T.C.
DOKUZ EYLÜL UNIVERSITY
IZMIR INTERNATIONAL
BIOMEDICINE AND GENOME
INSTITUTE

Transcriptional Impact of TGFβ on Epithelial-Mesenchymal Transition

HAZAL AYTEKİN

DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

MASTER OF SCIENCE THESIS

IZMIR-2020

THESIS CODE: DEU.IBG.MSc.2017850038

T.C.
DOKUZ EYLÜL UNIVERSITY
IZMIR INTERNATIONAL
BIOMEDICINE AND GENOME
INSTITUTE

Transcriptional Impact of TGFß on Epithelial Mesenchymal Transition

DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

MASTER OF SCIENCE THESIS

HAZAL AYTEKİN

Supervisor: Asst. Prof. Hani Alotaibi

THESIS CODE: DEU.IBG.MSc.2017850038

Dokuz Eylül University Izmir International Biomedicine and Genome Institute Department of Genomics and Molecular Biotechnology, Molecular Biology and Genetics graduate program Master of Science student Hazal Aytekin has successfully completed her Master of Science thesis titled "**TRANSCRIPTIONAL IMPACT OF TGFβ ON EPITHELIAL MESENCHYMAL TRANSITION**" on the date of 15/09/2020.

Asst. Prof. Hani Alotaibi
Dokuz Eylul University
CHAIR

Asst. Prof. Cihangir Yandım
Izmir University of Economics
MEMBER

Asst. Prof. Yavuz Oktay
Dokuz Eylul University
MEMBER

Assoc. Prof. Gökhan Karakülah
Dokuz Eylül University
SUBSTITUTE MEMBER

Asst. Prof. Banu Demir
Izmir University of Economics
SUBSTITUTE MEMBER

Dokuz Eylül Üniversitesi İzmir Uluslararası Biyotıp ve Genom Enstitüsü Genom Bilimleri ve Moleküler Biyoteknoloji Anabilim Dalı, Moleküler Biyoloji ve Genetik Yüksek Lisans programı öğrencisi, Hazal Aytekin "**TGFß'NIN EPİTELYAL MEZANKIMAL DÖNÜŞÜM ÜZERİNDEKİ TRANSKRİPSİYONEL ETKİLERİ**" konulu Yüksek Lisans tezini 15/09/2020 tarihinde başarılı olarak tamamlamıştır.

Dr. Öğr. Üye. Hani Alotaibi
Dokuz Eylül Üniversitesi
BAŞKAN

Dr. Öğr. Üye. Cihangir Yandım
İzmir Ekonomi Üniversitesi
ÜYE

Dr. Öğr. Üye. Yavuz Oktay
Dokuz Eylül Üniversitesi
ÜYE

Doç. Dr. Gökhan Karakülah
Dokuz Eylül Üniversitesi
YEDEK ÜYE

Dr. Öğr. Üye. Banu Demir
İzmir Ekonomi Üniversitesi
YEDEK ÜYE

Table of Contents

Index of Figures

Index of Tables

List of Abbreviations

EMT- Epithelial-Mesenchymal Transition

MET- Mesenchymal-Epithelial Transition

TGFβ- Transforming Growth Factor Beta

TGFBRI/TGFBRII- Transforming Growth Factor Beta Receptor I/II

GSEA- Gene Set Enrichment Analysis

TFs- Transcription Factors

ECM- Extracellular Matrix

CTCs- Circulating Tumor Cells

miRNA- Micro RNA

LncRNA- Long Noncoding RNA

HDAC- Histone Deacetylase

NMuMG- Normal Murine Mammary Gland Cell Line

BMP- Bone Morphogenic Protein

RTK- Receptor Tyrosine Kinase

ALK- Anaplastic Lymphoma Kinase

CDKs- Cyclin-Dependent Kinases

GEO- Gene Expression Omnibus

RMA- Robust Multi-array Average

PCA- Principal Component Analysis

GSEA- Gene Set Enrichment Analysis

ES- Enrichment Score

FDR- False Discovery Rate

GO- Gene Ontology

TCGA- The Cancer Genome Atlas

QC- Quality Control

RLE- Relative Logarithmic Expression Value

DEGs- Differentially Expressed Genes

Acknowledgments

First, I would like to express my gratitude to my thesis supervisor Hani Alotaibi for his guidance and patience during the three years of my academic journey, for continuous support throughout my master's, planning and writing my thesis.

I would also like to thank all past and present Alotaibi Lab. members for being there for each other whenever it is needed.

And to members of Izmir International Biomedicine and Genome Institute, Dokuz Eylul University for allowing me to conduct my studies and adding more to my interest in science.

I am grateful for my parents Metin Aytekin, Türkan Aytekin and my brother Savaş Aytekin for always supporting me in every field of my life, for guiding me through the right way. If it were not for them, I would not be where I am now.

Finally, I would like to thank TÜBİTAK for supporting the project by grant no:117Z223.

Abstract

An epithelial-mesenchymal transition (EMT) is a process that leads an epithelial cell to change by gaining a mesenchymal phenotype. During organogenesis, embryogenesis and metastasis of tumor cells, cells differentiate into epithelial or mesenchymal states via EMT and MET processes. Transforming Growth Factor Beta (TGFβ) and the signaling pathway is the most important EMT stimulator. TGFβ plays a prominent role as a tumor suppressor and a metastatic agent in cancer progression. From embryonic development to homeostasis, TGFβ and its downstream regulators in the signaling pathway are crucial that deficiencies and mutations in the gene or on the pathway cause developmental defects, immunological diseases and poor prognosis in cancer patients. TGFβ has three isoforms that are involved in these processes; TGFβ1, TGFβ2, and TGFβ3. Analyses in the thesis were performed computationally by R programming and tools like GSEA and GO Analysis among TGFβ1, TGFβ2 and TGFβ3 isoform treated public datasets of human and mouse organisms. In line with the thesis's hypothesis, significant differences in transcriptional level, gene sets, cancer types, mutations, pathways and molecular functions were identified between the isoforms of the TGFβ. These results of the thesis can illuminate further research in the TGFβ and EMT area and obtain new therapeutic aspects particularly in cancer diagnosis and treatment.

Keywords: EMT, MET, TGFβ, Differential expression analysis.

Özet

Epitelel- Mezenkimal dönüşüm (EMT), epitel bir hücrenin mezenkimal bir fenotip kazanarak dönüşümüne neden olduğu bir süreçtir. Bu olay transkripsiyon faktörlerinin aktivasyonu ile gerçekleşir. Organogenez, embriyonik gelişim ve tümör hücrelerinin metastazı sırasında, hücreler EMT ve MET süreçleri yoluyla epitelyal veya mezenkimal durumlara farklılaşır. Dönüştürücü Büyüme Faktörü Beta (TGFβ) ve sinyal yolağı, EMT'nin en önemli uyarıcısıdır. TGFβ, kanser ilerlemesinde tümör baskılayıcı ve metastatik ajan olarak önemli bir rol oynar. Embriyonik gelişimden homeostaza kadar, TGFβ ve sinyal yolağının diğer regülatörleri çok önemlidir. TGFβ'nin bu süreçlerde görevli TGFβ1, TGFβ2 ve TGFβ3 olmak üzere üç izoformu vardır. Bu araştırma çeşitli biyoinformatik analizler, R programlama dili ve GSEA ve GO analizi gibi araçlar kullanılarak gerçekleştirildi. Analizde TGFβ1, TGFβ2 ve TGFβ3 izoformlarıyla muamele edilmiş insan ve fare hücrelerinin kullanıldığı genel veri kümeleri kullanılarak sonuçlar elde edildi. Tezin hipotezi doğrultusundaki bu analiz sonuçlarıyla, TGFβ izoformları arasında transkripsiyonel düzeyde, gen setleri, kanser tipleri, mutasyonlar, yolaklar ve moleküler fonksiyonlarda önemli farklılıklar tespit edildi. Araştırmanın bu sonuçları, TGFβ ve EMT alanındaki gelecek araştırmaları aydınlatma ve özellikle kanser tanı ve tedavisinde yeni terapötik bakış açıları geliştirilmesine önemli katkılar sağlayacaktır.

Anahtar Kelimeler: EMT, MET, TGFβ, Diferansiyal ekspresyon analizi.

1 Introduction and Aim

1.1 Aim of the Study

Despite the similarities of the isoforms, differences among them, and how they affect gene expression, signaling pathways, transcriptional factors may play a critical role in their contribution to significant processes. The objective of this thesis is to identify the target genes, transcriptional factors and the gene sets that are common or showing substantial differences between TGFβ1, TGFβ2, and TGFβ3 isoforms by bioinformatic analysis to provide leading information for the better understanding of organogenesis, embryogenesis, and cell migration processes which will lead to metastasis and invasiveness in epithelial cancers. And will help further research in EMT/MET field that may provide new aspects into the possible therapeutic approaches.

2 Hypothesis of the Study

The study hypothesizes that there are differences among TGFβ isoforms at the transcriptional, functional and tumorigenic levels affecting EMT, leading to possible therapeutic approaches and further research.

3 General Information

3.1 EMT

An epithelial-mesenchymal transition (EMT) is a process that causes an epithelial cell, which is embedded on the basement membrane, via ECM components, to undergo a transition by gaining a mesenchymal phenotype. EMT and MET processes play an important role in organogenesis, embryogenesis and metastasis of tumor cells. In addition to morphogenic changes in embryonic programming, a broader understanding of the EMT-MET processes is essential in the controlling metastatic diseases. During EMT, tight and adherens junctions are disrupted and reconstructed. Cells become motile by changing the cytoskeletal arrangements like ECM proteins (Gandalovičová et al., 2016). EMT is classified into three subtypes. Type 1 EMT is associated with development, especially in embryogenesis. From gastrulation to germ layer formation, as cells migrate and differentiate between epithelial and mesenchymal states, EMT is crucial in developing works via Nodal and Wnt pathways. Type 2 EMT is related to inflammation, trauma caused organ fibrosis, and wound healing. After an inflammatory response, the cell's basement membrane is disrupted by cell differentiation leading to EMT.

Finally, Type 3 EMT is associated with invasiveness, metastasis of cancer cells. Invasive cancer cells extravasate from the primary tumor site by gaining a mesenchymal phenotype and form a secondary tumor in a different site. Type 3 transformation occurs on cells that have undergone genetic and epigenetic changes for tumor development, and these changes have effects on oncogenes and tumor suppressor genes (Kalluri & Weinberg, 2009).

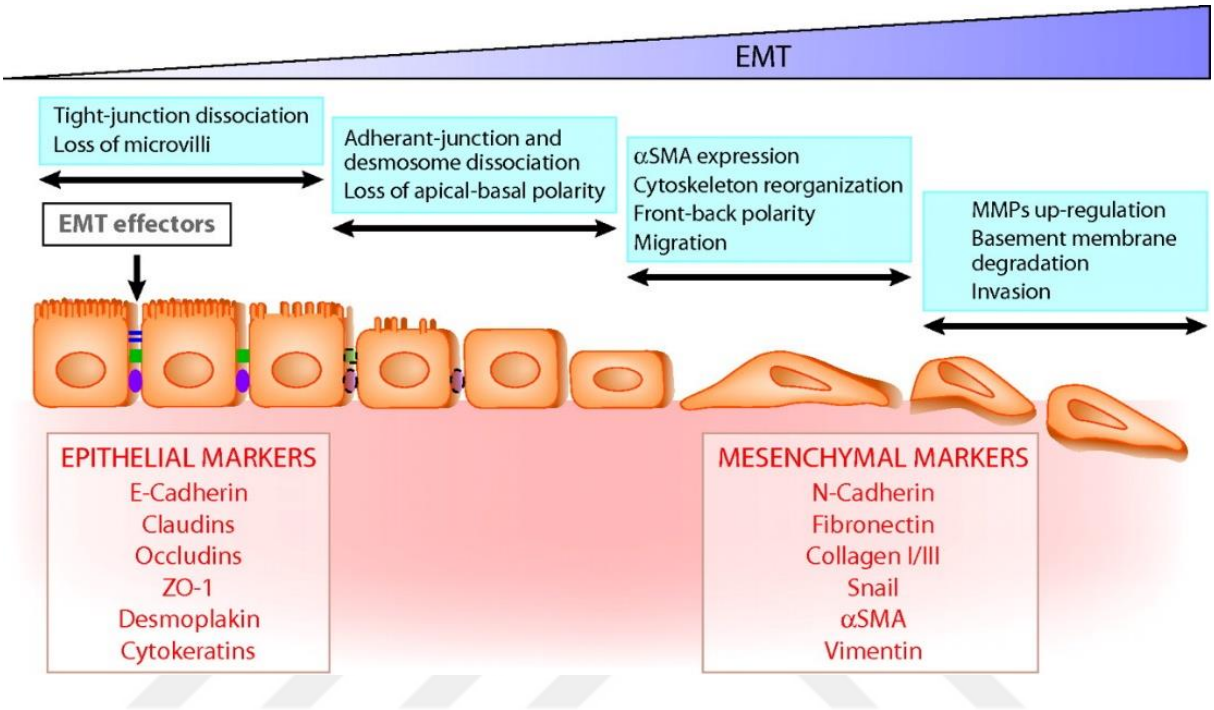


Figure 1: Events during EMT with epithelial and mesenchymal marker genes (Aroeira et al., 2007).

Epithelial-Mesenchymal Transition (EMT) plays critical roles in various processes including embryonic development, tumor progression, cancer metastasis, cell reprogramming. Transforming Growth Factor Beta (TGFβ) is the most recognized and studied driver of the EMT (Moustakas & Heldin, 2016). EMT is recognized by transcriptional changes at the gene level and, by the morphological changes cells undergo. Cells that are induced by TGFβ change their structure to elongated fibroblast-like spindles. During epithelial to mesenchymal differentiation, cells lose their attachments, and interactions with each other and alterations in extracellular matrix proteins gain migratory features to become a fibroblast-like mesenchymal phenotype. This switch is caused by transcriptional changes in genes like E-cadherin and N-cadherin (cell adhesion molecules). TGFβ signaling pathway activates the transcription of those marker genes via transcription factors (TFs) like Zeb1,2, Twist1,2, Snail1,2. These EMT-inducing transcription factors, induced by TGFβ, facilitate the acquisition of mesenchymal phenotype and suppress E-cadherin expression and increasing N-cadherin levels (Loh et al.,

2019). Genes involved in EMT affect cell cycle also cause epigenetic changes like histone modifications. Environmental factors are also included in the changes.

EMT is not only regulated during the transcription. Genes involved can be post-transcriptionally activated or repressed at translation like repression of ZEB molecules by miRNAs and alternative splicing regulates the EMT (Janakiraman et al., 2018).

3.1.1 Epithelial Cell Polarity

During epithelial to mesenchymal transition, change in the cellular polarity is highly required. In the phenotypic manner, while cells gain new mobile abilities, they lose their adhesion to each other and their apical-basal polarity via disruption of tight and adherens junctions in apical and basolateral surfaces. Proteins that are building blocks for junctions and maintaining polarity are claudins, occludin, PAR, CRB and SCRIB. Key inducers of the change in polarity are cadherins. The transcription factors responsible for E-cadherin downregulation (SNAI, ZEB) also promote claudin and occludin interactions. Besides the gene expression level, cellular polarity differences due to epithelial to mesenchymal changes are also affected by post-transcriptional changes mediated by miRNAs and non-coding RNAs (Moreno-Bueno et al., 2008).

3.1.2 Cellular Plasticity

During the transition between epithelial and mesenchymal states, cells preserve their plasticity as the processes are reversible. Throughout the transition, cells gain a hybrid stage which they co-express epithelial and mesenchymal markers. This mixed epithelial and mesenchymal stage of the cells is present in development, wound healing, especially in cancer metastasis (Liao & Yang, 2017). The hybrid phenotype of the circulating tumor cells (CTCs) express both markers of the two phenotypes and they strengthen the metastatic ability of the cells to form secondary tumors and increase drug resistance risk (Chin & Lim, 2019). Some of the TGFβ induced transcription factors associated with EMT (ZEB1, TWIST1, SNAI1) are also involved in cell stemness by effecting their epithelial plasticity and causing recurrence of the cancer in patients. Although the processes are involved with the same TFs, EMT and cancer cell stemness are independent from each other (Ye & Weinberg, 2015). Besides the transcriptional effects, epigenetic regulations like alternative splicing, chromatin modifications, are associated with cell plasticity.

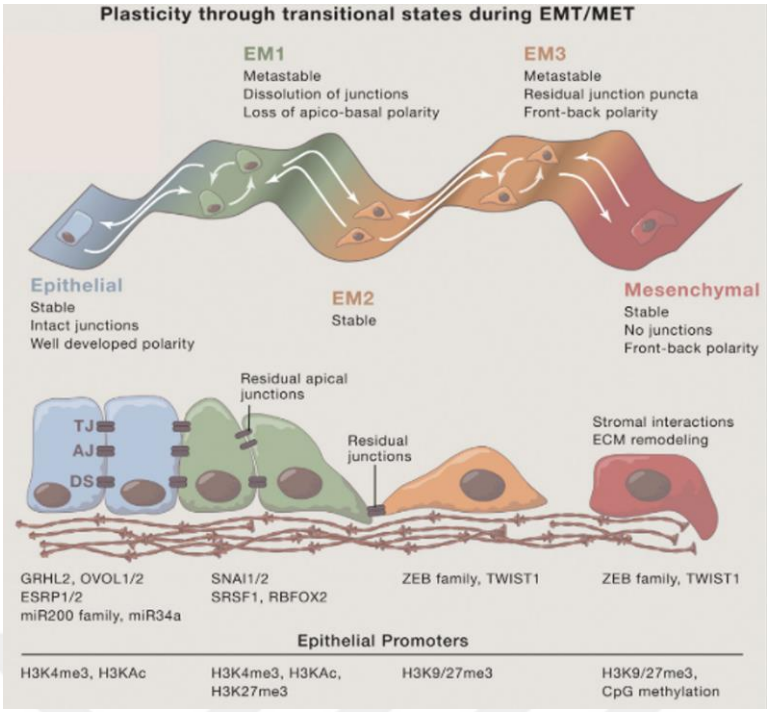


Figure 2: Epithelial cell plasticity states during EMT (EM: Epithelial and mesenchymal hybrid state.) (obtained from Nieto et al., 2016).

3.1.3 MET

Although Mesenchymal Epithelial Transition (MET) is defined as the inverse process of epithelial-mesenchymal transformation, recent studies and data from our lab researches emphasize the complex and dynamic nature of the MET process (Sengez et al., 2019). This process also implies a fundamental transformation mechanism of the same importance for embryonic programming. During MET, mesenchymal cells undergo transcriptional differences such as loss of N-cadherin and activation of E-cadherin expression, resulting in epithelial character. During the metastasis of cancer cells MET is also needed for secondary tumor formation in a distant site (Alotaibi et al., 2015).

3.1.4 EMT/MET Marker Genes

SNAI1 TF binds to DNA at the E-box element of the E-cadherin promoter region and suppresses its expression (Miyazono, 2009). TWIST1-2 increases the cell's invasive profile downregulating E-cadherin and upregulating vimentin, N-cadherin levels (Xu J., Lamouille S., 2016). The primary step of the EMT process is the downregulation of the levels of the E-cadherin. E-cadherin, organizes epithelial cell adhesion complexes, suppressing cell proliferation, resulting in the loss of cells' mesenchymal phenotypes. Claudin and occludin are downregulated for apical junction disruption (W. Guo et al., 2018).

ZEB family transcription factors act as a switch in transcription and are affected by miRNAs (Bendoraite et al., 2010). SNAI1, TWIST, ZEB downregulate epithelial genes and upregulate mesenchymal ones by binding to their E-boxes (Garg, 2013).

3.2 Epigenetic Regulation

GRHL2, ZEB1, TWIST1 are known epigenetic modifiers during EMT. Chromatin modifiers like HDAC1, HDAC2, LSD1 are recruited by SNAI1 and ZEB1 and repress CDH1 expression (Serrano-Gomez et al., 2016). During TGFβ induced EMT, CDH1 promoter hypermethylation at H3K9me2 and H3K27me3 by SNAI1 is a marker for epithelial cancers (Lin & Wu, 2020). TWIST causes H4K20me1 at CDH1 promoter and lead to EMT (Skrypek et al., 2017). Epigenetic regulator inhibitors are drug targets for circulating tumor cells of epithelial cancers. Targets include; BRD4770 for H3K9 methylation, Entinostat for HDACs, and GSK-126 for H3K27 methylation (Ning et al., 2015). However, the reversible role of EMT is causing some questions on the inhibitor therapies. Promoting MET by inhibiting the epigenetic repressions causes tumor invasion in some cancer types, which may be used in the tumor's early stages (Mishra & Johnsen, 2014).

3.3 TGFβ Signaling Pathway

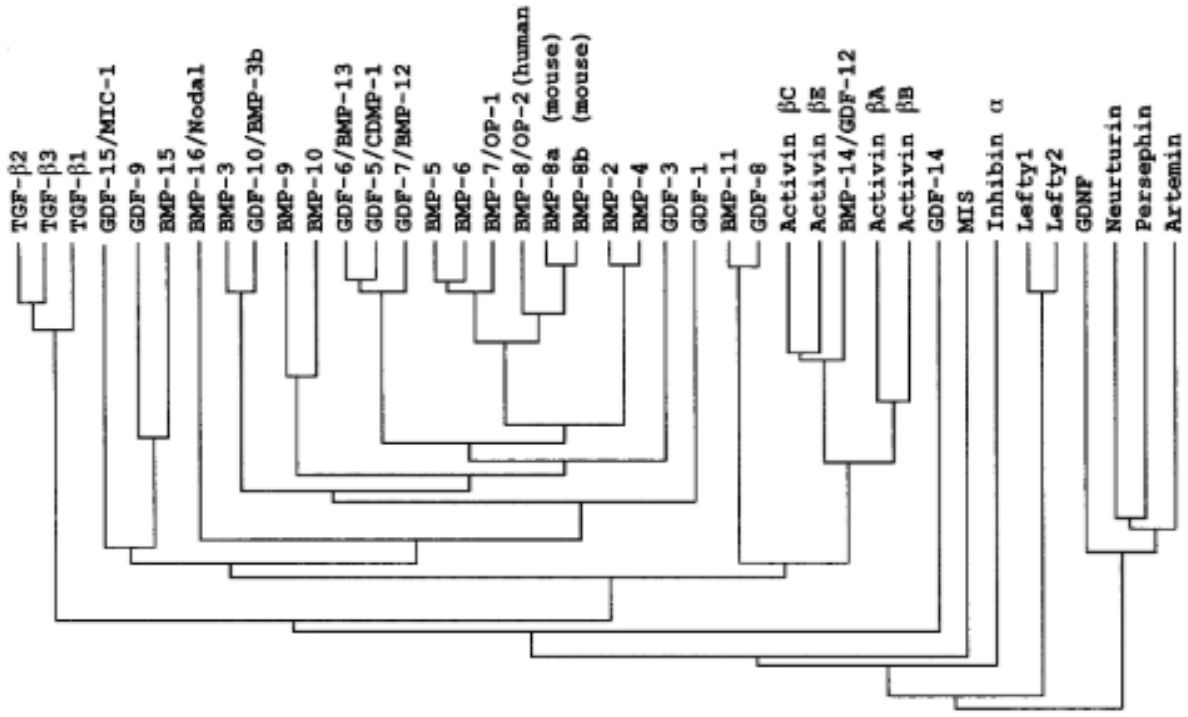


Figure 3: TGFβ superfamily tree based on their amino acid sequences (obtained from Chang et al., 2002).

Transforming Growth Factor Beta (TGFβ) signaling pathway is an important stimulator of the EMT. It is also known that in in-vitro studies, it induces EMT in certain cell lines and cancer cells like NMuMG (normal murine mammary gland), A549 (non-small cell lung cancer). In both adult and embryonic cells, TGFβ has a critical role for important cellular functions such as proliferation, differentiation and apoptosis. It is used as a differentiation factor, which transforms epithelial cells into mesenchymal cells in our in vitro studies with NMuMG cell line. Following treatment, TGFβ increases the expression of N-cadherin and vimentin by suppressing the expression of E-cadherin, efficiently returning from the epithelial phenotype to the mesenchymal phenotype. TGFβ ligands are secreted and they are directly activated or embedded in ECM to be activated when needed.

TGFβ was first discovered as a growth inducer factor in normal rat kidney cells (NRK) in 1978 (De Larco & Todaro, 1978). TGFβ superfamily has more than 35 members (Figure 3)(Chang et al., 2002). TGFβ signaling stars with two serine/threonine kinase receptors, Type I and Type II. Autophosphorylated Type II receptors phosphorylate Type I. TGFβ activates receptor complex, leads to the activation of Smad2-Smad3, which form trimers with Co-Smad, Smad4 (Hata & Chen, 2016).

That trimer complex translocates in the nucleus to associate and cooperate with DNA binding transcription factors, which will activate or repress target gene transcription. Besides the Smad dependent pathway, other signaling pathways are activated due to TGFβ response (Figure 4), for example MAPK signaling pathway, PI3K/AKT pathway, JNK, and p38 pathway (Y. E. Zhang, 2009). Erk-MAPK signaling increase the activity of TGFβ by decreasing E-cadherin levels. Receptor tyrosine kinase (RTK) signaling takes part in Smad signaling while they work as antagonists to BMP signaling by Smad1 ubiquitination (X. Guo & Wang, 2009). RTK signaling via Ras molecule leads to p53 phosphorylation. Liu et al. (2008) showed Ras signaling induces TGFβ pathway while TGFβ upregulates Ras signaling as well in NMuMG cells thus, inducing tumor invasion and EMT (Janda et al., 2002).

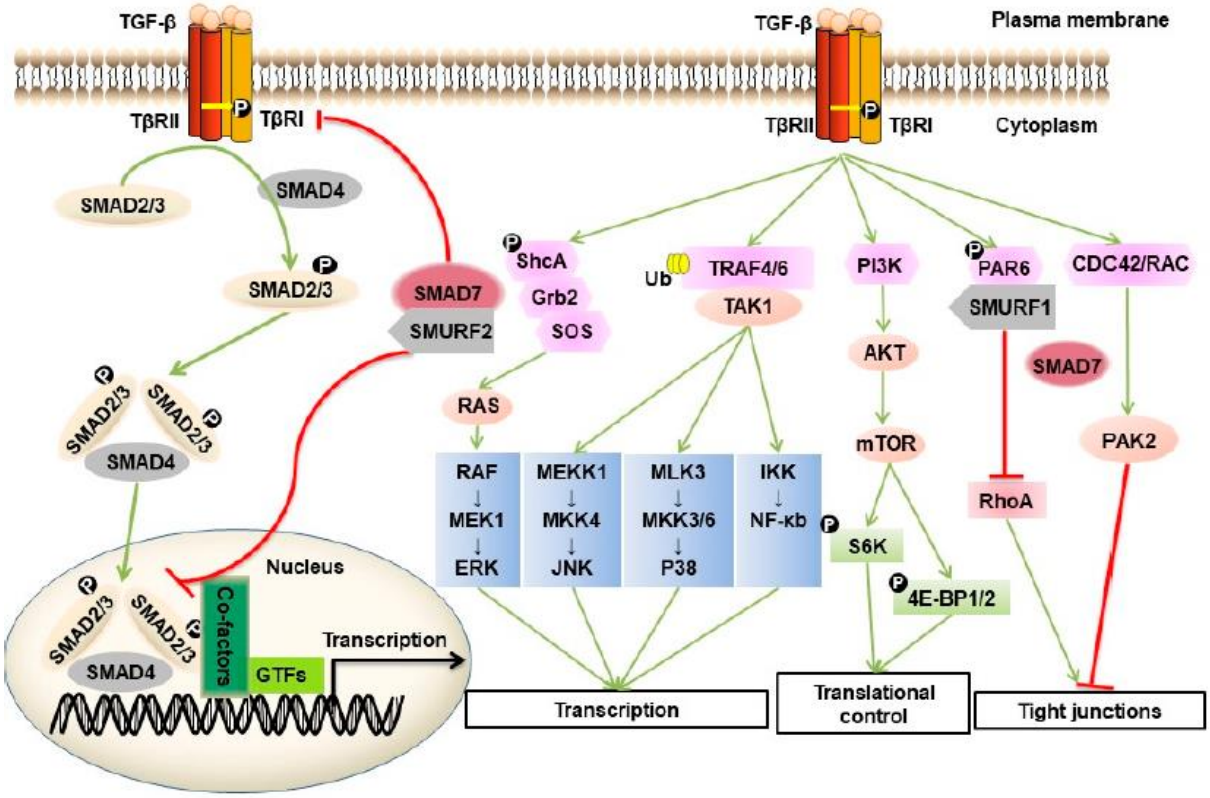


Figure 4: TGFβ SMAD and non-SMAD signaling pathway (obtained from Hao et al., 2019).

