP). *The Scatter plot shows that THBS1, STC2, PDE1C, and LIMCH1 genes are up-regulated in the Vertical Growth Phase. In the Radial Growth Phase, the H1-3, SHROOM3, COL6A3 SNCG, and DNAJB are up-regulated.*

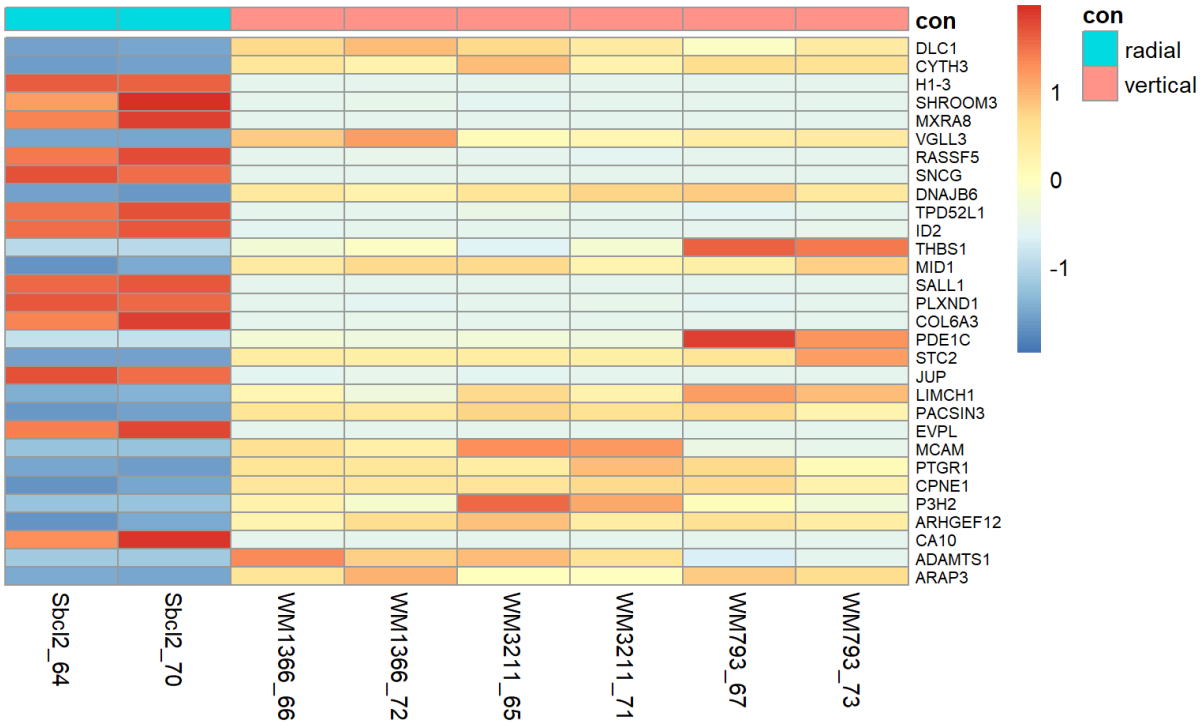


Figure 21: Heatmap of Top 30 genes (RGV vs. VGP). *The expression level of genes in dark blue indicates very low expressed genes, and red indicates very highly expressed genes.*

In the last step (VGP vs. MET), out of 2199, there were 1793 genes up-regulated and 1308 genes down-regulated.

The Differentially expressed genes are illustrated in Table 6 and Figures 22,23.

Ensgene	symbol	log2FoldChange	Description
ENSG00000124575	H1-3	9.85845990100365	H1.3 linker histone cluster member
ENSG00000206538	VGLL3	-4.26853276158157	vestigial like family member 3
ENSG00000164692	COL1A2	10.4952472001362	collagen type I alpha 2 chain
ENSG00000148154	UGCG	-3.18453285549245	UDP-glucose ceramide glucosyltransferase
ENSG00000124092	CTCFL	9.27524958606592	CCCTC-binding factor like
ENSG00000111907	TPD52L1	3.32781266334325	TPD52 like 1
ENSG00000133110	POSTN	6.80578004663601	periostin
"ENSG00000171388	APLN	-5.39524045193369	apelin"
"ENSG00000154122	ANKH	1.6298644326666	ANKH inorganic pyrophosphate transport regulator
ENSG00000136531	SCN2A	4.71432425527813	sodium voltage-gated channel alpha subunit 2
ENSG00000160886	LY6K	-6.79698273247152	lymphocyte antigen 6 family member K
ENSG00000184867	ARMCX2	6.23598348985909	armadillo repeat containing X-linked 2
ENSG00000230426	LINC01036	6.34422664171101	long intergenic non-protein coding RNA 1036
ENSG00000104611	SH2D4A	8.47891518910983	SH2 domain containing 4A
ENSG00000189060	H1-0	-2.68143789114609	H1.0 linker histone
ENSG00000091656	ZFHX4	5.68172810881639	zinc finger homeobox 4
ENSG00000145526	CDH18	7.68271589375278	cadherin 18
ENSG00000155962	CLIC2	5.78214604530213	chloride intracellular channel 2
ENSG00000164308	ERAP2	-4.12100196561962	endoplasmic reticulum aminopeptidase 2
ENSG00000147180	ZNF711	-6.88112325471719	zinc finger protein 711

Table 6: Top 20 differentially expressed genes (VGP vs. MET). *Ranking the genes was done by ordering the genes in ascending according to the p-adjusted value of (padj < 0.05).*

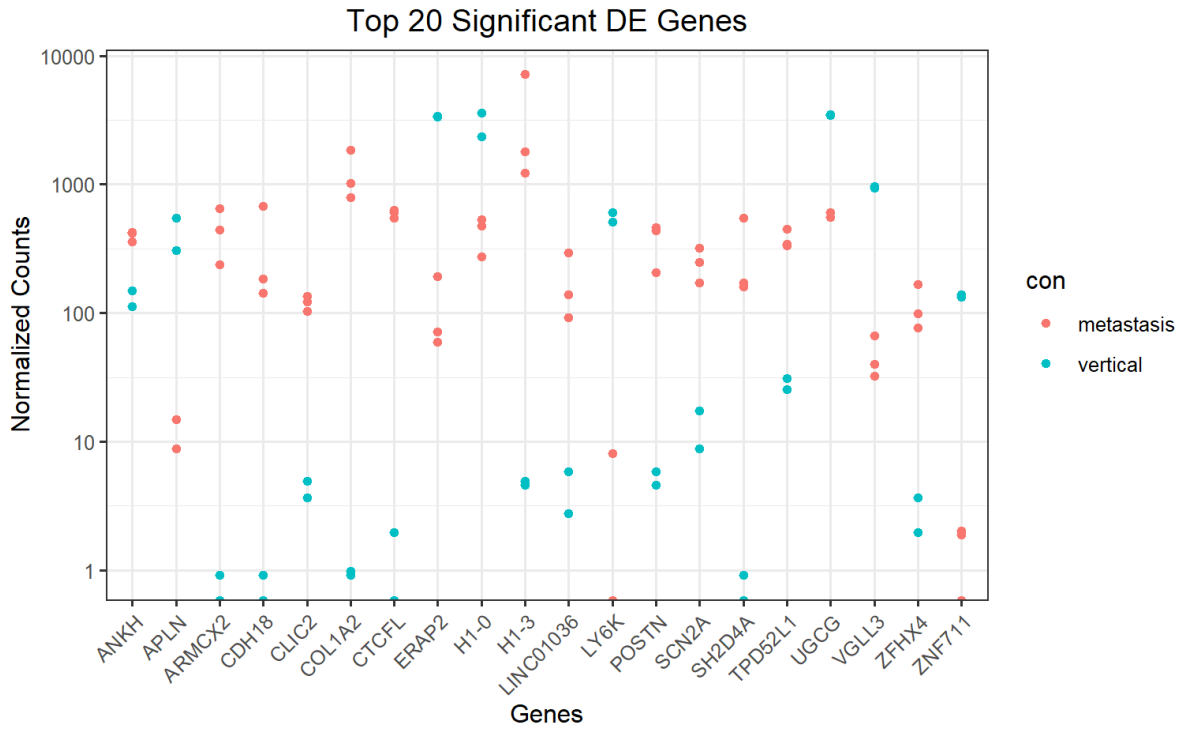


Figure 22: A scatter plot represents the top 20 significant differentially expressed genes (VGP vs. MET). *The scatterplot shows that in the Metastasis phase, the most common up-regulated genes are H1-3, COL1A2, CTCFL, and H1-0, while the most common down-regulated genes are ZNF711, APLN, and LY6K.*

3.1.5 Extract unique genes in every phase-comparison

The differential expression results of the previous section were used to identify the specific DEGs relevant for the transition from NHEMs to RGP, RGP to VGP, and VGP to MET.

3.1.5.1 Specific DE genes in the developmental step from NHEMs to RGP

In order to identify the differentially expressed genes specifically for the development step from NHEMs to RGP melanoma cells, we subtracted differentially expressed genes discovered in this stage (from NHEMs to RGP) from the other two phases (RGP to. VGP and VGP to MET). Table 7 below depicts the specific genes in this developmental step.

Ensgene	**Symbol**	**log2FoldChange**	**Description**
ENSG00000162409	PRKAA2	7.37075793645473	protein kinase AMP-activated catalytic subunit alpha 2
ENSG00000039319	ZFYVE16	-3.10466335917462	zinc finger FYVE-type containing 16
ENSG00000161544	CYGB	-6.49206892355045	cytoglobin
ENSG00000104044	OCA2	-10.3571394730959	OCA2 melanosomal transmembrane protein
ENSG00000130775	THEMIS2	-4.24888524871474	thymocyte selection associated family member 2

ENSG00000105048	TNNT1	5.07364396236536	troponin T1 slow skeletal type
ENSG00000119986	AVPI1	-4.52479482201962	arginine vasopressin induced 1
ENSG00000069974	RAB27A	-4.54591856773995	RAB27A member RAS oncogene family
ENSG00000160213	CSTB	-2.31594193678879	cystatin B
ENSG00000146072	TNFRSF21	6.11619848771848	TNF receptor superfamily member 21
ENSG00000111057	KRT18	9.85771240575672	keratin 18
ENSG00000140497	SCAMP2	-1.85809322419535	secretory carrier membrane protein 2
ENSG00000083544	TDRD3	-4.2157095117544	tudor domain containing 3
ENSG00000136854	STXBP1	-2.55975634644584	syntaxin binding protein 1
ENSG00000129493	HEATR5A	-2.28264778387864	HEAT repeat containing 5A
ENSG00000100364	KIAA0930	-3.17552891958289	KIAA0930"
ENSG00000132669	RIN2	5.1638000550275	Ras and Rab interactor 2
ENSG00000174738	NR1D2	2.46159035781227	nuclear receptor subfamily 1 group D member 2
ENSG00000113594	LIFr	5.99210170463055	LIF receptor subunit alpha
ENSG00000103249	CLCN7	-2.51580597658842	chloride voltage-gated channel 7

Table 7: Top 20 differentially expressed genes specific for the developmental step (NHEMs to RGP).

3.1.5.2 Specific DE genes in the developmental step from RGP to VGP

In the next step, we aimed to identify differentially expressed genes specific for the development from RGP to VGP melanoma cells. Therefore, we filtered the identified DE genes comparing the aforementioned conditions by subtracting the overlapping DE genes of the other two comparisons RGP vs. NHEMs and VGP vs. MET (Table 8).

Ensgene	**Symbol**	**log2FoldChange**	**Description**
ENSG00000154678	PDE1C	8.31000228654636	phosphodiesterase 1C
ENSG00000111371	SLC38A1	9.53455534567463	solute carrier family 38 member 1
ENSG00000180190	TDRP	-7.93547538974539	testis development related protein
ENSG00000270816	LINC00221	11.5755169554157	long intergenic non-protein coding RNA 221
ENSG00000158639	PAGE5	10.8167575210078	PAGE family member 5
ENSG00000187172	BAGE2	11.4417467476335	BAGE family member 2 pseudogene
ENSG00000099864	PALM	-5.82317120624576	paralemmin
ENSG00000154096	THY1	9.09634633404037	Thy-1 cell surface antigen

ENSG00000204382	XAGE1B	10.9216937035196	X antigen family member 1B
ENSG00000143727	ACP1	-1.06354823688948	acid phosphatase 1
ENSG00000060656	PTPRU	-5.17187040150217	protein tyrosine phosphatase receptor type U
ENSG00000119125	GDA	6.24926732673203	guanine deaminase
ENSG00000106034	CPED1	6.60498457640781	cadherin like and PC-esterase domain containing 1
ENSG00000060688	SNRNP40	1.19580721946153	small nuclear ribonucleoprotein U5 subunit 40
ENSG00000188452	CERKL	4.41998536214519	ceramide kinase like
ENSG00000101473	ACOT8	1.12027001332049	acyl-CoA thioesterase 8
ENSG00000223345	H2BP1	-4.36107820467876	H2B histone pseudogene 1
ENSG00000134363	FST	6.70344897153454	follistatin
ENSG00000122133	PAEP	-6.16014820739095	progestagen associated endometrial protein
ENSG00000170689	HOXB9	4.95357990137729	homeobox B9

Table 8: Top 20 differential expressed genes specific for the developmental step (RGP to VGP)*. PDE1C, SLC38A1, LINC00221, PAGE5, and BAGE2 are examples of the up-regulated genes in VGP melanoma cells.*

3.1.5.3 Specific DE genes in the developmental step (from VGP to MET)

Finally, in order to determine the differentially expressed genes specific for the development from VGP to MET melanoma cells, we eliminated the overlapping DE genes from the other two comparisons, RGP to NHEMs and VGP to MET (Table 9).

Ensgene	**symbol**	**log2FoldChange**	**Description**
ENSG00000124092	CTCFL	9.27524958606592	CCCTC-binding factor like
ENSG00000133110	POSTN	6.80578004663601	periostin
ENSG00000145526	CDH18	7.68271589375278	cadherin 18
ENSG00000155962	CLIC2	5.78214604530213	chloride intracellular channel 2
ENSG00000164308	ERAP2	-4.12100196561962	endoplasmic reticulum aminopeptidase 2
ENSG00000138031	ADCY3	-1.40467308545235	adenylate cyclase 3
ENSG00000124610	H1-1	-6.60766198065765	"H1.1 linker histone cluster member
ENSG00000101596	SMCHD1	-0.820124456333652	structural maintenance of chromosomes flexible hinge domain containing 1
ENSG00000167380	ZNF226	1.61471325785994	zinc finger protein 226
ENSG00000125772	GPCPD1	1.4563783559896	glycerophosphocholine phosphodiesterase 1

ENSG00000151532	VTI1A	-1.28268746117246	vesicle transport through interaction with t-SNAREs 1A
ENSG00000215256	DHRS4-AS1	1.51478877476327	DHRS4 antisense RNA 1
ENSG00000171469	ZNF561	1.12712399890873	zinc finger protein 561
ENSG00000118785	SPP1	8.13870296876909	secreted phosphoprotein 1
ENSG00000250337	PURPL	5.80251868525893	p53 upregulated regulator of p53 levels
ENSG00000119969	HELLS	-1.44895049354104	helicase lymphoid specific
ENSG00000241015	TPM3P9	1.54917133253993	tropomyosin 3 pseudogene 9
ENSG00000165071	TMEM71	-4.10598877127346	transmembrane protein 71
ENSG00000138182	KIF20B	-1.54797516016868	kinesin family member 20B
ENSG00000112379	ARFGEF3	-3.63576713756108	ARFGEF family member 3

Table 9: Top 20 differential expressed genes specific for the developmental step (VGP to MET). *For instance, CTCFL, POSTN, CDH18, and CLIC2 are genes that up-regulated in MET melanoma cells.*

3.1.6 Venn Analysis

In order to figure out the overlapping genes between all developmental steps, a Venn analysis was done. The package gplots v3.1.3 and the function venn were used to generate Venn diagrams and to get intersections between all developmental stages (Figure 24).

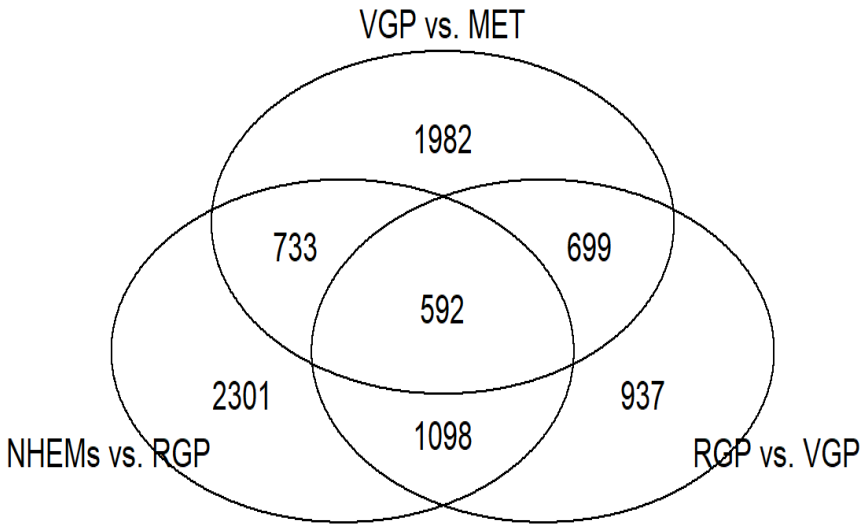


Figure 23: Venn diagram representing the differentially expressed genes in all stages and depicts the overlapping DEGs of all stages. *The diagram shows that 592 genes are differentially expressed in all investigated stages. 2301 genes are specific DE genes from NHEMs to RGP, 937 from RGP to VGP and 1982 from VGP to MET melanoma cells.*

3.2 Functional annotation analysis

3.2.1 Over representation analysis (ORA)

For a deeper understanding of the functional relevance and interaction of the identified DE genes and to perform a comparison between the generated gene sets associated to the respective developmental steps, the over-representation analysis with the clusterProfiler package (T. Wu et al., 2021) v4.2.2 was conducted. A list of background genes and a list of significant genes were provided. All genes tested for differential expression (all genes in our results table) were used as the background dataset for gene ontology (GO term) analysis focusing on Biological Processes (BP) (Ashburner et al., 2000; Gene Ontology Consortium, 2021). Differentially expressed genes with a padj < 0.05 were regarded as statistically significant.