3.3.1 Smads

Smads are intracellular signaling molecules that are downstream of the TGFβ pathway. There are eight types of Smads in three categories. R Smads are receptor-regulated, Smad 2,4,5,8; Inhibitory Smads 6,7; and Co-Smad Smad4. Smad1,5, and 8 signals through receptors ALK1,2,3 and 6. Smad2 and 3 signals through ALKs 4,5 and 7. TGFβ activated Smad2,3,4 expression directly affects EMT. It is shown that blocked Smad2 or Smad3 activation causes EMT processes to fail in *in vitro* experiments with NMuMG cell lines (Jian Xu et al., 2009). Depending on the context, some cell lines show adverse effects with Smads. While Smad3 negative hepatocytes are showing no changes in phenotype, Smad2 negative keratinocytes and hepatocytes gained mesenchymal features with faster tumor development (Hoot et al., 2008). On the other hand, increased Smad7 levels cause EMT to stall in several types (Xu J., Lamouille S., 2016). Smad7 performs its inhibition effect by recruiting SMURF1,2 ubiquitin ligases for degradation (Figure 4). Smads may act differently in different organisms; an inhibitory complex in Drosophila works as an activation in mammalian cells (Upadhyay et al., 2017). Smads move between the nucleus and the cytoplasm constantly to maintain the signal transduction. This movement is achieved by Smad phosphorylation and dephosphorylation (L. Xu, 2006). Some Smads bind DNA directly, and some need interaction with DNA binding factors like CBP, p300

(Janknecht et al., 1998). Some transcription factors like FoxH1, Mixer, Milk, and Bix3 also help Smad2 to interact with DNA. Smad2-3 and Smad1-5-8 complexes interact with co-molecule Smad4 for transcription regulation. But, without Smad4, these complexes find alternative regulators such as IKKα, TRIM33 (Itoh et al., 2018). Smads also recruit histone modifier molecules like p300 and Brg1 (M. Y. Wu & Hill, 2009).

3.3.2 Receptors

TGFβ RII receptor has a size of 567 amino acids with 298 serine/threonine kinase domain. The core binding site of the receptor has a cysteine-rich region. PRE1 and PRE2 are related to the transcriptional regulation of TGFβ RII (Kim et al., 2000).

It is known that most of the metastatic tumors overexpress TGFβ; besides, they are resistant to growth inhibition mediated by it, for the purpose of invasiveness. When dominant-negative TGFBRII is used, it naturally blocks TGFβ signaling and, thus, EMT. In fact, in mesenchymal spindle tumor cells, blocked TGFBR signaling induces MET (Ishigame et al., 2013). Depending on the EMT blockage and MET induction, metastasis and invasiveness of the colon carcinoma cells (CT26) are directly inhibited. Even when primary tumors are excised from the site, CT26 cells remain their metastatic phenotype. But when the cells are expressing dominant-negative TGFβRII, they don't form metastasis. Therefore, TGFβ signaling is required even so a metastatic environment is present when the primary tumor is removed (Oft et al., 1998b). Invasiveness of mouse colon cancer cells is checked with collagen gel when TGFβRII signaling is blocked, and they again didn't manage to invade (Erdogan & Webb, 2017).

In conclusion, TGFβRII signaling is necessary for metastatic tumors to form invasiveness via EMT regardless of the cell cycle (Oft et al., 1998a). Structural defects on the receptors cause gastric cancer cells to be resistant to growth inhibitory effects of TGFβ. DNA repair errors in human cervical carcinoma colon cancer establish a connection with a frameshift mutation in the TGFβ-RII coding region (De Miranda et al., 2015). These mutations are also detected in biliary tract tumors and sporadic cecum cancers. Defects and mutations on TGFβ-RI gene are observed in primary breast cancer carcinoma, human chronic lymphocytic leukemia (CLL), and human cervical carcinoma (Kim et al., 2000).

3.3.3 MiRNAs

Besides the TGFβ, the most known and studied driver of the EMT, microRNAs, noncoding RNAs are also involved in the process. ZEB and miR200 family have an adverse effect on each other; TGFβ downregulates miR200 family members. They function to maintain the epithelial state and inhibit invasive genes, just like miR34 and SNAIL1 act together as regulators of EMT

(Bendoraite et al., 2010). On the other hand, miR181 and miR182 work contrary to SMAD7 to increase invasiveness in some cancer cells (Suzuki, 2018). Long noncoding RNAs (LncRNA) are also related to TGFβ induced EMT in cancer. LncRNA activated by TGFβ cooperating with IL11, STAT3 act as an EMT inducer (Hao et al., 2019). From a developmental aspect of the TGFB superfamily members, Nodal signaling, which is one of the key players in development, is regulated by microRNAs in some organisms. For example, miR-430 is required for germ layer specification in Zebrafish and MiR-15/16 in Xenopus (M. Y. Wu & Hill, 2009).

3.4 TGFβ and EMT in Cancer

3.4.1 Tumor Suppressor and Metastatic Dual Role of TGFβ in Cancer

TGFβ induced EMT is a key indicator of metastasis and invasiveness in many cancer types. Somatic mutations and deletions on TGFβ and TGFβ receptors cause defects in various cell types (Massagué, 2008).

TGFβ acts as a tumor suppressor in early-stage tumors with epithelial cells. TGFβ leads to cell cycle arrest by suppressing cell growth and proliferation in the G1 phase. This process is done by inhibiting cyclin-dependent kinases (CDKs) like c-Myc (Iordanskaia & Nawshad, 2011). TGFβ causes cell death of tumor cells at an early stage by expressing TGFβ-SMAD dependent apoptotic factors like BIM, BIK, BCL-XL, and activating downstream pathways of the TGFβ signaling responsible for apoptosis (p38-JNK, MAPK) (Spender et al., 2009). TGFβ also leads to cell death through various events like reactive oxygen species (ROS) induction, autophagy, and epigenetic differences (Hao et al., 2019). In a recent study, the functional role change of TGFβ from tumor suppressor to enhancer has revealed as conserved14-3-3ζ molecule by changing the p53 levels in breast cancer cells (Jia Xu et al., 2016). In later stages, tumor cells become resistant to the growth inhibition caused by TGFβ. Thus, cancer cells start to produce TGFβ to enhance their invasive and migratory profile. TGFβ increases tumor invasiveness by activating factors for angiogenesis to supply oxygen for tumors and help for vessel formation (Pickup et al., 2013). Cytokines secreted for bone metastasis like IL11, CTGF, MMP1 and PMEPA1 are induce by TGFβ (Fournier et al., 2015). At the early stages of the cancer, high levels of PMEPA1 cause a decrease in the TGFβ leading a tumor suppressor function for itself. At the later stages, low levels of PMEPA1 is used as a bad prognosis factor (Xie et al., 2018).

Some breast cancer cells share the same regulated genes as EMT, such as INHBB, ELF3, FBN1, and TGFβ3. There is a high correlation between EMT levels and survival rates of brain, kidney,

and body cavity cancers and not as high but significant in thyroid, eye, and bladder cancers (Figure 5). From the TGFβ perspective, brain, bone, and skin tumors show high signaling levels of TGFβ (Foroutan et al., 2017). Gordian et al., created an EMT signature in cancer with a survival aspect. Genes like CDH1, LAG3, CTLA4, and PDL1 showed high levels of association with TGFβ induced EMT and metastasis-free survival rates (Gordian et al., 2019). Genes associated with EMT like PDGFA, COL6A1, IGFBP5, and IDB2 are used as bad prognosis markers in breast cancer (Valcourt et al., 2005).

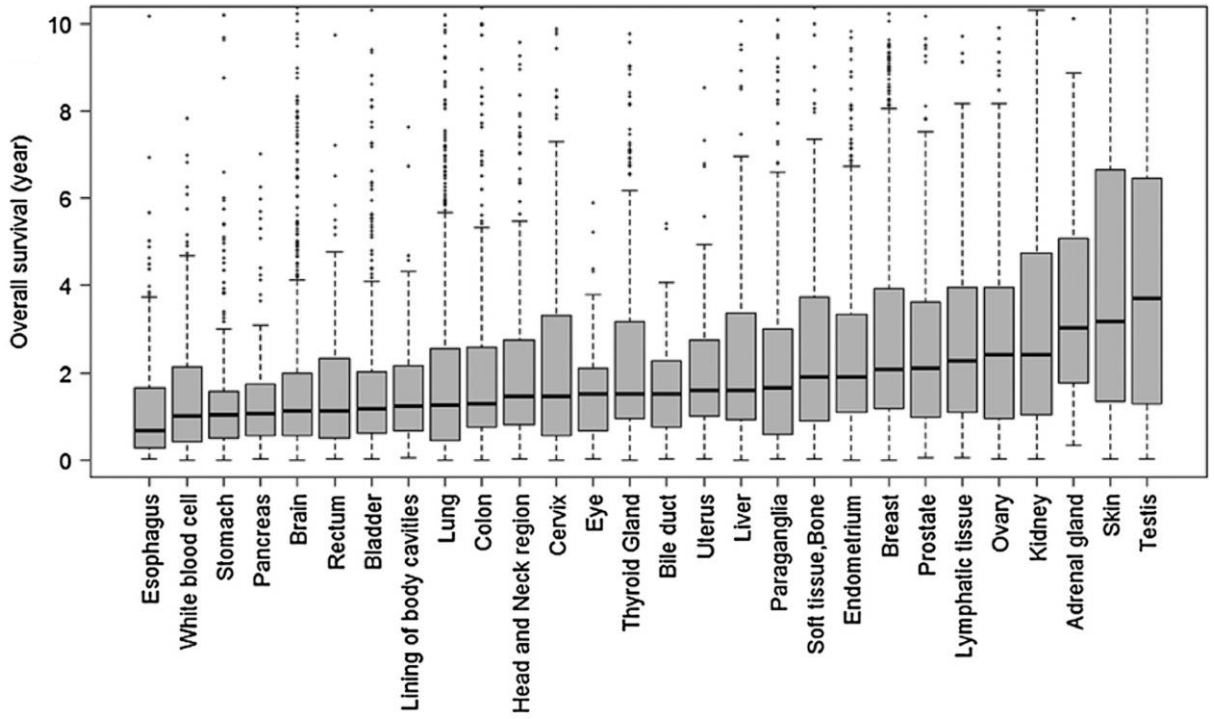


Figure 5: Overall survival rates of different cancer types with 10-year censoring (obtained from Foroutan et al., 2017).

Increasing cell migration and morphological changes caused by EMT are TGFβ dose-dependent on A549 cells (adenocarcinoma) in vitro. Keshamouni et al. (2006) concluded that; differentially expressed genes in Adenocarcinoma cells from TGFβ induction are also upregulated in different cancer cell lines and metastatic affect processes like adhesion, migration and invasion (Batlle & Massagué, 2019). Since pancreatic cancers have a poor survival rate and glycosylation may effect differentiation and migration processes with EMT, Maupin et al. (2010) investigated glycosylation alterations within the TGFβ induced EMT. Glycoproteins like cadherin 15 (CDH15), integrin alpha 2, and 5 (ITGA2, ITGA5) showed increased expression rates. Thus, glycans play a role in EMT and mesenchymal state maintenance (Maupin et al., 2010).

3.5 Drug Targets for TGFβ

TGFβ, its downstream signaling molecules and receptors are used in clinical trials as a drug target by itself or combined with other cancer therapy methods. Targeting is done by inhibiting receptor binding, blocking the expression of the ligand, and interfering with the translation. Targets such as; TGFβ2 inhibiting lung cancer vaccine Belagenpumatucel-L (Lucanix), TGFβ2 mRNA inhibits antisense oligonucleotide Trabedersen(AP 12009) for pancreatic cancer, TGFβRI inhibitors Galunisertib and SM16, neutralizing antibodies Fresolimumab, Metelimumab, and Lerdelimumab (Nemunaitis, 2013). Vactosertib is being used for immune checkpoint chemotherapies (Hao et al., 2019). There are some drugs in the market or in trials that use TGFβ1 and TGFβ receptors for various indications, such as Hyaluronidase as a TGFβ1 inhibitor used for increasing the effectiveness of other drugs. Fostamatinib is used for chronic immune thrombocytopenia (ITP) treatment, and it's an inhibitor of the TGFβ receptor (Table 1) (Wishart et al., 2018).

Table 1: Drug relations of TGFβ1 and TGFβ receptors

Drug Relations		
TGFβ1		
Drug Name	Actions	Indication
Hyaluronidase (ovine)	Inhibitor	For an increase of absorption and distribution of other injected drugs and for rehydration.
Vorhyaluronidase alfa	Inhibitor	an adjunct in subcutaneous urography for improving resorption of radiopaque agents.
Foreskin fibroblast (neonatal)	Agonist	Indicated for use for the treatment of full-thickness, diabetic foot ulcers
Foreskin keratinocyte (neonatal)	Agonist	For chronic leg ulcers and diabetic foot
Terazosin	Agonist	For use in treating symptomatic benign prostatic hyperplasia and hypertension
TGFβ Receptor Type I, II		
Fostamatinib	Inhibitor	For use in the treatment of chronic immune thrombocytopenia (ITP)

3.6 Role of TGFβ on Immunology

Treg cells activate TGFβ derived by tumor cells to enhance tumor growth. Those tumors induce TGFβ, block T Cell differentiation to Th1 cells and lead them to differentiate into Tregs to block CD8+ activity. It represses the activity of neutrophils and natural killer (NK) cells,

inhibits dendritic cell proliferation and their antigen-presenting functions (Hao et al., 2019). TGFβ1 notably causes immune evasion of cancer cells. TGFβ signaling also regulates T cell tolerance, which restrain possible autoimmunity. TGFβ affects the tumor microenvironment in an immune suppressive way to the tumors favor. TGFβ drives immune evasion in cancer cells. TGFβ inhibits the antigen presentation by dendritic cells (Batlle & Massagué, 2019).

3.7 Developmental Effects of TGFβ

Besides TGFβ1, TGFβ2, and TGFβ3 isoforms, the TGFβ superfamily has more than 30 superfamily members as bone morphogenetic proteins (BMPs), growth and differentiation factors (GDFs), and activins, which have prominent duties in developmental processes like embryonic development, differentiation, migration and axes generation (Namwanje & Brown, 2016). For instance, Nodal is responsible for the induction of endoderm, mesoderm, and ectoderm germ layers, and other superfamily ligands have roles in tissue organogenesis and homeostasis.

3.8 Diseases Caused by EMT and TGFβ Signaling Pathway Mutations

TGFβ isoform deficiencies cause various developmental defects in different organisms. TGFβ is an inducer for EMT that plays a crucial role in cardiogenesis. It enhances angiogenesis to increase intravasation to the circulatory system during development. TGFβ induced EMT may also prompt cardiac fibrosis. TGFβ isoforms (1,2,3) are upregulated during different stages of heart valve development in mice and chicken. While TGFβ2 initiates EMT, TGFβ3 controls the migration during the process. (M. Y. Wu & Hill, 2009) In mice, TGFβ1 null embryos show defects in cardiogenesis, such as in atrioventricular junctions. Also, with TGFβ2 deficiency and, the atrioventricular parts, defects of the outflow tract occur in mice (Cui et al., 1996). There are different studies on all three isoforms and their developmental effects when deficient. TGFβ3 causes palatal shelve fusion error in null mice (Nawshad et al., 2007). As in cardiac fibrosis, TGFβ also acts in fibrosis of different organs due to tissue injury such as in kidney, pulmonary, and hepatic tissues, TGFβ1 knockout mice mostly die from embryonic developmental defects. TGFβ2 knockouts have defects in the skeleton, heart, and retina (Xu J., Lamouille S., 2016).

Signaling failures in developmental steps can cause diseases. As an example, Polycystic ovary syndrome and Persistent Müllerian Duct Syndrome (PMDS) are caused by Anti-Müllerian

hormone (AMH) deficiency, which is a TGFβ family member (Korkmaz et al., 2017). Mutations in ligands like BMPR1B may cause loss or shortened limbs and chondrodysplasia (Stange et al., 2015). Hereditary Hemorrhagic Telangiectasia (HHT) and Rendu-Osler-Weber syndrome are caused by the defects in ALK1, a TGFβ receptor (Fernández-L. et al., 2006). Mutations in Smad4 are linked to Juvenile Polyposis. Excessive TGFβ induction and EMT causes fibrotic diseases due to ECM components like tubular atrophy (Johansson et al., 2015). Marfan Syndrome (MFS) is caused by connective tissue defects which have a mutation in the Fibrillin-1 gene. MFS type 2 occurs due to TGFβ2 mutations (Martínez-Quintana et al., 2014). Germline mutations on TGFβ1 cause Camurati-Engelmann disease (Chang et al., 2002).

3.9 Apoptosis and Stress

TGFβ induced apoptosis is one of the tumor suppressor roles of TGFβ in early-stage tumors via the Smad pathway by regulating pro and anti-apoptotic genes (Song, 2007). In leukemia cells apoptotic genes are induced by decreasing Bcl-2 levels (Spender et al., 2009). Reduced levels of NFKβ causes cell death by inducing C-myc expression (Liao & Yang, 2017). Activation of the caspases by TGFβ also causes apoptosis (Ramesh et al., 2009).

EMT takes a role in various stress types such as; mechanical stress, altered extracellular matrix composition, acidosis, genotoxic stress, extracellular mediators, and ER stress (Gjorevski et al., 2012). Lipolysis products activate stress response in TGFβ signaling via Smad and non-Smad pathways, which cause apoptosis. Cell death by TGFβ downregulation is also caused by SAPK (Stress Activated Protein Kinase) pathways (Han et al., 2018). Oxidative stress causes TGFβ downregulation and leading cancer cells to undergo apoptosis. TGFβ rapidly activates RhoA-dependent signaling pathways to induce stress fiber formation (Lee et al., 2019). Proteotoxic stress downregulates the expression of TGFβ RII and suppresses activation of the downstream effectors, thus inhibiting EMT (Tan et al., 2017). A mechanical type of stress, fluid shear stress, causes cells to change into spindle-like cells morphologically and increases TGFβ levels (Kunnen et al., 2017).

4 Materials and Methods

4.1 Type of The Study

The study was performed computationally.

4.2 Time and Place of The Study

The study took place at Izmir International Biomedicine and Genome Institute, Dokuz Eylul University in 2019-2020.

4.3 Ethical Committee

No ethical committee approval was required, as the study was only performed by computational analysis.

4.4 Variables of The Study

Publicly available datasets from the GEO (Gene Expression Omnibus) database of NCBI (The National Center for Biotechnology Information) portal were used. Human and mouse cell lines treated with TGFβ isoforms with EMT signs were analyzed.

4.5 Computational Analysis

All statistical and computational analyses were done by R version 3.6.0 (R Core Team, 2019) and Bioconductor package of R (Huber et al., 2015).

4.6 Microarray Data Selection

Public microarray data were collected from the GEO (Gene Expression Omnibus) database in the NCBI portal. Chosen data were collected depending on (1) only from the Affymetrix platform, (2) various cancer and normal cell lines from only human and mouse organisms. (3) valid expression levels of epithelial and mesenchymal marker genes like CDH1, VIM, ZEB1, SNAI2 for TGFβ induced EMT presence. (4) at least three repeats of each treatment. (5) clear knowledge of the TGFβ isoform (1-2-3) used.

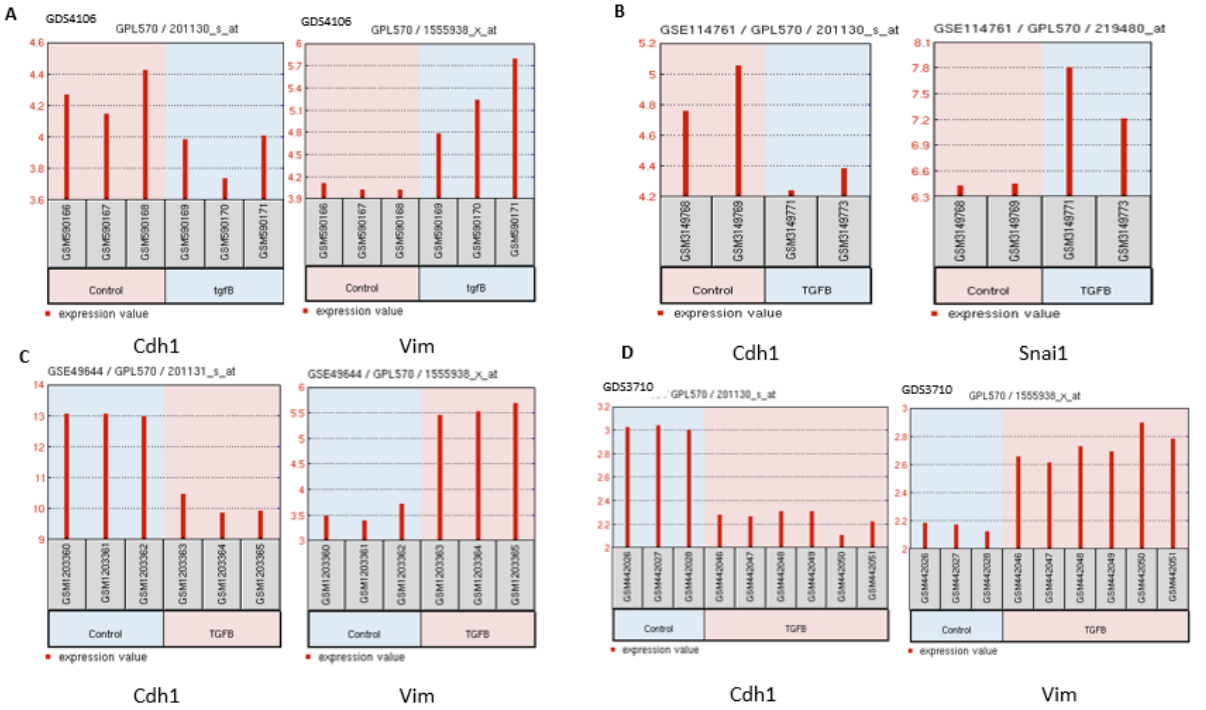


Figure 6: Examples of valid expression levels of epithelial marker Cdh1 and mesenchymal markers Vim or Snail checked for data selection. GEO2R web tool of NCBI (which uses the GEOquery and limma R packages) was used. (A) Dataset GEO: GDS4106, (B) Dataset GEO: GSE114761, (C) Dataset GEO:GSE49644 (D) Dataset GEO: GDS3710.

4.7 Population and Samples of The Study: TGFβ Induced Datasets

In GEO: GDS3710 dataset, Human A549 (Adenocarcinoma) cell line was induced with five ng/mL of TGFβ1 for 72 hours. GPL570 Affymetrix platform was used for microarray data (Keshamouni et al., 2006). In GEO: GDS4106 dataset, Human Panc-1 (Pancreas) cell line was treated with five ng/mL of TGFβ1 for 48 hours. GPL570 Affymetrix platform was used for microarray data.(Maupin et al., 2010) In GEO: GSE13986 dataset, murine mammary gland cells (NMuMG) were treated with five ng/mL of TGFβ1 and TGFβ3 for 24 hours. GPL81 Affymetrix platform was used for microarray data (Liu et al., 2008). In GEO: GSE114761 dataset, human lung cancer cell lines A549, H1944, H292, and H358 were treated with five ng/mL of TGFβ1 for 48 hours. GPL570 Affymetrix platform was used for microarray data (Gordian et al., 2019). In GEO: GSE35830 dataset, human ectocervical epithelial cells were treated with 15 ng/mL of TGFβ3 for 10 hours. GPL570 Affymetrix platform was used for microarray data (Sharkey et al., 2012). In GEO: GSE49644 dataset, human non-small cell lung cancer cell lines (NSCLC) as; A549, HCC827, and H358 were treated with two ng/mL of TGFβ1 for three weeks. GPL570 Affymetrix platform was used for microarray data (Sun et al.,

2014). In GEO: GSE40266 dataset, human normal ovarian fibroblast (NOF) line NOF151 was treated with five ng/mL of TGFβ1 and TGFβ2 for 48 hours. GPL570 Affymetrix platform was used for microarray data (Yeung et al., 2013). In GEO: GSE40466 dataset, murine mammary gland cells (NMuMG) were treated with five ng/mL of TGFβ2 for 24 hours. GPL1261 Affymetrix mouse platform was used for microarray data (Hussey et al., 2012).

Table 2: Public datasets used from GEO Database.

GEO Accession Number	Organism	Cell Line	Tissue	Platform	TGFB Isoform	Treatment Time	Reference
GDS3710	Homo Sapiens	A549	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	72h	PMID:20007254
GDS4106	Homo Sapiens	Panc-1	Pancreas	Affymetrix - HG-U133_Plus_2 GPL570	B1	48h	PMID:20885998
GSE13986	Mus Musculus	NMuMG	Mammary Gland	Affymetrix - MG_U74Av2 GPL81	B1	24h	PMID: 19094238
GSE13986	Mus Musculus	NMuMG	Mammary Gland	Affymetrix - MG_U74Av2 GPL81	B3	24h	PMID: 19094238
GSE35830	Homo Sapiens	Ect1	Ectocervical Epithelial	Affymetrix - HG-U133_Plus_2 GPL570	B3	10h	PMID: 22706080
GSE49644	Homo Sapiens	H358	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	3w	PMID: 25379179
GSE49644	Homo Sapiens	HCC827	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	3w	PMID: 25379179
GSE49644	Homo Sapiens	A549	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	3w	PMID: 25379179
GSE114761	Homo Sapiens	A549	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	48h	PMID: 30783512
GSE114761	Homo Sapiens	H358	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	48h	PMID: 30783512
GSE114761	Homo Sapiens	H292	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	48h	PMID: 30783512
GSE114761	Homo Sapiens	H1944	Lung	Affymetrix - HG-U133_Plus_2 GPL570	B1	48h	PMID: 30783512
GSE40266	Homo Sapiens	NOF151	Ovary	Affymetrix - HG-U133_Plus_2 GPL570	B1	48h	PMID: 23824740
GSE40266	Homo Sapiens	NOF151	Ovary	Affymetrix - HG-U133_Plus_2 GPL570	B2	48h	PMID: 23824740
GSE40466	Mus Musculus	NMuMG	Mammary Gland	Affymetrix -Mouse430_2 GPL1261	B2	24h	PMID: 23285117

4.8 Differential Expression Analysis

Differential expression analysis was performed using R programming. (Version 3.6.0 for Windows (32/64 bit)) R is a programming language developed for statistical and computational analysis, graphical visualization. Bioconductor is an open-source software for high throughput single-cell data analysis. The data analysis workflow in Bioconductor begins with quality control, which will eliminate low-quality cells, normalization, and remove batch effects and

biases. Then, cells are clustered based on their expression profile. The next step is differential expression analysis to identify genes in the separate clusters (Amezquita et al., 2020).

Data were normalized by the RMA Method using the Affy package of Bioconductor (Gautier et al., 2004). Quality control charts of the Affymetrix microarray datasets were created by using the Simpleaffy package (Miller, 2019). Beta-actin and Gapdh genes were used as QC markers; the expression threshold was determined as 3. Using pOverA(0.25, log2(16)), filterfun and genefilter functions of Genefilter package, data were filtered (Gentleman et al., 2019).
NA values, repeated genes were removed during filtration. Non-significant values based on their P-values were excluded during the analysis. With the Limma package, differential expression analysis was performed (Ritchie et al., 2015). Genes were ranked by their logarithmic fold change values (log2±0.6 and log2±1). Upregulated and downregulated genes from TGFβ1, TGFβ2 and TGFβ3 treated samples were intersected in Venn schemes using R to determine common and isotype-specific regulated genes. Additional PCA analysis, dendrograms, and quality control charts were created by the Chipster web tool (Kallio et al., 2011). Volcano plot figures were plotted with an enhanced volcano package of Bioconductor/R (Blighe, 2019).

4.9 GSEA and Go Enrichment Analysis

Gene set enrichment analysis (GSEA) was performed with GSEA software, and The Molecular Signature Database (MSigDB) gene sets. Data were prepared in GenePattern modules (Reich et al., 2006). Gene set enrichment analysis is a software for analyzing microarray data to detect differentially expressed genes in certain phenotypes (Control or treated). It correlates the gene sets, which are “Groups of genes that share common biological function, chromosomal location, or regulation.” (Subramanian et al., 2005) with differentially expressed genes. This method is used to avoid the limitations of single-gene analysis. If the genes in a gene set are highly correlated with the given phenotype, they show higher correlation scores. (t-test) Enrichment score (ES) is defined as the gene’s expression difference between two phenotypes. A higher enrichment score means the gene is highly correlated with the phenotype. P-values and false discovery rates are calculated to maintain significance (Shi & Walker, 2008).

50 Hallmark (H) gene sets from MSigDB collections were used as gene families. Regulated genes were ranked based on their enrichment scores (ES) on the assigned phenotype. Gene sets were filtered with parameters like minimum (50) and maximum gene set sizes (150). False discovery rate (FDR) was determined as less than 25% and P-value as less than 0.01. Gene sets

were correlated with the enriched regulated genes within the phenotype. Upregulated and downregulated genes, those genes with hallmarks of specific functions or pathways in the gene sets, were revealed (Mootha et al., 2003)(Subramanian et al., 2005).

GO(Gene Ontology) enrichment analysis was performed via G: profiler (Raudvere et al., 2019) online tool. Gene Ontology (GO) enrichment analysis is a software developed to investigate the roles of the genes in the cell as; what are their molecular functions, which biological processes those genes are involved in and the cellular compartments where they actively reside (Gene Ontology Consortium et al., 2000). G: profiler is a web server where GO enrichment analysis can be performed and available online (https://biit.cs.ut.ee/gprofiler).

Tools available within the g: profiler are KEGG, Reactome, Wikipathways, Transfacs. The server uses Ensembl as the data source.

Overrepresented genes in the query are analyzed by hypergeometric tests. Significant information based on the logarithmic P values, about biological processes, molecular functions and cellular components are determined. Upregulated genes belonging to each isoform phenotype (TGFβ1, TGFβ2, TGFβ3) from differential expression analysis were used as input queries.

4.10 Cancer Genomics Data Analysis

Cbio portal (Gao et al., 2013) was used for cancer genomic data analysis. Cbio portal is an open source available online. (http:// cbioportal.org) It is developed to access the cancer genomic data collected from The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC). Portal's primary purpose for users to interpret data quickly and visualize it. Entered genes as a query are classified as altered or not. Altered genes are investigated in the cancer studies if they have mutations like amplifications, deletions (Cerami et al., 2012).

Cancer types that TGFβ1, TGFβ2, and TGFβ3 genes are altered; alteration types as mutation, deletion, amplification; expression rates of the TGFβ genes in the cancer types were determined using the portal.

5 Results

5.1 Quality Control Charts

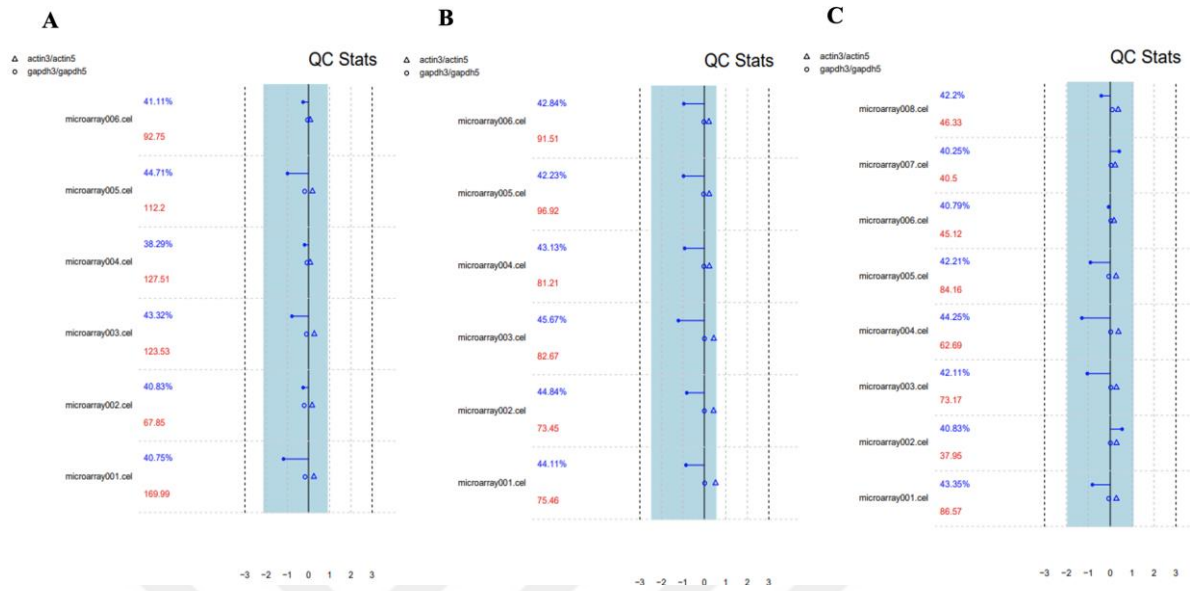


Figure 7: Quality Control Charts of the analyzed human Affymetrix microarray datasets. Percentage: the average background on the chip. Red numbers: number of probesets. Blue region: scaling factors less than 3-fold of the mean scale factors of all chips. Bars: scaling factors for the chips. Triangles: beta-actin 3':5' ratio, Circles: GADPH 3':5' ratios. (A) GEO:GD4106 (B) GEO: GDS3710 (C) GEO: GSE35830.

Affymetrix microarray datasets used in the analyses were checked for quality control. QC Stats of human (Figure 7, Figure 8) and mouse (Figure 10) datasets showed that samples were qualified for analysis as Actin and Gapdh genes were within the threshold fold. Relative log expression value (RLE) graphs of human (Figure 9) and mouse (Figure 11) dataset samples show the overall quality of the expression levels. β2 samples of mouse dataset GEO: GSE40466 was excluded from the quality control analysis due to 2 sample sets.

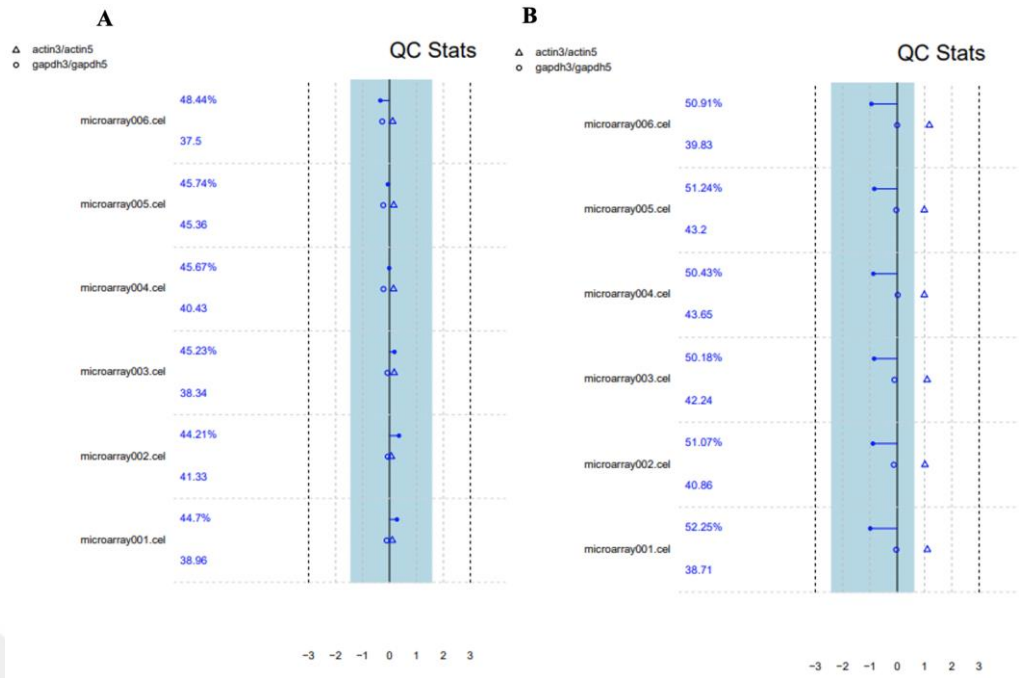


Figure 8: Quality Control Charts of the analyzed human Affymetrix microarray datasets. Percentage: the average background on the chip. Red numbers: number of probesets. Blue region: scaling factors less than 3-fold of the mean scale factors of all chips. Bars: scaling factors for the chips. Triangles: beta-actin 3':5' ratio, Circles: GADPH 3':5' ratios. (A) GEO: GSE40266 (B) GEO: GSE49644.

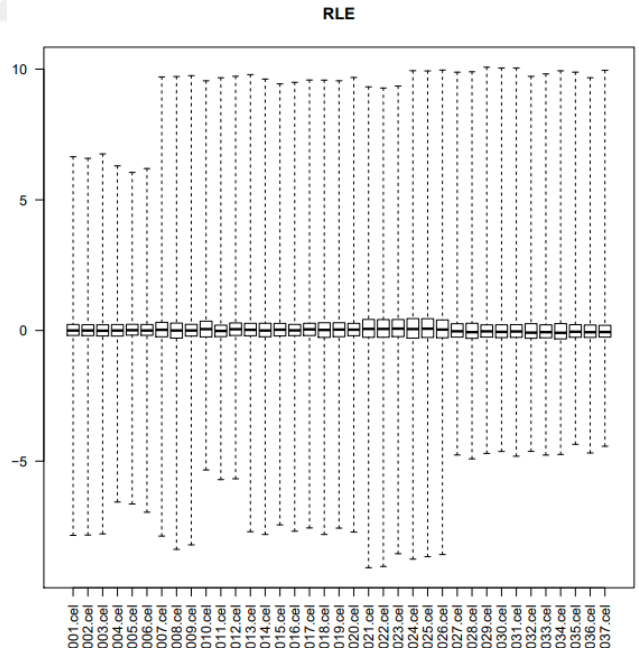


Figure 9: Relative log expression values (RLE) of all samples of human Affymetrix datasets used for analysis. From left to right; GEO:GD4106, GEO: GDS3710, GEO: GSE35830, GEO: GSE40266, GEO: GSE49644, GEO: GSE114761.

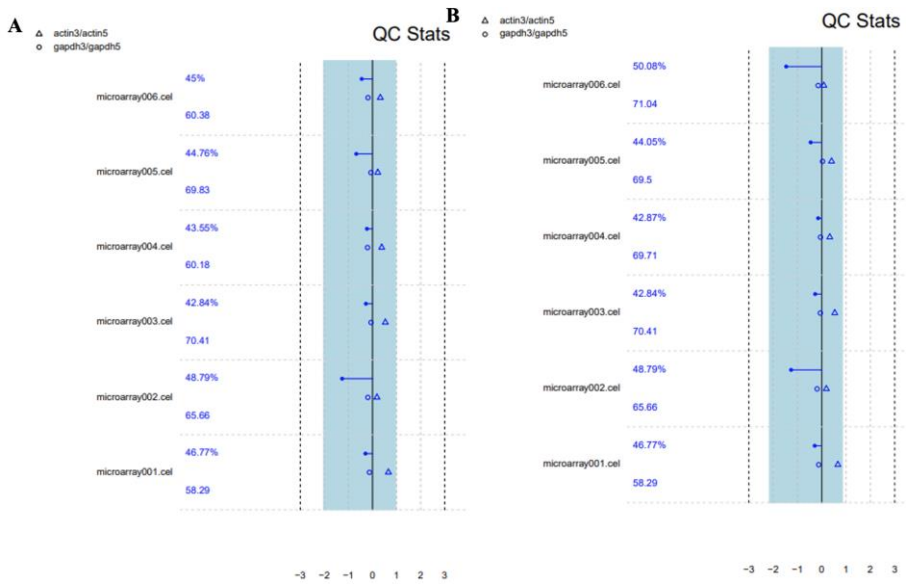


Figure 10: Quality Control Charts of the analyzed mouse Affymetrix microarray datasets. Percentage: the average background on the chip. Red numbers: number of probesets. Blue region: scaling factors less than 3-fold of the mean scale factors of all chips. Bars: scaling factors less than 3-fold of the mean scale factors of all chips. Bars: scaling factors for the chips. Triangles: beta-actin 3':5' ratio, Circles: GADPH 3':5' ratios. (A) GEO: GSE13986- β1 (B) GEO: GSE13986- β3.

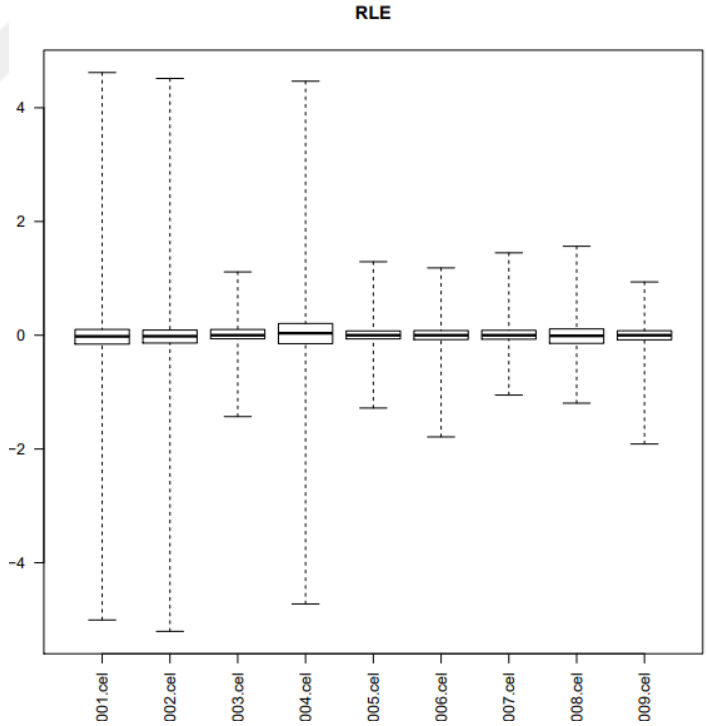


Figure 11: Relative log expression values (RLE) of β1 and β3 samples of mouse Affymetrix datasets used for analysis. From left to right; GEO: GSE13986- β1, GEO: GSE13986- β3.

5.2 Differential Expression Analysis

Differential expression analysis was performed in R version 3.6.0 (R Core Team, 2019) and Bioconductor package of R (Huber et al., 2015). Public datasets from the GEO database (Table 2) of TGFβ1, TGFβ2, TGFβ3 induced human, and mouse cell lines were analyzed, normalized and filtered. Differentially expressed genes (DEGs) were ranked by their logarithmic fold change values (log2) for each dataset and cell type. Genes were filtered using the Genefilter package of Bioconductor, and the number of genes after normalization and filtration were determined.

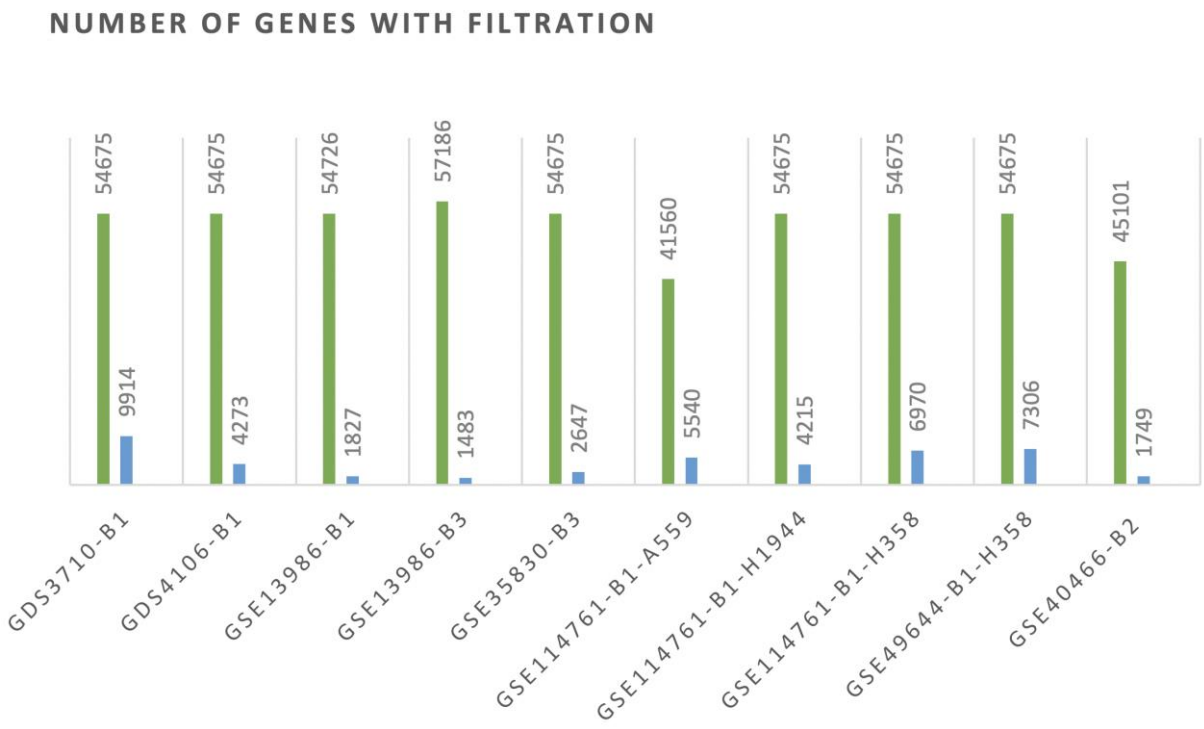


Figure 12: Number of genes used in differential expression analysis before and after filtrated.

Table 3: Logarithmic fold change(log2) values of epithelial and mesenchymal marker genes from differential expression analysis.

Data Set	GRHL3	CDH1	OVOL2	ELF3	EHF	SNAI1	SNAI2	ZEB1	ZEB2	VIM	CDH2	TWIST2
GDS3710		-4,424		-4,553	-5,536		7,188	2,210		1,338	1,906	-1,278
GDS4106	-0,944	-2,344				1,483	2,420			2,336		-1,632
GSE13986				1,076	1,229			-1,002			-0,994	
GSE13986				0,999	1,089			-1,106				
GSE49644		-4,031	-3,379	-3,503	-2,575	1,047	1,143	3,878	3,116	1,310	2,078	
GSE49644	-1,261	-6,741		-3,701			3,817	0,781		0,777	1,994	-0,671
GSE90954									0,704		0,633	
GSE114761	-1,475	-1,205	-0,8469	-2,725	-3,383	1,332	4,109	0,626	0,679	1,465	1,820	
GSE114761		-0,631			-1,178		1,630		0,943	1,094	1,207	
GSE114761					1,021			0,641				
GSE114761			-0,8639	-2,378	-0,885	1,840	1,118	1,010		1,130	1,101	
GSE35830							1,486			0,677	0,859	
GSE40266		2,208									0,692	
GSE40266		2,412							-0,605		0,924	
GSE40466					1,528	0,929	-1,292					1,339
GSE40466					0,839				0,984			-0,654

Epithelial and mesenchymal marker genes and their log2 fold change values were analyzed. (Table 3) GRHL3, CDH1, OVOL2, ELF3, and EHF were checked for epithelial markers where SNAI1, SNAI2, ZEB1, ZEB2, VIM, TWIST1, and CDH2 were marked as mesenchymal genes. Except for a few values, results were as expected.

Values may differ in each gene set due to treatment time with TGFβ and cell type.

CDH1 changing levels were around -2 (log2) in most of the datasets where it is even -6.7(log2) in GEO: GSE49644 dataset because of the three weeks long treatment time. The same difference can be seen in ZEB2 levels. When upregulated (+log2) (Table 4) and downregulated (-log2) (Table 5) genes were ranked based on their fold change values, top of the lists were visualized. Some of the genes were common between the dataset and will be analyzed in other parts.

In the cluster dendrograms from Chipster (Kallio et al., 2011) of (Figure 13 A) human, from left to right; GEO:GD4106, GDS3710, GSE35830, GSE40266, GSE49644, and GSE114761 (Figure 13 B) mouse, from left to right; GSE13986- β1, GSE13986- β3 datasets; control and treated samples were clustered within each other.

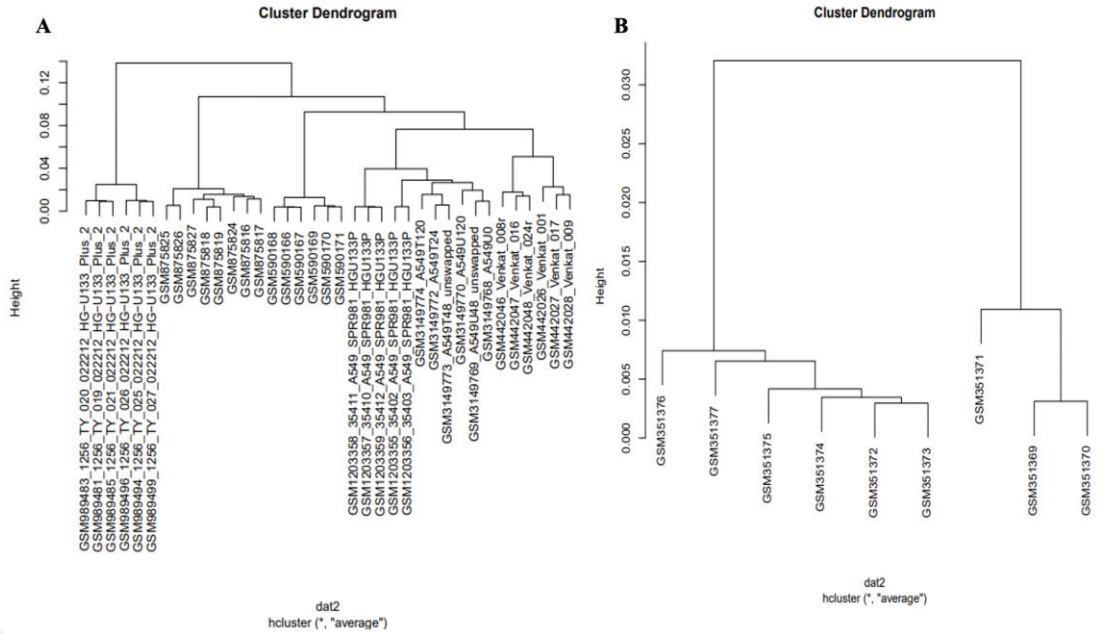


Figure 13: Cluster dendrograms of all samples from (A)Human, from left to right; GEO:GD4106, GEO: GDS3710, GEO: GSE35830, GEO: GSE40266, GEO: GSE49644, GEO:GSE114761. (B)Mouse, from left to right; GSE13986- β1, GEO: GSE13986- β3 datasets.

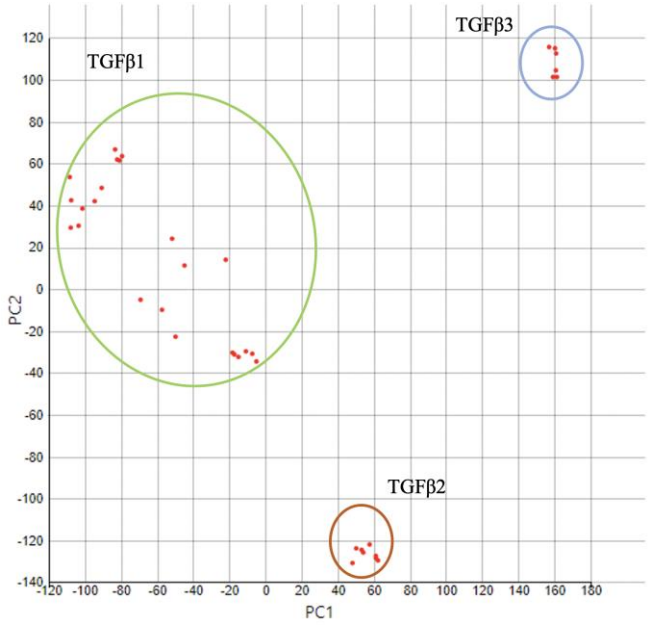


Figure 14: Principal Component Analysis (PCA) of samples from human datasets. TGFβ1; GEO:GD4106, GEO: GDS3710, GEO: GSE49644, GEO: GSE114761. TGFβ2; GEO: GSE40266. TGFβ3; GEO: GSE35830.

The principal component analysis (PCA) of human datasets showed that different TGFβ isoform treated samples were clustered separately (Figure 14).

Within the differential expression analysis in R, regulated genes were intersected to investigate the common genes with β1, β2, and β3 isotypes and genes regulated only in one isoform. Regulated genes were selected based on their logarithmic fold change levels (log2+ for upregulated, log2- for downregulated genes).

Table 4: Top 10 upregulated genes from differential expression analysis. Genes were ranked based on their logarithmic fold change (+log2).

Data Set	Cell LIne	TGFB	Top 10 Upregulated Genes
GDS3710	A549	B1	MAF, IGFBP5,IGFL1,SPOCK1,IL11,DHRS2,INHBA,LINC01279,CDH11
GDS4106	Panc-1	B1	CDH19,COL1A1,FOXS1,LTBP2,THBS1,TAGLIN,OXTR,MYO10
GSE13986	NMuMG	B1	TTR,UGT2B34,SERPINA1E, ALDH1A1, GSTA4,ITIH2,CYP3A13,CYPCL5,VNN,FGB
GSE13986	NMuMG	B3	RNASE4,UGT2B34,SERPINA1C, APOC2, ALDH1A1, SERPINA1E, HPX,TFF2,GSTA2,VNN1
GSE49644	H358	B1	IGFBP5,COL14A1,HS3STSB1,ALPK2,F2RL2,APCDD1L-DT,L1CAM,SPARC, ARHGEF40
GSE49644	A549	B1	INHBA,DNER,CDH11,IL33,VIP,BMP2,IL1A,BHLHE41,ANO4,LCE3D
GSE114761	A549	B1	INHBA,LCE3D,COL15A1,MAF,TNFAIP6,IGHL1,KANK4,COL1A1,SLN,MMP1
GSE114761	H1944	B1	NEDD9,RBP1,MAF,COL1A1,TSPAN2,LUM,PPBP,PTPRR,PMEPA1,LBH
GSE114761	H292	B1	HMGA2,CXCL8,SCEL,PHLDA1,TMCC3,PTPN22,KRT6A,ADGRF1,DNER,FST,SOX2
GSE114761	H358	B1	SERPINE1,MAF,ARHGEF40,MSC-AS1,ACKR3,AL11,LTBP2,LBH,CST6,PDLIM4
GSE35830	Ect1	B3	FN1,SERPINE1,LINC02551,NEDD9,PMEPA1,KANK4,IGFBP3,AKAP12,JUN,IGFBP3
GSE40266	Nof151	B1	COMP,ID4,TSPAN2,NEDD9,PLN,PMEPA1,ITGA4,DHRS2,MYOZ1,LDB3
GSE40266	Nof151	B2	COMP,TSPAN2,ID4,ATP10A,PLN,DHRS2,ITGA4,PMEPA1,RASGRP1,MYOZ1
GSE40466	E1kd	B2	BC080696, TCSTV1,GM20767,PDGFRB,GM12800,AF067061,CEP55,SOX11,EGR2
GSE40466	NMuMG	B2	KLHL30, AK1, LCE1F, CTLA2A, FN1, GJA1, CTSW, ACKR3, NCAM1,CLDN4

Table 5: Top 10 downregulated genes from differential expression analysis. Genes were ranked based on their logarithmic fold change (+log2).

Data Set	Cell LIne	TGFB	Top 10 Downregulated Genes
GDS3710	A549	B1	CP,TM4SF20,SMIM31,ALDH3A1, TRIM31, PDK4,CYP4F11,STS8SIA4,SOX2,RARRES1
GDS4106	Panc-1	B1	TXNIP, SCIN,SSTR5-AS1,KCNB1,CADM2,KLRC3,ARRDC4,TGFBR3,KLRC4,SORBS1
GSE13986	NMuMG	B1	SERPINE1,PTGS2,FBLN2, IL11, GJA1, CTLA2B, CYR61, VEGFC, F2R, GJA1
GSE13986	NMuMG	B3	SERPINE1, IL11, PTGS2, GJA1, WISP1, CTLA2B, CYR61, VEGFC, F2R, A1
GSE49644	H358	B1	ADGRF1, ESRP1, TMEM30B, ATP8B1, ERBB3, S100P, CYP4X1, PRRG4,MPZL2,GRHL2
GSE49644	A549	B1	TM4SF18,HPGD,MTUS1,CAVIN2,PDK4,PDE1C,CDH1,HPGF,AREG,CYP1B1
GSE114761	A549	B1	SCARA5,CYP4F3,CP,AGR2,SPX,MUC5AC,AGR2,FGA,S100P,AGR3
GSE114761	H1944	B1	GC,ALB,HAL,ADH6,ADH1C,DDC,S100A8
GSE114761	H292	B1	RASGRP1,ADAMTS15,SULT1E1,COL5A1,CAPN6,LGSN,ACKR3,GREM1,NEBL
GSE114761	H358	B1	HPGD,RGCC,EVI2B,ZNF750,INHBB,SHISA3,SMIM31,SLC27A2
GSE35830	Ect1	B3	CYP1A1,GPNMB,MYLIP,PTPRZ1,ALDH5A1,XK,GLRB,ETV1,ANKRD22,HSPA6
GSE40266	Nof151	B1	SERPINB2,IL33,CXCL6,SERPIND1,SLC14A1,ANXA10,ALDH1A3,CXCL8,ADAMTS1,TSPAN8
GSE40266	Nof151	B2	SERPINB2,IL33,CXCL6, CXCL1,IL24,SERPIND1,CXCL8,SLC14A1,ANXA10, ALDH1A3
GSE40466	E1kd	B2	NDN,PRR32,MAMDC2,CAR6,AVİL,AKR1B7, MMP13, KRT7
GSE40466	NMuMG	B2	MEOX2, CRİSP1, TTR, RANBP3L, S100A9, ARG2, CYP4A12A, ANXA10

For *Mus musculus* (Mouse) cell line NMuMG, Datasets GSE13986 for β1 and β3; GSE40466 for β2 were used. Intersection was performed out of 1590 genes in β1, 1175 genes in β2, and 1279 genes in β3. In Log2 fold change ±0.6 Mus musculus datasets (NMuMG cells), β1, β2, and β3 shared 246 genes. β2 phenotype had 859 regulated genes significant for itself. Even though β1 and β3 isoforms were from the same dataset and shared 1117 genes in common, β1 had 377, and B3 had 96 genes differentially expressed only with this phenotype. Those 96 genes may represent significant results for mouse cell line research (Figure 15 A). When the threshold is set to Log2 fold change ±1 number of shared genes among all isotypes became 6.

These genes were; SERPINE1, THBS1, SPSB1, LAMB3, GADD45B, and F3. SERPINE1 or plasminogen activator inhibitor 1 (PAI- 1) plays a role in wound healing and upregulated in EMT induced cancer cases like glioblastoma and gastric cancer (Seker et al., 2019). It is thought that SERPINE1 is strongly correlated with the TGFβ pathway and may be downstream of it (B. Xu et al., 2019). THBS1 is an ECM protein that appears in the tumor microenvironment and there are few studies about it is induced by TGFβ1 and expression of the gene activates TGFβ production (Radeke et al., 2015). Although SPSB1 is a negative regulator of TGFβ by decreasing the TGFβ receptor II levels, it was upregulated when treated with all three isoforms. LAMB3 plays a role in various cancer types, especially in pancreas cancer (H. Zhang et al., 2019).