Gene Ontology ID	Biological Process Description	Gene Ratio	p.adjusted value
GO:0030335	positive regulation of cell migration	168/3609	0.000246913665919288
GO:0040017	positive regulation of locomotion	174/3609	0.000246913665919288
GO:2000147	positive regulation of cell motility	171/3609	0.000246913665919288
GO:0051272	positive regulation of cellular component movement	174/3609	0.000246913665919288
GO:0042330	taxis	166/3609	0.000246913665919288
GO:0001667	ameboidal-type cell migration"	138/3609	0.000246913665919288
GO:0006935	chemotaxis	165/3609	0.000273954978316588
GO:0048732	gland development	137/3609	0.000552553076307141
GO:1901617	organic hydroxy compound biosynthetic process	78/3609	0.00107134062451136
GO:0001525	angiogenesis	160/3609	0.00107134062451136
GO:0048568	embryonic organ development	131/3609	0.00107134062451136
GO:0060485	mesenchyme development	96/3609	0.00122038000240647
GO:0044057	regulation of system process	151/3609	0.00125695267241402
GO:0006470	protein dephosphorylation	97/3609	0.00141737214039833
GO:0010631	epithelial cell migration	101/3609	0.00141737214039833
GO:0090132	epithelium migration	101/3609	0.00151728063819727
GO:0030855	epithelial cell differentiation	168/3609	0.00151728063819727
GO:0061564	axon development	146/3609	0.00160984599339525

GO:0007264	small GTPase mediated signal transduction	159/3609	0.00172317983312376
GO:0090130	tissue migration	102/3609	0.00177714600525882

Table 10: Gene Ontology Enrichment Analysis (NHEMs vs. RGP). *The table represents the enriched biological processes that DEGs are involved in. Gene ratio represents the number of genes in the list of differentially expressed genes compared to the GO background list.*

Gene Ontology ID	Biological Process Description	Gene Ratio	p.adjust
GO:0001525	angiogenesis	117/2552	0.0303741042627947
GO:0046135	pyrimidine nucleoside catabolic process	9/2552	0.0303741042627947
GO:0006216	cytidine catabolic process	8/2552	0.0303741042627947
GO:0009972	cytidine deamination	8/2552	0.0303741042627947
GO:0046087	cytidine metabolic process	8/2552	0.0303741042627947
GO:0072529	pyrimidine-containing compound catabolic process	16/2552	0.0303741042627947
GO:0007264	small GTPase mediated signal transduction	116/2552	0.0404321137646536

Table 11: Gene Ontology Enrichment Analysis (RGP vs. VGP). *In this developmental stage, the Gene Ontology dropped to only seven biological pathways, with the angiogenesis biological pathway likely to have the most gene count.*

ONTOLOGY	ID	Description	Gene Ratio	p.adjust
BP	GO:0000302	response to reactive oxygen species	73/3092	0.00145166122056834
CC	GO:0098644	complex of collagen trimers	13/3198	0.0220068773637677
CC	GO:0031965	nuclear membrane	85/3198	0.0389288815450656

Table 12: Gene Ontology Enrichment Analysis (VGP vs. MET*). In this developmental stage, the analysis was done to all Gene Ontology terms (Biological Process BP, Molecular Function, and cellular component CC).*

A widespread technique for visualizing enriched GO biological process is bar plot. It shows the gene ratio and enrichment scores as bar height and color (Figure 25).

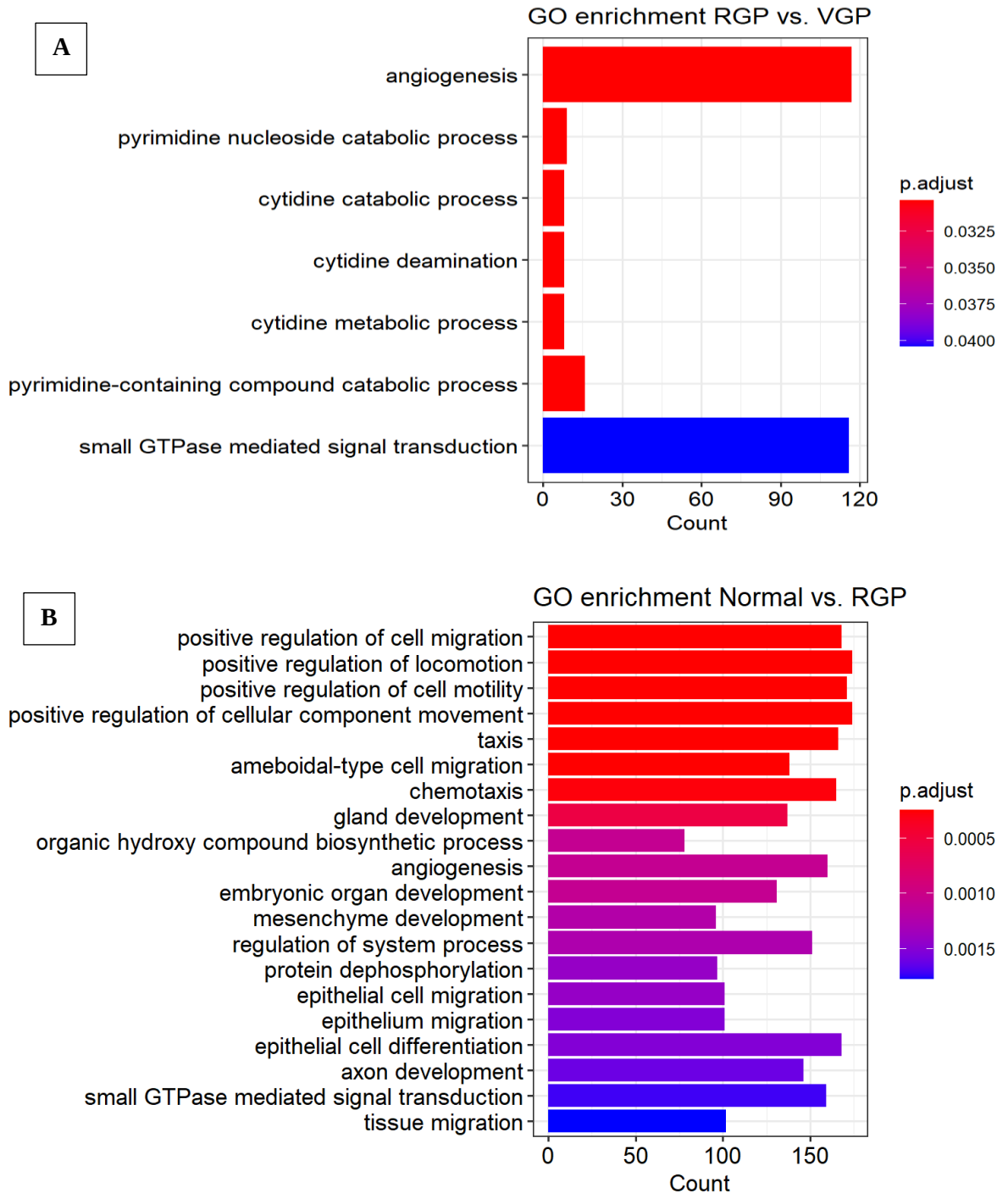


Figure 24: Bar plot of gene ontology enrichment. *(A) The bar plot illustrates the Gene Ontology in the Biological Process BP in the second developmental step from RGP to VGP. (B) represents the GO enrichment from NHEMs to RGP*

The dot plot is similar to the bar blot to visualize the enrichment results. Dot plots have the ability to add another parameter. In the figures below, the dot plot shows the Count and the p-adjusted value, and the gene ratio.

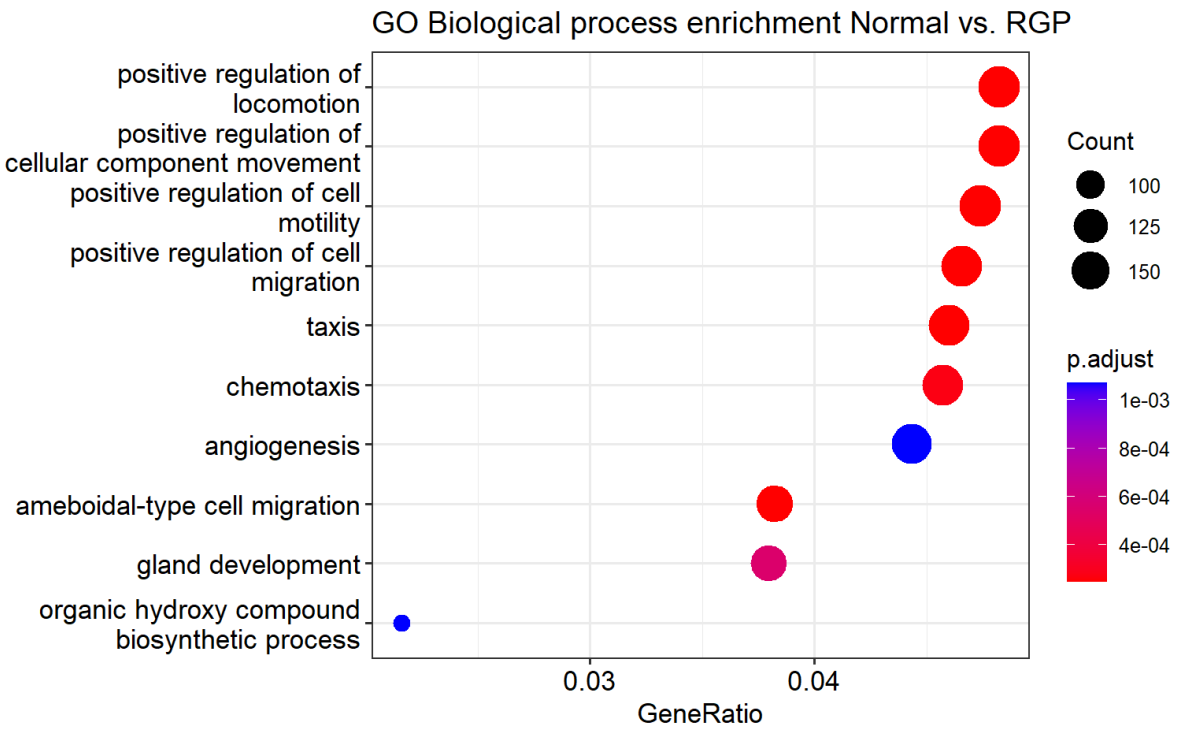


Figure 25: Dot plot shows the Gene Ontology Enrichment regarding the biological process (NHEMs vs. RGP). *The plot illustrates that the positive regulation of (locomotion, cellular component movements, cell motility, and cell migration) is the most common biological pathway in that DEGs are involved. The size of the circles represents the count of differentially expressed genes that appear in the Gene Ontology list. The color of the circle depicts the p-adjusted value.*

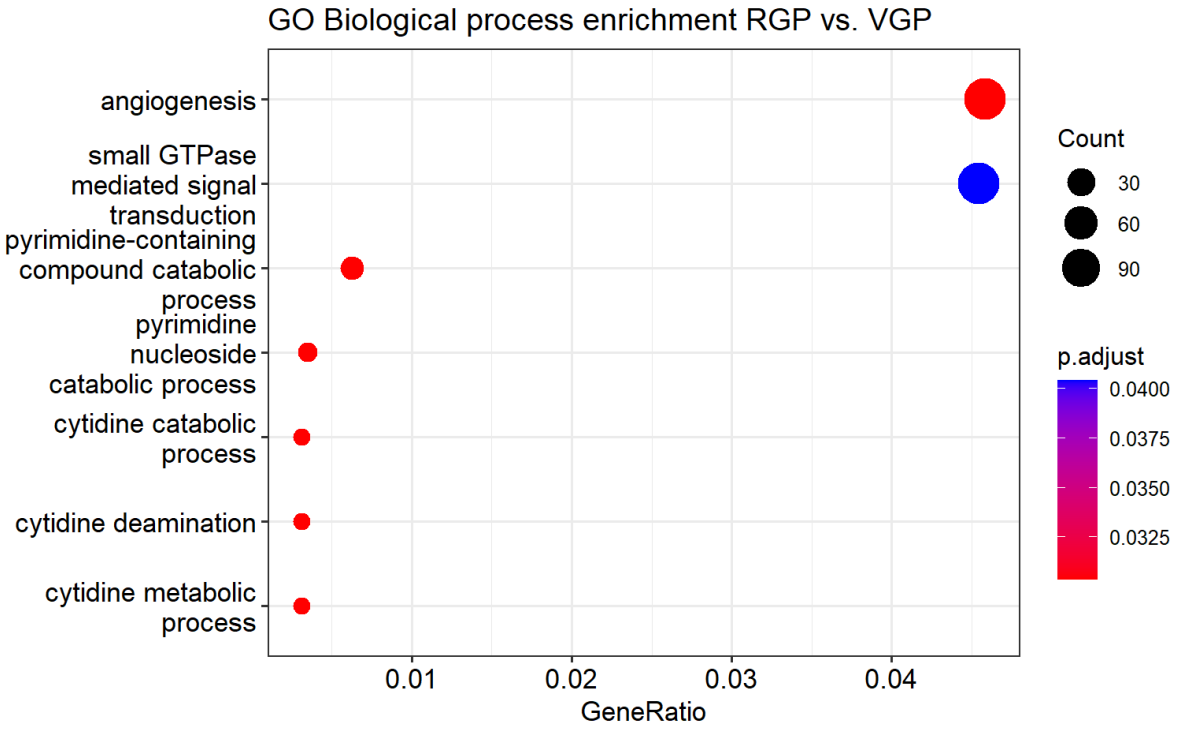


Figure 26: Dot plot shows the Gene Ontology Enrichment regarding the biological process (RGP vs. VGP). *Angiogenesis is the most common biological process the DEGs are involved in, which explains the behavior of cancer cells in this developmental stage.*

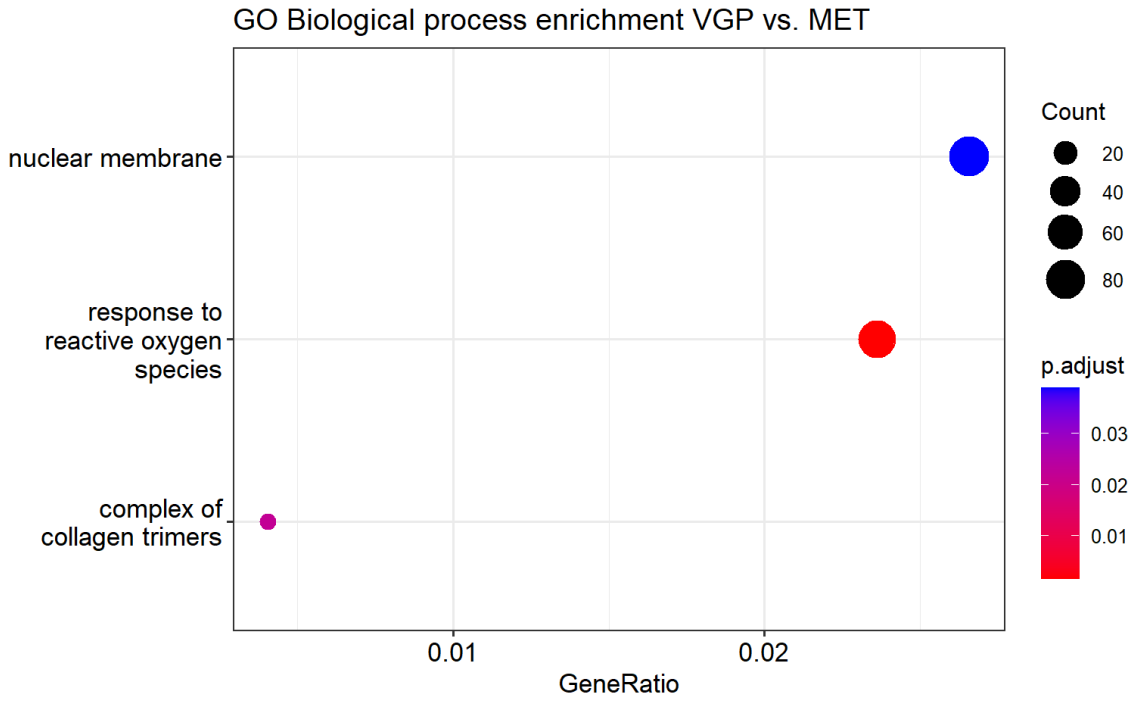


Figure 27: Dot plot shows the Gene Ontology Enrichment regarding the biological process and cellular component (VGP vs. MET). *In this developmental stage, all the Gene Ontology terms were involved in the analysis including Biological Process BP, Molecular Function, and cellular component CC.*

3.2.2 Gene Set Enrichment Analysis (GSEA)

GSEA is used for interpreting the genome-wide expression profiles. It differs from gene ontology because it doesn't consider only the significant changes in gene expression. Instead, it considers the change in gene expression of all genes in the dataset.

The gene sets used in this study were Hallmarks (Liberzon et al., 2015) gene sets and C5BP (ontology gene sets), and were downloaded from the Molecular Signature Database (MSigDB) (Liberzon et al., 2015; Subramanian et al., 2005). The metric used for ranking the genes was the log2 ratio of classes, and the permutation type was the gene set.