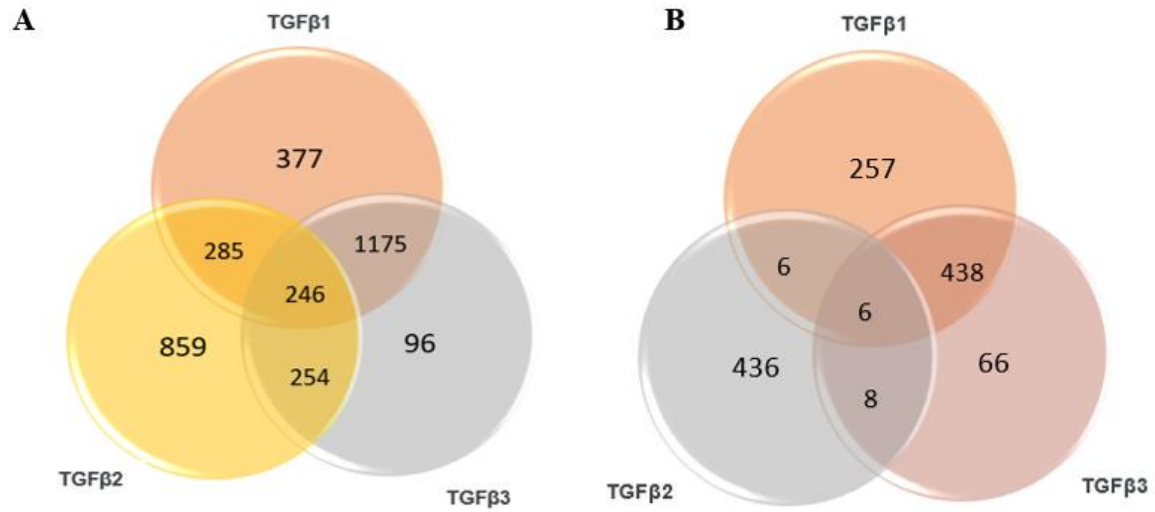


Figure 15: Differentially expressed genes shared in Mus musculus (NMuMG) datasets treated with TGFβ1, TGFβ2, TGFβ3. Datasets with GEO Accession numbers; GSE13986 for β1 and β3; GSE40466 for β2 were used from Gene Expression Omnibus (GEO), NCBI portal. (A) (log2 f old change ± 0.6) (B) (log2 fold change ± 1)

It contributes to the invasive and metastatic properties of the tumor cells. It may contribute to apoptotic decisions in the cell cycle by regulating Akt pathway, and the expression level of the gene is correlated with the N-cadherin and Vimentin (Jung et al., 2018). GADD45B is a Smad dependent target gene for TGFβ. It is a growth arrest and DNA damage repair protein that may be activated in stress conditions by activating the p38/Jnk pathway.

It may be activated due to cells that are not able transform into mesenchymal phenotype and end up in apoptosis. Finally, F3 is a cell surface glycoprotein and responsible for cell surface assembly which is a feature expected to be regulated with TGFβ treatment. With lower gene numbers, genes specific for certain isotypes were 257 for β1, 436 for B2 and 66 for β3 (Figure 15 B).

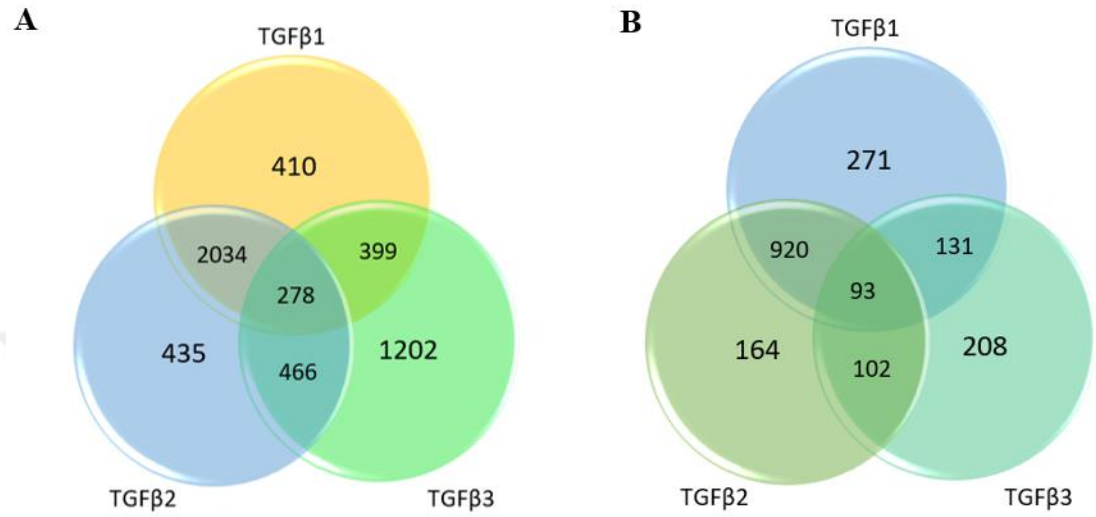


Figure 16: Differentially expressed genes shared in Homo sapiens (Human) datasets treated with TGFβ1, TGFβ2, TGFβ3. (A) log2 fold change ± 0.6 (B) log2 fold change ± 1. TGFβ treated datasets with GEO Accession numbers; GSE35830 (ECT1) for β3, GSE40266 (NOF151) for β1 and β2 and GSE114761 (A549, H358), GSE49644 (A549, H358) for β1 were used from Gene Expression Omnibus (GEO), NCBI portal.

For Homo sapiens (human) data, datasets GSE35830 (ECT1) for B3, GSE40266 (NOF151) for β1 and β2 and GSE114761 (A549, H358), GSE49644 (A549, H358) for β1 were used. When genes were ranked based on log2 fold change (±0.6), even though TGFβ1 and TGFβ2 had the same datasets with different isotype treatment and they share 2034 genes in common, the TGFβ1 phenotype had 410, TGFβ2 had 435 specific genes only regulated in each phenotype. The highest number of specific genes are within the TGFβ3 phenotype (1202). Three isoforms have shared 278 genes in common (Figure 16 A). By changing the fold change levels from ± 0.6 to ± 1, gene numbers decreased accordingly. Ninety-three genes shared by three isoforms, 271 genes specific for TGFβ1, 164 for TGFβ2, and 208 for TGFβ3 (Figure 16 B).

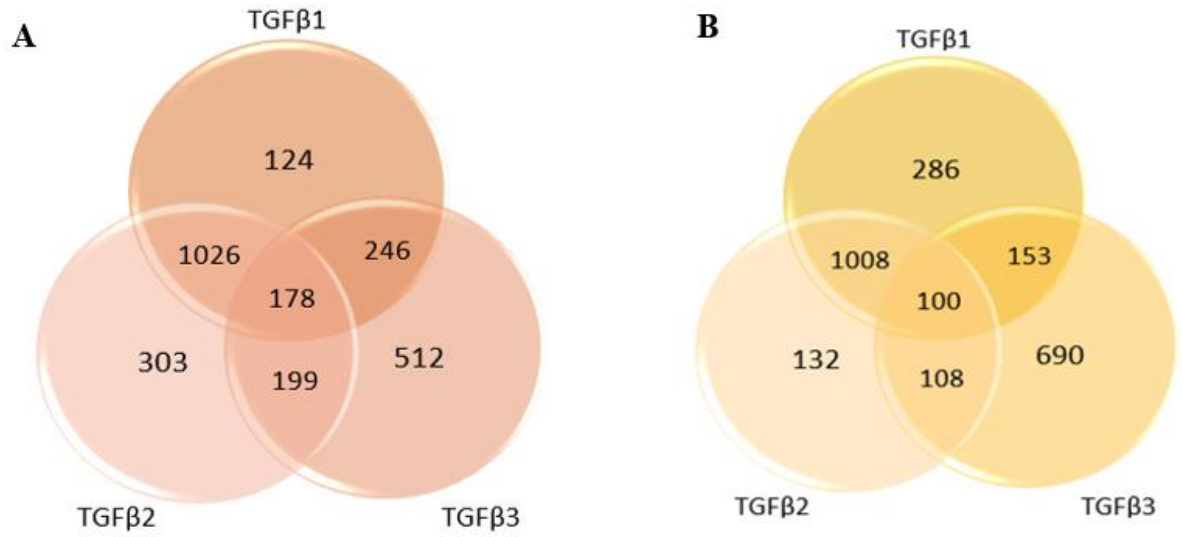


Figure 17: (A) Upregulated and (B) Downregulated (log2 fold change +0.6) genes shared in Homo sapiens (Human) datasets treated with TGFβ1, TGFβ2, TGFβ3. TGFβ treated datasets with GEO Accession numbers; GSE35830 (ECT1) for β3, GSE40266 (NOF151) for β1 and β2 and GSE114761 (A549, H358), GSE49644 (A549, H358) were used from Gene Expression Omnibus (GEO), NCBI portal.

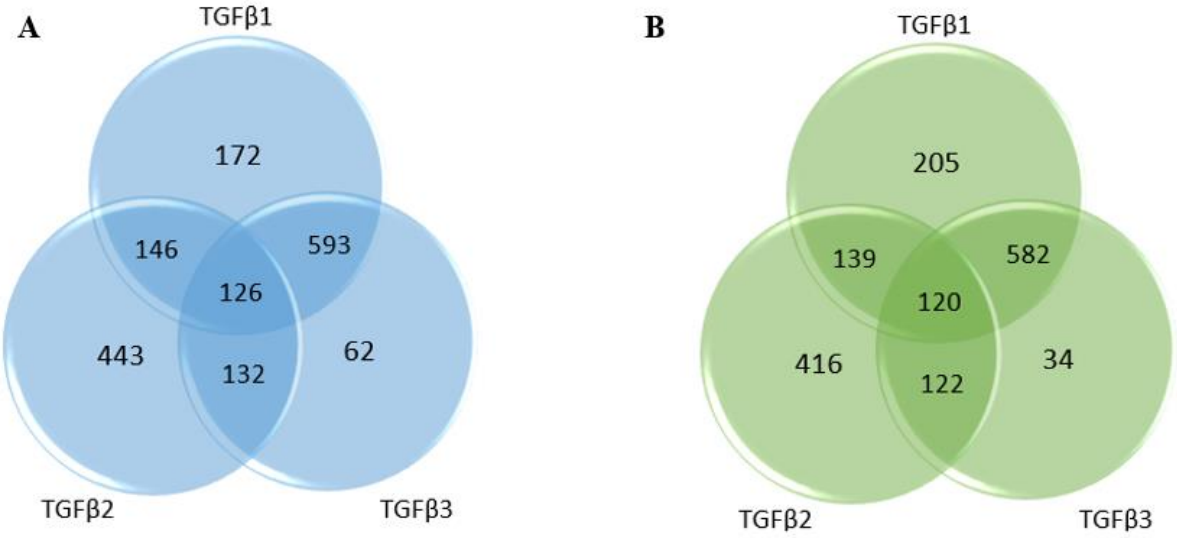


Figure 18: (A) Upregulated and (B) Downregulated (log2 fold change +0.6) genes shared in Mus musculus (NMuMG) datasets treated with TGFβ1, TGFβ2, TGFβ3 from differential expression analysis. TGFβ treated datasets with GEO Accession numbers; GSE13986 for β1, β3 and GSE40466 for β2 were used from Gene Expression Omnibus (GEO), NCBI portal.

Besides the total regulated genes treated with all three isoforms and in both organisms, upregulated (log_2 +0.6) and downregulated (log_2 -0.6) genes were separately intersected for mouse and human datasets. In human data, common genes in three isoforms are higher in upregulated values. One hundred seventy-eight upregulated genes, and 100 downregulated genes were shared among all isotypes.

One hundred twenty-four genes were specific for TGFβ1 like; ID2, CCPG1, FOXS1, ITGB3, DOCK2, ZEB1, NGF, and LOXL1. Three hundred three genes were specific for TGFβ2 such as; IDH2, DAP, MALAT1, MAPK11, NES, and TNNT2. Some of the 512 TGFβ3 specific genes were; SOX4, IL6, LAMA3, SMURF2, VEGFA, TUBA1A, GATA6, ITGA2, MAFF, CDK6, and OVOL1 (Figure 17). Also, upregulated (log_2 +0.6) and downregulated (log_2 -0.6) genes were analyzed for mouse data. Shared gene numbers were nearly equal. Such as common genes in all isotypes were 126 in upregulated and 120 in downregulated. TGFβ2 treated datasets have the highest number of specific genes for the isoform (Figure 18).

In the lists obtained from differential expression analysis (log2 fold change ± 0.6), out of 1590 genes from TGFβ1 treated mouse dataset and 2602 genes from human datasets had 395 common genes. In TGFB2 treated datasets, out of 1151 genes from the mouse dataset and 1093 human dataset there were 146 shared genes. In TGFβ3 treated datasets, out of 1279 genes from the mouse dataset and 420 genes from the human dataset, 70 genes were common. Between these shared values, an intersection of 3 isoforms were checked. 13 genes were common for all isotypes obtained from common genes of mouse and human. Those 13 genes were IL11, PTGS2, PDLIM7, GADD45B, LAMB3, IFIT3, SKIL, ID3, CXCL1, PCDH7, DHRS3, FN1, and MAP3K5 (Figure 20). TGFβ3 had 19 genes specifically upregulated in that isoform, which are; TUBA1A, IGF1R, GATA6, FOXQ1, KLF6, and IRF6. TGFβ1 only had a higher number of genes, 305, compared to the other two isotypes. As an example of TGFβ1 only genes; VEGFC, TSPAN8, ANKRD1, ID2, FDFR2, IGFBP7, BMP1, ACTN1 and so on. TGFβ2 had 92 isotype-specific genes such as; CALD1, TNS1, SLCO2A1, and MAP6 (Figure 19)

Later, all cancer cells within the TGFβ1 phenotype (GDS3710, GDS4106, GSE49644, GSE114761) were intersected. They shared 97 genes from differential expression analysis, such as IL11, COL1A1, TAGLN, THBS1, PMEPAI, JUN, and ADAM19.

A

TGFB Isotype	Number of Common Genes
B1	395
B2	146
B3	70

B

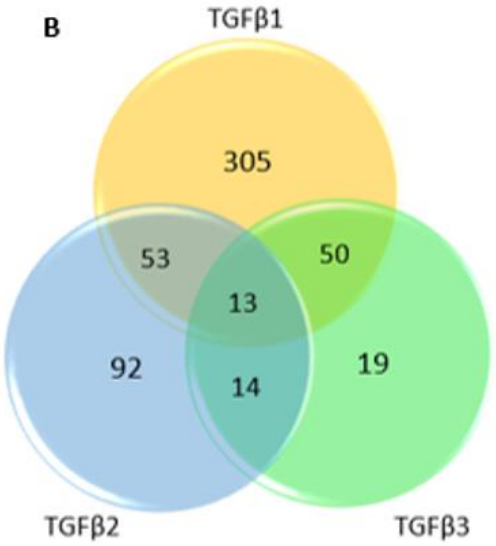


Figure 19: (A) Number of differentially expressed genes (log2 ± 0.6) common between TGFβ treated mouse and human datasets. (B) Intersection of differentially expressed genes that are common between human and mouse datasets from part A.

Also, upregulated genes with TGFβ induction were connected with related chromatin remodeler genes (Figure 21) based on their STRING interaction networks (Szklarczyk et al., 2019).

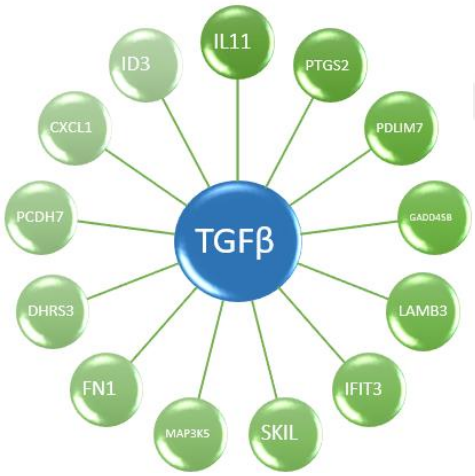


Figure 20: Commonly regulated genes for all isotypes obtained from intersected mouse and human datasets. (Figure 19 B)

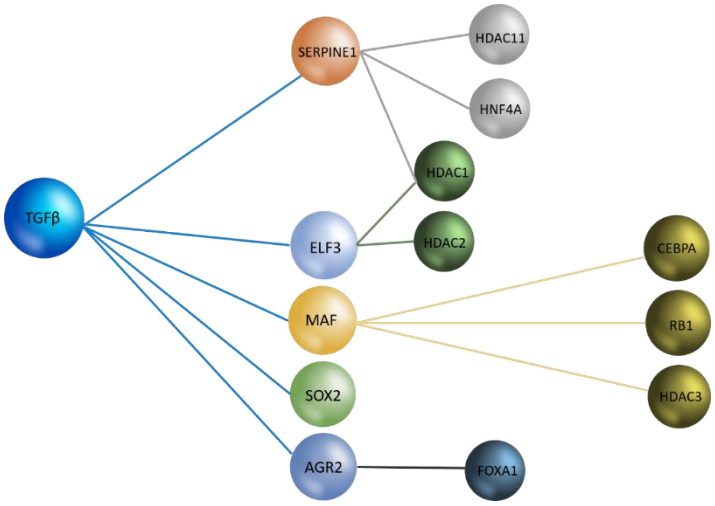


Figure 21: Interaction network between upregulated genes with TGFβ induction and chromatin remodeling genes.

Upregulated genes intersected in all isoforms of TGFβ from human (Figure 17 A) and mouse (Figure 18 A) datasets were investigated in the STRING (Szklarczyk et al., 2019) interaction networks. (Figure 22) Genes in the network showed high interaction with each other as they are associated with similar functions and pathways based on the curated databases and experimental results.

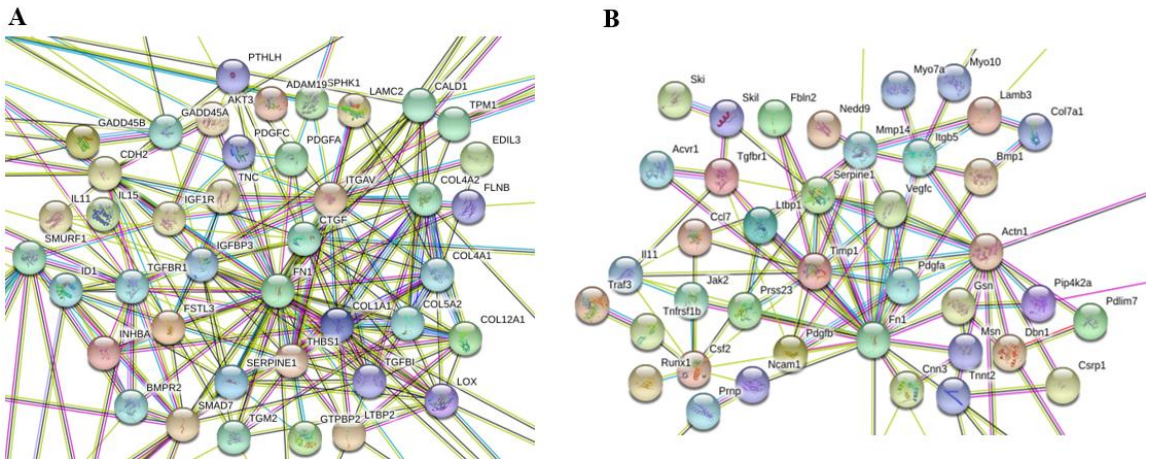


Figure 22: STRING interaction networks of upregulated genes intersected in all isoforms. (A)Human (B)Mouse.

Results of the differential expression analysis of the mouse (Figure 23) and human datasets (Figure 24) (Figure 25). Plots were based on logarithmic fold change values and adjusted p-values. Significantly differentiated genes can be seen in their gene IDs with red color code. Some datasets were excluded whereat they did now show significant results using the current plot thresholds. Gene names of the IDs highlighted in the plots by showing a significant correlation in mouse (Table 6) and human (Table 7) datasets can be seen in the tables.

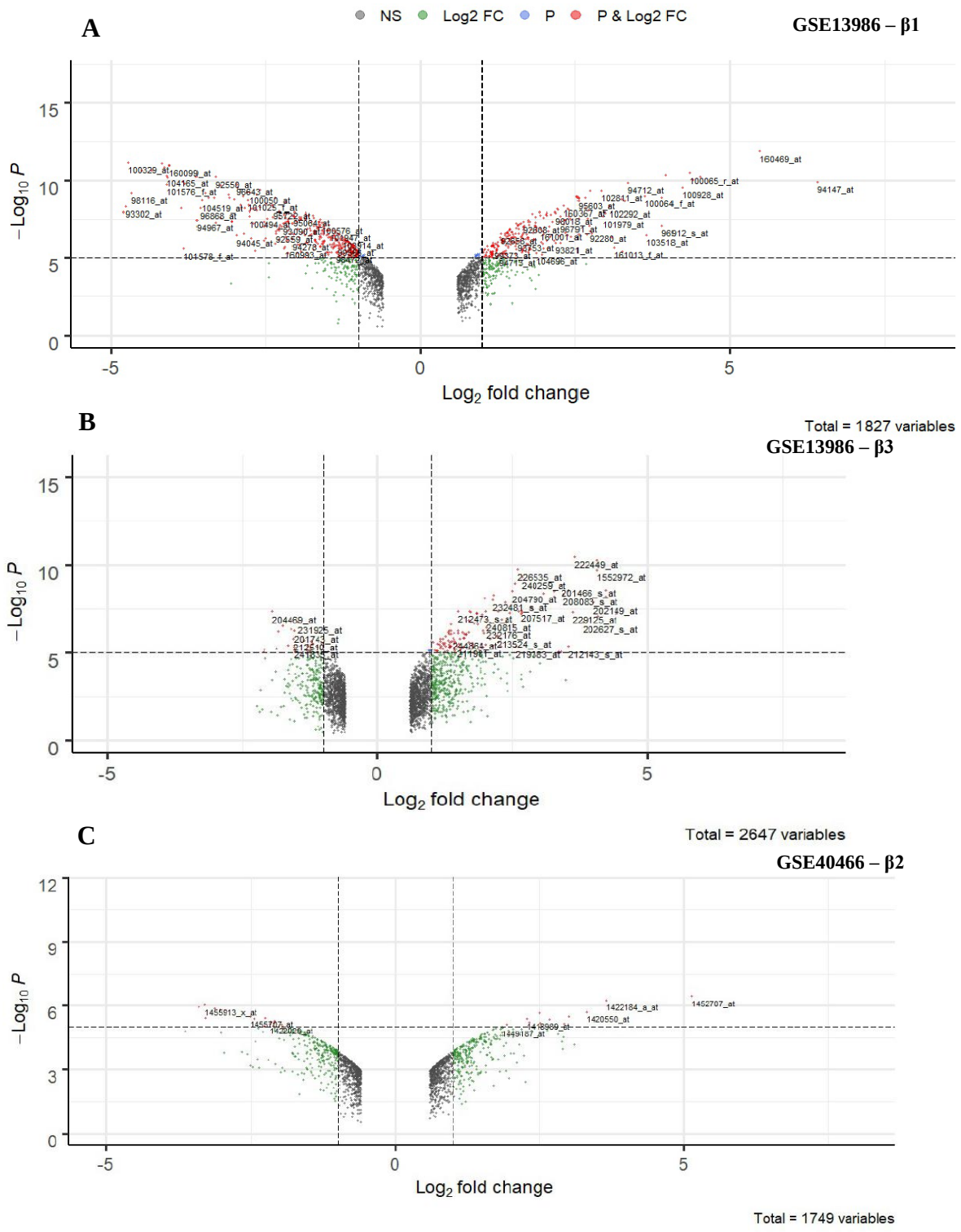


Figure 23: Volcano plots of differentially expressed gene sets in Mus musculus (NMuMG) datasets. X-axis Log2 fold change threshold levels between -5, +5, and Y-axis for log10 P values are differentially expressed and plotted. Genes showing high correlation are marked with their gene IDs. (A) with GEO Accession Number GSE13986, treated with TGFβ1. (B) with GEO Accession Number GSE13986, treated with TGFβ3. (C) with GEO Accession Number GSE40466, treated with TGFβ2.

Table 6: Gene names showing a higher correlation Mus musculus (Mouse) datasets in volcano plots in Figure 23.

GDS3710-B1	GDS4106-B1	GSE49644-B1	GSE114761-B1	GSE40266-B2	GSE35830-B3
MAF	CDH19	INHBA	INHBA	POSTN	PMEPA1
RUNX2	COL1A1	SCG2	LCE3D	KRT14	LOC100507431
INHBA	TUBA4A	SLC6A15	TNFAIP6	PLN	ITGB6
SCG2	SPOCK1	NOX4	KANK4	SULF1	FLRT2
CDH11	HEY1	DNER	DOCK2	OPCML	JUN
TSPAN2	ANGPTL4	VIP	ADAM19	NALCN	SMAD7
COL5A1	S100A2	BHLHE41	SLCO5A1	RBP1	SLITRK6
COL4A1	HRH1	ADAMTS6	LPAR5	LDB3	NEDD9
DSE	KIAA1549L	BMP2	ALOX5AP	CNN1	MICAL2
COL1A1	PROC	CXCL8	MMP10	ADAM12	LAMC2
BMP2	TMEM51-AS1	COL22A1	ANGPTL4	CDKN2B	KANK4
DACT1	MYO10	COL15A1	TNS1	NEDD9	SERPINE1
IGFL2	CLN8	MMP1	DOCK2	SLC16A2	SLITRK6
FAM43A	ENAH	PDGFC	COL4A1	FBXO32	G0S2
SLN	BCAT1	GPAM	ADAMTS15	COL4A1	LOC101929475
XYLT1	TRIM9	PID1	VGLL3	SEPT11	COL4A1
GPR87	MCL1	FAM198B		DHRS2	PRR5L
KANK4		ST3GAL1		GLS	IGFBP3
HOXD10		PPP1R14C		ITIH3	
BEAN1		ATP8A2		RHOJ	
TAGLN		DCLK1		CAVIN4	
LINC01943		MCC		PMEPA1	
SERPINE2				OPCML	
SLC26A2				POSTN	
FBXO32				JADE1	
				SYTL5	
				INHBE	

Table 7: Gene names showing a higher correlation in volcano plots of Homo sapiens (Human) datasets in Figure 24 and Figure 25.

GSE13986-B1	GSE13986-B3	GSE40466-B2
Pald1	Serpine1	Klhl30
Serpine1	Pald1	Ak1
Gja1	Wisp1	Lce1f
Fbln2	Fbln2	Ctse
Ext1	Gja1	Pdgfa
Gldc	Vegfc	
Spsb1	Zc3h14	
Ctla2b	F2r	
Pdlim7	Ctla2a /// Ctla2b	
Gadd45a	Sox11	
Procr	Spsb1	
Fam101b	Ncam1	
Csrp1	Myzap	
Myzap	Gldc	
Actn1	Ccrn4l	
Foxq1	Actn1	
Litaf	Frmd6	
Wwc2	Tnk2	
Myo7a	Foxq1	
	Glce	

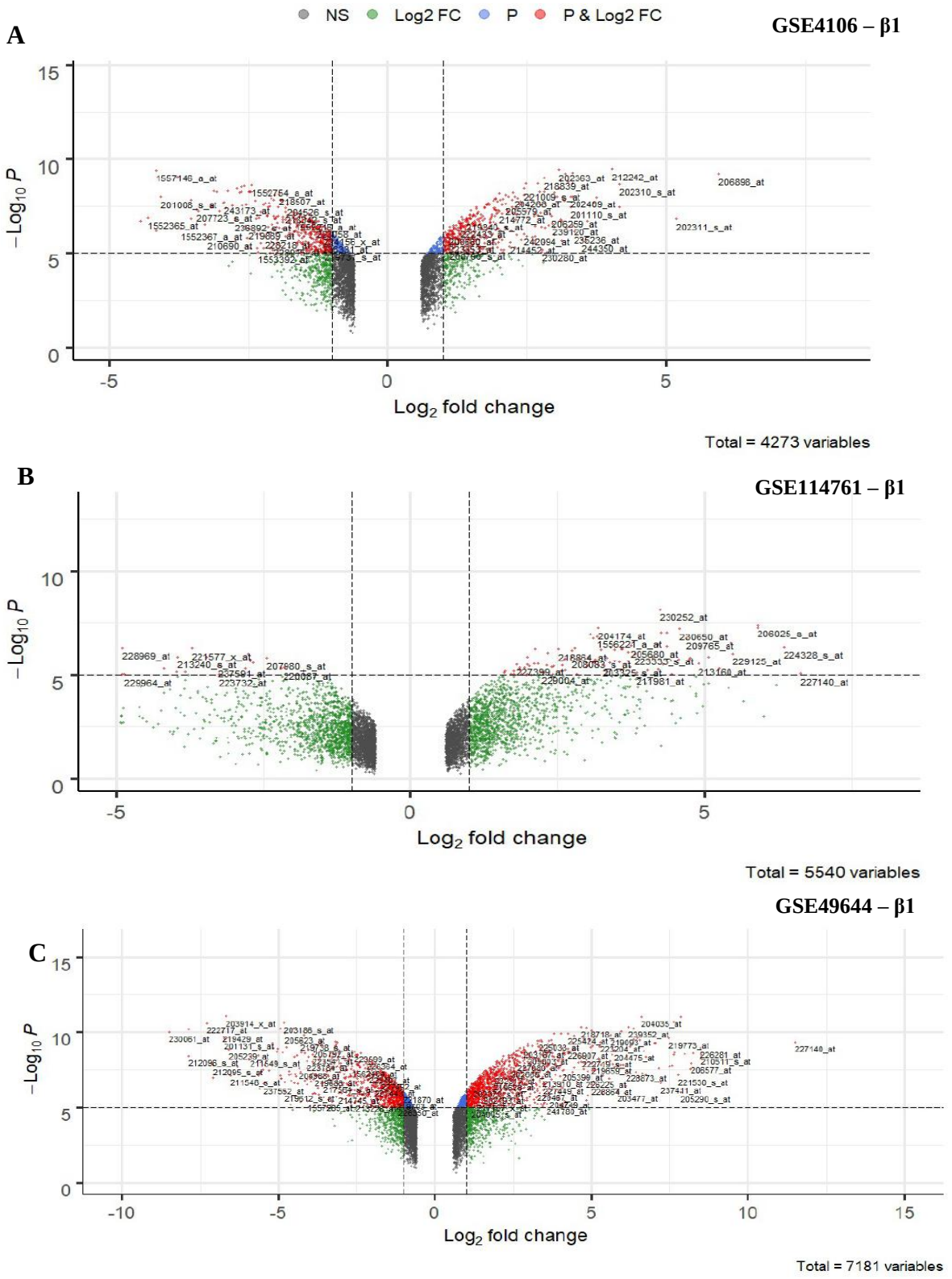


Figure 24: Volcano plots of differentially expressed gene sets in Homo sapiens (Human) datasets. X-axis Log2 fold change threshold levels between -5, +5 or -10, +10 and Y-axis for log10 P values are differentially expressed and plotted. Genes showing high correlation are marked with their gene IDs. (A) with GEO Accession Number GSE4106, treated with TGFβ1. (B) with GEO Accession Number GSE114761, treated with TGFβ1. (C) with GEO Accession Number GSE49644, treated with TGFβ1.

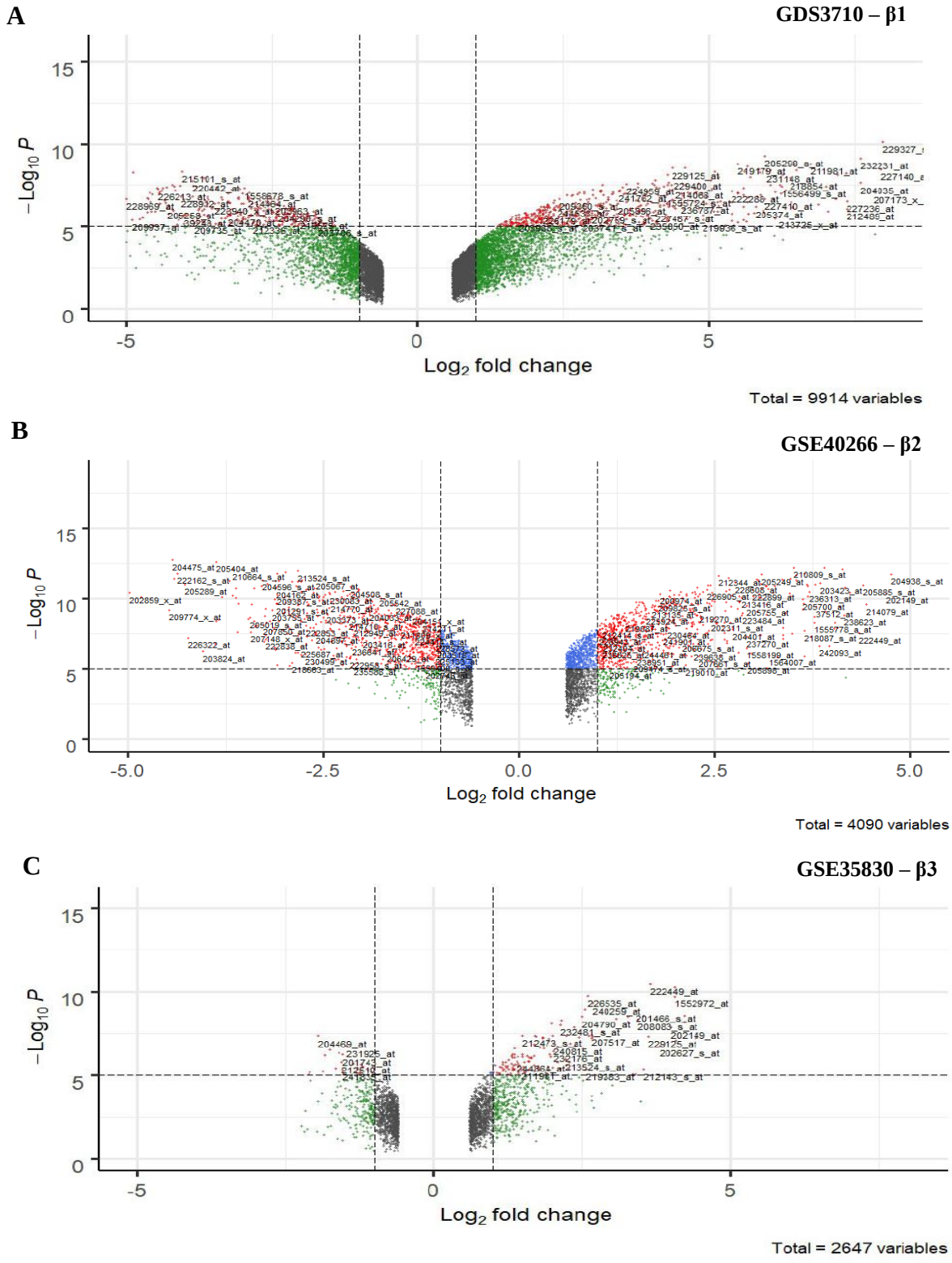


Figure 25: Volcano plots of differentially expressed gene sets in Homo sapiens (Human) datasets. X-axis Log2 fold change threshold levels between -5, +5, and Y-axis for log10 P values are differentially expressed and plotted. Genes showing high correlation are ma marked with their gene IDs. (A) with GEO Accession Number GDS3710, treated with TGFβ1. (B) with GEO Accession Number GSE40266, treated with TGFβ2. (C) with GEO Accession Number GSE35830, treated with TGFβ3.

5.3 GSEA Analysis

Table 8: (A) GSEA: Top 20 upregulated genes in control and TGFβ treated phenotype. Genes were ranked based on their enrichment score (ES). (B) Genes were color-coded as unique for only one dataset or presented in more than one dataset from yellow to dark green.

A

GDS3710	A549	GDS4105	Panc 1	GSE13886	NMuMG	GSE13986	NMuMG	GSE35830	Ect1	GSE40266	NOF151	GSE40266	NOF151
Control	TGFB1	Control	TGFB1	Control	TGFB1	Control	TGFB3	Control	TGFB3	Control	TGFB1	Control	TGFB2
ALDH3A1	DHRS2	FLJ32252	HTR1D	RNASE4	GJB3	TTR	GJB3	PTPRZ1	PGM2L1	CXCL6	SLCO2A1	CXCL6	NOX4
CLDN2	DOCK4	KCNB1	IGFBP3	TTR	PDGFA	RNASE4	GLDC	CD14	ZBED2	CXCL1	KRT14		KRT14
SLC16A7	LAMC2	PDK4	COL1A1	UGT2B10	GLDC	APOC2	GJA1	PPFIBP2	CTGF	C9ORF26	MGC16121	C9ORF26	CILP
CYP4F11	CXCR7	SLC4A11	TMEM158	HPX	RAMP1	HPX	IL11	PIK3R3	JUN	IL24	PSAT1	IL24	MGC16121
GPX2	CORO2B	TXNIP	CCND2	APOC2	LAMB3	UGT2B10	TNNT2	GPD1L	LOC284262	ANXA10	POSTN	ANXA10	RASGRP1
SOX2	COL4A1	FAM46C	TMEPAI	VNN1	ANKRD1	CYP3A7	NCAM1	SDPR	COL4A1	ALDH1A3	HSD17B6	ALDH1A3	PSAT1
CDH1	KCNMA1	KLRC3	OXTR	CYP3A7	CTGF	TFF2	AK1	FGFR3	GOS2	MMP1	RASGRP1	MMP1	ASS1
NR0B1	SERPINE1	ARRDC4	MYO10	SERPINA1	GADD45G	SERPINA1	RHOB	EPN3	SMAD7	BMP2	NEDD9	BMP2	LRRC32
ERBB3	NOX4	ARID5B	C20ORF160	RBP4	AK1	VNN1	GADD45G	GLUL	IVNS1ABP	ADAMTS1	FBXO32	ADAMTS1	HSD17B6
DDC	MARCH4	AGR2	HEY1	TFF2	EXT1	CYP2S1	ANKRD1	GLOXD1	PTHLH	IL8	CILP	IL8	RBP1
PPFIBP2	IGFL2	RDHE2	PIK3AP1	ITIH2	F2R	RBP4	SERPINE1	BLNK	PTPRK	SERPIND1	LMCD1	SERPIND1	POSTN
ELF3	COL5A1	RAB26	SPOCK1	UGT1A3 ///	VEGFC	CXCL6	EXT1	LOC153364	LOC399978	HSD11B1	ASS1	HSD11B1	FBXO32
INHBB	TSPAN2	TPPP	ITGB3	CYP2S1	GADD45A	UGT1A3 ///	LAMB3	GLS2	SLITRK6	TMTC1 ///	OPCML	TMTC1 ///	OPCML
TM4SF4	IGFBP7	CA8P7	PLEK2	CXCL6	PTGS2	ID1	VEGFC	THBS2	TGM2	CXCL2	LRRC32	CXCL2	LMCD1
METTL7A	SNAI2	TFF3	SNF1LK	TSPAN8	GJA1	C4BPA	GADD45A	C8ORF55	LOC646819 ///	LOC221091	RBP1	LOC221091	NEDD9
UPK1B	FRMD6	LOC643961	FKBL18	MUC1	CYR61	CALML4	F2R	GIMAP2	AMIGO2	RRM2	MFAP4	RRM2	MFAP4
SLC27A2	LCE3D	SUSD2	IGF2	ALDH1A1	WISP1	MUC1	CYR61	GLDC	MMP10	TFPI	TSPAN2	TFPI	ATP10A
SLC7A2	THBS1	IGSF4D	LTBP2	SPRR2A	FBLN2	ITIH2	PTGS2	IL17D	SERPINE1	SLC43A3	ATP10A	SLC43A3	TSPAN2
KITLG	IGFBP5	LGR4	CDH19	CALML4	IL11	APOA1	WISP1	SLITRK5	ITGB6	TSPAN8	COMP	TSPAN8	COMP
SELENBP1	MAF	TGFBR3	TUBA1	ID1	SERPINE1	HABP2	FBLN2	POLR1B	TMEPAI	DTL	LOC646324	DTL	LOC646324

GSE49644	H358	GSE49644	HCC827	GSE49644	A549	GSE114761	H1944	GSE114761	A549	GSE114761	H358
Control	TGFB1	Control	TGFB1	Control	TGFB1	Control	TGFB1	Control	TGFB1	Control	TGFB1
RBM35A	GAL	FGB	NAP1L3	MTUS1	COL22A1	GC	SPOCK1	MUC5AC	CRLF1	HPGD	FBN1
S100P	LOXL1	C9ORF84	EBI2	AREG /// I	ZBED2	LCN2	IGFL1	INHBB	HOXD11	ZNF750	LOC646919
AGR2	SRPX	C12ORF46	INHBA	HPGD	ADAMTS6	HAL	PGM2L1	SDPR	BEAN	INHBB	NOG
RAB25	COL14A1	SFTPG	EDNRA	GPX2	MAF	ORM1 /// C	PTPRR	AGR2	COL4A2	RGC32	VGLL3
TJP2	DSPG3	RAB25	C4ORF18	TM4SF18	IL1A	ADH6	IGFBP7	GPX2	MMP10	EVI2B	MMP9
MAL2	IGFBP5	LCN2	CXCR7	CYP1B1	BHLHB3	HBA1	EDIL3	TGFBR3	PLEK2	MCM10	CRLF1
EVA1	SPARC	MUC1	LOC283824	ABCG2	NOX4	HBA1 /// HI	C8ORF4	GDF15	PLCXD2	LOC388743	TSPAN2
CD9	SCG5	UCA1	CACHD1	S100A4	VIP	HBA2	CHRNA9	VGCNL1	ICHTHYIN	HOP	CTGF
ALDH1A3	CXCL1	PDZK1IP1	SRPX	ARHGAP2	SLC6A15	DDC	CTGF	CITED2	LOC646591	FGF19	MAMDC2
C12ORF46	CCL20	ELF3	HTRA1	ALDH3A1	PPAPDC1A	CCNA1	ANKRD38	ADORA2B	FRMD6	FA2H	LRRN1
C10ORF11	MGC4294	EVA1	WNT5A	NTS	IL8	TXNIP	FN1	RDHE2	CHRNA9	ST6GALNAC	DOC1
TMPRSS4	CHRNA9	EHF	COL5A1	FA2H	LOC283824	CAMP	TSPAN2	SOX2	COL5A1	GINS2	LOXL1
EHF	GNG11	IGFBP3	SERPINE1	ABCC2	MMP2	S100A8	TMEPAI	ABCC2	CCDC80	SLC27A2	LTBP2
SLC16A14	IL11	RBM35A	IL11	CDH1	SCG2	ALB	MAF	CYP4F11	ALOX5AP	KRT4	LBH /// LO
ARBG /// LC	4ORF18	MAL2	C4ORF26	SDPR	IL6	HLA-DMA	NEDD9	RERG	COL4A1	GDF15	CXCR7
TMEM30B	HS3ST3A1	TMEM30B	AYTL1	EREG	BMP2	ADH1A ///	LBH /// LC	FA2H	IL11	LOC152573	IL11
FGFBP1	ALPK2	NR5A2	MMP2	ADM	C9ORF26	ORM1	LAMC2	NR0B1	TAGLN	STS	C4ORF26
TACSTD2	SERPINE1	TMEM27	PPAPDC1A	NTN4	LCE3D	NR0B1	PPBP	RNF43	GPR92	LAMP3	MAF
KRT19	CGB /// CGB	RNF183	COL3A1	IGSF11	INHBA	OSTBETA	RBP1	RARRES3	ADAM19	KIAA1199	CGB /// CG
INHBB	F2RL2	ARL14	ANKRD38	DCDC2	DNER	HPSE	LUM	STK32B	TNFAIP6	ATAD2	SERPINE1

B

Gene set enrichment analysis (GSEA) was performed to reveal the missing points from the single gene expression analysis like affected pathways and cellular processes. Overrepresented genes were ranked based on their expression differences and correlations with two different phenotypes (control vs. TGFβ treated). Twenty genes that are on the top of the ranked list based on their enrichment score (ES) in control and TGFβ treated phenotype were listed. Genes that are present in more than one dataset and present in only one dataset were color-coded (Table 8). Genes like SERPINE1, IL11, TSPAN2, MAF, and CTGF appeared in more than three datasets in TGFβ treated phenotype as upregulated and in all isotypes regardless of the organism.

Out of 13 datasets, SERPINE1 is in the seven datasets of the top 20 upregulated genes. It has the highest number of datasets to be upregulated, including TGFβ1, TGFβ3 phenotypes, lung cancer cells, mammary gland cells, and ovarian cells. It is also upregulated in other datasets, even if it is not in the top 20 list. Regardless of the organism or the isoform, SERPINE1 can be used as a TGFβ induced EMT marker. IL11 is correlated with non-small cell lung cancer and bad prognosis (Zhao et al., 2018). It is upregulated along with the TGFβ induced EMT through the PI3K/AKT/STAT pathway and increases the invasiveness of lung and thyroid carcinoma cells (Zhong et al., 2016). It is in the six datasets top 20 upregulated genes. Also, it was one of the 13 genes in the intersected common genes among isotypes and both organisms (Figure 19).

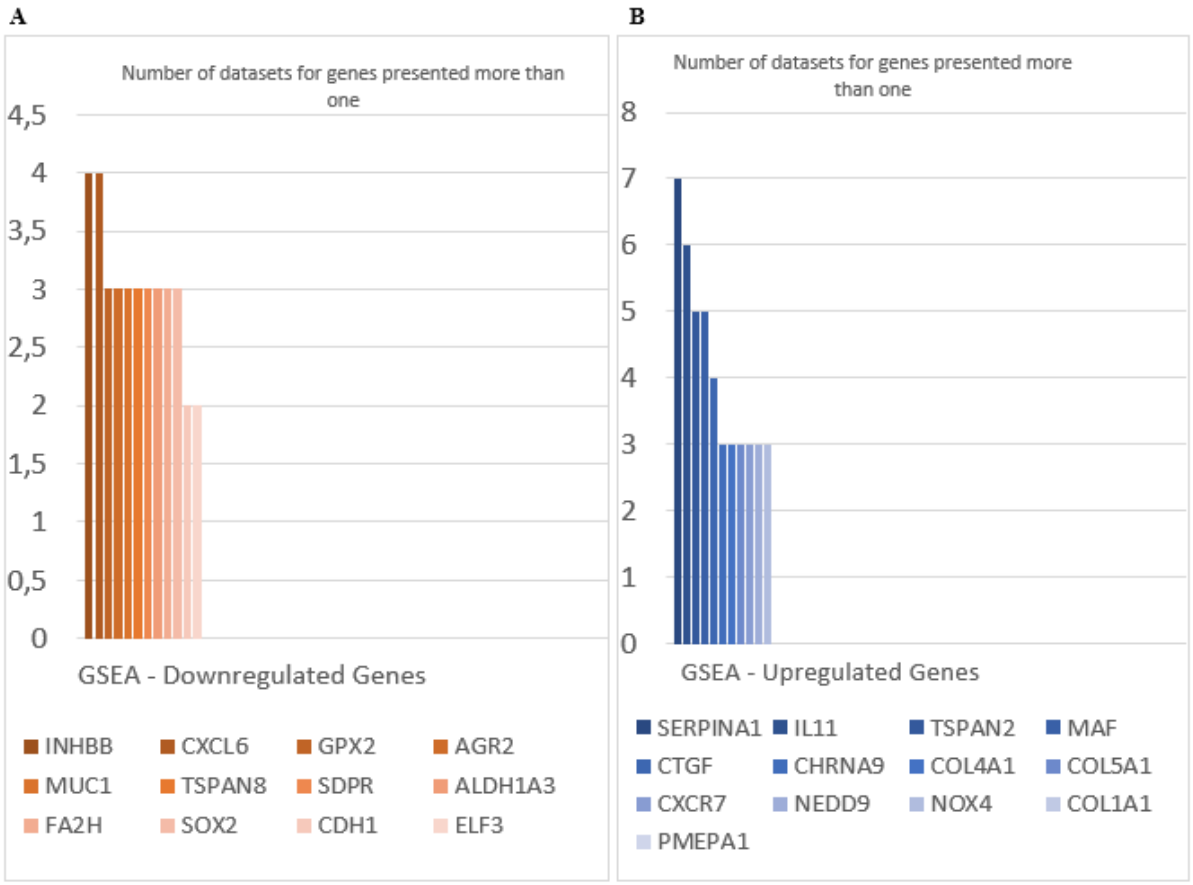


Figure 26: Gene set enrichment analysis; Regulated genes common in more than one dataset in the TGFβ phenotype. The X-axis represents the number of datasets of each gene where they are regulated. (A) Downregulated common genes. (B) Upregulated common genes.

Collagens responsible for fibrosis and myofibroblast formation are upregulated in various EMT cases, and cancers like COL4A1, COL5A1 and COL1A1 (Sayarlioglu et al., 2016) were in the top 20 upregulated list with more than two datasets. CXCR7 was upregulated in three datasets with adenocarcinoma cell line in the top 20 list. This chemokine receptor has a significantly high upregulation levels in cancers like bladder, ovarian, head and neck with TGFβ induced

EMT and associated with late-stage lung cancers when upregulated with TGFβ1 (Y. C. Wu et al., 2016). PMEPA1 or TMEPA1 which is a bad prognosis sign in TGFβ induced cancers as mentioned before, was upregulated in TGFβ3 treated endocervical cell line, TGFβ1 induced lung cancer, and pancreas cancer cell line.

NEDD9, which regulates TGFβ induced EMT in cancer and used as a marker gene for various cancer types like prostate and breast cancer (Kong et al., 2011) was upregulated in the top 20 genes in 3 datasets, including ovarian and lung cancer cell lines with TGFβ1 and TGFβ2 phenotype. In the downregulated top 20 genes; epithelial marker CDH1, pluripotency-associated transcription factor SOX2, which helps to maintain the epithelial phenotype (Thierauf et al., 2017), and ELF3 which is an essential keeper for the epithelial state (Sengez et al., 2019) were present as expected (Figure 26).

Upregulated genes of GSEA treated with TGFβ1 and TGFβ3 had differences among them. As mentioned before, SERPINE1, PMEPA1, CTGF, INHBA and COL4A1 were common in both phenotype where TSPAN2, MAF, NEDD9, CXCR7 were specific for TGFβ1 while GADD45, LAMB3, SKIL, and PTGS2 were upregulated in TGFβ3 (Figure 27). GADD45 is a

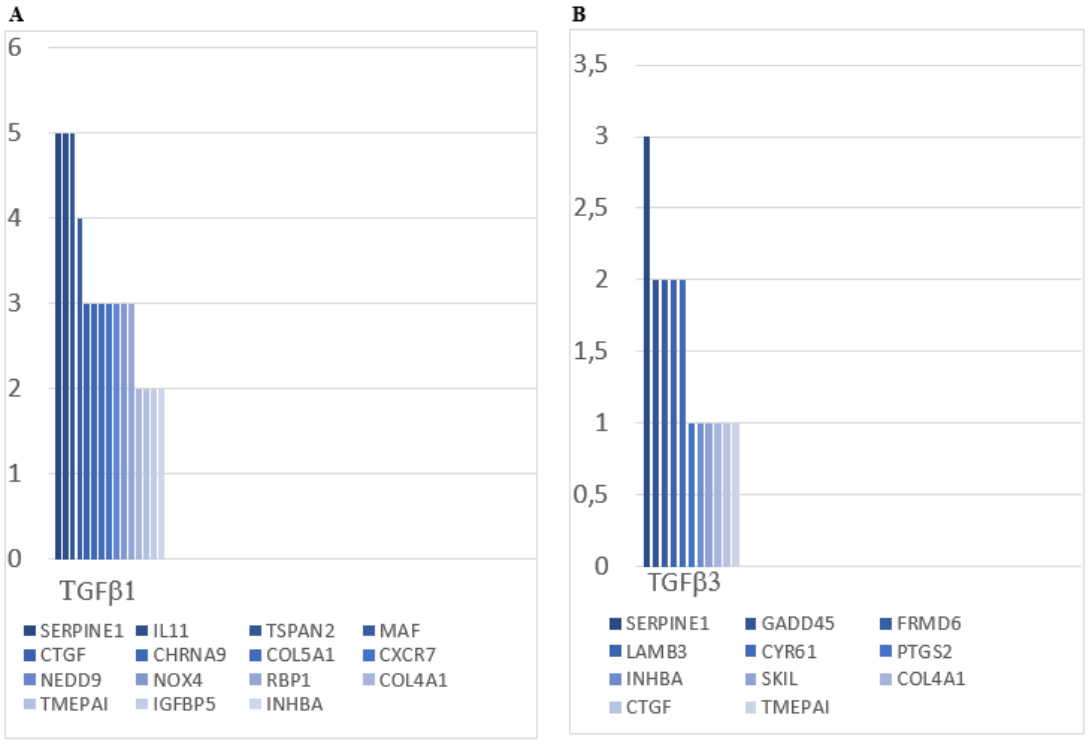


Figure 27: Gene set enrichment analysis; Upregulated genes common in more than one dataset in the TGFβ phenotype. (A) TGFβ1 common genes. (B) TGFβ3 common genes.

downstream agent of the MAPK pathway associated with growth arrest, and LAMB3 regulates the Akt pathway (Levy & Hill, 2005).

Upregulated genes correlated with certain gene sets and functions were listed and separated by the TGFβ phenotype. Out of 13 datasets (11 TGFβ1 and 2 TGFβ3), all the genes were associated with myogenesis, EMT, and TGFβ signaling, which is accurate as all the datasets were treated with TGFβ to induce EMT. Although gene sets are almost the same, with some differences, some of them were not highly correlated with the other phenotype. For example, apoptosis was seen in all the TGβ3 phenotype genes but only 5 out of 11 TGFβ1 phenotype genes were correlated with apoptosis. Also, the ROS pathway and complement system were correlated with the TGFβ1 phenotype genes while they were not present in TGFβ3.

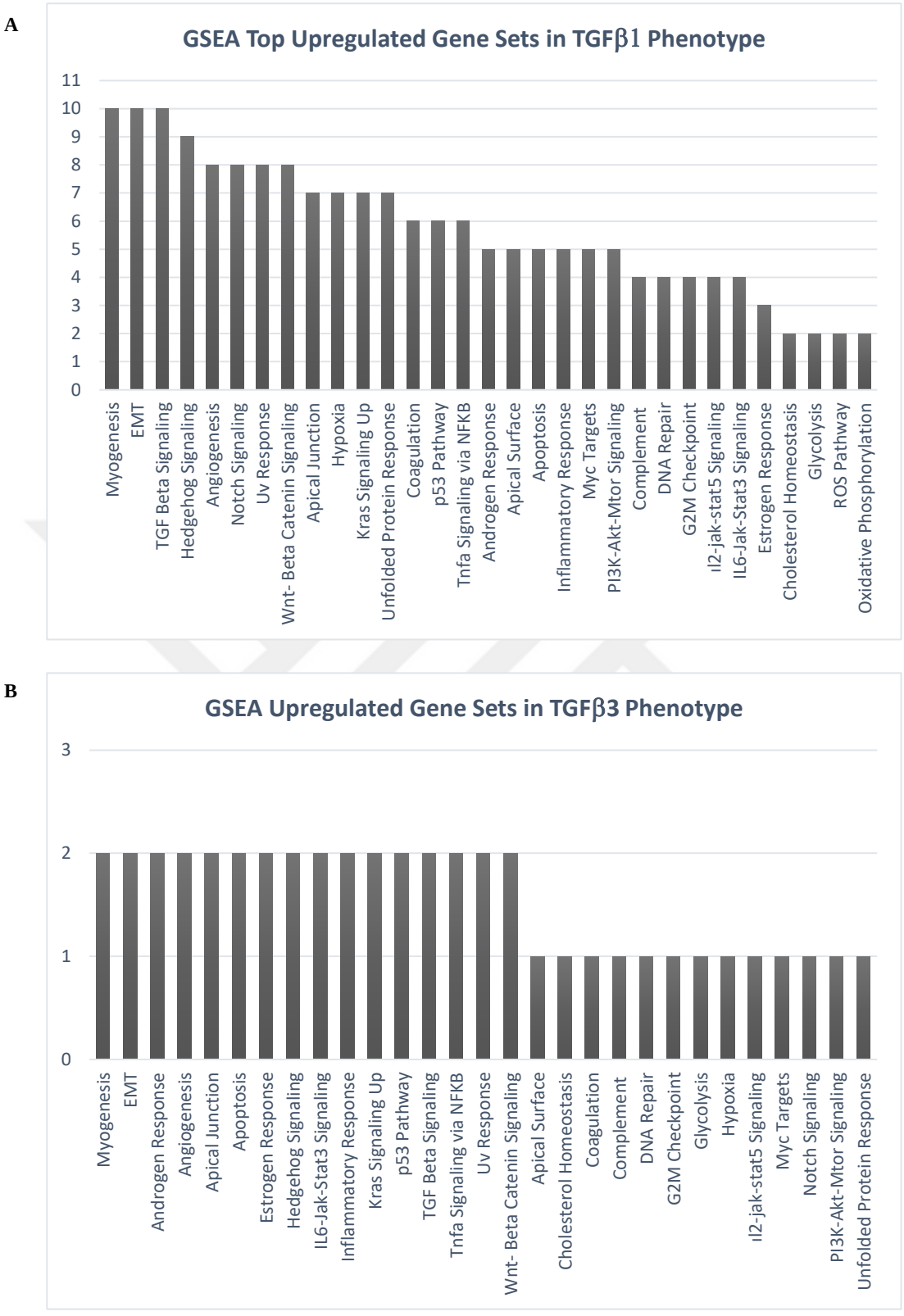


Figure 28: Gene set enrichment analysis (GSEA): Gene sets correlated with upregulated genes. (A) TGFβ1 phenotype (B) TGFβ3 phenotype.

Reactive oxygen species (ROS) induce EMT via inducing SNAIL and are activated in various cancers (Lee et al., 2019). ROS pathway is correlated in two datasets of TGFβ1. In one study, crosstalk between TGFβ1 and complement system was investigated in fibrosis (Gu et al., 2014). Signaling pathways that are downstream of TGFβ were expected to be correlated. Such as; Hedgehog, Notch, Wnt- Beta Catenin, PI3K/Akt, p53, and Jak/Stat signaling pathways. Also, some of those pathways play an important role in development by EMT. Androgen and estrogen responses are important regulators of prostate, ovarian, and breast cancers. As can be seen, TGFβ induced genes and pathways are correlated with those responses.

As mentioned before, glycosylation is affected by EMT and maintains the mesenchymal phenotype. Both phenotypes were correlated with glycolysis. As angiogenesis is one of the key functions of EMT of the tumor cells, it is correlated with both phenotypes.