3.2.2.1 Using Hallmark gene sets

Hallmarks exhibits meaningful expression and summarizes well-defined biological processes (Liberzon et al., 2015).

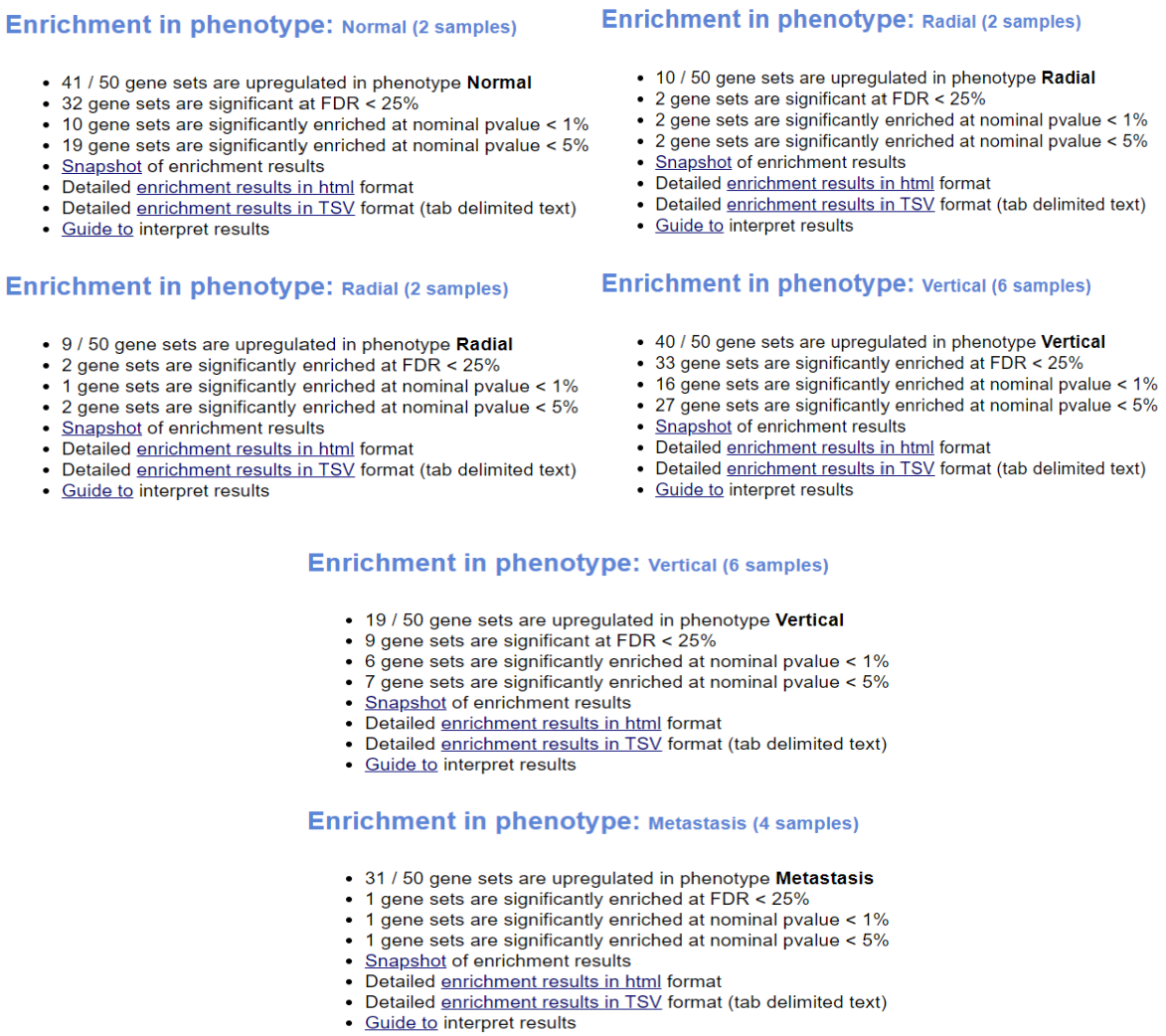


Figure 28: GSEA results using Hallmark gene sets. *In the RGP Phenotype cell line, 2 gene sets are significantly enriched at a false discovery rate of less than 0.25, whereas in the Metastasis phase, there is only one gene set.*

Table: Gene sets enriched in phenotype Normal (2 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	HALLMARK_MTORC1_SIGNALING	Details ...	200	0.41	1.94	0.000	0.022	0.010
2	HALLMARK_CHOLESTEROL_HOMEOSTASIS	Details ...	74	0.47	1.87	0.000	0.019	0.018
3	HALLMARK_ADIPOGENESIS	Details ...	192	0.38	1.79	0.000	0.024	0.034
4	HALLMARK_FATTY_ACID_METABOLISM	Details ...	142	0.40	1.79	0.000	0.020	0.038
5	HALLMARK_ANDROGEN_RESPONSE	Details ...	98	0.39	1.67	0.003	0.043	0.101
6	HALLMARK_COMPLEMENT	Details ...	177	0.35	1.64	0.000	0.044	0.125
7	HALLMARK_OXIDATIVE_PHOSPHORYLATION	Details ...	199	0.34	1.63	0.000	0.039	0.129
8	HALLMARK_PANCREAS_BETA_CELLS	Details ...	25	0.51	1.60	0.041	0.048	0.178
9	HALLMARK_PROTEIN_SECRETION	Details ...	95	0.37	1.57	0.010	0.052	0.210
10	HALLMARK_UNFOLDED_PROTEIN_RESPONSE	Details ...	112	0.36	1.54	0.011	0.057	0.246
11	HALLMARK_HEME_METABOLISM	Details ...	180	0.32	1.51	0.008	0.064	0.290
12	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	Details ...	196	0.31	1.47	0.005	0.089	0.413
13	HALLMARK_INTERFERON_ALPHA_RESPONSE	Details ...	97	0.35	1.45	0.037	0.095	0.463
14	HALLMARK_COAGULATION	Details ...	112	0.34	1.44	0.027	0.091	0.474
15	HALLMARK_PEROXISOME	Details ...	96	0.34	1.42	0.028	0.099	0.531
16	HALLMARK_IL6_JAK_STAT3_SIGNALING	Details ...	73	0.35	1.42	0.043	0.095	0.540
17	HALLMARK_XENOBIOTIC_METABOLISM	Details ...	166	0.30	1.36	0.040	0.131	0.676
18	HALLMARK_UV_RESPONSE_DN	Details ...	144	0.30	1.34	0.045	0.151	0.741
19	HALLMARK_INTERFERON_GAMMA_RESPONSE	Details ...	191	0.28	1.33	0.040	0.147	0.755
20	HALLMARK_APICAL_SURFACE	Details ...	41	0.37	1.30	0.122	0.173	0.820
21	HALLMARK_APOPTOSIS	Details ...	154	0.28	1.29	0.060	0.172	0.833
22	HALLMARK_KRAS_SIGNALING_DN	Details ...	134	0.29	1.29	0.064	0.169	0.837
23	HALLMARK_BILE_ACID_METABOLISM	Details ...	101	0.30	1.28	0.097	0.174	0.861
24	HALLMARK_MYOGENESIS	Details ...	174	0.27	1.28	0.072	0.167	0.861
25	HALLMARK_E2F_TARGETS	Details ...	200	0.26	1.25	0.080	0.184	0.891
26	HALLMARK_G2M_CHECKPOINT	Details ...	200	0.26	1.24	0.085	0.189	0.907
27	HALLMARK_IL2_STAT5_SIGNALING	Details ...	182	0.27	1.24	0.115	0.184	0.912
28	HALLMARK_HEDGEHOG_SIGNALING	Details ...	33	0.36	1.20	0.238	0.223	0.960
29	HALLMARK_MITOTIC_SPINDLE	Details ...	198	0.26	1.20	0.115	0.220	0.961
30	HALLMARK_TNFA_SIGNALING_VIA_NFKB	Details ...	196	0.25	1.19	0.141	0.222	0.967
31	HALLMARK_GLYCOLYSIS	Details ...	193	0.25	1.19	0.135	0.224	0.971
32	HALLMARK_UV_RESPONSE_UP	Details ...	154	0.25	1.16	0.169	0.248	0.988

Table: Gene sets enriched in phenotype Radial (2 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	HALLMARK_KRAS_SIGNALING_UP	Details ...	175	-0.38	-1.69	0.002	0.025	0.083
2	HALLMARK_APICAL_JUNCTION	Details ...	184	-0.31	-1.42	0.021	0.140	0.610

Table 13: Hallmark gene sets enriched in the developmental step (NHEMs vs. RGP). *The gene sets were selected with a false discovery rate of less than 0.25. In the RGP phenotype cell line, 175 genes are enriched in response to KRAS signaling _up, and 184 genes in Apical Junction.*

Table: Gene sets enriched in phenotype Radial (2 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	HALLMARK_ESTROGEN_RESPONSE_LATE	Details ...	177	0.40	1.55	0.003	0.071	0.234
2	HALLMARK_ESTROGEN_RESPONSE_EARLY	Details ...	190	0.39	1.53	0.004	0.046	0.290

Table: Gene sets enriched in phenotype Vertical (6 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	Details ...	196	-0.57	-2.52	0.000	0.000	0.000
2	HALLMARK_TNFA_SIGNALING_VIA_NFKB	Details ...	196	-0.51	-2.23	0.000	0.000	0.000
3	HALLMARK_INFLAMMATORY_RESPONSE	Details ...	172	-0.46	-1.98	0.000	0.001	0.001
4	HALLMARK_IL6_JAK_STAT3_SIGNALING	Details ...	73	-0.50	-1.91	0.000	0.001	0.002
5	HALLMARK_G2M_CHECKPOINT	Details ...	200	-0.43	-1.88	0.000	0.001	0.002
6	HALLMARK_E2F_TARGETS	Details ...	200	-0.41	-1.80	0.000	0.005	0.010
7	HALLMARK_COMPLEMENT	Details ...	177	-0.41	-1.80	0.000	0.004	0.010
8	HALLMARK_ALLOGRAFT_REJECTION	Details ...	153	-0.41	-1.78	0.000	0.004	0.010
9	HALLMARK_COAGULATION	Details ...	112	-0.43	-1.74	0.000	0.004	0.014
10	HALLMARK_UNFOLDED_PROTEIN_RESPONSE	Details ...	112	-0.41	-1.65	0.003	0.012	0.042
11	HALLMARK_HYPOXIA	Details ...	187	-0.38	-1.63	0.005	0.013	0.051
12	HALLMARK_PROTEIN_SECRETION	Details ...	95	-0.41	-1.60	0.000	0.018	0.070
13	HALLMARK_MTORC1_SIGNALING	Details ...	200	-0.36	-1.59	0.000	0.018	0.078
14	HALLMARK_UV_RESPONSE_DN	Details ...	144	-0.37	-1.54	0.012	0.027	0.124
15	HALLMARK_MYC_TARGETS_V1	Details ...	200	-0.34	-1.52	0.000	0.030	0.143
16	HALLMARK_KRAS_SIGNALING_UP	Details ...	175	-0.34	-1.50	0.000	0.033	0.167
17	HALLMARK_MYC_TARGETS_V2	Details ...	58	-0.41	-1.47	0.033	0.040	0.209

Table 14: Hallmark gene sets enriched in the developmental step (RGP vs. VGP).

Table: Gene sets enriched in phenotype Vertical (6 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	HALLMARK_E2F_TARGETS	Details ...	200	0.54	2.39	0.000	0.000	0.000
2	HALLMARK_G2M_CHECKPOINT	Details ...	200	0.52	2.33	0.000	0.000	0.000
3	HALLMARK_TNFA_SIGNALING_VIA_NFKB	Details ...	196	0.44	1.95	0.000	0.001	0.001
4	HALLMARK_MYC_TARGETS_V1	Details ...	200	0.39	1.69	0.000	0.012	0.014
5	HALLMARK_MYC_TARGETS_V2	Details ...	58	0.44	1.64	0.006	0.017	0.024
6	HALLMARK_HYPOXIA	Details ...	187	0.33	1.42	0.019	0.085	0.128
7	HALLMARK_MITOTIC_SPINDLE	Details ...	198	0.30	1.33	0.000	0.133	0.229
8	HALLMARK_INFLAMMATORY_RESPONSE	Details ...	172	0.29	1.26	0.052	0.188	0.354
9	HALLMARK_IL6_JAK_STAT3_SIGNALING	Details ...	73	0.33	1.24	0.127	0.194	0.395

Table: Gene sets enriched in phenotype Metastasis (4 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	HALLMARK_ANGIOGENESIS	Details ...	32	-0.61	-1.64	0.004	0.027	0.038

Table 15: Hallmark gene sets enriched in the developmental step (VGP vs. MET).

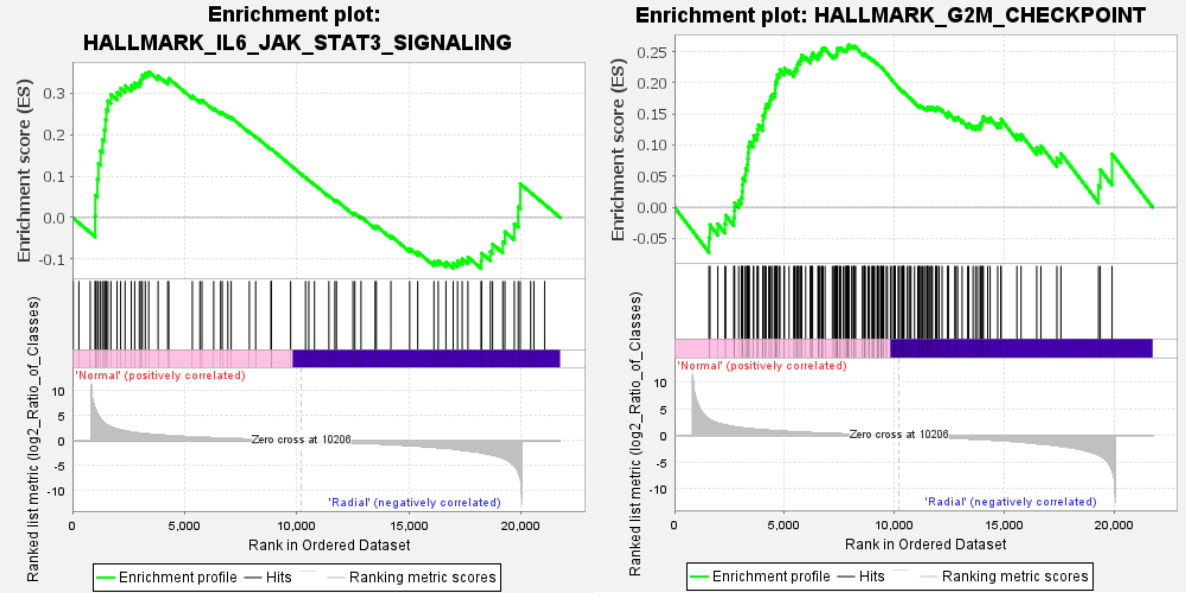


Figure 29: Enrichment plot Hallmark G2M checkpoint and Il6 JAK-STAT signalling (NHEMs vs. RGP). *Positive enrichment score indicates that the genes associated with G2M checkpoint and Il6 JAK-STAT signaling are up-regulated in the Normal Cell line. The vertical bars depict overlapping between the genes in the Hallmark gene sets and the ranked gene set.*

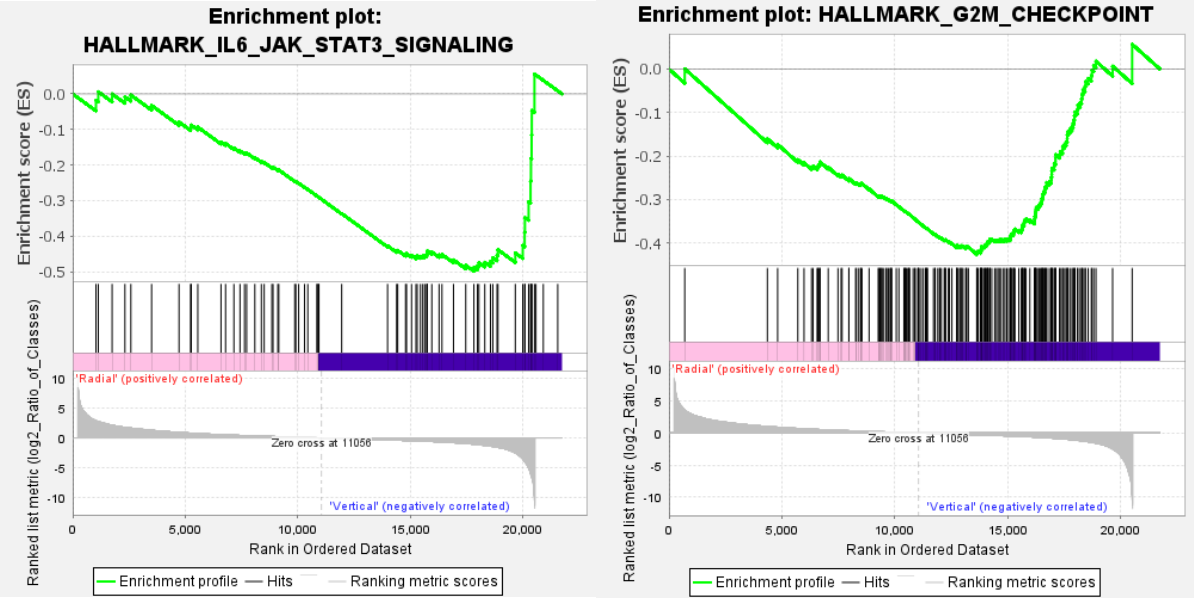


Figure 30: Enrichment plot Hallmark G2M checkpoint and Il6 JAK-STAT signalling (RGP vs. VGP). *A negative enrichment score shows that the genes associated with G2M checkpoint and Il6 JAK-STAT signaling are up-regulated in the Vertical Cell line and down-regulated in the Radial cell line.*

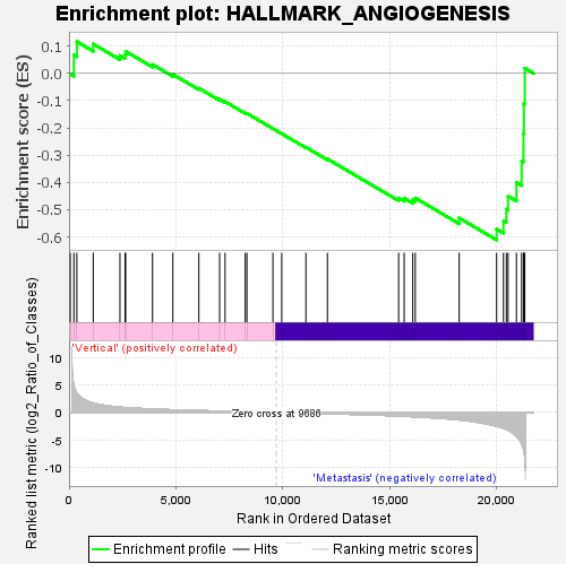


Figure 31: Enrichment plot Hallmark Angiogenesis (VGP vs. MET). *The genes associated with angiogenesis are enriched/up-regulated in the Metastasis. This process is crucial for cancer cells to form blood vessels.*

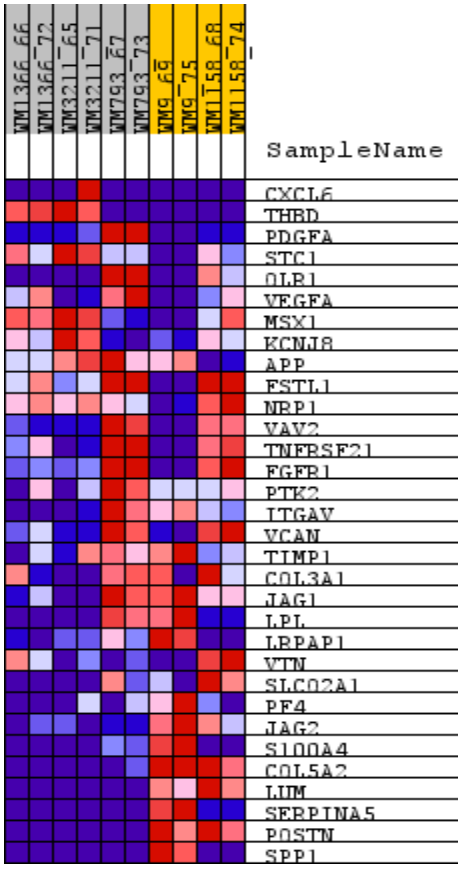


Figure 32: Heatmap of the enriched genes in the Angiogenesis in the Metastasis phase. *JAG1, JAG2, Col5A2 LUM, and POSTN show high expression in NHEMs and Metastasis cell lines.*

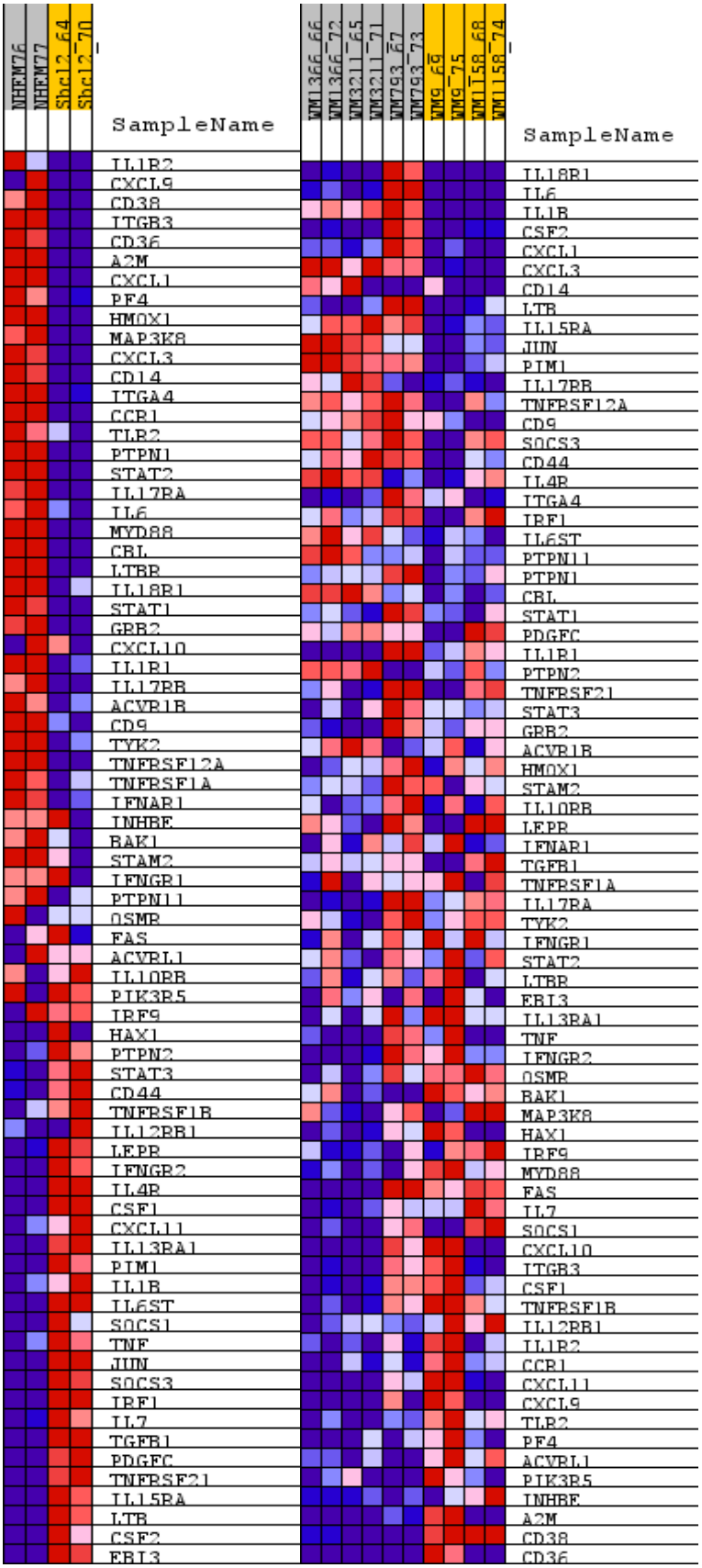


Figure 33: Heatmap of the most common genes enriched in the IL6 JAK-STAT3 in the NHEMs and VGP cell lines. *Interleukins IL6, IL7, IL6ST, IL18R1, and IL17RA show high expression in vertical growth phase cell lines.*

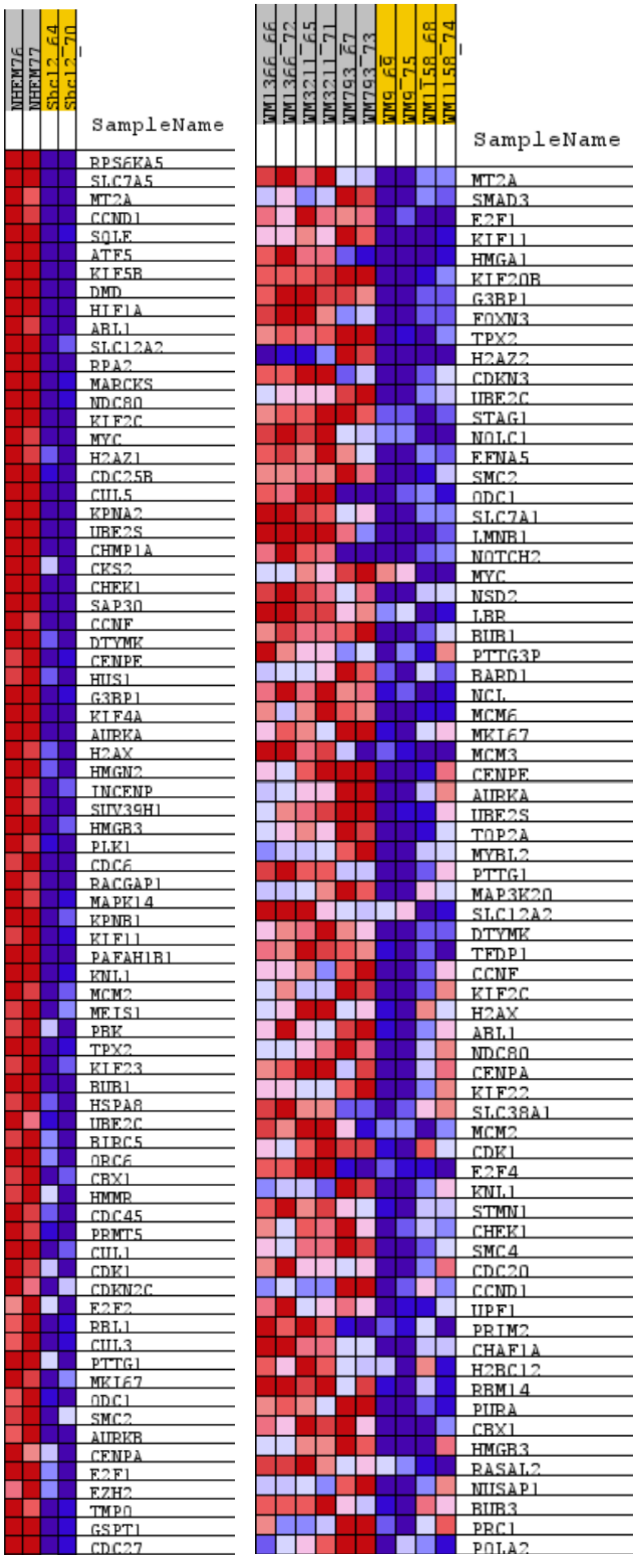


Figure 34: Heatmap of the most common genes enriched in the G2M Checkpoint in the NHEMs and VGP cell lines. *MT2A and MYC genes show high expression in both NHEMs and vertical cells.*

3.2.2.2 Using gene ontology sets C5BP

The C5:GO sub-collection is composed of three main components that represent the three root GO ontologies: biological process (BP), cellular component (CC), or molecular function (MF) (Liberzon et al., 2015; Subramanian et al., 2005).

In this analysis, the sub-collection Gene ontology biological process (C5BP) was used to figure out meaningful biological annotation of genes and their products.

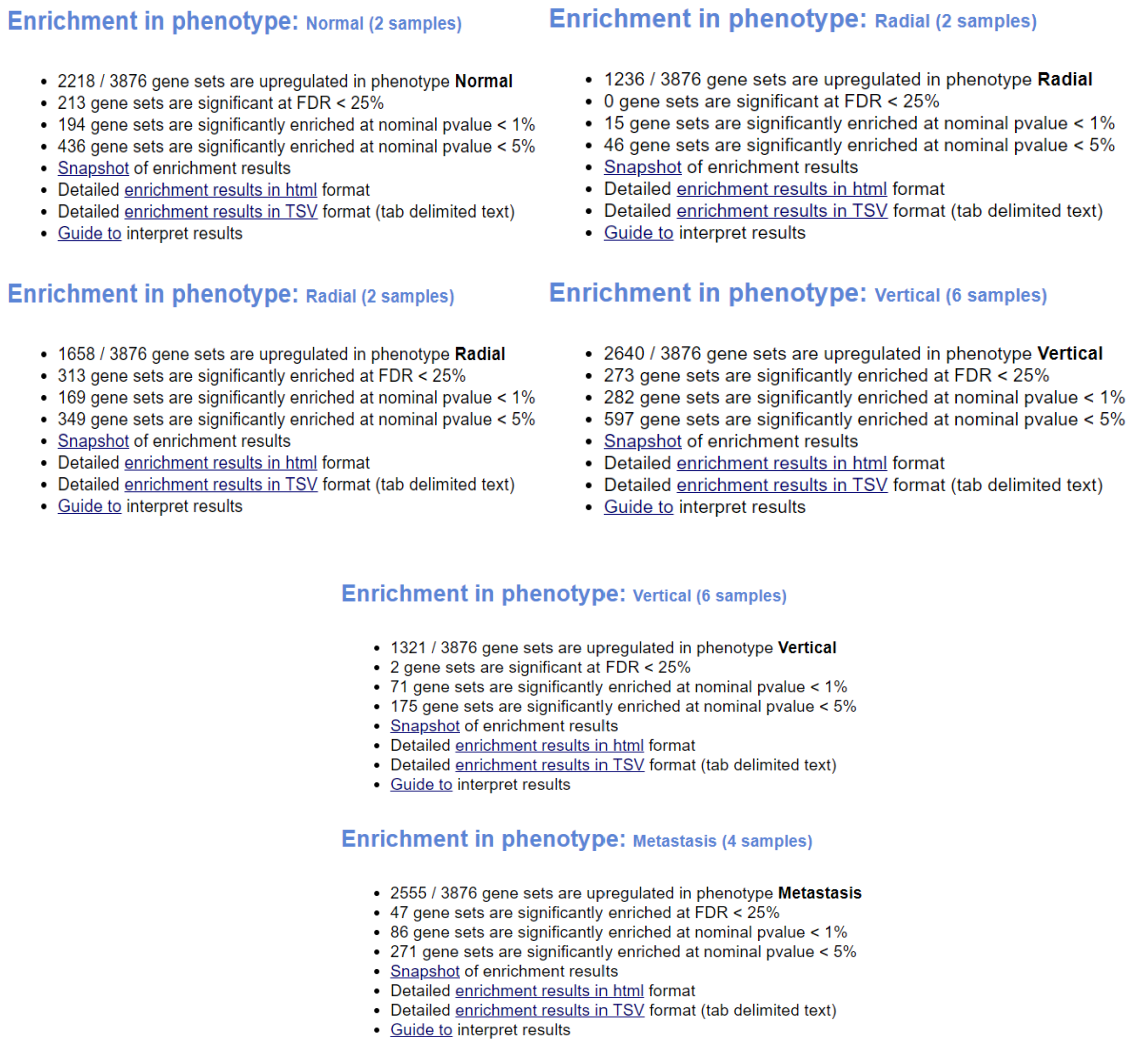


Figure 35: GSEA results using C5BP Biological Process Ontology gene sets. *In the last developmental phase (VGP vs. MET), two gene sets are significantly enriched at a false discovery rate of less than 0.25 in the Vertical cell lines. In contrast, in the Metastasis phase, 47 gene sets are significantly enriched.*

Table: Gene sets enriched in phenotype Normal (2 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	GOBP_PIGMENT_METABOLIC_PROCESS	Details ...	78	0.67	2.72	0.000	0.000	0.000
2	GOBP_PHENOL_CONTAINING_COMPOUND_BIOSYNTHETIC_PROCESS	Details ...	37	0.77	2.64	0.000	0.000	0.000
3	GOBP_PIGMENT_BIOSYNTHETIC_PROCESS	Details ...	66	0.67	2.61	0.000	0.000	0.000
4	GOBP_DEVELOPMENTAL_PIGMENTATION	Details ...	43	0.71	2.58	0.000	0.000	0.000
5	GOBP_PIGMENTATION	Details ...	95	0.59	2.49	0.000	0.000	0.000
6	GOBP_SECONDARY_METABOLITE_BIOSYNTHETIC_PROCESS	Details ...	23	0.81	2.43	0.000	0.000	0.002
7	GOBP_SECONDARY_METABOLIC_PROCESS	Details ...	43	0.70	2.43	0.000	0.000	0.002
8	GOBP_PHENOL_CONTAINING_COMPOUND_METABOLIC_PROCESS	Details ...	82	0.58	2.37	0.000	0.001	0.005
9	GOBP_PIGMENT_CELL_DIFFERENTIATION	Details ...	33	0.69	2.28	0.000	0.004	0.028
10	GOBP_NEUTROPHIL_CHEMOTAXIS	Details ...	68	0.56	2.22	0.000	0.009	0.061
11	GOBP_CGMP_MEDIATED_SIGNALING	Details ...	26	0.69	2.21	0.000	0.008	0.064
12	GOBP_PIGMENT_GRANULE_ORGANIZATION	Details ...	29	0.70	2.19	0.000	0.011	0.093
13	GOBP_REGULATION_OF_MACROAUTOPHAGY	Details ...	155	0.48	2.18	0.000	0.014	0.124
14	GOBP_CYCLIC_NUCLEOTIDE_BIOSYNTHETIC_PROCESS	Details ...	21	0.73	2.17	0.000	0.014	0.134
15	GOBP_MEMBRANE_LIPID_METABOLIC_PROCESS	Details ...	180	0.46	2.17	0.000	0.013	0.136
16	GOBP_ORGANELLE_DISASSEMBLY	Details ...	119	0.50	2.17	0.000	0.012	0.139
17	GOBP_AUTOPHAGY_OF_MITOCHONDRION	Details ...	88	0.52	2.16	0.000	0.013	0.153
18	GOBP_MELANOCYTE_DIFFERENTIATION	Details ...	23	0.69	2.15	0.002	0.013	0.158
19	GOBP_ANTIMICROBIAL_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_ANTIMICROBIAL_PEPTIDE	Details ...	34	0.64	2.14	0.000	0.016	0.202
20	GOBP_ORGANIC_HYDROXY_COMPOUND_BIOSYNTHETIC_PROCESS	Details ...	199	0.45	2.13	0.000	0.016	0.216

Table: Gene sets enriched in phenotype Radial (2 samples) [plain text format]