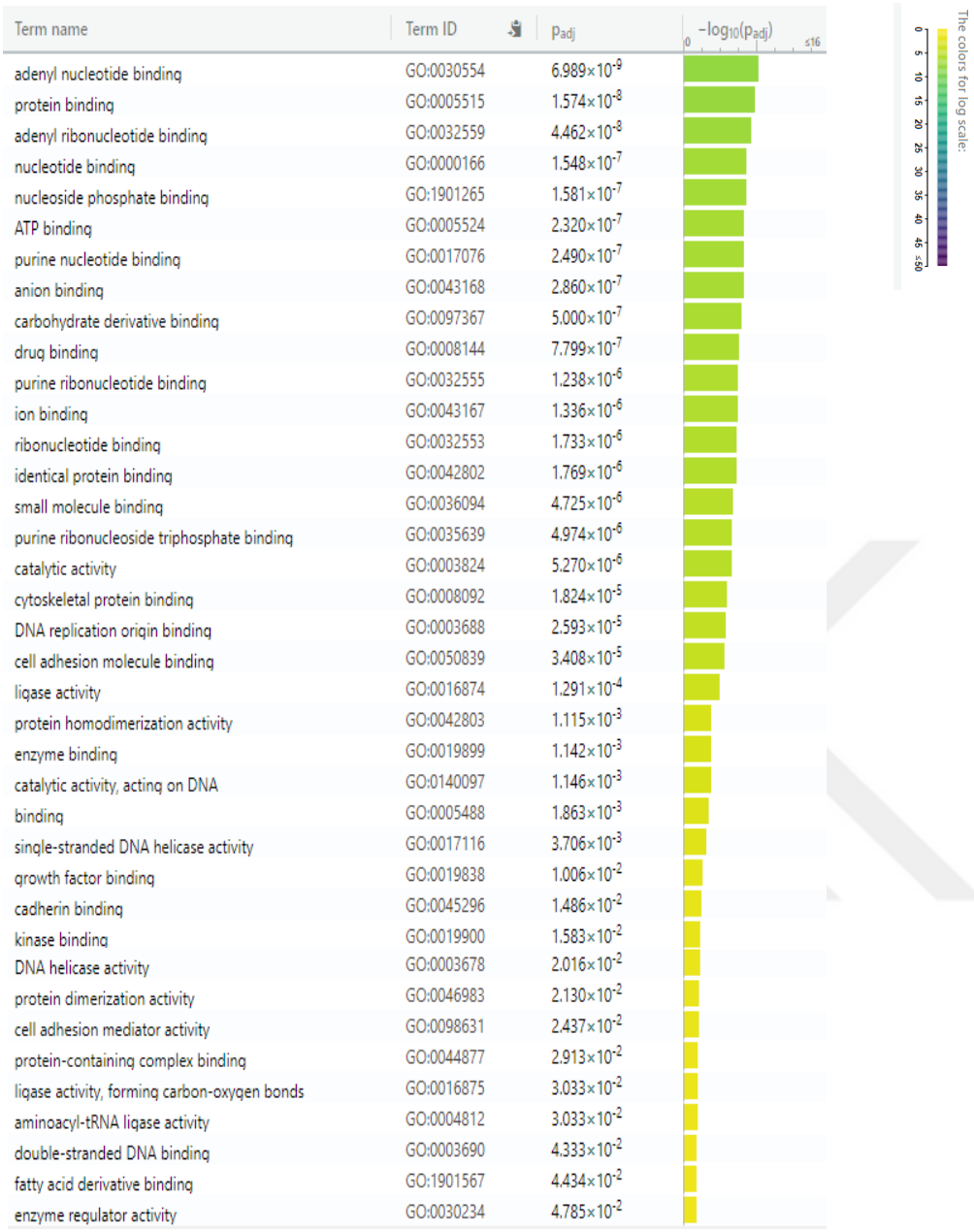
In the cell cycle, EMT occurs in the G1/S phase while apoptosis occurs in the G2/M phase (Song, 2007). G2/M checkpoint here is correlated with the upregulated genes where DNA damage repair occurs. Not all the epithelial cells survive the EMT process when induced with TGFβ, as it may lead to DNA damage during the cell cycle. The survived cells manage to transform into the mesenchymal phenotype in the G1/S phase (Figure 28).

5.1 GO Enrichment Analysis

GO (Gene Ontology) analysis was performed via G: profiler. The platform's data source is based on Ensembl. Genes that are upregulated only in certain isoform were used as query and analyzed. Each gene set belonging to a certain isotype (TGFβ1, TGFβ2, and TGFβ3) was revealed based on their molecular functions, biological processes that they take part in, and cellular components they belong to. Lists were ranked based on their logarithmic P-values. Adjusted P-values were color-coded from yellow to dark blue as significance increases. Genes that are correlated with certain terms can be seen under term IDs as GO annotations. For the TGFβ1 phenotype, the most significant molecular functions are adenyl nucleotide, adenyl ribonucleotide, nucleotide, nucleoside phosphate, ATP, and protein binding (Figure 29). TGFβ1 is mostly found in the cytoplasm, then membrane-bound organelle and membrane-enclosed lumen. Beside the organelles and organelle lumens, TGFβ1 can also be found in extracellular exosomes and vesicles, nucleus, cytoskeleton, as revealed by the cellular component analysis (Figure 30).

Biological processes highly correlated with TGFβ1 genes are; mitotic cell cycle, cell division, cell cycle phase transition, regulation of cell cycle, and cellular component organization (Figure 31). TGFβ2 had molecular functions with high logarithmic significance as; enzyme binding, protein binding, RNA binding and heterocyclic compound binding (Figure 32). For cellular components, TGFβ2 is highly present in intracellular parts; intracellular membrane bound organelles, cytoplasm, and intracellular organelle lumen (Figure 33). In biological processes, TGFβ2 functions in metabolic cellular processes, organic substance, macromolecule and nitrogen compound processes (Figure 34). Finally, TGFβ3 has similar molecular functions with TGFβ1, like protein binding, ATP binding, and nucleotide binding. TGFβ3 took part in distinct biological processes like cellular response to chemical stimulus, regulation of localization and positive regulation of cell death. TGFβ3 resides in the cell's intracellular parts, membrane-bound organelle, cytoplasm, and nucleus (Figure 35).

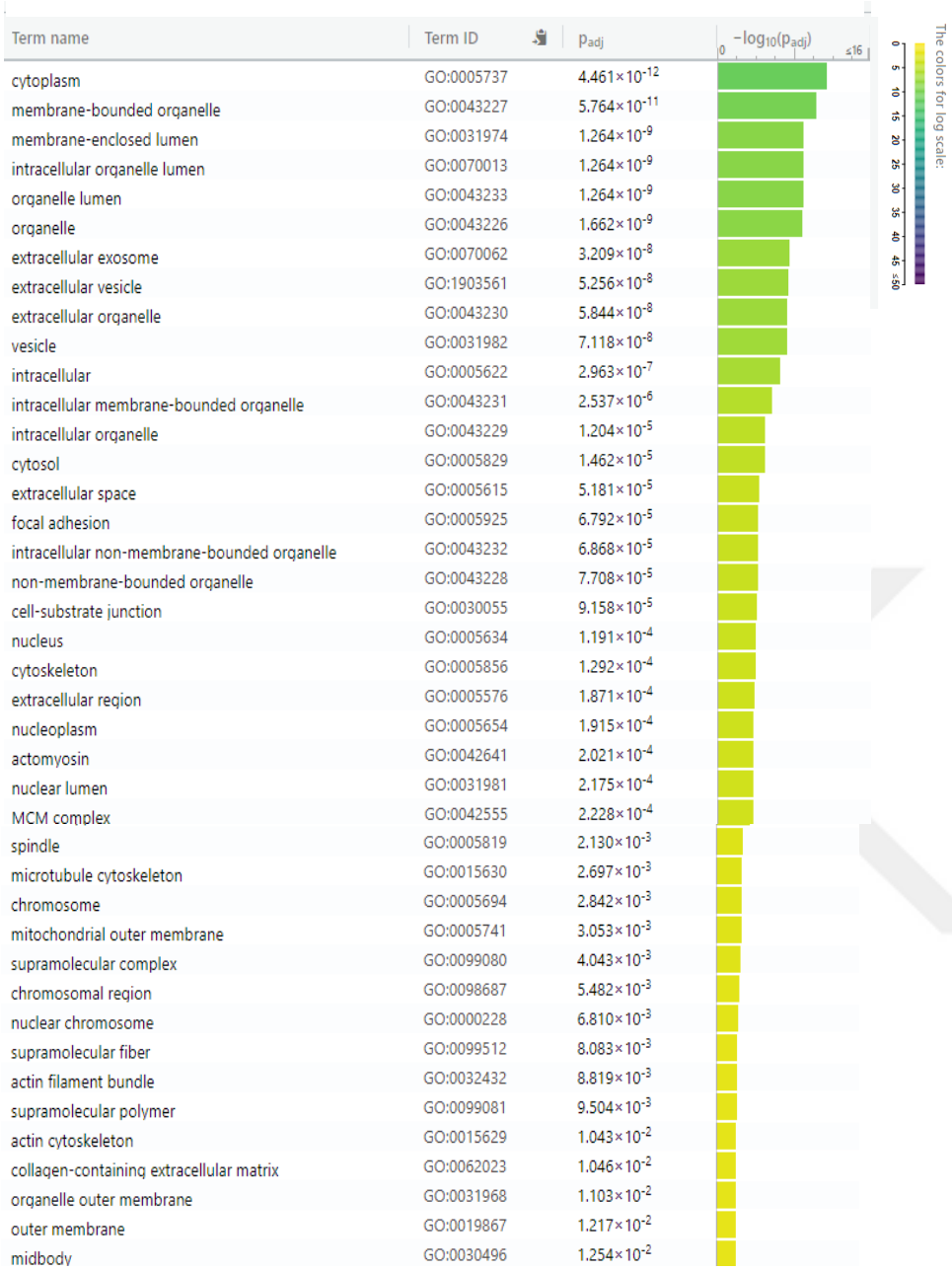
GO: Molecular Function - TGβ1



Term name	Term ID	p_{adj}
adenyl nucleotide binding	GO:0030554	6.989×10^{-9}
protein binding	GO:0005515	1.574×10^{-8}
adenyl ribonucleotide binding	GO:0032559	4.462×10^{-8}
nucleotide binding	GO:0000166	1.548×10^{-7}
nucleoside phosphate binding	GO:1901265	1.581×10^{-7}
ATP binding	GO:0005524	2.320×10^{-7}
purine nucleotide binding	GO:0017076	2.490×10^{-7}
anion binding	GO:0043168	2.860×10^{-7}
carbohydrate derivative binding	GO:0097367	5.000×10^{-7}
drug binding	GO:0008144	7.799×10^{-7}
purine ribonucleotide binding	GO:0032555	1.238×10^{-6}
ion binding	GO:0043167	1.336×10^{-6}
ribonucleotide binding	GO:0032553	1.733×10^{-6}
identical protein binding	GO:0042802	1.769×10^{-6}
small molecule binding	GO:0036094	4.725×10^{-6}
purine ribonucleoside triphosphate binding	GO:0035639	4.974×10^{-6}
catalytic activity	GO:0003824	5.270×10^{-6}
cytoskeletal protein binding	GO:0008092	1.824×10^{-5}
DNA replication origin binding	GO:0003688	2.593×10^{-5}
cell adhesion molecule binding	GO:0050839	3.408×10^{-5}
ligase activity	GO:0016874	1.291×10^{-4}
protein homodimerization activity	GO:0042803	1.115×10^{-3}
enzyme binding	GO:0019899	1.142×10^{-3}
catalytic activity, acting on DNA	GO:0140097	1.146×10^{-3}
binding	GO:0005488	1.863×10^{-3}
single-stranded DNA helicase activity	GO:0017116	3.706×10^{-3}
growth factor binding	GO:0019838	1.006×10^{-2}
cadherin binding	GO:0045296	1.486×10^{-2}
kinase binding	GO:0019900	1.583×10^{-2}
DNA helicase activity	GO:0003678	2.016×10^{-2}
protein dimerization activity	GO:0046983	2.130×10^{-2}
cell adhesion mediator activity	GO:0098631	2.437×10^{-2}
protein-containing complex binding	GO:0044877	2.913×10^{-2}
ligase activity, forming carbon-oxygen bonds	GO:0016875	3.033×10^{-2}
aminoacyl-tRNA ligase activity	GO:0004812	3.033×10^{-2}
double-stranded DNA binding	GO:0003690	4.333×10^{-2}
fatty acid derivative binding	GO:1901567	4.434×10^{-2}
enzyme regulator activity	GO:0030234	4.785×10^{-2}

Figure 29: GO Analysis: Molecular functions of TGFβ1 upregulated genes. Terms were ranked based on their logarithmic P values. Genes involved in the terms can be checked from term IDs.

GO: Cellular Component –TGβ1



Term name	Term ID	p_{adj}
cytoplasm	GO:0005737	4.461×10^{-12}
membrane-bounded organelle	GO:0043227	5.764×10^{-11}
membrane-enclosed lumen	GO:0031974	1.264×10^{-9}
intracellular organelle lumen	GO:0070013	1.264×10^{-9}
organelle lumen	GO:0043233	1.264×10^{-9}
organelle	GO:0043226	1.662×10^{-9}
extracellular exosome	GO:0070062	3.209×10^{-8}
extracellular vesicle	GO:1903561	5.256×10^{-8}
extracellular organelle	GO:0043230	5.844×10^{-8}
vesicle	GO:0031982	7.118×10^{-8}
intracellular	GO:0005622	2.963×10^{-7}
intracellular membrane-bounded organelle	GO:0043231	2.537×10^{-6}
intracellular organelle	GO:0043229	1.204×10^{-5}
cytosol	GO:0005829	1.462×10^{-5}
extracellular space	GO:0005615	5.181×10^{-5}
focal adhesion	GO:0005925	6.792×10^{-5}
intracellular non-membrane-bounded organelle	GO:0043232	6.868×10^{-5}
non-membrane-bounded organelle	GO:0043228	7.708×10^{-5}
cell-substrate junction	GO:0030055	9.158×10^{-5}
nucleus	GO:0005634	1.191×10^{-4}
cytoskeleton	GO:0005856	1.292×10^{-4}
extracellular region	GO:0005576	1.871×10^{-4}
nucleoplasm	GO:0005654	1.915×10^{-4}
actomyosin	GO:0042641	2.021×10^{-4}
nuclear lumen	GO:0031981	2.175×10^{-4}
MCM complex	GO:0042555	2.228×10^{-4}
spindle	GO:0005819	2.130×10^{-3}
microtubule cytoskeleton	GO:0015630	2.697×10^{-3}
chromosome	GO:0005694	2.842×10^{-3}
mitochondrial outer membrane	GO:0005741	3.053×10^{-3}
supramolecular complex	GO:0099080	4.043×10^{-3}
chromosomal region	GO:0098687	5.482×10^{-3}
nuclear chromosome	GO:0000228	6.810×10^{-3}
supramolecular fiber	GO:0099512	8.083×10^{-3}
actin filament bundle	GO:0032432	8.819×10^{-3}
supramolecular polymer	GO:0099081	9.504×10^{-3}
actin cytoskeleton	GO:0015629	1.043×10^{-2}
collagen-containing extracellular matrix	GO:0062023	1.046×10^{-2}
organelle outer membrane	GO:0031968	1.103×10^{-2}
outer membrane	GO:0019867	1.217×10^{-2}
midbody	GO:0030496	1.254×10^{-2}

Figure 30: GO Analysis: Cellular components of TGFβ1 upregulated genes. Terms were ranked based on their logarithmic P values. Genes involved in the terms can be checked from term IDs.

GO: Biological Process – TGβ1

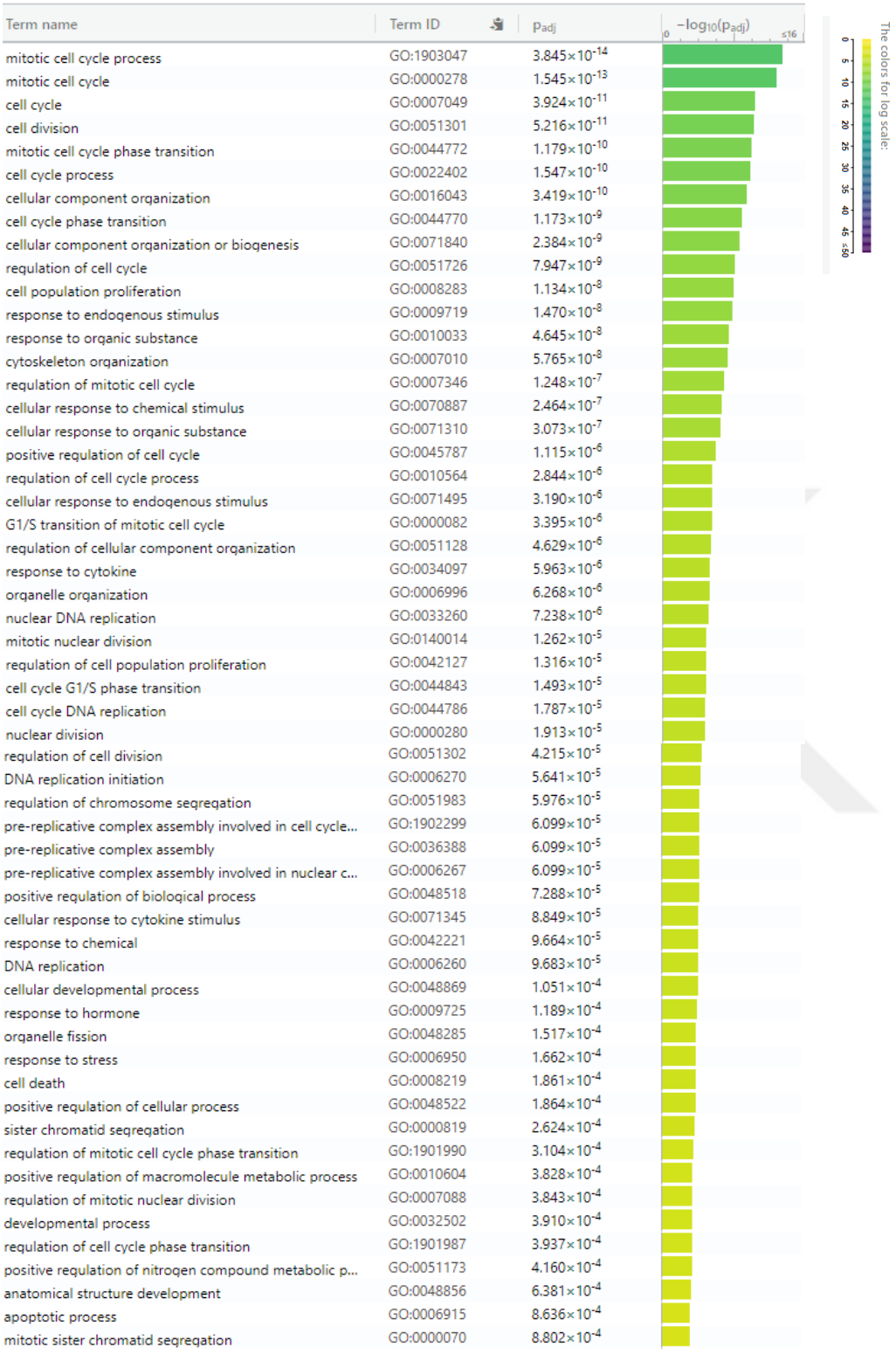


Figure 31: GO Analysis: Biological processes of TGFβ1 upregulated genes. Terms were ranked based on their logarithmic P values. Genes involved in the terms can be checked from term IDs.

GO: Molecular Function – TGβ2

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$
enzyme binding	GO:0019899	6.286×10^{-6}	
protein binding	GO:0005515	1.446×10^{-5}	
binding	GO:0005488	2.249×10^{-5}	
RNA binding	GO:0003723	3.081×10^{-5}	
heterocyclic compound binding	GO:1901363	6.385×10^{-4}	
organic cyclic compound binding	GO:0097159	1.038×10^{-3}	
ubiquitin protein ligase binding	GO:0031625	1.573×10^{-3}	
protein-containing complex binding	GO:0044877	2.916×10^{-3}	
ubiquitin-like protein ligase binding	GO:0044389	3.304×10^{-3}	
identical protein binding	GO:0042802	3.402×10^{-3}	
nucleic acid binding	GO:0003676	3.546×10^{-3}	
anion binding	GO:0043168	4.885×10^{-3}	
protein domain specific binding	GO:0019904	7.956×10^{-3}	

Figure 32: GO Analysis: Molecular functions of TGFβ2 upregulated genes. Terms were ranked based on their logarithmic P values. Genes involved in the terms can be checked from term IDs.

GO: Cellular Component – TGβ2

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$
intracellular	GO:0005622	4.936×10^{-30}	
intracellular membrane-bounded organelle	GO:0043231	2.950×10^{-28}	
intracellular organelle	GO:0043229	3.055×10^{-27}	
organelle	GO:0043226	1.407×10^{-25}	
membrane-bounded organelle	GO:0043227	1.786×10^{-24}	
cytoplasm	GO:0005737	1.720×10^{-16}	
intracellular organelle lumen	GO:0070013	1.831×10^{-16}	
membrane-enclosed lumen	GO:0031974	1.870×10^{-16}	
organelle lumen	GO:0043233	1.870×10^{-16}	
nucleus	GO:0005634	1.085×10^{-14}	
nuclear lumen	GO:0031981	2.585×10^{-13}	
nucleoplasm	GO:0005654	1.806×10^{-12}	
cytosol	GO:0005829	2.322×10^{-10}	
protein-containing complex	GO:0032991	1.967×10^{-9}	
intracellular non-membrane-bounded organelle	GO:0043232	1.408×10^{-6}	
non-membrane-bounded organelle	GO:0043228	1.732×10^{-6}	
mitochondrion	GO:0005739	2.446×10^{-6}	
organelle envelope	GO:0031967	3.173×10^{-6}	
envelope	GO:0031975	3.259×10^{-6}	
ribonucleoprotein complex	GO:1990904	4.003×10^{-5}	
catalytic complex	GO:1902494	2.626×10^{-3}	
nuclear envelope	GO:0005635	3.374×10^{-3}	
organelle membrane	GO:0031090	4.167×10^{-3}	
lytic vacuole	GO:0000323	8.221×10^{-3}	

Figure 33: GO Analysis: Cellular components of TGFβ2 upregulated genes. Terms were ranked based on their logarithmic P-values. Genes involved in the terms can be checked from term IDs

GO: Biological Process – TGβ2

Term name	Term ID	p_{adj}
metabolic process	GO:0008152	9.762×10^{-12}
cellular metabolic process	GO:0044237	1.346×10^{-11}
organic substance metabolic process	GO:0071704	1.884×10^{-9}
primary metabolic process	GO:0044238	6.835×10^{-9}
nitrogen compound metabolic process	GO:0006807	1.581×10^{-8}
regulation of cellular metabolic process	GO:0031323	2.816×10^{-8}
regulation of metabolic process	GO:0019222	3.028×10^{-8}
macromolecule metabolic process	GO:0043170	4.298×10^{-8}
regulation of macromolecule metabolic process	GO:0060255	1.583×10^{-7}
organonitrogen compound metabolic process	GO:1901564	2.167×10^{-7}
protein localization to organelle	GO:0033365	2.439×10^{-7}
regulation of cellular protein metabolic process	GO:0032268	2.734×10^{-7}
positive regulation of cellular metabolic process	GO:0031325	3.490×10^{-7}
cellular response to stress	GO:0033554	6.557×10^{-7}
regulation of protein metabolic process	GO:0051246	6.585×10^{-7}
cellular protein metabolic process	GO:0044267	9.715×10^{-7}
protein metabolic process	GO:0019538	1.591×10^{-6}
regulation of nitrogen compound metabolic process	GO:0051171	1.732×10^{-6}
cellular catabolic process	GO:0044248	2.575×10^{-6}
regulation of primary metabolic process	GO:0080090	3.298×10^{-6}
cellular macromolecule metabolic process	GO:0044260	3.356×10^{-6}
positive regulation of metabolic process	GO:0009893	4.208×10^{-6}
catabolic process	GO:0009056	9.887×10^{-6}
cellular nitrogen compound metabolic process	GO:0034641	2.014×10^{-5}
DNA replication	GO:0006260	3.713×10^{-5}
organic substance biosynthetic process	GO:1901576	5.228×10^{-5}
amide biosynthetic process	GO:0043604	6.069×10^{-5}
positive regulation of macromolecule metabolic process	GO:0010604	6.829×10^{-5}
cellular biosynthetic process	GO:0044249	6.908×10^{-5}
cellular component organization or biogenesis	GO:0071840	7.274×10^{-5}
biosynthetic process	GO:0009058	1.039×10^{-4}
positive regulation of nitrogen compound metabolic p...	GO:0051173	1.695×10^{-4}
cellular macromolecule catabolic process	GO:0044265	2.660×10^{-4}
cellular process	GO:0009987	2.792×10^{-4}
positive regulation of proteolysis	GO:0045862	3.197×10^{-4}
regulation of DNA replication	GO:0006275	3.997×10^{-4}
cellular macromolecule localization	GO:0070727	4.198×10^{-4}
positive regulation of cellular protein metabolic process	GO:0032270	4.285×10^{-4}
cellular amide metabolic process	GO:0043603	4.817×10^{-4}
peptide biosynthetic process	GO:0043043	4.847×10^{-4}
regulation of cellular catabolic process	GO:0031329	5.247×10^{-4}
positive regulation of cellular process	GO:0048522	5.585×10^{-4}
organic cyclic compound metabolic process	GO:1901360	6.257×10^{-4}
cellular protein localization	GO:0034613	7.960×10^{-4}
translation	GO:0006412	8.030×10^{-4}
programmed cell death	GO:0012501	8.979×10^{-4}
apoptotic process	GO:0006915	9.124×10^{-4}
macromolecule biosynthetic process	GO:0009059	1.020×10^{-3}
cellular macromolecule biosynthetic process	GO:0034645	1.249×10^{-3}
response to organic substance	GO:0010033	1.369×10^{-3}
organic substance catabolic process	GO:1901575	1.387×10^{-3}
positive regulation of protein metabolic process	GO:0051247	1.415×10^{-3}

Figure 34: GO Analysis: Biological processes of TGFβ2 upregulated genes. Terms were ranked based on their logarithmic P values. Genes involved in the terms can be checked from term IDs.

A GO: Molecular Function – TGβ3

Term name	Term ID	p_{adj}
protein binding	GO:0005515	1.148×10^{-3}
nucleotide binding	GO:0000166	4.698×10^{-3}
nucleoside phosphate binding	GO:1901265	4.698×10^{-3}
heterocyclic compound binding	GO:1901363	1.206×10^{-2}
adenyl ribonucleotide binding	GO:0032559	1.794×10^{-2}
adenyl nucleotide binding	GO:0030554	2.005×10^{-2}
organic cyclic compound binding	GO:0097159	2.014×10^{-2}
anion binding	GO:0043168	2.694×10^{-2}
ATP binding	GO:0005524	3.129×10^{-2}
enzyme binding	GO:0019899	3.585×10^{-2}
binding	GO:0005488	4.038×10^{-2}
purine ribonucleotide binding	GO:0032555	4.588×10^{-2}

B GO: Biological Process– TGβ3

Term name	Term ID	p_{adj}
cellular response to chemical stimulus	GO:0070887	2.455×10^{-5}
cellular response to organic substance	GO:0071310	6.406×10^{-4}
cellular metabolic process	GO:0044237	4.227×10^{-3}
regulation of localization	GO:0032879	6.284×10^{-3}
response to organic substance	GO:0010033	6.681×10^{-3}
positive regulation of nitrogen compound metabolic p...	GO:0051173	1.400×10^{-2}
positive regulation of cell death	GO:0010942	1.808×10^{-2}
positive regulation of biological process	GO:0048518	1.829×10^{-2}
positive regulation of cellular process	GO:0048522	2.386×10^{-2}
organic substance metabolic process	GO:0071704	2.824×10^{-2}
nitrogen compound metabolic process	GO:0006807	2.901×10^{-2}
cellular process	GO:0009987	4.413×10^{-2}

C GO: Cellular Component – TGβ3

Term name	Term ID	p_{adj}
intracellular	GO:0005622	6.056×10^{-9}
membrane-bounded organelle	GO:0043227	4.843×10^{-8}
organelle	GO:0043226	1.427×10^{-6}
intracellular membrane-bounded organelle	GO:0043231	2.265×10^{-6}
intracellular organelle	GO:0043229	3.202×10^{-6}
cytoplasm	GO:0005737	5.340×10^{-5}
intracellular organelle lumen	GO:0070013	7.366×10^{-4}
organelle lumen	GO:0043233	7.411×10^{-4}
membrane-enclosed lumen	GO:0031974	7.411×10^{-4}
nucleus	GO:0005634	1.147×10^{-3}
nuclear lumen	GO:0031981	1.107×10^{-2}
cytosol	GO:0005829	1.682×10^{-2}
nucleoplasm	GO:0005654	2.501×10^{-2}
cellular anatomical entity	GO:0110165	3.725×10^{-2}

Figure 35: GO Analysis of TGFβ3 upregulated genes. Terms were ranked based on their logarithmic P values. Genes involved in the terms can be checked from term IDs. (A) Molecular function (B) Biological Process (C) Cellular Component.

5.1 CbioPortal Analysis

Cancer genomic analysis was performed via CbioPortal. 46760 samples from 176 studies in TCGA PanCancer Atlas were used in the analysis. Out of 46760 samples almost 30000 samples were profiled. Alteration frequencies and cancer types are shown. 1503 (3%) of samples were altered in the TGFβ1 gene, and 1079 (2%) were altered in TGFβ2. Bladder/Urinary tract cancer has the highest alteration frequency in the gene; TGFβ1 is amplified 9.7% of this cancer's cases. In sarcoma, TGFβ1 is 0.39% mutated, 3.53% amplified, and 0.39% had multiple alterations. TGFβ1 is altered in 4.31% of sarcoma cases. Prostate cancer, small cell lung cancer, endometrial carcinoma, and ovarian tumor have more than 3% alteration in the TGFβ1 gene. TGFβ2 is highly altered in breast cancer. 18.9% of the breast cancer cases have TGFβ2 amplification. Also, in invasive breast carcinoma, 7.57% of the cases have a mutated TGFβ2 gene. TGFβ2 is altered in more than 5% of the cases of Hepatocellular carcinoma, endometrial cancer, and cholangiocarcinoma. Endometrial Carcinoma has the highest alteration in the TGFβ3 gene. 4.27% of the cases have alterations in TGFβ3. TGFβ3 gene is mutated, amplified, and deleted in 3.19% of the cervical squamous cell carcinoma cases. Prostate cancer and colon adenocarcinoma have high alteration levels in TGFβ3 (Figure 36).