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX
1	GOBP_AXIS_ELONGATION	Details ...	23	-0.77	-2.24	0.000	0.000	0.000	2632
2	GOBP_REGULATION_OF_MORPHOGENESIS_OF_A_BRANCHING_STRUCTURE	Details ...	45	-0.61	-2.15	0.000	0.003	0.009	3039
3	GOBP_POSITIVE_REGULATION_OF_ERBB_SIGNALING_PATHWAY	Details ...	33	-0.65	-2.12	0.000	0.006	0.025	4351
4	GOBP_REGULATION_OF_MORPHOGENESIS_OF_AN_EPITHELIUM	Details ...	58	-0.57	-2.11	0.000	0.006	0.030	3724
5	GOBP_LUNG_EPITHELIUM_DEVELOPMENT	Details ...	27	-0.69	-2.07	0.000	0.008	0.049	2696
6	GOBP_NEURON_FATE_COMMITMENT	Details ...	45	-0.60	-2.07	0.000	0.007	0.055	2813
7	GOBP_BRANCH_ELONGATION_OF_AN_EPITHELIUM	Details ...	15	-0.78	-2.07	0.000	0.006	0.055	2632
8	GOBP_GLANDULAR_EPITHELIAL_CELL_DIFFERENTIATION	Details ...	42	-0.60	-2.06	0.002	0.007	0.065	3170
9	GOBP_REGULATION_OF_ANIMAL_ORGAN_MORPHOGENESIS	Details ...	118	-0.49	-2.04	0.000	0.008	0.089	4295
10	GOBP_MESONEPHROS_DEVELOPMENT	Details ...	84	-0.51	-2.02	0.000	0.013	0.152	5268
11	GOBP_NEPHRON_MORPHOGENESIS	Details ...	65	-0.54	-2.01	0.000	0.012	0.155	4986
12	GOBP_REGULATION_OF_MESENCHYMAL_CELL_PROLIFERATION	Details ...	25	-0.66	-2.01	0.000	0.012	0.164	2632
13	GOBP_RETINOIC_ACID_RECEPTOR_SIGNALING_PATHWAY	Details ...	28	-0.64	-2.00	0.002	0.014	0.205	2396
14	GOBP_GLAND_MORPHOGENESIS	Details ...	113	-0.48	-1.99	0.002	0.013	0.207	4124
15	GOBP_MORPHOGENESIS_OF_A_BRANCHING_STRUCTURE	Details ...	177	-0.44	-1.98	0.000	0.015	0.250	4656
16	GOBP_CELLULAR_RESPONSE_TO_RETINOIC_ACID	Details ...	56	-0.55	-1.98	0.000	0.014	0.256	3864
17	GOBP_LUNG_CELL_DIFFERENTIATION	Details ...	17	-0.74	-1.98	0.000	0.014	0.271	2696
18	GOBP_NEPHRON_DEVELOPMENT	Details ...	128	-0.46	-1.97	0.000	0.016	0.308	4562
19	GOBP_STEM_CELL_PROLIFERATION	Details ...	61	-0.53	-1.96	0.000	0.016	0.330	5642
20	GOBP_LYMPHANGIOGENESIS	Details ...	15	-0.73	-1.96	0.000	0.016	0.354	3774

Table 16: Top biological process ontology gene sets in the first developmental stage (NHEMs vs. RGP). *The gene sets were selected with a false discovery rate of less than 0.25. In the RGP phenotype cell line, 78 genes are enriched in Pigment metabolic process and 66 in Pigment biosynthetic process. While in the RGP , 25 genes are enriched in Mesenchymal cell proliferation.*

Table: Gene sets enriched in phenotype Vertical (6 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
1	GOBP_POSITIVE_REGULATION_OF_VASCULAR_ENDOTHELIAL_GROWTH_FACTOR_PRODUCTION	Details ...	25	-0.73	-2.16	0.000	0.014
2	GOBP_NEUTROPHIL_MIGRATION	Details ...	82	-0.55	-2.11	0.000	0.031
3	GOBP_NEUTROPHIL_CHEMOTAXIS	Details ...	68	-0.56	-2.08	0.000	0.036
4	GOBP_REGULATION_OF_CELL_SUBSTRATE_JUNCTION_ORGANIZATION	Details ...	65	-0.57	-2.07	0.000	0.031
5	GOBP_VASCULAR_ENDOTHELIAL_GROWTH_FACTOR_PRODUCTION	Details ...	31	-0.65	-2.07	0.000	0.028
6	GOBP_INTERLEUKIN_1_MEDIATED_SIGNALING_PATHWAY	Details ...	25	-0.68	-2.04	0.000	0.040
7	GOBP_APOPTOTIC_CELL_CLEARANCE	Details ...	37	-0.63	-2.02	0.000	0.048
8	GOBP_NEGATIVE_REGULATION_OF_CELL_JUNCTION_ASSEMBLY	Details ...	29	-0.65	-2.01	0.003	0.049
9	GOBP_REGULATION_OF_CELL_JUNCTION_ASSEMBLY	Details ...	173	-0.48	-2.01	0.000	0.045
10	GOBP_GRANULOCYTE_MIGRATION	Details ...	101	-0.50	-2.01	0.000	0.042
11	GOBP_GRANULOCYTE_CHEMOTAXIS	Details ...	85	-0.52	-2.00	0.000	0.043
12	GOBP_REGULATION_OF_EXECUTION_PHASE_OF_APOPTOSIS	Details ...	18	-0.73	-1.99	0.000	0.049
13	GOBP_REGULATION_OF_EPITHELIAL_CELL_APOPTOTIC_PROCESS	Details ...	71	-0.53	-1.98	0.000	0.054
14	GOBP_POSITIVE_REGULATION_OF_CHEMOTAXIS	Details ...	106	-0.49	-1.97	0.000	0.054
15	GOBP_POSITIVE_CHEMOTAXIS	Details ...	49	-0.58	-1.97	0.000	0.055
16	GOBP_HEART_FIELD_SPECIFICATION	Details ...	15	-0.73	-1.96	0.000	0.057
17	GOBP_NEURON_MIGRATION	Details ...	142	-0.47	-1.95	0.000	0.063
18	GOBP_MYELOID_LEUKOCYTE_MIGRATION	Details ...	155	-0.46	-1.95	0.000	0.061
19	GOBP_NEGATIVE_REGULATION_OF_EPITHELIAL_CELL_APOPTOTIC_PROCESS	Details ...	41	-0.59	-1.95	0.003	0.059
20	GOBP_PERIPHERAL_NERVOUS_SYSTEM_DEVELOPMENT	Details ...	69	-0.53	-1.95	0.000	0.057

Table 17: Top biological process ontology gene sets in the second developmental stage (RGP vs. VGP). There is no enriched gene set in the RGP cell line with a false discovery rate of less than 0.25, whereas in the VGP phenotype cell line showed 273 enriched biological processes.

Table: Gene sets enriched in phenotype Metastasis (4 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	GOBP_SECONDARY_METABOLIC_PROCESS	Details ...	43	-0.65	-1.82	0.000	0.453	0.416
2	GOBP_DIGESTION	Details ...	81	-0.59	-1.82	0.000	0.232	0.428
3	GOBP_DETECTION_OF_MECHANICAL_STIMULUS	Details ...	43	-0.65	-1.81	0.000	0.183	0.482
4	GOBP_ISOPRENOID_BIOSYNTHETIC_PROCESS	Details ...	29	-0.68	-1.79	0.000	0.209	0.642
5	GOBP_PLASMA_MEMBRANE_PHOSPHOLIPID_SCRAMBLING	Details ...	18	-0.76	-1.79	0.003	0.182	0.676
6	GOBP_CELLULAR_RESPONSE_TO_RETINOIC_ACID	Details ...	56	-0.61	-1.77	0.001	0.194	0.765
7	GOBP_INNATE_IMMUNE_RESPONSE_ACTIVATING_SIGNAL_TRANSDUCTION	Details ...	18	-0.74	-1.77	0.003	0.167	0.769
8	GOBP_NATURAL_KILLER_CELL_MEDIATED_IMMUNITY	Details ...	56	-0.62	-1.77	0.000	0.155	0.791
9	GOBP_INTERMEDIATE_FILAMENT_ORGANIZATION	Details ...	23	-0.70	-1.77	0.003	0.146	0.812
10	GOBP_DETECTION_OF_STIMULUS_INVOLVED_IN_SENSORY_PERCEPTION	Details ...	139	-0.54	-1.76	0.000	0.146	0.838
11	GOBP_FC_EPSILON_RECEPTOR_SIGNALING_PATHWAY	Details ...	21	-0.71	-1.76	0.001	0.133	0.839
12	GOBP_DETECTION_OF_MECHANICAL_STIMULUS_INVOLVED_IN_SENSORY_PERCEPTION	Details ...	26	-0.68	-1.76	0.000	0.131	0.855
13	GOBP_NEGATIVE_REGULATION_OF_CELL_KILLING	Details ...	21	-0.71	-1.75	0.003	0.123	0.861
14	GOBP_REGULATION_OF_NATURAL_KILLER_CELL_MEDIATED_IMMUNITY	Details ...	36	-0.64	-1.75	0.000	0.120	0.868
15	GOBP_MRNA_5_SPLICE_SITE_RECOGNITION	Details ...	17	-0.74	-1.75	0.001	0.112	0.869
16	GOBP_POSITIVE_REGULATION_OF_INTERLEUKIN_12_PRODUCTION	Details ...	32	-0.66	-1.75	0.000	0.113	0.882
17	GOBP_TERPENOID_BIOSYNTHETIC_PROCESS	Details ...	16	-0.75	-1.74	0.001	0.119	0.906
18	GOBP_GAMMA_DELTA_T_CELL_ACTIVATION	Details ...	17	-0.73	-1.72	0.001	0.151	0.969
19	GOBP_SECONDARY_METABOLITE_BIOSYNTHETIC_PROCESS	Details ...	23	-0.69	-1.72	0.001	0.147	0.973
20	GOBP_POSITIVE_REGULATION_OF_NATURAL_KILLER_CELL_MEDIATED_IMMUNITY	Details ...	23	-0.69	-1.72	0.000	0.154	0.977

Table: Gene sets enriched in phenotype Vertical (6 samples)

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
1	GOBP_POSITIVE_REGULATION_OF_CYTOSOLIC_CALCIUM_ION_CONCENTRATION_INVOLVED_IN_PHOSPHOLIPASE_C_ACTIVATING_G_PROTEIN_COUPLED_SIGNALING_PATHWAY	Details ...	24	0.70	2.11	0.000	0.140	0.079
2	GOBP_METAPHASE_PLATE_CONGRESSION	Details ...	66	0.54	2.03	0.000	0.222	0.227

Table 18: Top biological process ontology gene sets in the last developmental stage (VGP vs. MET). *There is only two enriched gene set in the VGP cell line with a false discovery rate of less than 0.25, whereas, in the Metastasis phenotype cell line, there are 47 GO biological processes enriched gene sets.*

Table: Snapshot of enrichment results

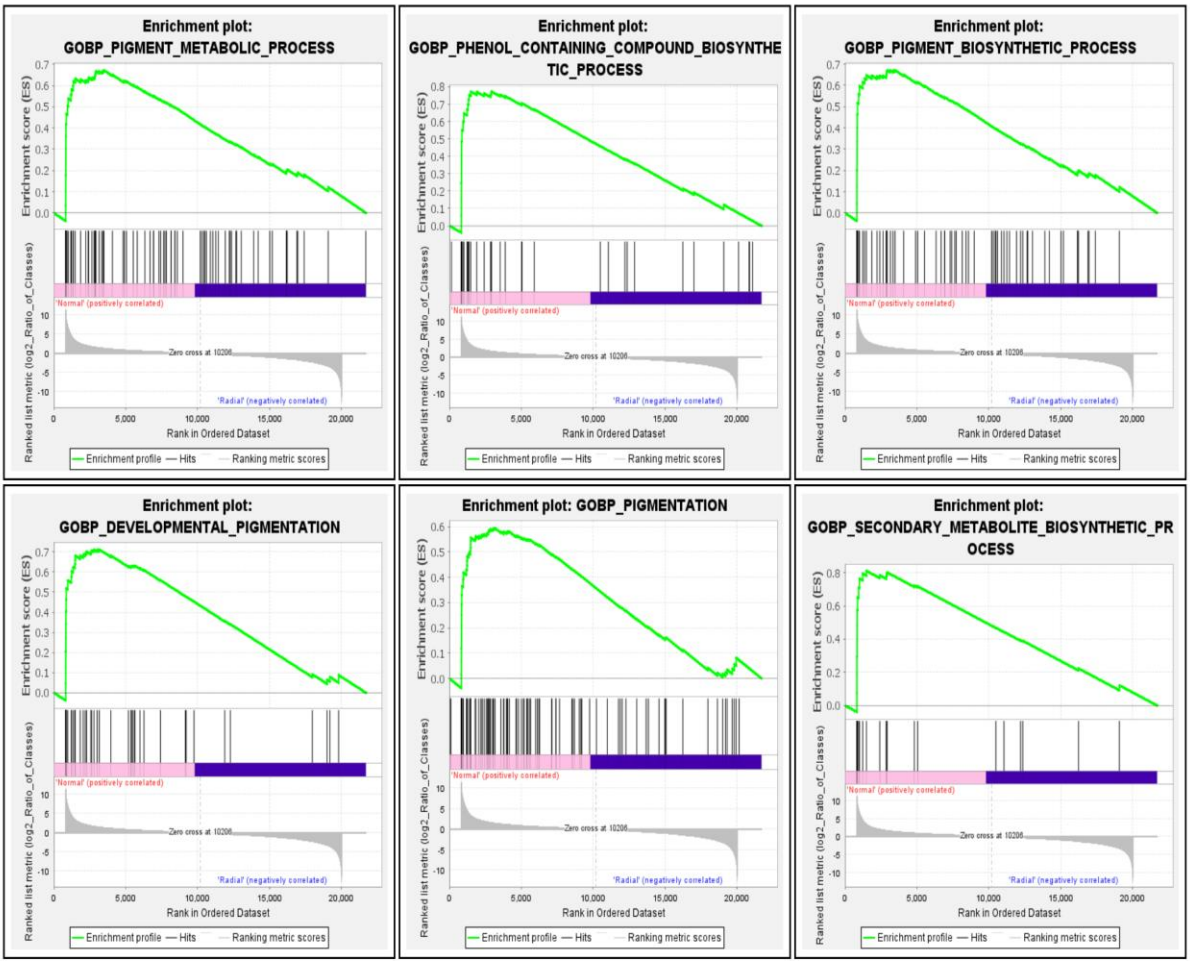


Figure 36: Snapshot of top enriched biological processes in the first developmental stage (NHEMs vs. RGP). *In NHEMs the genes associated with pigmentation, developmental pigmentation, pigment metabolic, and pigment biosynthetic processes exhibit up-regulated results.*

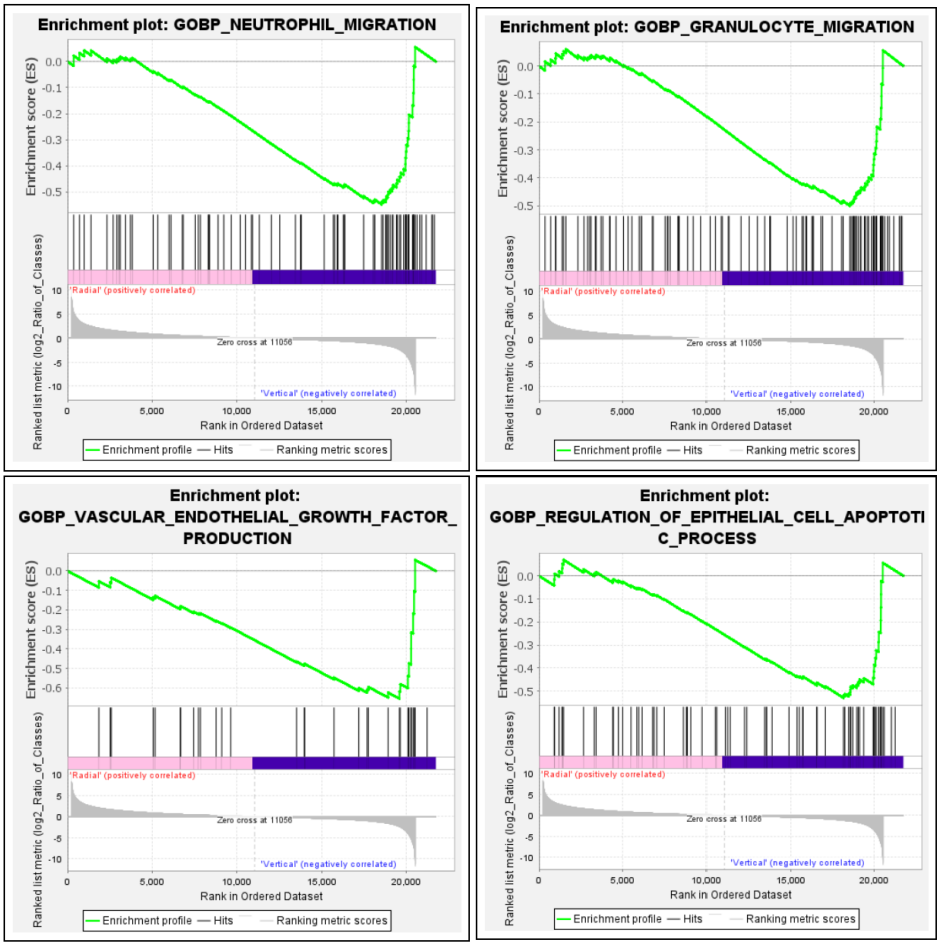


Figure 37: Snapshot of crucial enriched biological processes in the second developmental step (RGP vs. VGP). *In VGP cell line, The genes associated with neutrophil and granulocyte migration, endothelial growth factor production, and epithelial cell apoptosis processes show up-regulated results.*

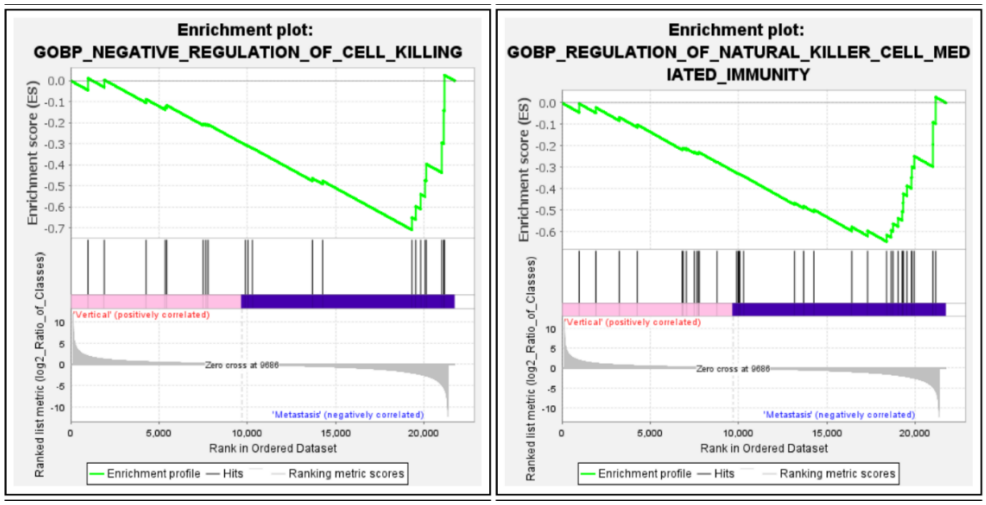


Figure 38: Snapshot of two enriched biological processes in the last developmental step (VGP vs. MET). *In MET cell line the genes associated with negative regulation of cell killing and regulation of the natural killer cell-mediated immunity show up-regulated results.*

Chapter 4: Discussion

In this chapter, we will discuss the novelty of our approach in studying melanoma progression showing how the genes expressed during the 4 steps of melanoma (NHEMs, RGP, VGP, and MET). And compare our findings with previous studies.