Table 9 : Number of cases that are different or co-occurred in each phenotype

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value
TGFB1	TGFB2	23754	232	488	35	2.876	<0.001
TGFB2	TGFB3	23805	497	181	26	2.782	<0.001
TGFB1	TGFB3	24054	248	188	19	>3	<0.001

Some of the cases have alterations in both genes or none of the genes. Thirty-five of cases have alterations in both TGFβ1 and TGFβ2. TGFβ1 and TGFβ3 are altered together in 19 of the cases. TGFβ1 and TGFβ2 have more co-occurrence in alteration (Table 9).

There were 96 missense, 16 truncating, 2 inframe and total 120 mutations on the TGFβ1 gene. Eight mutations are caused by a missense G29R/E amino acid change in non-small cell lung cancer, invasive breast carcinoma, and hepatocellular carcinoma. In the TGFβ2 gene, there were 138 missense, 35 truncating, and 176 total mutations. The highest number of mutations were a R131*/Q amino acid change truncating mutation, which affects colon and colorectal adenocarcinoma. 86 missense, 19 truncating, 4 inframe, and 109 total mutations on the TGFβ3 gene were spotted. In three of the inframe mutations, there was a K38del/N amino acid change on the gene (Figure 38).

TGFβ1, TGFβ2, and TGFβ3 expression levels in various cancer studies were analyzed. In TGFβ1, head and neck cancer, sarcoma, bladder cancer, and clear cell renal cell carcinoma (CCRCC) had the gene's highest expression levels with mostly truncating alterations. In TGFβ2, expression levels were higher in invasive breast carcinoma, low-grade glioma and glioblastoma. On the other hand, TGFβ3 levels in mesothelioma, sarcoma, prostate, and breast invasive carcinoma were the highest (Figure 37). These findings were consistent with the mutation data shared (Figure 36). Different isoforms of the TGFβ gene had different expression levels and mutations on other cancer types.

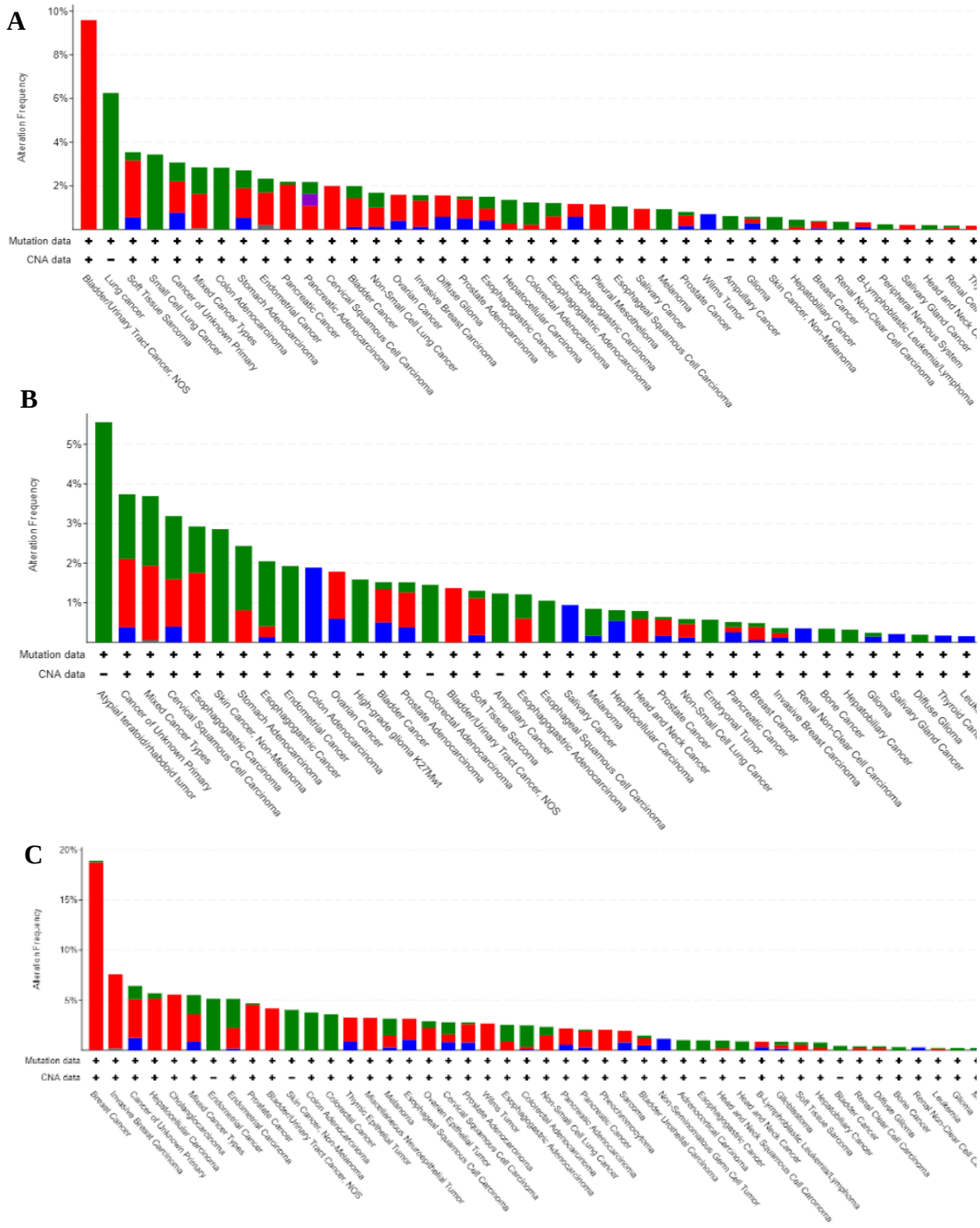


Figure 36: TGFβ gene alterations of cancer types. Y axis represents the percentages of alteration frequencies. X axis shows the cancer types. Color coding shows the mutation type; red: amplification, green: mutation, blue: deletion, purple: fusion. (A)TGFβ1 (B) TGFβ1 (C)TGFβ3

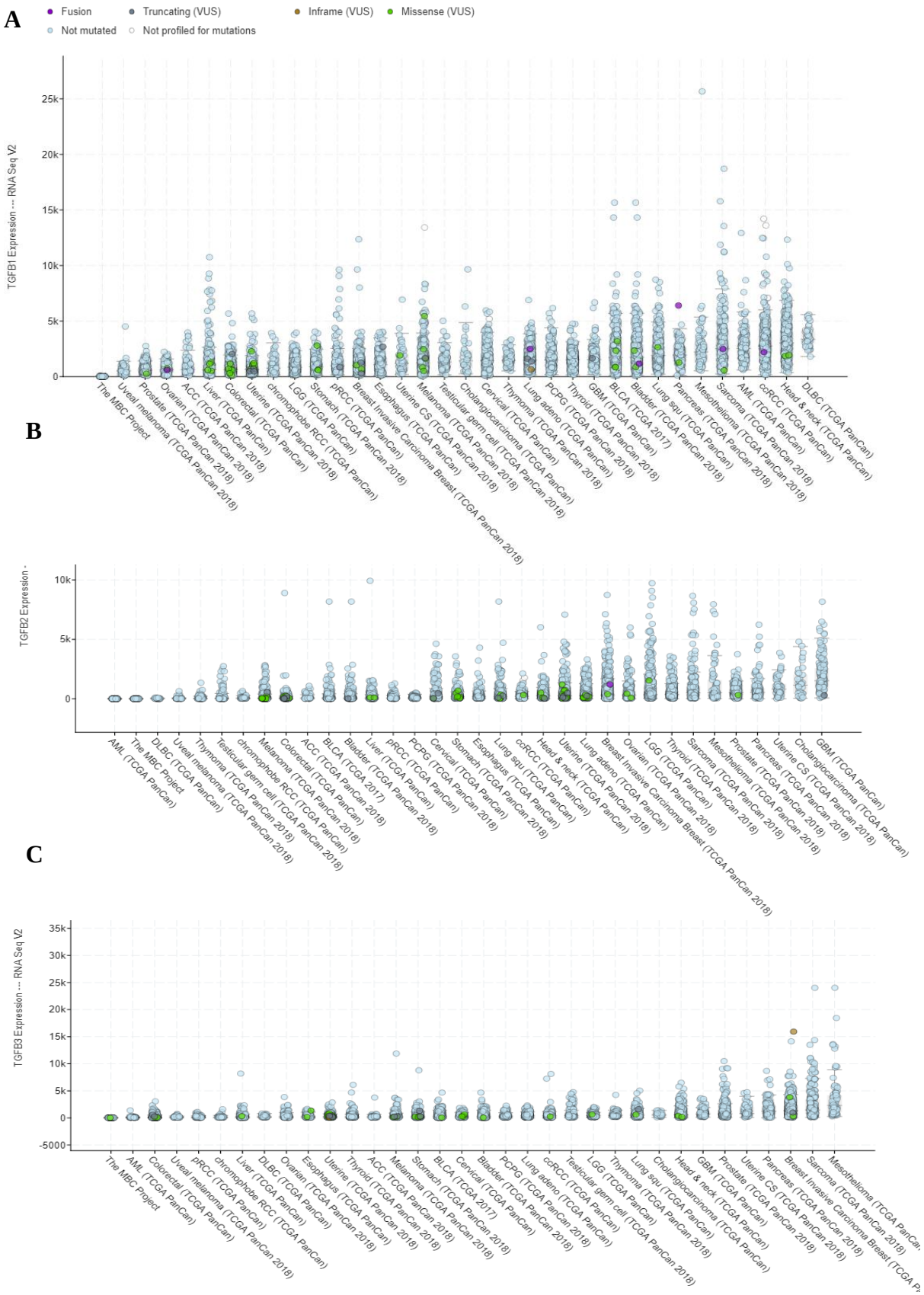


Figure 37: TGFβ gene expression levels in various cancer types. (A) TGFβ1, (B) TGFβ2, (C) TGFβ.

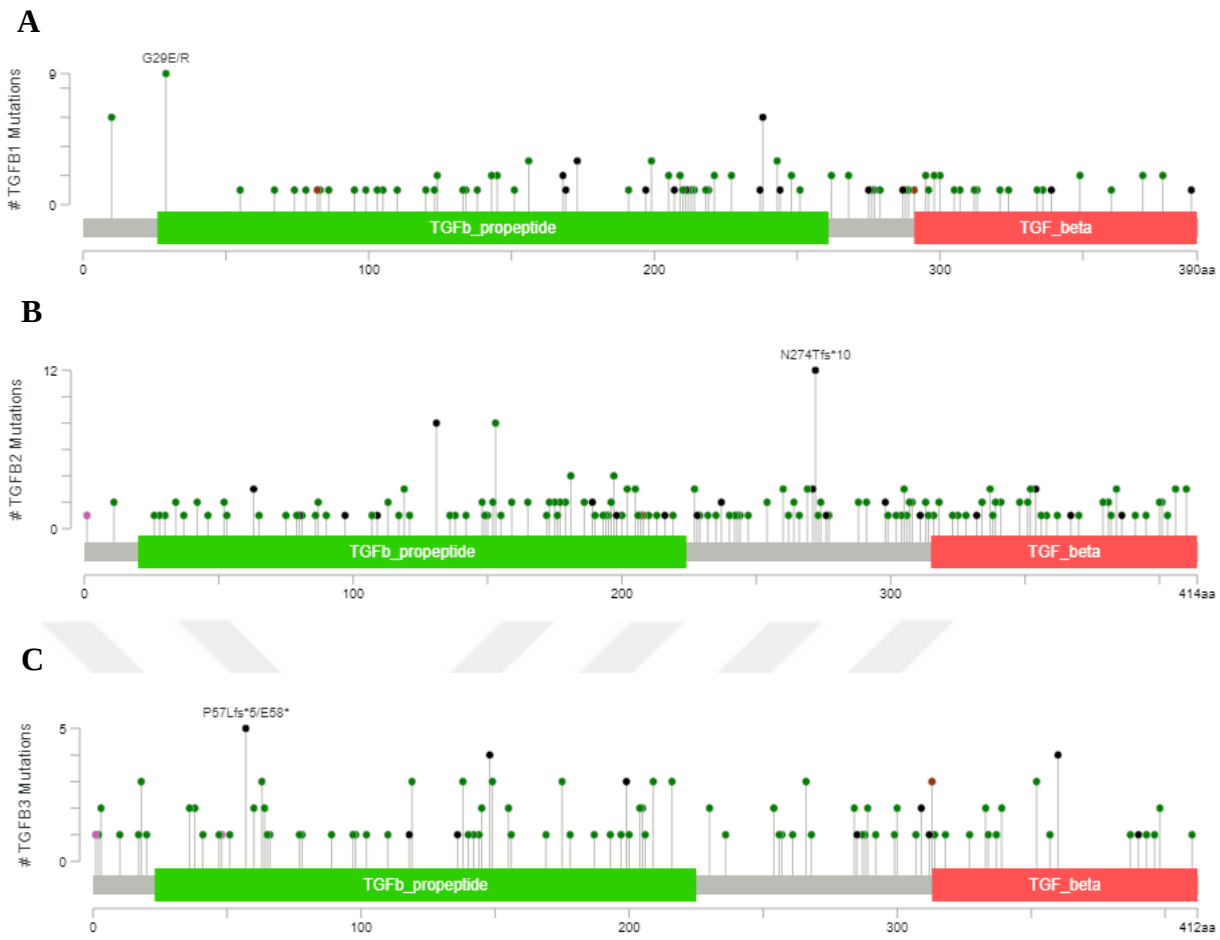


Figure 38: Number and amino acid points of the mutations in the TGFβ gene. The highest number of changes in the amino acid is shown. (A)TGFβ1, (B)TGFβ2, (C) TGFβ3.

6 Discussion

EMT is involved in plenty of processes from development to wound healing and cancer metastasis (Nieto et al., 2016). Since the TGFβ molecule is the key regulator of this transition, due to its central role in many processes and therapeutic possibilities, a large amount of data already exists in the literature about TGFβ and EMT. However, studies in the literature are mostly carried out with only one isoform of the TGFβ. Suppose the comparison among the isoforms is the case. In that case, it is only investigated in a cell line of interest like breast cancer (Hachim et al., 2018) (Chen et al., 2015) or in a specific topic like gonadal development (Memon et al., 2008). Thus, there is a lack in the literature of a broad range research on TGFβ isoforms and its transcriptional effects on many aspects, including datasets from different cell lines and different organisms. As indicated in this thesis's aim, differences among the isoforms of such a critical molecule are noteworthy for future research and therapeutic approaches.

To evaluate the research questions in a broad spectrum, 15 public datasets from 2 organisms of different ranges were analyzed with Bioconductor, R programming language, and a variety of online research tools (GO, GSEA). On the other hand, the study results are limited as it is performed only computationally, and further experimental validation is needed. As the significance and quality control measures were relevant, datasets are from different cell lines and have different treatment times.

Results are consistent with the hypothesis; the differential expression and gene set enrichment analysis results indicate a significant difference in transcriptional level between TGFβ isoforms. Withal, there are common genes and functions among all three isoforms, some regardless of the organism. Mutations on particular isotype of TGFβ are upfront in some cancer types; they differ in molecular functions; they reside in different cellular components and are associated with different pathways in some processes. The data contributes a better understanding of the TGFβ isoforms and their transcriptional effects.

Thirteen genes were shared among all three isoforms of intersected human and mouse differentially expressed genes. PDLIM7 a cytoskeletal protein and is expected to be in a TGFβ induced EMT gene list. CXCL1 cytokine is associated with metastasis of various types of cancers by activating the NF-κB pathway and related as a tumor-associated macrophages cytokine (Xiang et al., 2015). IL11 and PTGS2 will be mentioned in the later parts as they show some relevant results in further analysis. GADD45B and LAMB3 are two common genes among all isotypes in mouse datasets (Figure 15 B). Id proteins are one of the critical regulators

of TGFβ response to cell differentiation and so cell fate. ID3 is repressed by TGFβ induction and directly affects the E-cadherin levels to decrease (Kowanetz et al., 2004). Studies on Id proteins are with TGFβ1, but It can be seen inferentially that all isoforms raise the same interactions with ID3. Although it is studied as a cancer marker especially for lung cells, besides the adenocarcinoma cell lines, it has one of the highest enrichment scores in NMuMG cells both in TGFβ1 and TGFβ3 phenotypes of GSEA results. These genes can be considered as TGFβ induced EMT markers as they do not depend on isoform, organism and tumorigenic or normal cell type.

TSPAN2 is associated with invasiveness and bad prognosis especially when there is a p53 mutation in lung cancer cells (Otsubo et al., 2014). It is present in 5 datasets in the list, including ovarian fibroblast cells besides the adenocarcinoma cell lines. CTGF is responsible for tissue fibrosis and differentiation, also associated with poor prognosis in cancer (Zhu et al., 2015). As it is related with EMT induction, poor outcomes may be caused by EMT. It is upregulated in four datasets within the lung and NMuMG cells. MAF transcription factor is a proto-oncogene involved in development. There are no further studies about its relationship with TGFβ or EMT. It appears in 5 datasets of top 20 upregulated genes with high enrichment scores, all adenocarcinoma cell lines with TGFβ1 phenotype. Regarding this result, it may be used as a possible marker for metastatic lung cancer.

Whereas genes selected for the query were specific for each phenotype, similar results in the GO analysis may show that different genes can be correlated with the same biological process or molecular function. Numbers in the intersected Venn schemes (Figure 16) are now worthy of attention with the biological processes they participate in, molecular functions they regulate, and cellular components they reside.

As in further research besides the experimental validations, these results can be considered when planning a target-oriented new research on TGFβ and EMT. They may affect the choice of isoform in the practice. Common genes of all TGFβ isotypes from differential expression analysis results can be considered as TGFβ induced EMT markers. It can lead to a more specific way for possible drug targets and predictive markers. Especially in cancer research, since TGFβ has a dual role in tumor suppression and tumor invasion, different therapeutic approaches can develop for early and late-stage cancers based on the possible future research that may be carried out with this thesis's findings.

7 Conclusion and Suggestions

- There are significant differences and similarities between TGFβ isoforms at the transcriptional level even within the same dataset. Diversity of the analyzed datasets and methods let the research have a broad range of perspective.
- In upregulated genes from human datasets, 124 genes for TGFβ1 like ID2, CCPG1, FOXS1, ITGB3, DOCK2, ZEB1, NGF, LOXL1; 303 genes for TGFβ2 like IDH2, DAP, MALAT1, MAPK11, NES, TNNT2 and 512 genes for TGFβ3 like SOX4, IL6, LAMA3, SMURF2, VEGFA, TUBA1A, GATA6, ITGA2, MAFF, CDK6, OVOL1 were specific to isotype. In upregulated genes from mouse datasets, 172 genes for TGFβ1, 443 genes for TGFβ2 and 62 genes for TGFβ3 were specific to isotype.
- These isoform-specific upregulated genes are associated with different, thus important molecular functions and biological processes.
- All cancer cell lines within the TGFβ1 phenotype datasets shared 97 upregulated genes like; IL11, COL1A1, TAGLN, THBS1, PMEPAI, JUN, ADAM19.
- In the intersected datasets from human and mouse organisms, TGFβ3 had 19 genes specifically upregulated in that isoform such as; TUBA1A, IGF1R, GATA6, FOXQ1, KLF6, IRF6. TGFβ1 only genes like; VEGFC, TSPAN8, ANKRD1, ID2, FDFR2, IGFBP7, BMP1, ACTN1.
- All isoforms of human and mouse dataset intersection shared 13 genes regardless of the organism such as; IL11, PTGS2, PDLIM7, GADD45B, LAMB3, IFIT3, SKIL, ID3, CXCL1, PCDH7, DHRS3, FN1 and MAP3K5, which can be considered as TGFβ induced EMT markers.
- Each isoform of TGFβ has specific levels and types of mutations for a certain type of cancer.
- Common regulated genes for all isotypes are associated with histone modification factors.
- These results should be considered for target-specific further research and therapeutic approaches particularly in the treatment and diagnoses of early and late-stage cancers.

8 References

Alotaibi, H., Basilicata, M. F., Shehwana, H., Kosowan, T., Schreck, I., Braeutigam, C., Konu, O., Brabletz, T., & Stemmler, M. P. (2015). Enhancer cooperativity as a novel mechanism underlying the transcriptional regulation of E-cadherin during mesenchymal to epithelial transition. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2015.01.005

Amezquita, R. A., Lun, A. T. L., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., Waldron, L., Pagès, H., Smith, M. L., Huber, W., Morgan, M., Gottardo, R., & Hicks, S. C. (2020). Orchestrating single-cell analysis with Bioconductor. *Nature Methods*, *17*(2), 137–145. https://doi.org/10.1038/s41592-019-0654-x

Aroeira, L. S., Aguilera, A., Sánchez-Tomero, J. A., Bajo, M. A., Del Peso, G., Jiménez-Heffernan, J. A., Selgas, R., & López-Cabrera, M. (2007). Epithelial to mesenchymal transition and peritoneal membrane failure in peritoneal dialysis patients: Pathologic significance and potential therapeutic interventions. In *Journal of the American Society of Nephrology* (Vol. 18, Issue 7, pp. 2004–2013). American Society of Nephrology. https://doi.org/10.1681/ASN.2006111292

Batlle, E., & Massagué, J. (2019). Transforming Growth Factor-β Signaling in Immunity and Cancer. In *Immunity* (Vol. 50, Issue 4, pp. 924–940). Cell Press. https://doi.org/10.1016/j.immuni.2019.03.024

Bendoraite, A., Knouf, E. C., Garg, K. S., Parkin, R. K., Kroh, E. M., O'Briant, K. C., Ventura, A. P., Godwin, A. K., Karlan, B. Y., Drescher, C. W., Urban, N., Knudsen, B. S., & Tewari, M. (2010). Regulation of miR-200 family microRNAs and ZEB transcription factors in ovarian cancer: Evidence supporting a mesothelial-to-epithelial transition. *Gynecologic Oncology*, *116*(1), 117–125. https://doi.org/10.1016/j.ygyno.2009.08.009

Blighe, K. (2019). *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling*. https://github.com/kevinblighe/EnhancedVolcano

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

Chang, H., Brown, C. W., & Matzuk, M. M. (2002). Genetic Analysis of the Mammalian Transforming Growth fFactor-β Superfamily. *Endocrine Reviews*, *23*(6), 787–823. https://doi.org/10.1210/er.2002-0003

Chen, C., Zhao, K. N., Masci, P. P., Lakhani, S. R., Antonsson, A., Simpson, P. T., & Vitetta, L. (2015). TGFβ isoforms and receptors mRNA expression in breast tumours: Prognostic value and clinical implications. *BMC Cancer*, *15*(1), 1–12. https://doi.org/10.1186/s12885-015-1993-3

Chin, V. L., & Lim, C. L. (2019). Epithelial-mesenchymal plasticity — Engaging stemness in an interplay of phenotypes. In *Stem Cell Investigation* (Vol. 6, Issue August). AME Publishing Company. https://doi.org/10.21037/sci.2019.08.08

Cui, W., Fowlis, D. J., Bryson, S., Duffie, E., Ireland, H., Balmain, A., & Akhurst, R. J. (1996).

TGFβ1 inhibits the formation of benign skin tumors, but enhances progression to invasive spindle carcinomas in transgenic mice. *Cell*, *86*(4), 531–542. https://doi.org/10.1016/S0092-8674(00)80127-0

De Larco, J. E., & Todaro, G. J. (1978). Growth factors from murine sarcoma virus-transformed cells. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(8), 4001–4005. https://doi.org/10.1073/pnas.75.8.4001

De Miranda, N. F. C. C., Van Dinther, M., Van Den Akker, B. E. W. M., Van Wezel, T., Ten Dijke, P., & Morreau, H. (2015). Transforming growth factor β signaling in colorectal cancer cells with microsatellite instability despite biallelic mutations in TGFBR2. *Gastroenterology*, *148*(7), 1427-1437.e8. https://doi.org/10.1053/j.gastro.2015.02.052

Erdogan, B., & Webb, D. J. (2017). Cancer-associated fibroblasts modulate growth factor signaling and extracellular matrix remodeling to regulate tumor metastasis. In *Biochemical Society Transactions* (Vol. 45, Issue 1, pp. 229–236). Portland Press Ltd. https://doi.org/10.1042/BST20160387

Fernández-L., A., Sanz-Rodriguez, F., Blanco, F. J., Bernabéu, C., & Botella, L. M. (2006). Hereditary hemorrhagic telangiectasia, a vascular dysplasia affecting the TGF-β signaling pathway. In *Clinical Medicine and Research* (Vol. 4, Issue 1, pp. 66–78). Marshfield Clinic. https://doi.org/10.3121/cmr.4.1.66

Foroutan, M., Cursons, J., Hediyeh-Zadeh, S., Thompson, E. W., & Davis, M. J. (2017). A transcriptional program for detecting TGFβ-induced EMT in cancer. *Molecular Cancer Research*, *15*(5), 619–631. https://doi.org/10.1158/1541-7786.MCR-16-0313

Fournier, P. G. J., Juárez, P., Jiang, G., Clines, G. A., Niewolna, M., Kim, H. S., Walton, H. W., Peng, X. H., Liu, Y., Mohammad, K. S., Wells, C. D., Chirgwin, J. M., & Guise, T. A. (2015). The TGF-β Signaling Regulator PMEPA1 Suppresses Prostate Cancer Metastases to Bone. *Cancer Cell*, *27*(6), 809–821. https://doi.org/10.1016/j.ccell.2015.04.009

Gandalovičová, A., Vomastek, T., Rosel, D., & Brábek, J. (2016). Cell polarity signaling in the plasticity of cancer cell invasiveness. *Oncotarget*, *7*(18), 25022–25049. https://doi.org/10.18632/oncotarget.7214

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, *6*(269), pl1. https://doi.org/10.1126/scisignal.2004088

Garg, M. (2013). Epithelial-mesenchymal transition - activating transcription factors - multifunctional regulators in cancer. *World Journal of Stem Cells*, *5*(4), 188. https://doi.org/10.4252/wjsc.v5.i4.188

Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). affy-analysis of Affymetrix GeneChip data at the probe level. *BIOINFORMATICS*, *20*(3), 307–315. https://doi.org/10.1093/bioinformatics/btg405

Gene Ontology Consortium, T., Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Michael Cherry, J., 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., … Genet Author manuscript, N. (2000). Gene Ontology: tool for the

unification of biology NIH Public Access Author Manuscript. *Nat Genet*, *25*(1), 25–29. https://doi.org/10.1038/75556

Gjorevski, N., Boghaert, E., & Nelson, C. M. (2012). Regulation of epithelial-mesenchymal transition by transmission of mechanical stress through epithelial tissues. In *Cancer Microenvironment* (Vol. 5, Issue 1, pp. 29–38). Cancer Microenviron. https://doi.org/10.1007/s12307-011-0076-5

Gordian, E., Welsh, E. A., Gimbrone, N., Siegel, E. M., Shibata, D., Creelan, B. C., Cress, W. D., Eschrich, S. A., Haura, E. B., & Muñoz-Antonia, T. (2019). Transforming growth factor β-induced epithelial-to-mesenchymal signature predicts metastasis-free survival in non-small cell lung cancer. *Oncotarget*, *10*(8), 810–824. https://doi.org/10.18632/oncotarget.26574

Gu, H., Mickler, E. A., Cummings, O. W., Sandusky, G. E., Weber, D. J., Gracon, A., Woodruff, T., Wilkes, D. S., & Vittal, R. (2014). Crosstalk between TGF-β1 and complement activation augments epithelial injury in pulmonary fibrosis. *FASEB Journal*, *28*(10), 4223–4234. https://doi.org/10.1096/fj.13-247650

Guo, W., Wang, P., Liu, Z. H., & Ye, P. (2018). Analysis of differential expression of tight junction proteins in cultured oral epithelial cells altered by Porphyromonas gingivalis, Porphyromonas gingivalis lipopolysaccharide, and extracellular adenosine triphosphate. *International Journal of Oral Science*, *10*(1), e8–e8. https://doi.org/10.1038/ijos.2017.51

Guo, X., & Wang, X. F. (2009). Signaling cross-talk between TGF-β/BMP and other pathways. In *Cell Research* (Vol. 19, Issue 1, pp. 71–88). NIH Public Access. https://doi.org/10.1038/cr.2008.302

Hachim, M. Y., Hachim, I. Y., Dai, M., Ali, S., & Lebrun, J. J. (2018). Differential expression of TGFβ isoforms in breast cancer highlights different roles during breast cancer progression. *Tumor Biology*, *40*(1), 1–12. https://doi.org/10.1177/1010428317748254

Han, Z., Kang, D., Joo, Y., Lee, J., Oh, G. H., Choi, S., Ko, S., Je, S., Choi, H. J., & Song, J. J. (2018). TGF-β downregulation-induced cancer cell death is finely regulated by the SAPK signaling cascade. *Experimental and Molecular Medicine*, *50*(12), 1–19. https://doi.org/10.1038/s12276-018-0189-8