Melanoma is identified at late stages and has a low survival rate. Melanoma is the most aggressive form of skin cancer and the detailed molecular mechanisms of melanoma progression, development and metastasis still remain elusive (Siegel et al., 2016).

Melanoma develops when melanocytes undergo malignant change. Genetic instability and mutations play a role in the progression of melanomas. The MAPK, PI3K/PTEN/AKT, and MITF signaling pathways are some of the most significant signaling pathways implicated in the pathogenesis of melanoma (Vuković et al., 2020). BRAF and NRAS hyperactivation is found in 65% and 20% of melanomas, respectively (Davies et al., 2002). The deletion of the CDKN2A locus, which codes for the tumor suppressor genes INK4a and ARF, is another frequent genetic mutation in melanoma. Lack of INK4a results in improper cyclinD/cdk4 activation, which then promotes cell cycle progression (Daniotti et al., 2004). The MITF gene has an oncogenic function, and it is highly expressed in many melanomas (Garraway et al., 2005).

RNA-Seq data analysis allows the detection of differentially expressed genes by comparing two or more conditions like cancerous versus normal healthy cells. This might help to determine new potential predictive or diagnostic biomarkers and therapeutical targets.

In this study, we investigated 14 samples derived from either NHEMs or RGP, VGP, and MET melanoma cells. Differential expression analysis using a stage-specific approach resulted in revealing the highly expressed genes that might involve in the development of melanoma.

In the first developmental step (Normal vs. RGP), out of 26574 genes, there were 2308 genes up-regulated and 2401 genes down-regulated. An example of the DEGs with higher log2FoldChange values was PRKAA2, TNNT1, KRT18, and LIFr.

In the present study, the protein kinase AMP-activated catalytic subunit alpha 2 (PRKAA2) shows a high expression in the RGP melanoma cells. PRKAA2 is the catalytic subunit of AMP-activated protein kinase which is implicated in the control of cellular and organismal metabolism (Fisslthaler & Fleming, 2009; Kim et al., 2012). A previous study by Kim et al. figured that the expression level of PRKAA2 has a large impact on key signaling nodes regulating tumor metabolism in gastric cancer (Kim et al., 2012).

We found that TNNT1 exhibits a higher expression in the RGP melanoma cells. Troponin T1 (TNNT1), is a subunit of troponin that has been connected to neuromuscular disorders (Y. Chen et al., 2020). It

was recently discovered that TNNT1 promotes the growth of breast cancer cells (Tutar, 2014). A study by Y. Chen et al. found that TNNT1 is highly expressed in colorectal cancer cell lines and associated with poor survival (Y. Chen et al., 2020). Additionally, the knockdown of the TNNT1 gene significantly slowed down invasion, migration, and proliferation while also inducing apoptosis (Y. Chen et al., 2020).

In RGP melanoma cells, we also found a higher expression of Keratin 18. Moreover, previous studies showed that Keratin 18 (KRT18) is crucial for several cellular functions, including sustaining the cytoplasmic and mitochondrial structural integrity and tolerating external stress (Weng et al., 2012). Lately, it was revealed that Keratin 18 works as an oncogene and exhibits abnormal expression in several human malignancies (P. Wang et al., 2021; Zhang et al., 2019). As well, another study by T. Liu et al. indicates that KRT18 is significantly expressed in melanoma tissues (T. Liu et al., 2020). Honokiol is a naturally occurring biphenolic substance that is derived from *Magnolia Officinalis* leaves and bark, and it has been shown to have many pharmacological benefits, including anti-aging, anti-oxidant, and anti-cancer actions (Rauf et al., 2018; Sarrica et al., 2018). T. Liu et al. demonstrated that honokiol therapy effectively reduced KRT18 protein level and inhibited tumor growth in mouse models using melanoma cell-derived xenografts. Therefore, honokiol has the potential to be a KRT18 inhibitor and KRT18 plays an oncogenic function in melanoma (T. Liu et al., 2020).

Furthermore, we have revealed that LIFr shows a high expression in RGP melanoma cells. The leukemia inhibitory factor (LIF), a member of the IL-6 family, is a multifunctional cytokine that impacts cell division and proliferation (Kamohara et al., 2007; Kellokumpu-Lehtinen et al., 1996) . Additionally, adult skin homeostasis and hyperproliferative skin diseases are significantly influenced by LIF (Szepietowski et al., 2004). LIF receptor (LIFr) activation controls metastatic behavior, and LIFr expression gradually increases from primary melanoma to metastatic melanoma (Wysoczynski et al., 2007). Gou et al. indicate that, through the STAT3 pathway, LIFr activation might promote melanoma cell migration and higher expression of LIFr is connected with poor survival (Guo et al., 2015). In addition, LIFr knockdown was associated with STAT3 suppression and inhibited melanoma cell migration, and reduced stress fiber formation. They suggested that LIFr could be a potential target for the development of early intervention treatment (Guo et al., 2015).

In the second developmental step (RGP vs. VGP), there were 1903 genes up-regulated and 1308 genes down-regulated. We will discuss the DLC1, SLC38A1, and MCAM genes which were up-regulated in the VGP melanoma cells. In the vertical growth phase, the melanoma cells grow vertically (Urso, 2004). When melanoma enters the vertical growth phase, it gained the ability to metastasize, and the melanoma cells display as a pigmented nodule supervening on a pre-existing lesion (Barnhill et al., 1996; Clark Jr et al., 1984, 1989).

In this study, we found that the Sodium-coupled neutral amino acid transporter 1 (SLC38A1) was highly expressed in the VGP melanoma cells. A previous study by Böhme-Schäfer et al. demonstrated a high

increase of SLC38A1 expression in human melanoma tissue compared to the healthy epidermis, as well as in melanoma cell lines compared to healthy human melanocytes (NHEMs). They depicted that (methylamino) isobutyric acid (MeAIB) and siRNA-mediated downregulation of SLC38A1 functionally inhibit cancer cell growth, cellular migration, and invasion, and induces aging process in melanoma cells. Taken together, SLC38A1 seems to be necessary for the development of cancer and thus a potential cancer treatment target (Böhme-Schäfer et al., 2022).

Additionally, we demonstrated that DLC1 shows a high expression in the VGP melanoma cells. The Rho GTPase-activating protein (RhoGAP), DLC1, has been recognized as a tumor suppressor, it is commonly downregulated in a variety of cancer types and restoring its expression in cancer cell lines inhibited tumorigenic growth (Y.-C. Liao & Lo, 2008; Popescu & Goodison, 2014). As well, a prior study revealed that a decrease in DLC1 expression in primary cutaneous and metastatic melanomas is connected to poor survival, indicating that DLC1 might prevent the growth and spread of melanoma (Sjoestroem et al., 2014). Also, Yang et al. found that most melanoma tissues showed high DLC1 expression, which was found in both the cytoplasm and nucleus of the cells (Yang et al., 2020). In addition, Lieu et.al showed that the asymmetric location of the DLC1 is crucial for the migration of trunk neural crest cells in a directional manner (J. A. Liu et al., 2017). So, it is still unclear whether DLC1 is an oncogene or a tumor-suppressor gene in melanoma (Yang et al., 2020).

Interestingly, we revealed that the melanoma cell adhesion molecule (MCAM) shows an increased expression in the VGP melanoma cells in this study and it was down-regulated in the RGP melanoma cells, which indicates that MCAM could be involved in the progression of melanoma from the RGP to VGP. CD146 and MUC18 are also referred to MCAM, a newly found crucial member of the cell adhesion molecule (CAM) family (Z. Wang & Yan, 2013). The MCAM/CD146 was first identified as a marker of melanoma progression and has been suspected of being directly associated with the metastatic process of melanoma (Schlagbauer-Wadl et al., 1999). Also, studies demonstrated that overexpressed CD146 is directly linked to a poor patient prognosis and has a significant role in promoting the progression of metastatic melanomas (Ishikawa, Wondimu, Oikawa, Gentilcore, et al., 2014; Ishikawa, Wondimu, Oikawa, Ingerpuu, et al., 2014). In addition, recent research has shown that inflammation can accelerate the progression of melanoma through a variety of mechanisms, including tumor initiation, angiogenesis, and metastasis (Coffelt & de Visser, 2014; Maru et al., 2014). Lei et al., 2015 suggested that CD146 might be a useful potential biomarker for inflammation, and a factor that influences the progression of melanoma by controlling the level of inflammation (Lei et al., 2015). Since the MCAM could be involved in the transition from the RGP to VGP and involved in the metastatic processes of melanoma we could conclude that MCAM could be a potential biomarker, and the knockout of MCAM could inhibit the progression of melanoma.

Signalling pathway	Progression	Mechanism
TNF-α-NF-κB-CD146 CD146-NF-κB	inflammation	promotion proinflammatory leukocyte extravasations
CD146-PI3K–AKT-CD146	survival	inhibition of the pro-apoptotic protein BAD, resistance to staurosporine-induced cell death, and the cleavage of caspase 3
PAR1-PAFR-CD146	metastasis	promotion of heterotypic adhesion, diapedesis, and retention of the ability for metastasis
CD146-ATF-3-Id-1-MMP2	invasion	cleavage or degradation of the extracellular matrix to invade surrounding tissues
CD146/moesin/RhoGDI1- RhoA-PI4P5K-PIP2-CD146/moesin/RhoGDI1/PIP2-actin CD146-IL-6-p38α-MAPK-Wnt5a-CD146/DVL2/Fz3-WRAMP	motility	direction of tail-end membrane retraction, and the forward translocation of the cell body; degradation of focal adhesions and disassembly of stress fibers
CD146-NF-κB p50-IL-6-VEGF	angiogenesis	Promotion of endothelial proliferation and the development of capillary-like structures

Table 19: Pathways influencing melanoma progression and CD146-correlated signals (Lei et al., 2015)**.** *The CD146 is involved in signalling pathways that induce the progression of melanoma. For instance, the CD146-ATF-3-Id-1-MMP2 signalling pathway impacts the degradation of the extracellular matrix to invade surrounding tissues. The CD146-NF-κB p50-IL-6-VEGF signalling pathway induces the proliferation and the development of capillary-like structures in the angiogenesis process* (Lei et al., 2015)**.**

In our investigation, there were 1793 up-regulated genes and 1308 down-regulated genes in the last developmental step (VGP vs. MET). The signs of malignant melanoma are varied, and it is known to metastasize to all organs of the human body. It is nearly impossible to forecast which organ system melanoma will invade from a specific source location (Y.-T. N. Lee, 1980). In the present study, we depicted up-regulated genes in MET melanoma cells such as POSTN, CTCFL, ZFHX4, and COL1A2.

Our results demonstrated that periostin (POSTN) exhibits a high expression in MET melanoma cells. Periostin (POSTN) is an extracellular matrix (ECM) protein that interacts with other ECM proteins and involves in the control of intercellular adhesion (Jia et al., 2021). Numerous cancer types, including malignant pleural mesothelioma, renal cell carcinoma, and non-small cell lung cancer, have been linked to periostin overexpression (Dahinden et al., 2010; Soltermann et al., 2008). Periostin binding to integrins increases the signaling pathways regulated by Akt/PKB, enhancing cell survival, angiogenesis, invasion, metastasis, and critically, the epithelial-mesenchymal transition of cancer cells (Morra & Moch, 2011). EMT, which happens when epithelial cells lose their epithelial properties and gain mesenchymal characteristics, is frequently linked to the development of invasive properties, the potential for metastasis, and drug resistance in cancer cells (Polyak & Weinberg, 2009; Tian et al., 2018). Furthermore, a study demonstrated that POSTN induces melanoma progression by altering the melanoma microenvironment and could be a potential therapeutic target for malignant melanoma (Kotobuki et al., 2014). In line with previous studies, POSTN is linked with oncogenesis and progression of cancer cells and could be a potential biomarker for tumor identification and therapy.

Another promising finding in our study is the CTCFL which shows high expression in MET melanoma cells. The CCCTC binding factor-Like (CTCFL) refers also to the transcriptional modulator called Brother of Regulator of Imprinted Sites (BORIS) (Janssen et al., 2020). BORIS has the power to change transcription by serving as a transcriptional activator, invoking a transcriptional activator, changing DNA methylation and histone modifications, or employing proteins that modify chromatin (Janssen et al., 2020). The processes through which BORIS can change transcription depend on its ability to bind the DNA at particular binding sites (Bergmaier et al., 2018; Sleutels et al., 2012). Janssen et al. figured that Boris can change melanoma cells' gene expression program to promote a more invasive phenotype (Janssen et al., 2020). In accordance with these findings, they demonstrated that BORIS expression decreased melanoma cell proliferation while increasing their capacity for migration and invasion m, which means that BORIS impacts melanoma cells' transition from a proliferative to an invasive state at both the transcriptional and phenotypic levels (Janssen et al., 2020).

In the present study, ZFHX4 displays a higher expression in metastasis cell lines. ZFHX4 is expected to activate transcription factors that bind to DNA in the nucleus (Zong et al., 2022). Confirming our findings, Qing et al. figured that ZFHX4 was overexpressed in tumor tissues when compared to healthy controls (Qing et al., 2017). They found that ZFHX4 knockdown significantly reduced cell invasion and migration in vitro. In addition, they depicted that mutations in this gene are closely linked to poor survival and down-regulation slows the development of esophageal squamous cell carcinoma (Qing et al., 2017). Furthermore, Wang et al. discovered that patients with metastatic colorectal cancer had higher ZFHX4 amplifications than those without metastases (Z. Wang et al., 2021). Following these

results, ZFHX4 might be a potential biomarker target to decrease the invasion and migration of cancer cells.

In the current research, MET melanoma cells have higher expression of COL1A2. The collagen type I alpha 2 chain (COL1A2) is present in the majority of connective tissue and embryonic tissue (Cole, 1994). The abnormal expression of COL1A2 has been discovered in numerous cancers (Bonazzi et al., 2011; Ibanez de Caceres et al., 2006). In agreement with our findings, Li et al. found that COL1A2 mRNA expression was considerably higher in malignant tissues and was associated with a lower overall survival rate in gastric cancer (Li et al., 2016). Additionally, they figured that Tumor size and depth of invasion were both correlated with COL1A2 mRNA expression (Li et al., 2016).

Functional annotation is the matching of biological data to gene sequences and describing a gene's biological activity (Berardini et al., 2004; Mudge & Harrow, 2016). In our analysis, we used two approaches, over-representation analysis (ORA) and gene set enrichment analysis (GSEA).

Beginning with over-representation analysis, we found that the biological processes of angiogenesis, positive regulation of cell migration, positive regulation of cell motility, epithelial cell migration, and epithelial cell differentiation were observed in the first developmental step (Normal vs. RGP).

Interestingly, in the over-representation analysis, the angiogenesis process was figured in the next developmental stage VGP. There were 117 from our DEGs list Out of 2550 genes from the GO list involved in the angiogenesis biological process. The angiogenesis process is the formation of new blood vessels from pre-existing blood vessels which is highly controlled and involved in several physiological and pathological processes (Cho et al., 2019). Angiogenesis is a crucial marker of tumor aggressiveness and a poor clinical prognosis in a variety of solid tumors, including melanoma and plays a significant role in tumor proliferation, survival, and distant metastasis (Cho et al., 2019).

In gene set enrichment analysis (GSEA), we demonstrated an exciting finding from the results of comparing our gene list to the Hallmark database which showed that the genes associated with angiogenesis were enriched in the metastatic phase. We found that JAG1, JAG2, Col5A2, LUM, and POSTN genes were involved in the angiogenesis process. In agreement with our result, a previous study demonstrated that JAG2 might play a crucial role in inducing Notch activity, growth, and metastasis in uveal melanoma (Asnaghi et al., 2013). Another study showed that JAG1 promotes adrenocortical carcinoma cell proliferation and tumor aggressiveness via activation of Notch signaling in adjacent cells (Simon et al., 2012). POSTN is thought to have tumor-progressive properties because of how it interacts with cell surface receptors, promotes tumor angiogenesis, and modifies signal transduction pathways (Baril et al., 2007; Choi et al., 2011).

Finally, we demonstrated that genes involved in the IL-6/JAK/STAT3 pathway are significantly deregulated in VGP melanoma cells compared to RGP and thus indicate to be involve in the gain of

invasive properties and the interleukins IL6, IL7, IL6ST, IL18R1, and IL17RA show high expression VGP melanoma cells. In agreement with our finding, a previous study demonstrated that IL-6 is known to induce tumor progression by preventing apoptosis and enhancing tumor angiogenesis (Hoejberg et al., 2012). The high levels of IL-6 serum concentration are associated with a poor prognosis in individuals with many malignancies, including melanoma (Hoejberg et al., 2012). Drugs that target IL-6, the IL-6 receptor, or JAKs have already received FDA approval (Johnson et al., 2018). An investigation found that the inactivation of IL-17RA by small hairpin RNA (shRNA) decreased cell proliferation, migration, and invasion (Y.-S. Chen et al., 2019).

Furthermore, previous studies demonstrated that, in many cancer types, the IL-6/JAK/STAT3 pathway is abnormally hyperactivated, and this hyperactivation is typically correlated with a poor clinical prognosis (Johnson et al., 2018). IL-6/JAK/STAT3 signaling drives the growth, survival, invasiveness, and metastasis of tumor cells in the tumor microenvironment while severely inhibiting the antitumor immune response. The IL-6/JAK/STAT3 pathway is therefore set to offer therapeutic advantage by directly suppressing tumor cell development and by inducing antitumor immunity in cancer patients who receive medications that target it (Johnson et al., 2018).

Chapter 5: Conclusion

In this thesis, we have successfully analyzed RNA-Seq data for primary melanoma and metastatic melanoma cell lines and compared the gene expression patterns in a melanoma- stage dependent manner. Therefore, stage dependent genesets of the identified DE genes were created comparing NHEMs vs. RGP, RGP vs. VGP, and finally VGP vs. MET. After identifying DEGs we further determined the overlapping genes of the 3 developmental steps and specified the specific DE for each developmental step.

Further, GO over representation analysis showed that the biological processes such as positive regulation of cell migration, locomotion, cell motility and cellular component movement, mesenchyme development, epithelial cell migration, and epithelial cell differentiation were enriched in the RGP cell line. Whereas, the angiogenesis process was enriched in the VGP cell line.

In the gene set enrichment analysis (GSEA), the genes involved in IL-6/JAK/STAT3 signaling pathway such as interleukins IL6, IL7, IL6ST, IL18R1, and IL17RA were enriched in the VGP cell line, which promotes tumor cell proliferation, survival, invasiveness, and metastasis.

The genes associated with the angiogenesis process like JAG1, JAG2, Col5A2 LUM, and POSTN were enriched in the MET melanoma cells which have associations in melanoma progression and proliferation.

Finally, further investigation is required to figure out the value of these genes for a more comprehensive annotation of the identified enriched genes and the biological process they are involved.

Chapter 6: Bibliography

Akalin, A. (2020). *Computational Genomics with R*. CRC Press. https://books.google.de/books?id=fK0PEAAAQBAJ

American Cancer Society, Inc. (2022, February 14). *What Is Cancer?* Cancer.Org. https://www.cancer.org/treatment/understanding-your-diagnosis/what-is-cancer.html#written_by

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Nature Precedings*, *5*. https://doi.org/10.1038/npre.2010.4282.2

Andrews, S. (2014). FastQC A Quality Control tool for High Throughput Sequence Data. *Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/*.

Annie Stuart. (2021, July 25). *Benign Tumors*. WebMD. https://www.webmd.com/a-to-z-guides/benign-tumors-causes-treatments

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Asnaghi, L., Handa, J. T., Merbs, S. L., Harbour, J. W., & Eberhart, C. G. (2013). A Role for Jag2 in Promoting Uveal Melanoma Dissemination and Growth. *Investigative Opthalmology & Visual Science*, *54*(1), 295. https://doi.org/10.1167/iovs.12-10209

Baril, P., Gangeswaran, R., Mahon, P. C., Caulee, K., Kocher, H. M., Harada, T., Zhu, M., Kalthoff, H., Crnogorac-Jurcevic, T., & Lemoine, N. R. (2007). Periostin promotes invasiveness and resistance of pancreatic cancer cells to hypoxia-induced cell death: role of the β4 integrin and the PI3k pathway. *Oncogene*, *26*(14), 2082–2094.

Barnhill, R. L., Fine, J. A., Roush, G. C., & Berwick, M. (1996). Predicting five-year outcome for patients with cutaneous melanoma in a population-based study. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *78*(3), 427–432.

Berardini, T. Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L. A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miller, N., Weems, D., & Rhee, S. Y. (2004). Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiology*, *135*(2), 745–755. https://doi.org/10.1104/pp.104.040071

Bergmaier, P., Weth, O., Dienstbach, S., Boettger, T., Galjart, N., Mernberger, M., Bartkuhn, M., & Renkawitz, R. (2018). Choice of binding sites for CTCFL compared to CTCF is driven by chromatin and by sequence preference. *Nucleic Acids Research*, *46*(14), 7097–7107.

Bertolotto, C. (2013). Melanoma: from melanocyte to genetic alterations and clinical options. *Scientifica*, *2013*.

BG, E., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, *8*, 186–194. https://doi.org/10.1101/gr.8.3.186

Böhme-Schäfer, I., Lörentz, S., & Bosserhoff, A. K. (2022). Role of Amino Acid Transporter SNAT1/SLC38A1 in Human Melanoma. *Cancers*, *14*(9), 2151. https://doi.org/10.3390/cancers14092151

Bonazzi, V. F., Nancarrow, D. J., Stark, M. S., Moser, R. J., Boyle, G. M., Aoude, L. G., Schmidt, C., & Hayward, N. K. (2011). Cross-platform array screening identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as genes frequently silenced by methylation in melanoma. *PloS One*, *6*(10), e26121.

Breitkreutz, D., Koxholt, I., Thiemann, K., & Nischt, R. (2013). Skin Basement Membrane: The Foundation of Epidermal Integrity—BM Functions and Diverse Roles of Bridging Molecules Nidogen and Perlecan. *BioMed Research International*, *2013*, 1–16. https://doi.org/10.1155/2013/179784

Buescher, J. M., & Driggers, E. M. (2016). Integration of omics: more than the sum of its parts. *Cancer & Metabolism*, *4*(1), 4. https://doi.org/10.1186/s40170-016-0143-y

Burns, T., Breathnach, S., Cox, N., & Griffiths, C. (2004). *Rook's Textbook of Dermatology: 4 Volume Set*. Wiley. https://books.google.de/books?id=tBrjzAEACAAJ

Carr, S., Smith, C., & Wernberg, J. (2020). Epidemiology and risk factors of melanoma. *Surgical Clinics*, *100*(1), 1–12.

Chen, J., Wu, F., Shi, Y., Yang, D., Xu, M., Lai, Y., & Liu, Y. (2019). Identification of key candidate genes involved in melanoma metastasis. *Molecular Medicine Reports*, *20*(2), 903–914.

Chen, Y., Wang, J., Wang, D., Kang, T., Du, J., Yan, Z., & Chen, M. (2020). TNNT1, negatively regulated by miR-873, promotes the progression of colorectal cancer. *The Journal of Gene Medicine*, *22*(2), e3152. https://doi.org/10.1002/jgm.3152

Chen, Y.-S., Huang, T.-H., Liu, C.-L., Chen, H.-S., Lee, M.-H., Chen, H.-W., & Shen, C.-R. (2019). Locally Targeting the IL-17/IL-17RA Axis Reduced Tumor Growth in a Murine B16F10 Melanoma Model. *Human Gene Therapy*, *30*(3), 273–285. https://doi.org/10.1089/hum.2018.104

Cho, W. C., Jour, G., & Aung, P. P. (2019). Role of angiogenesis in melanoma progression: Update on key angiogenic mechanisms and other associated components. *Seminars in Cancer Biology*, *59*, 175–186. https://doi.org/10.1016/j.semcancer.2019.06.015

Choi, K. U., Yun, J. S., Lee, I. H., Heo, S. C., Shin, S. H., Jeon, E. S., Choi, Y. J., Suh, D., Yoon, M., & Kim, J. H. (2011). Lysophosphatidic acid-induced expression of periostin in stromal cells: Prognoistic relevance of periostin expression in epithelial ovarian cancer. *International Journal of Cancer*, *128*(2), 332–342.

Cichorek, M., Wachulska, M., Stasiewicz, A., & Tymińska, A. (2013). Skin melanocytes: biology and development. *Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii*, *30*(1), 30–41.

Clark Jr, W. H., Elder, D. E., Guerry IV, D., Braitman, L. E., Trock, B. J., Schultz, D., Synnestvedt, M., & Halpern, A. C. (1989). Model predicting survival in stage I melanoma based on tumor progression. *JNCI: Journal of the National Cancer Institute*, *81*(24), 1893–1904.

Clark Jr, W. H., Elder, D. E., Guerry IV, D., Epstein, M. N., Greene, M. H., & van Horn, M. (1984). A study of tumor progression: the precursor lesions of superficial spreading and nodular melanoma. *Human Pathology*, *15*(12), 1147–1165.

Coffelt, S. B., & de Visser, K. E. (2014). Inflammation lights the way to metastasis. *Nature*, *507*(7490), 48–49.

Cole, W. G. (1994). Collagen genes: mutations affecting collagen structure and expression. *Progress in Nucleic Acid Research and Molecular Biology*, *47*, 29–80.

Conesa, A., & Beck, S. (2019). Making multi-omics data accessible to researchers. *Scientific Data*, *6*(1), 251. https://doi.org/10.1038/s41597-019-0258-4

Consortium, G. O. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(suppl_1), D258–D261. https://doi.org/10.1093/nar/gkh036

Cooper, G. M. (2000). *The Cell: A Molecular Approach. 2nd edition*. Sinauer Associates 2000. https://www.ncbi.nlm.nih.gov/books/NBK9963/

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLOS ONE*, *12*(12), e0190152. https://doi.org/10.1371/journal.pone.0190152

Cui, R., Widlund, H. R., Feige, E., Lin, J. Y., Wilensky, D. L., Igras, V. E., D'Orazio, J., Fung, C. Y., Schanbacher, C. F., & Granter, S. R. (2007). Central role of p53 in the suntan response and pathologic hyperpigmentation. *Cell*, *128*(5), 853–864.

Dahinden, C., Ingold, B., Wild, P., Boysen, G., Luu, V.-D., Montani, M., Kristiansen, G., Sulser, T., Bühlmann, P., & Moch, H. (2010). Mining Tissue Microarray Data to Uncover Combinations of Biomarker Expression Patterns that Improve Intermediate Staging and Grading of Clear Cell Renal Cell CancerBiomarker Combinations in Staging and Grading of ccRCC. *Clinical Cancer Research*, *16*(1), 88–98.

Daniotti, M., Oggionni, M., Ranzani, T., Vallacchi, V., Campi, V., di Stasi, D., Torre, G. della, Perrone, F., Luoni, C., & Suardi, S. (2004). BRAF alterations are associated with complex mutational profiles in malignant melanoma. *Oncogene*, *23*(35), 5968–5977.

Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., & Bottomley, W. (2002). Mutations of the BRAF gene in human cancer. *Nature*, *417*(6892), 949–954.

Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin cancer detection: a review using deep learning techniques. *International Journal of Environmental Research and Public Health*, *18*(10), 5479.

Dobin, A., Davis, C., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*. https://doi.org/10.1093/bioinformatics/bts635

Elder, D. E., Guerry, D. T., Epstein, M. N., Zehngebot, L., Lusk, E., van Horn, M., & Clark Jr, W. H. (1984). Invasive malignant melanomas lacking competence for metastasis. *The American Journal of Dermatopathology*, *6*, 55–61.

Elio Campitelli. (2022). *ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'* (0.4.7). R package .

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, *149*(4), 778–789. https://doi.org/10.1002/ijc.33588

Fey, M. F., & Wainscoat, J. S. (1988). Molecular diagnosis of haematological neoplasms. *Blood Reviews*, *2*(2), 78–87.

Fisslthaler, B., & Fleming, I. (2009). Activation and signaling by the AMP-activated protein kinase in endothelial cells. *Circulation Research*, *105*(2), 114–127.

Flies, E. J., Mavoa, S., Zosky, G. R., Mantzioris, E., Williams, C., Eri, R., Brook, B. W., & Buettel, J. C. (2019). Urban-associated diseases: Candidate diseases, environmental risk factors, and a path forward. *Environment International*, *133*, 105187.

Funck-Brentano, E., Malissen, N., Roger, A., Lebbé, C., Deilhes, F., Frénard, C., Dréno, B., Meyer, N., Grob, J., & Tétu, P. (2021). Which adjuvant treatment for patients with BRAFV600-mutant cutaneous melanoma? *Annales de Dermatologie et de Vénéréologie*, *148*(3), 145–155.

Garland, J. (2017). Unravelling the complexity of signalling networks in cancer: A review of the increasing role for computational modelling. *Critical Reviews in Oncology/Hematology*, *117*, 73–113.

Garraway, L. A., Widlund, H. R., Rubin, M. A., Getz, G., Berger, A. J., Ramaswamy, S., Beroukhim, R., Milner, D. A., Granter, S. R., & Du, J. (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature*, *436*(7047), 117–122.

Gene Ontology Consortium. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Ghannam, M. G., & Varacallo, M. (2018). *Biochemistry, Polymerase Chain Reaction*.

Guangchuang Yu. (2022). *enrichplot: Visualization of Functional Enrichment Result* (1.14.2). R package .

Guo, H., Cheng, Y., Martinka, M., & McElwee, K. (2015). High LIFr expression stimulates melanoma cell migration and is associated with unfavorable prognosis in melanoma. *Oncotarget*, *6*(28), 25484–25498. https://doi.org/10.18632/oncotarget.4688

Heidi Chial. (2008). Proto-oncogenes to Oncogenes to Cancer . *Nature Education* .

Hoejberg, L., Bastholt, L., & Schmidt, H. (2012). Interleukin-6 and melanoma. *Melanoma Research*, *22*(5), 327–333. https://doi.org/10.1097/CMR.0b013e3283543d72

Hong, M., Tao, S., Zhang, L., Diao, L.-T., Huang, X., Huang, S., Xie, S.-J., Xiao, Z.-D., & Zhang, H. (2020). RNA sequencing: new technologies and applications in cancer research. *Journal of Hematology & Oncology*, *13*(1), 1–16.

Hotelling, H. (1936). Relations Between Two Sets of Variates. *Biometrika*, *28*(3/4), 321–377. https://doi.org/10.2307/2333955

Ibanez de Caceres, I., Dulaimi, E., Hoffman, A. M., Al-Saleem, T., Uzzo, R. G., & Cairns, P. (2006). Identification of novel target genes by an epigenetic reactivation screen of renal cancer. *Cancer Research*, *66*(10), 5021–5028.

Igarashi, T., Nishino, K., & Nayar, S. K. (2007). The Appearance of Human Skin: A Survey. *Foundations and Trends® in Computer Graphics and Vision*, *3*(1), 1–95. https://doi.org/10.1561/0600000013

InformedHealth.org. (2006). Institute for Quality and Efficiency in Health Care (IQWiG). https://www.ncbi.nlm.nih.gov/books/NBK279410/

Ishikawa, T., Wondimu, Z., Oikawa, Y., Gentilcore, G., Kiessling, R., Brage, S. E., Hansson, J., & Patarroyo, M. (2014). Laminins 411 and 421 differentially promote tumor cell migration via α6β1 integrin and MCAM (CD146). *Matrix Biology*, *38*, 69–83.

Ishikawa, T., Wondimu, Z., Oikawa, Y., Ingerpuu, S., Virtanen, I., & Patarroyo, M. (2014). Monoclonal antibodies to human laminin α4 chain globular domain inhibit tumor cell adhesion and migration on laminins 411 and 421, and binding of α6β1 integrin and MCAM to α4-laminins. *Matrix Biology*, *36*, 5–14.

Janssen, S. M., Moscona, R., Elchebly, M., Papadakis, A. I., Redpath, M., Wang, H., Rubin, E., van Kempen, L. C., & Spatz, A. (2020). BORIS/CTCFL promotes a switch from a proliferative towards an invasive phenotype in melanoma cells. *Cell Death Discovery*, *6*(1), 1. https://doi.org/10.1038/s41420-019-0235-x

Jia, Y.-Y., Yu, Y., & Li, H.-J. (2021). POSTN promotes proliferation and epithelial-mesenchymal transition in renal cell carcinoma through ILK/AKT/mTOR pathway. *Journal of Cancer*, *12*(14), 4183–4195. https://doi.org/10.7150/jca.51253

Johnson, D. E., O'Keefe, R. A., & Grandis, J. R. (2018). Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nature Reviews. Clinical Oncology*, *15*(4), 234–248. https://doi.org/10.1038/nrclinonc.2018.8

Kamohara, H., Ogawa, M., Ishiko, T., Sakamoto, K., & Baba, H. (2007). Leukemia inhibitory factor functions as a growth factor in pancreas carcinoma cells: Involvement of regulation of LIF and its receptor expression. *International Journal of Oncology*, *30*(4), 977–983.

Kappelmann-Fenzl, M. (2021). *Next Generation Sequencing and Data Analysis*. https://doi.org/10.1007/978-3-030-62490-3

Kappelmann-Fenzl, M., Gebhard, C., Matthies, A., Kuphal, S., Rehli, M., & Bosserhoff, A. (2019). C-Jun drives melanoma progression in PTEN wild type melanoma cells. *Cell Death & Disease*, *10*, 1–16. https://doi.org/10.1038/s41419-019-1821-9

Kellokumpu-Lehtinen, P., Talpaz, M., Harris, D., Van, Q., Kurzrock, R., & Estrov, Z. (1996). Leukemia-inhibitory factor stimulates breast, kidney and prostate cancer cell proliferation by paracrine and autocrine pathways. *International Journal of Cancer*, *66*(4), 515–519.

Khan, N. H., Mir, M., Qian, L., Baloch, M., Khan, M. F. A., Ngowi, E. E., Wu, D.-D., & Ji, X.-Y. (2021). Skin cancer biology and barriers to treatment: Recent applications of polymeric micro/nanostructures. *Journal of Advanced Research*.

Kim, Y. H., Liang, H., Liu, X., Lee, J.-S., Cho, J. Y., Cheong, J.-H., Kim, H., Li, M., Downey, T. J., Dyer, M. D., Sun, Y., Sun, J., Beasley, E. M., Chung, H. C., Noh, S. H., Weinstein, J. N., Liu, C.-G., & Powis, G. (2012). AMPKα modulation in cancer progression: multilayer integrative analysis of the whole transcriptome in Asian gastric cancer. *Cancer Research*, *72*(10), 2512–2521. https://doi.org/10.1158/0008-5472.CAN-11-3870

Kolde, R., & Kolde, M. R. (2018). Package 'pheatmap'. *R Package*, *1*.

Kotobuki, Y., Yang, L., Serada, S., Tanemura, A., Yang, F., Nomura, S., Kudo, A., Izuhara, K., Murota, H., Fujimoto, M., Katayama, I., & Naka, T. (2014). Periostin accelerates human malignant melanoma progression by modifying the melanoma microenvironment. *Pigment Cell & Melanoma Research*, *27*(4), 630–639. https://doi.org/10.1111/pcmr.12245

Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, *2015*(11), pdb.top084970. https://doi.org/10.1101/pdb.top084970

Lee, C., Collichio, F., Ollila, D., & Moschos, S. (2013). Historical review of melanoma treatment and outcomes. *Clinics in Dermatology*, *31*(2), 141–147.

Lee, J.-H., Miele, M. E., Hicks, D. J., Phillips, K. K., Trent, J. M., Weissman, B. E., & Welch, D. R. (1996). KiSS-1, a novel human malignant melanoma metastasis-suppressor gene. *JNCI: Journal of the National Cancer Institute*, *88*(23), 1731–1737.

Lee, Y.-T. N. (1980). Malignant Melanoma: Pattern of Metastasis. *CA: A Cancer Journal for Clinicians*, *30*(3), 137–142. https://doi.org/10.3322/canjclin.30.3.137

Lei, X., Guan, C.-W., Song, Y., & Wang, H. (2015). The multifaceted role of CD146/MCAM in the promotion of melanoma progression. *Cancer Cell International*, *15*(1), 3. https://doi.org/10.1186/s12935-014-0147-z

Li, J., Ding, Y., & Li, A. (2016). Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World Journal of Surgical Oncology*, *14*(1), 297. https://doi.org/10.1186/s12957-016-1056-5

Liao, J. B. (2006). Viruses and human cancer. *The Yale Journal of Biology and Medicine*, *79*(3–4), 115–122.

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.

Liao, Y.-C., & Lo, S. H. (2008). Deleted in liver cancer-1 (DLC-1): a tumor suppressor not just for liver. *The International Journal of Biochemistry & Cell Biology*, *40*(5), 843–847.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, *1*(6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Liu, J. A., Rao, Y., Cheung, M. P. L., Hui, M.-N., Wu, M.-H., Chan, L.-K., Ng, I. O.-L., Niu, B., Cheah, K. S. E., & Sharma, R. (2017). Asymmetric localization of DLC1 defines avian trunk neural crest polarity for directional delamination and migration. *Nature Communications*, *8*(1), 1–17.

Liu, T., Liu, H., Wang, P., Hu, Y., Yang, R., Liu, F., Kim, H. G., Dong, Z., & Liu, K. (2020). Honokiol Inhibits Melanoma Growth by Targeting Keratin 18 in vitro and in vivo. *Frontiers in Cell and Developmental Biology*, *8*, 603472. https://doi.org/10.3389/fcell.2020.603472

Love, M., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/PREACCEPT-8897612761307401

Luo, C., Lim, J.-H., Lee, Y., Granter, S. R., Thomas, A., Vazquez, F., Widlund, H. R., & Puigserver, P. (2016). A PGC1α-mediated transcriptional axis suppresses melanoma metastasis. *Nature*, *537*(7620), 422–426.

Luo, W., & Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, *29*(14), 1830–1831.

Marc Carlson. (2021). *org.Hs.eg.db: Genome wide annotation for Human*. R package version 3.14.0.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., & Chen, Z. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376–380.

Martin Morgan and Lori Shepherd. (2022). *AnnotationHub: Client to access AnnotationHub resources* (3.2.2). R package.

Maru, G. B., Gandhi, K., Ramchandani, A., & Kumar, G. (2014). The role of inflammation in skin cancer. *Inflammation and Cancer*, 437–469.

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, *74*(2), 560–564.

McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019a). Interpretation of differential gene expression results of RNA-seq data: review and integration. *Briefings in Bioinformatics*, *20*(6), 2044–2054.

McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019b). Interpretation of differential gene expression results of RNA-seq data: review and integration. *Briefings in Bioinformatics*, *20*(6), 2044–2054.

McKenzie, J. A., Liu, T., Jung, J. Y., Jones, B. B., Ekiz, H. A., Welm, A. L., & Grossman, D. (2013). Survivin promotion of melanoma metastasis requires upregulation of α 5 integrin. *Carcinogenesis*, *34*(9), 2137–2144.

Mery, B., Vallard, A., Rowinski, E., & Magne, N. (2019). High-throughput sequencing in clinical oncology: from past to present. *Swiss Medical Weekly*, *13*.

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., … Groop, L. C. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, *34*(3), 267–273. https://doi.org/10.1038/ng1180

Morra, L., & Moch, H. (2011). Periostin expression and epithelial-mesenchymal transition in cancer: a review and an update. *Virchows Archiv : An International Journal of Pathology*, *459*(5), 465–475. https://doi.org/10.1007/s00428-011-1151-5

Moscow, J. A., Fojo, T., & Schilsky, R. L. (2018). The evidence framework for precision cancer medicine. *Nature Reviews Clinical Oncology*, *15*(3), 183–192.

Mudge, J. M., & Harrow, J. (2016). The state of play in higher eukaryote gene annotation. *Nature Reviews Genetics*, *17*(12), 758–772.

Ng, K. W., & Lau, W. M. (2015). Skin Deep: The Basics of Human Skin Structure and Drug Penetration. In *Percutaneous Penetration Enhancers Chemical Methods in Penetration Enhancement* (pp. 3–11). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45013-0_1

Palumbo, G., di Lorenzo, G., Ottaviano, M., & Damiano, V. (2016). The future of melanoma therapy: Developing new drugs and improving the use of old ones. In *Future Oncology* (Vol. 12, Issue 22, pp. 2531–2534). Future Medicine.

Parrado, C., Mercado-Saenz, S., Perez-Davo, A., Gilaberte, Y., Gonzalez, S., & Juarranz, A. (2019). Environmental stressors on skin aging. Mechanistic insights. *Frontiers in Pharmacology*, *10*, 759.

Patel, A. (2020). Benign vs Malignant Tumors. *JAMA Oncology*, *6*(9), 1488. https://doi.org/10.1001/jamaoncol.2020.2592

Plonka, P. M., Passeron, T., Brenner, M., Tobin, D. J., Shibahara, S., Thomas, A., Slominski, A., Kadekaro, A. L., Hershkovitz, D., Peters, E., Nordlund, J. J., Abdel-Malek, Z., Takeda, K., Paus, R., Ortonne, J. P., Hearing, V. J., & Schallreuter, K. U. (2009). What are melanocytes really doing all day long...? *Experimental Dermatology*, *18*(9), 799–819. https://doi.org/10.1111/j.1600-0625.2009.00912.x

Polyak, K., & Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nature Reviews Cancer*, *9*(4), 265–273.

Pomyen, Y., Segura, M., Ebbels, T. M. D., & Keun, H. C. (2015). Over-representation of correlation analysis (ORCA): a method for identifying associations between variable sets. *Bioinformatics*, *31*(1), 102–108. https://doi.org/10.1093/bioinformatics/btu589

Popescu, N. C., & Goodison, S. (2014). Deleted in liver cancer-1 (DLC1): an emerging metastasis suppressor gene. *Molecular Diagnosis & Therapy*, *18*(3), 293–302.

Qing, T., Zhu, S., Suo, C., Zhang, L., Zheng, Y., & Shi, L. (2017). Somatic mutations in ZFHX4 gene are associated with poor overall survival of Chinese esophageal squamous cell carcinoma patients. *Scientific Reports*, *7*(1), 4951. https://doi.org/10.1038/s41598-017-04221-7

Qiu, T., Wang, H., Wang, Y., Zhang, Y., Hui, Q., & Tao, K. (2015). Identification of genes associated with melanoma metastasis. *The Kaohsiung Journal of Medical Sciences*, *31*(11), 553–561.

Rainer, J., Gatto, L., & Weichenberger, C. X. (2019). ensembldb: an R package to create and use Ensembl-based annotation resources. *Bioinformatics*, *35*(17), 3151–3153. https://doi.org/10.1093/bioinformatics/btz031

Rauf, A., Patel, S., Imran, M., Maalik, A., Arshad, M. U., Saeed, F., Mabkhot, Y. N., Al-Showiman, S. S., Ahmad, N., & Elsharkawy, E. (2018). Honokiol: An anticancer lignan. *Biomedicine & Pharmacotherapy*, *107*, 555–562.

Rebecca, V. W., Sondak, V. K., & Smalley, K. S. M. (2012). A brief history of melanoma: from mummies to mutations. *Melanoma Research*, *22*(2), 114.

Rehm, B. (2001). Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification. *Applied Microbiology and Biotechnology*, *57*(5), 579–592.

Russo, G., Zegar, C., & Giordano, A. (2003). Advantages and limitations of microarray technology in human cancer. *Oncogene*, *22*(42), 6497–6507.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467.

Sarrica, A., Kirika, N., Romeo, M., Salmona, M., & Diomede, L. (2018). Safety and toxicology of magnolol and honokiol. *Planta Medica*, *84*(16), 1151–1164.

Schadt, E. E., Turner, S., & Kasarskis, A. (2011). Corrigendum: A window into third generation sequencing. *Human Molecular Genetics*, *20*(4), 853.

Schlagbauer-Wadl, H., Jansen, B., M�ller, M., Polterauer, P., Wolff, K., Eichler, H.-G., Pehamberger, H., Konakand, E., & Johnson, J. P. (1999). Influence of MUC18/MCAM/CD146 expression on human melanoma growth and metastasis in SCID mice. *International Journal of Cancer*, *81*(6), 951–955. https://doi.org/10.1002/(SICI)1097-0215(19990611)81:6<951::AID-IJC18>3.0.CO;2-V

Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, *66*(1), 7–30. https://doi.org/10.3322/caac.21332

Silpa, S. R., & Chidvila, V. (2013). A review on skin cancer. *International Research Journal of Pharmacy*, *4*(8), 83–88.

Simon, D. P., Giordano, T. J., & Hammer, G. D. (2012). Upregulated JAG1 enhances cell proliferation in adrenocortical carcinoma. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *18*(9), 2452–2464. https://doi.org/10.1158/1078-0432.CCR-11-2371

Sjoestroem, C., Khosravi, S., Cheng, Y., Safaee Ardekani, G., Martinka, M., & Li, G. (2014). DLC1 expression is reduced in human cutaneous melanoma and correlates with patient survival. *Modern Pathology*, *27*(9), 1203–1211.

Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., van de Nobelen, S., & Caesar, L. (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics & Chromatin*, *5*(1), 1–21.

Sokolenko, A. P., & Imyanitov, E. N. (2018). Molecular diagnostics in clinical oncology. *Frontiers in Molecular Biosciences*, *5*, 76.

Soltermann, A., Tischler, V., Arbogast, S., Braun, J., Probst-Hensch, N., Weder, W., Moch, H., & Kristiansen, G. (2008). Prognostic significance of epithelial-mesenchymal and mesenchymal-epithelial transition protein expression in non–small cell lung cancer. *Clinical Cancer Research*, *14*(22), 7430–7437.

Stanford Medicine. (n.d.). *What Causes Cancer?* Stanfordhealthcare.Org. Retrieved 2 September 2022, from https://stanfordhealthcare.org/medical-conditions/cancer/cancer/cancer-causes.html#about

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, *20*(11), 631–656. https://doi.org/10.1038/s41576-019-0150-2

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Szepietowski, J. C., Reich, A., & McKenzie, R. C. (2004). The multifunctional role of leukaemia inhibitory factor in cutaneous biology. *Acta Dermatovenerol Alp Panonica Adriat*, *13*, 125–129.

The National Cancer Institute. (2021). *What Is Cancer?*

Tian, X., Zhou, D., Chen, L., Tian, Y., Zhong, B., Cao, Y., Dong, Q., Zhou, M., Yan, J., & Wang, Y. (2018). Polo-like kinase 4 mediates epithelial–mesenchymal transition in neuroblastoma via PI3K/Akt signaling pathway. *Cell Death & Disease*, *9*(2), 1–14.

Tsao, H., Atkins, M. B., & Sober, A. J. (2004). Management of cutaneous melanoma. *New England Journal of Medicine*, *351*(10), 998–1012.

Tutar, Y. (2014). Editorial (thematic issue:"miRNA and cancer; computational and experimental approaches"). *Current Pharmaceutical Biotechnology*, *15*(5), 429.

Ulahannan, D., Kovac, M. B., Mulholland, P. J., Cazier, J.-B., & Tomlinson, I. (2013). Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *British Journal of Cancer*, *109*(4), 827–835. https://doi.org/10.1038/bjc.2013.416

Uong, A., & Zon, L. I. (2010). Melanocytes in development and cancer. *Journal of Cellular Physiology*, *222*(1), 38–41.

Urso, C. (2004). Are growth phases exclusive to cutaneous melanoma? *Journal of Clinical Pathology*, *57*(5), 560. https://doi.org/10.1136/jcp.2003.014852

Vuković, P., Lugović-Mihić, L., Ćesić, D., Novak-Bilić, G., Šitum, M., & Spoljar, S. (2020). Melanoma development: current knowledge on melanoma pathogenesis. *Acta Dermatovenerologica Croatica*, *28*(2), 163.

Wang, P., Chen, Y., Ding, G., Du, H., & Fan, H. (2021). Keratin 18 induces proliferation, migration, and invasion in gastric cancer via the MAPK signalling pathway. *Clinical and Experimental Pharmacology and Physiology*, *48*(1), 147–156.

Wang, Z., & Yan, X. (2013). CD146, a multi-functional molecule beyond adhesion. *Cancer Letters*, *330*(2), 150–162.

Wang, Z., Zheng, X., Wang, X., Chen, Y., Li, Z., Yu, J., Yang, W., Mao, B., Zhang, H., & Li, J. (2021). Genetic differences between lung metastases and liver metastases from left-sided microsatellite stable colorectal cancer: Next generation sequencing and clinical implications. *Annals of Translational Medicine*, *9*(12).

Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Mächler, M., Magnusson, A., & Möller, S. (2005). gplots: Various R programming tools for plotting data. In *R package version* (Vol. 2).

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738.

Weinberg, R. A. (1996). How Cancer Arises. *Scientific American*, *275*(3), 62–70. https://doi.org/10.1038/scientificamerican0996-62

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wu, S., Cho, E., Li, W.-Q., Weinstock, M. A., Han, J., & Qureshi, A. A. (2016). History of severe sunburn and risk of skin cancer among women and men in 2 prospective cohort studies. *American Journal of Epidemiology*, *183*(9), 824–833.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141

Wysoczynski, M., Miekus, K., Jankowski, K., Wanzeck, J., Bertolone, S., Janowska-Wieczorek, A., Ratajczak, J., & Ratajczak, M. Z. (2007). Leukemia inhibitory factor: a newly identified metastatic factor in rhabdomyosarcomas. *Cancer Research*, *67*(5), 2131–2140.

Yanez, D. A., Lacher, R. K., Vidyarthi, A., & Colegio, O. R. (2017). The role of macrophages in skin homeostasis. *Pflügers Archiv - European Journal of Physiology*, *469*(3–4), 455–463. https://doi.org/10.1007/s00424-017-1953-7

Yang, X., Hu, F., Liu, J. A., Yu, S., Cheung, M. P. L., Liu, X., Ng, I. O.-L., Guan, X.-Y., Wong, K. K. W., Sharma, R., Lung, H. L., Jiao, Y., Lee, L. T. O., & Cheung, M. (2020). Nuclear DLC1 exerts oncogenic function through association with FOXK1 for cooperative activation of MMP9 expression in melanoma. *Oncogene*, *39*(20), 4061–4076. https://doi.org/10.1038/s41388-020-1274-8

Yu, G., Wang, L.-G., Yan, G.-R., & He, Q.-Y. (2015). DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*, *31*(4), 608–609. https://doi.org/10.1093/bioinformatics/btu684

Zhang, J., Hu, S., & Li, Y. (2019). KRT18 is correlated with the malignant status and acts as an oncogene in colorectal cancer. *Bioscience Reports*, *39*(8).

Zong, S., Xu, P., Xu, Y., & Guo, Y. (2022). A bioinformatics analysis: ZFHX4 is associated with metastasis and poor survival in ovarian cancer. *Journal of Ovarian Research*, *15*(1), 90. https://doi.org/10.1186/s13048-022-01024-x