Hao, Y., Baker, D., & Dijke, P. Ten. (2019). TGF-β-mediated epithelial-mesenchymal transition and cancer metastasis. *International Journal of Molecular Sciences*, *20*(11). https://doi.org/10.3390/ijms20112767

Hata, A., & Chen, Y. G. (2016). TGF-β signaling from receptors to smads. *Cold Spring Harbor Perspectives in Biology*, *8*(9). https://doi.org/10.1101/cshperspect.a022061

Hoot, K. E., Lighthall, J., Han, G., Lu, S. L., Li, A., Ju, W., Kulesz-Martin, M., Bottinger, E., & Wang, X. J. (2008). Keratinocyte-specific Smad2 ablation results in increased epithelial-mesenchymal transition during skin cancer formation and progression. *Journal of Clinical Investigation*, *118*(8), 2722–2732. https://doi.org/10.1172/JCI33713

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MaCdonald, J., Obenchain, V., Oleš, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Hussey, G. S., Link, L. A., Brown, A. S., Howley, B. V., Chaudhury, A., & Howe, P. H. (2012). Establishment of a TGFβ-Induced Post-Transcriptional EMT Gene Signature. *PLoS ONE*, *7*(12), 1–12. https://doi.org/10.1371/journal.pone.0052624

Iordanskaia, T., & Nawshad, A. (2011). Mechanisms of transforming growth factor β induced cell cycle arrest in palate development. *Journal of Cellular Physiology*, *226*(5), 1415–1424. https://doi.org/10.1002/jcp.22477

Ishigame, H., Mosaheb, M. M., Sanjabi, S., & Flavell, R. A. (2013). Truncated Form of TGF-βRII, But Not Its Absence, Induces Memory CD8 + T Cell Expansion and Lymphoproliferative Disorder in Mice . *The Journal of Immunology*, *190*(12), 6340–6350. https://doi.org/10.4049/jimmunol.1300397

Itoh, Y., Saitoh, M., & Miyazawa, K. (2018). Smad3-STAT3 crosstalk in pathophysiological contexts. In *Acta Biochimica et Biophysica Sinica* (Vol. 50, Issue 1, pp. 82–90). Oxford University Press. https://doi.org/10.1093/abbs/gmx118

Janakiraman, H., House, R. P., Gangaraju, V. K., Diehl, J. A., Howe, P. H., & Palanisamy, V. (2018). The long (IncRNA) and short (miRNA) of it: TGFb- mediated control of RNA-binding proteins and noncoding RNAs. In *Molecular Cancer Research* (Vol. 16, Issue 4, pp. 567–579). American Association for Cancer Research Inc. https://doi.org/10.1158/1541-7786.MCR-17-0547

Janda, E., Lehmann, K., Killisch, I., Jechlinger, M., Herzig, M., Downward, J., Beug, H., & Grünert, S. (2002). Ras and TGFβ cooperatively regulate epithelial cell plasticity and metastasis: Dissection of Ras signaling pathways. *Journal of Cell Biology*, *156*(2), 299–313. https://doi.org/10.1083/jcb.200109037

Janknecht, R., Wells, N. J., & Hunter, T. (1998). TGF-β-stimulated cooperation of Smad proteins with the coactivators CBP/p300. *Genes and Development*, *12*(14), 2114–2119. https://doi.org/10.1101/gad.12.14.2114

Johansson, J., Sahin, C., Pestoff, R., Ignatova, S., Forsberg, P., Edsjö, A., Ekstedt, M., & Stenmark Askmalm, M. (2015). A Novel SMAD4 Mutation Causing Severe Juvenile Polyposis Syndrome with Protein Losing Enteropathy, Immunodeficiency, and Hereditary Haemorrhagic Telangiectasia . *Case Reports in Gastrointestinal Medicine*, *2015*, 1–5. https://doi.org/10.1155/2015/140616

Jung, S. N., Lim, H. S., Liu, L., Chang, J. W., Lim, Y. C., Rha, K. S., & Koo, B. S. (2018). LAMB3 mediates metastatic tumor behavior in papillary thyroid cancer by regulating c-MET/Akt signals. *Scientific Reports*, *8*(1), 2718. https://doi.org/10.1038/s41598-018-21216-0

Kallio, M. A., Tuimala, J. T., Hupponen, T., Klemelä, P., Gentile, M., Scheinin, I., Koski, M., Käki, J., & Korpelainen, E. I. (2011). Chipster: User-friendly analysis software for microarray and other high-throughput data. *BMC Genomics*, *12*(1), 507. https://doi.org/10.1186/1471-2164-12-507

Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. In *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI39104

Keshamouni, V. G., Michailidis, G., Grasso, C. S., Anthwal, S., Strahler, J. R., Walker, A., Arenberg, D. A., Reddy, R. C., Akulapalli, S., Thannickal, V. J., Standiford, T. J., Andrews, P. C., & Omenn, G. S. (2006). Differential protein expression profiling by

iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *Journal of Proteome Research*, *5*(5), 1143–1154. https://doi.org/10.1021/pr050455t

Kim, S. J., Im, Y., Markowitz, S. D., & Bang, Y. (2000). Molecular mechanisms of inactivation of TGF- b receptors during carcinogenesis. *Cytokine & Growth Factor Reviews*, *11*(1–2), 159±168. https://doi.org/10.1016/S1359-6101(99)00039-8

Kong, C., Wang, C., Wang, L., Ma, M., Niu, C., Sun, X., Du, J., Dong, Z., Zhu, S., Lu, J., & Huang, B. (2011). NEDD9 is a positive regulator of epithelial-mesenchymal transition and promotes invasion in aggressive breast cancer. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0022666

Korkmaz, Ö., Özen, S., Özcan, N., Bayındır, P., Şen, S., Onay, H., Gökşen, D., Avanoğlu, A., Özkınay, F., & Darcan, Ş. (2017). Persistent müllerian duct syndrome with transverse testicular ectopia: A novel anti-müllerian hormone receptor mutation. *JCRPE Journal of Clinical Research in Pediatric Endocrinology*, *9*(2), 179–181. https://doi.org/10.4274/jcrpe.4058

Kowanetz, M., Valcourt, U., Bergström, R., Heldin, C.-H., & Moustakas, A. (2004). Id2 and Id3 Define the Potency of Cell Proliferation and Differentiation Responses to Transforming Growth Factor β and Bone Morphogenetic Protein. *Molecular and Cellular Biology*, *24*(10), 4241–4254. https://doi.org/10.1128/mcb.24.10.4241-4254.2004

Kunnen, S. J., Leonhard, W. N., Semeins, C., Hawinkels, L. J. A. C., Poelma, C., ten Dijke, P., Bakker, A., Hierck, B. P., & Peters, D. J. M. (2017). Fluid shear stress-induced TGF-β/ALK5 signaling in renal epithelial cells is modulated by MEK1/2. *Cellular and Molecular Life Sciences*, *74*(12), 2283–2298. https://doi.org/10.1007/s00018-017-2460-x

Lee, S. Y., Ju, M. K., Jeun, H. M., Lee, Y. J., Kim, C. H., Park, H. G., Han, S. I., & Kang, H. S. (2019). Reactive oxygen species induce epithelial-mesenchymal transition, glycolytic switch, and mitochondrial repression through the Dlx-2/Snail signaling pathways in MCF-7 cells. *Molecular Medicine Reports*, *20*(3), 2339–2346. https://doi.org/10.3892/mmr.2019.10466

Levy, L., & Hill, C. S. (2005). Smad4 Dependency Defines Two Classes of Transforming Growth Factor β (TGF-β) Target Genes and Distinguishes TGF-β-Induced Epithelial-Mesenchymal Transition from Its Antiproliferative and Migratory Responses. *Molecular and Cellular Biology*, *25*(18), 8108–8125. https://doi.org/10.1128/mcb.25.18.8108-8125.2005

Liao, T. T., & Yang, M. H. (2017). Revisiting epithelial-mesenchymal transition in cancer metastasis: the connection between epithelial plasticity and stemness. In *Molecular Oncology* (Vol. 11, Issue 7, pp. 792–804). Wiley Blackwell. https://doi.org/10.1002/1878-0261.12096

Lin, Y. Te, & Wu, K. J. (2020). Epigenetic regulation of epithelial-mesenchymal transition: Focusing on hypoxia and TGF-β signaling. In *Journal of Biomedical Science* (Vol. 27, Issue 1, pp. 1–10). BioMed Central Ltd. https://doi.org/10.1186/s12929-020-00632-3

Liu, Z., Wang, M., Alvarez, J. V., Bonney, M. E., Chen, C. C., D'Cruz, C., Pan, T. C., Tadesse, M. G., & Chodosh, L. A. (2008). Singular value decomposition-based regression identifies activation of endogenous signaling pathways in vivo. *Genome Biology*, *9*(12). https://doi.org/10.1186/gb-2008-9-12-r180

Loh, C. Y., Chai, J. Y., Tang, T. F., Wong, W. F., Sethi, G., Shanmugam, M. K., Chong, P. P., & Looi, C. Y. (2019). The E-Cadherin and N-Cadherin Switch in Epithelial-to-Mesenchymal Transition: Signaling, Therapeutic Implications, and Challenges. In *Cells* (Vol. 8, Issue 10). NLM (Medline). https://doi.org/10.3390/cells8101118

Martínez-Quintana, E., Rodríguez-González, F., Garay-Sánchez, P., & Tugores, A. (2014). A novel fibrillin 1 gene mutation leading to marfan syndrome with minimal cardiac features. *Molecular Syndromology*, *5*(5), 236–240. https://doi.org/10.1159/000358846

Massagué, J. (2008). TGFβ in Cancer. In *Cell* (Vol. 134, Issue 2, pp. 215–230). NIH Public Access. https://doi.org/10.1016/j.cell.2008.07.001

Maupin, K. A., Sinha, A., Eugster, E., Miller, J., Ross, J., Paulino, V., Keshamouni, V. G., Tran, N., Berens, M., Webb, C., & Haab, B. B. (2010). Glycogene expression alterations associated with pancreatic cancer Epithelial-mesenchymal transition in complementary model systems. *PLoS ONE*, *5*(9). https://doi.org/10.1371/journal.pone.0013002

Memon, M. A., Anway, M. D., Covert, T. R., Uzumcu, M., & Skinner, M. K. (2008). *Transforming Growth Factor Beta (TGFβ1, TGFβ2 and TGFβ3) Null-Mutant Phenotypes in Embryonic Gonadal Development*. https://doi.org/10.1016/j.mce.2008.08.017

Miller, C. (2019). *simpleaffy: Very simple high level analysis of Affymetrix data.*

Mishra, V. K., & Johnsen, S. A. (2014). Targeted therapy of epigenomic regulatory mechanisms controlling the epithelial to mesenchymal transition during tumor progression. In *Cell and Tissue Research* (Vol. 356, Issue 3, pp. 617–630). Springer Verlag. https://doi.org/10.1007/s00441-014-1912-y

Miyazono, K. (2009). Transforming growth factor-β signaling in epithelial-mesenchymal transition and progression of cancer. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, *85*(8), 314–323. https://doi.org/10.2183/pjab.85.314

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

Moreno-Bueno, G., Portillo, F., & Cano, A. (2008). Transcriptional regulation of cell polarity in EMT and cancer. In *Oncogene* (Vol. 27, Issue 55, pp. 6958–6969). Oncogene. https://doi.org/10.1038/onc.2008.346

Moustakas, A., & Heldin, C.-H. (2016). Mechanisms of TGFβ-Induced Epithelial–Mesenchymal Transition. *Journal of Clinical Medicine*, *5*(7), 63. https://doi.org/10.3390/jcm5070063

Namwanje, M., & Brown, C. W. (2016). Activins and inhibins: Roles in development, physiology, and disease. *Cold Spring Harbor Perspectives in Biology*, *8*(7). https://doi.org/10.1101/cshperspect.a021881

Nawshad, A., Medici, D., Liu, C. C., & Hay, E. D. (2007). TGFβ3 inhibits E-cadherin gene expression in palate medial-edge epithelial cells through a Smad2-Smad4-LEF1 transcription complex. *Journal of Cell Science*, *120*(9), 1646–1653. https://doi.org/10.1242/jcs.003129

Nemunaitis, J. (2013). Cancer targeting vaccines: Surrogate measures of activity. In *Human Vaccines and Immunotherapeutics* (Vol. 9, Issue 1, pp. 213–218). Taylor & Francis. https://doi.org/10.4161/hv.22091

Nieto, M. A., Huang, R. Y. Y. J., Jackson, R. A. A., & Thiery, J. P. P. (2016). EMT: 2016. In *Cell* (Vol. 166, Issue 1, pp. 21–45). Cell Press. https://doi.org/10.1016/j.cell.2016.06.028

Ning, B., Li, W., Zhao, W., & Wang, R. (2015). Targeting epigenetic regulations in cancer. In *Acta Biochimica et Biophysica Sinica* (Vol. 48, Issue 1, pp. 97–109). Oxford University Press. https://doi.org/10.1093/abbs/gmv116

Oft, M., Heider, K. H., & Beug, H. (1998a). TGFβ signaling is necessary for carcinoma cell invasiveness and metastasis. *Current Biology*, *8*(23), 1243–1252. https://doi.org/10.1016/s0960-9822(07)00533-7

Oft, M., Heider, K. H., & Beug, H. (1998b). TGFβ signaling is necessary for carcinoma cell invasiveness and metastasis. *Current Biology*, *8*(23), 1243–1252. https://doi.org/10.1016/s0960-9822(07)00533-7

Otsubo, C., Otomo, R., Miyazaki, M., Matsushima-Hibiya, Y., Kohno, T., Iwakawa, R., Takeshita, F., Okayama, H., Ichikawa, H., Saya, H., Kiyono, T., Ochiya, T., Tashiro, F., Nakagama, H., Yokota, J., & Enari, M. (2014). TSPAN2 Is Involved in Cell Invasion and Motility during Lung Cancer Progression. *Cell Reports*, *7*(2), 527–538. https://doi.org/10.1016/j.celrep.2014.03.027

Pickup, M., Novitskiy, S., & Moses, H. L. (2013). The roles of TGFβ in the tumour microenvironment. In *Nature Reviews Cancer* (Vol. 13, Issue 11, pp. 788–799). NIH Public Access. https://doi.org/10.1038/nrc3603

R. Gentleman V. Carey and W. Huber and F. Hahne. (2019). *Genefilter: genefilter: methods for filtering genes from high-throughput experiments* (R package version 1.66.0).

R Core Team. (2019). *R: A Language and Environment for Statistical Computing* (3.6.0). R Foundation for Statistical Computing. https://www.r-project.org/

Radeke, M. J., Radeke, C. M., Shih, Y.-H., Hu, J., Bok, D., Johnson, L. V., & Coffey, P. J. (2015). Restoration of mesenchymal retinal pigmented epithelial cells by TGFβ pathway inhibitors: implications for age-related macular degeneration. *Genome Medicine*, *7*(1), 58. https://doi.org/10.1186/s13073-015-0183-x

Ramesh, S., Wildey, G. M., & Howe, P. H. (2009). Transforming growth factor β (TGFβ)-induced apoptosis: The rise & fall of Bim. In *Cell Cycle* (Vol. 8, Issue 1, pp. 11–17). Taylor and Francis Inc. https://doi.org/10.4161/cc.8.1.7291

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*, 191–198. https://doi.org/10.1093/nar/gkz369

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., & Mesirov, J. P. (2006). GenePattern 2.0 [2]. *Nature Genetics*, *38*(5), 500–501. https://doi.org/10.1038/ng0506-500

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7). https://doi.org/10.1093/nar/gkv007

Sayarlioglu, H., Okuyucu, A., Bedir, A., Salis, O., Yenen, E., Bekfilavioglu, G., & Kaya, C.

(2016). Is there any role of epithelial to mesenchymal transition in the pathogenesis of contrast nephropathy? *Renal Failure*, *38*(8), 1249–1255. https://doi.org/10.1080/0886022X.2016.1209381

Seker, F., Cingoz, A., Sur-Erdem, İ., Erguder, N., Erkent, A., Uyulur, F., Selvan, M. E., Gümüş, Z. H., Gönen, M., Bayraktar, H., Wakimoto, H., & Bagci-Onder, T. (2019). Identification of SERPINE1 as a regulator of glioblastoma cell dispersal with transcriptome profiling. *Cancers*, *11*(11), 1–21. https://doi.org/10.3390/cancers11111651

Sengez, B., Aygün, I., Shehwana, H., Toyran, N., Tercan Avci, S., Konu, O., Stemmler, M. P., & Alotaibi, H. (2019). The Transcription Factor Elf3 Is Essential for a Successful Mesenchymal to Epithelial Transition. *Cells*, *8*(8), 8–10. https://doi.org/10.3390/cells8080858

Serrano-Gomez, S. J., Maziveyi, M., & Alahari, S. K. (2016). Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. In *Molecular Cancer* (Vol. 15, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12943-016-0502-x

Sharkey, D. J., Macpherson, A. M., Tremellen, K. P., Mottershead, D. G., Gilchrist, R. B., & Robertson, S. A. (2012). TGF-β Mediates Proinflammatory Seminal Fluid Signaling in Human Cervical Epithelial Cells. *The Journal of Immunology*, *189*(2), 1024–1035. https://doi.org/10.4049/jimmunol.1200005

Shi, J., & Walker, M. (2008). Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. *Current Bioinformatics*. https://doi.org/10.2174/157489307780618231

Skrypek, N., Goossens, S., De Smedt, E., Vandamme, N., & Berx, G. (2017). Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity. In *Trends in Genetics* (Vol. 33, Issue 12, pp. 943–959). Elsevier Ltd. https://doi.org/10.1016/j.tig.2017.08.004

Song, J. (2007). EMT or apoptosis: A decision for TGF-β. In *Cell Research*. https://doi.org/10.1038/cr.2007.25

Spender, L. C., O'Brien, D. I., Simpson, D., Dutt, D., Gregory, C. D., Allday, M. J., Clark, L. J., & Inman, G. J. (2009). TGF-β induces apoptosis in human B cells by transcriptional regulation of BIK and BCL-XL. *Cell Death and Differentiation*, *16*(4), 593–602. https://doi.org/10.1038/cdd.2008.183

Stange, K., Désir, J., Kakar, N., Mueller, T. D., Budde, B. S., Gordon, C. T., Horn, D., Seemann, P., & Borck, G. (2015). A hypomorphic BMPR1B mutation causes du Pan acromesomelic dysplasia. *Orphanet Journal of Rare Diseases*, *10*(1). https://doi.org/10.1186/s13023-015-0299-5

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

Sun, Y., Daemen, A., Hatzivassiliou, G., Arnott, D., Wilson, C., Zhuang, G., Gao, M., Liu, P., Boudreau, A., Johnson, L., & Settleman, J. (2014). Metabolic and transcriptional profiling

reveals pyruvate dehydrogenase kinase 4 as a mediator of epithelial-mesenchymal transition and drug resistance in tumor cells. *Cancer & Metabolism*, *2*(1), 20. https://doi.org/10.1186/2049-3002-2-20

Suzuki, H. I. (2018). MicroRNA control of TGF-β signaling. In *International Journal of Molecular Sciences* (Vol. 19, Issue 7). MDPI AG. https://doi.org/10.3390/ijms19071901

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. Von. (2019). *STRING v11 : protein – protein association networks with increased coverage , supporting functional discovery in genome-wide experimental datasets*. *47*(November 2018), 607–613. https://doi.org/10.1093/nar/gky1131

Tan, X., Chen, C., Zhu, Y., Deng, J., Qiu, X., Huang, S., Shang, F., Cheng, B., & Liu, Y. (2017). Proteotoxic Stress Desensitizes TGF-beta Signaling Through Receptor Downregulation in Retinal Pigment Epithelial Cells. *Current Molecular Medicine*, *17*(3), 189–199. https://doi.org/10.2174/1566524017666170619113435

Thierauf, J., Veit, J. A., & Hess, J. (2017). Epithelial-to-mesenchymal transition in the pathogenesis and therapy of head and neck cancer. In *Cancers* (Vol. 9, Issue 7). MDPI AG. https://doi.org/10.3390/cancers9070076

Upadhyay, A., Moss-Taylor, L., Kim, M. J., Ghosh, A. C., & O'Connor, M. B. (2017). TGF-β family signaling in drosophila. *Cold Spring Harbor Perspectives in Biology*, *9*(9). https://doi.org/10.1101/cshperspect.a022152

Valcourt, U., Kowanetz, M., Hideki, N., Heldin, C.-H., & Moustakas, A. (2005). TGF-B and the Smad Signaling Pathway Support Transcriptomic Reprogramming during Epithelial Mesenchymal Cell Transition. *Mol Biol Cell*, *16*(April), 1097–2002. https://doi.org/10.1091/mbc.e04-08-0658

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., MacIejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, Di., … Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Wu, M. Y., & Hill, C. S. (2009). TGF-β Superfamily Signaling in Embryonic Development and Homeostasis. *Developmental Cell*, *16*(3), 329–343. https://doi.org/10.1016/j.devcel.2009.02.012

Wu, Y. C., Tang, S. J., Sun, G. H., & Sun, K. H. (2016). CXCR7 mediates TGFβ1-promoted EMT and tumor-initiating features in lung cancer. *Oncogene*, *35*(16), 2123–2132. https://doi.org/10.1038/onc.2015.274

Xiang, Z., Jiang, D. P., Xia, G. G., Wei, Z. W., Chen, W., He, Y., & Zhang, C. H. (2015). CXCL1 expression is correlated with Snail expression and affects the prognosis of patients with gastric cancer. *Oncology Letters*, *10*(4), 2458–2464. https://doi.org/10.3892/ol.2015.3614

Xie, F., Ling, L., Van Dam, H., Zhou, F., & Zhang, L. (2018). TGF-β signaling in cancer metastasis. *Acta Biochimica et Biophysica Sinica*, *50*(1), 121–132. https://doi.org/10.1093/abbs/gmx123

Xu, B., Bai, Z., Yin, J., & Zhang, Z. (2019). Global transcriptomic analysis identifies

SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer . *PeerJ*, *7*, e7091. https://doi.org/10.7717/peerj.7091

Xu J., Lamouille S., D. R. (2016). TGF-β-induced epithelial to mesenchymal transition. *Cell Res.*, *176*(1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Xu, Jia, Acharya, S., Sahin, O., Zhang, Q., Saito, Y., Yao, J., Sahin, A. A., Zhang, X. H., Hung, M., & Zhang, J. D. (2016). 14-3-3ζ turns TGF-β's function from tumor suppressor to metastasis promoter in breast cancer by contextual changes of Smad partners from p53 to Gli2. *Cancer Cell*, *27*(2), 177–192. https://doi.org/10.1016/j.ccell.2014.11.025.14-3-3

Xu, Jian, Lamouille, S., & Derynck, R. (2009). TGF-B-induced epithelial to mesenchymal transition. In *Cell Research* (Vol. 19, Issue 2, pp. 156–172). NIH Public Access. https://doi.org/10.1038/cr.2009.5

Xu, L. (2006). Regulation of Smad activities. In *Biochimica et Biophysica Acta - Gene Structure and Expression* (Vol. 1759, Issues 11–12, pp. 503–513). NIH Public Access. https://doi.org/10.1016/j.bbaexp.2006.11.001

Ye, X., & Weinberg, R. A. (2015). Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression. In *Trends in Cell Biology* (Vol. 25, Issue 11, pp. 675–686). Elsevier Ltd. https://doi.org/10.1016/j.tcb.2015.07.012

Yeung TL, Leung CS, Wong KK, S. G. et al. (2013). TGF-β modulates ovarian cancer invasion by upregulating CAF- derived. *Cancer Research*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Zhang, H., Pan, Y. zhen, Cheung, M., Cao, M., Yu, C., Chen, L., Zhan, L., He, Z. wei, & Sun, C. yi. (2019). LAMB3 mediates apoptotic, proliferative, invasive, and metastatic behaviors in pancreatic cancer by regulating the PI3K/Akt signaling pathway. *Cell Death and Disease*, *10*(3), 1–14. https://doi.org/10.1038/s41419-019-1320-z

Zhang, Y. E. (n.d.). *Non-Smad pathways in TGF-β signaling*. https://doi.org/10.1038/cr.2008.328

Zhao, M., Liu, Y., Liu, R., Qi, J., Hou, Y., Chang, J., & Ren, L. (2018). Upregulation of IL-11, an IL-6 Family Cytokine, Promotes Tumor Progression and Correlates with Poor Prognosis in Non-Small Cell Lung Cancer. *Cellular Physiology and Biochemistry*, *45*(6), 2213–2224. https://doi.org/10.1159/000488166

Zhong, Z., Hu, Z., Jiang, Y., Sun, R., Chen, X., Chu, H., Zeng, M., & Sun, C. (2016). Interleukin-11 promotes epithelial-mesenchymal transition in anaplastic thyroid carcinoma cells through PI3K/Akt/GSK3β signaling pathway activation. *Oncotarget*, *7*(37), 59652–59663. https://doi.org/10.18632/oncotarget.10831

Zhu, X., Zhong, J., Zhao, Z., Sheng, J., Wang, J., Liu, J., Cui, K., Chang, J., Zhao, H., & Wong, S. (2015). Epithelial derived CTGF promotes breast tumor progression via inducing EMT and collagen I fibers deposition. *Oncotarget*, *6*(28), 25320–25338. https://doi.org/10.18632/oncotarget.4659

9 Appendix

9.1 List of differentially expressed (log2 fold change +0.6) genes shared in Homo sapiens (human) datasets treated with TGFβ1, TGFβ2, TGFβ3

CXCL8 FBXO32 G0S2 CXCL1 FSTL3 BMP2 STC1 CLDN11 ASS1 TGFBR3 UNC5B RFLNB CDKN2B GPRC5B FN1 NEDD9 CEBPG COL12A1 SLC46A3 CLEC2B PTGS2 SLC43A3 GLS SLC1A4 COL4A1 PLIN2 DLC1 LAMB3 PHLDA1 GPRC5A TFPI2 RAB27B GADD45B SLIT2 CALD1 HMGA2 KCTD11 ID1 OPN3 IL11 TOP2A TPM1 DHRS2 MAMDC2 SLC7A5 OXTR SMAD3 LPXN PAWR SERPINB2 NUAK1 TP53I3 LAMC2 TSPAN2 TGM2 CHST11 AKAP12 CTSC SEMA7A SEC14L2 TUFT1 COL8A2 PCDH18 LTBP2 PDGFA CHAC1 SOCS2 PMEPA1 IQGAP3 C15of48 TJP2 FST FAM107B PRR15 O GFRL1 PLA2G4A IGFBP3 SPHK1 PTPRG STAMBPL1 HMGCS1 ENC1 PXDC1 TMEM97 CCND1 FZD7 IRAK2 FHL1 FERMT1 HPCAL1 CASP1 COL4A2 CTGF COL1A1 KITLG PLAU FGD4 SESN2 GOPC MAP3K5 AMIGO2 ERRFI1 HECW2 THBS1 TMEM200A CD44 SGPL1 GPAM CXCL3 MBOAT2 TFRC PDGFC JUNB TNFAIP3 SMAD7 SKIL PRKAB2 FADS3 SEL1L3 MIR181A2HG KCNK2 LBH TMEM117 WNT5B ATF3 SLCO2A1 PCDH7 CSF2 SOX4 ABLIM3 CORO6 PLCB4 MAPRE2 SARS ID3 TXNRD1 PTPN21 TBC1D4 MAP4K4 TMCC1 RAPH1 CCL20 KLF12 AKR1C3 BMPR2 IL1A DCBLD1 AQP3 ANTXR2 SPECC1 SLC8A1 PRKCA DHRS3 LOX CDH2 LAMP3 PPIF APBB2 LDLR FERMT2 C9orf3 LAYN ATXN1 COL5A2 OSR1 EDIL3 TNIP1 LRIG1 EXOC6 ANKLE2 ABHD17C HMGB2 INHBA PGM3 PAG1 HIVEP1 MICAL3 GLIPR2 GM2A PDLIM4 BAG2 VAV2 STK17B ST6GALNAC5 DPP4 CEMIP KLHL5 NCF2 SYNC NDRG1 CCDC50 MTSS1 IL15 GLIS2 ACSL1 FGD6 TGFBI LOC101928076 PPP1R13L ETS2 MLKL RTTN ADAM19 ZNF281 ETS1 MKNK2 CDH10 S1PR3 ADAMTS15 DUSP10 TFEB DLEU1 NR4A2 ANAPC13 SNX24 AGPS ITGAV NLRP1 LINC00842 CASC10 SLC25A37 JUN SLC41A2 ETV1 CTSB RBM24 CRYAB CITED2 MCC TEAD1 PPP4R4 FUCA1 DNAJB2 ARHGEF40 F2RL1 RNF19B SLC25A19 C1S SERPINE1 IFIT3 FOXC1 PDLIM7 HK2 GLMP ATF4 SLC6A6 RNF217 SPRY1 JDP2 KLF10 BHLHE40 F3 TTC39C VEGFC ETV6 TRERF1

9.2 List of differentially expressed (log2 fold change +0.6) genes shared in Mus musculus (mouse) datasets treated with TGFβ1, TGFβ2, TGFβ3

SERPINA1C LGALS4 APOC2 UGT2B34 GSTA4 GJA1 VNN1 IL11 KRT19 AW112010 PTGS2 RBP4 SERPINE1 VEGFC SERPINA1B TTR MUC1 FBLN2 ALDOC GCLC HPX ANXA10 TSPAN8 PDGFA CYP2S1 GLDC C4BP SLFN2 ITIH2 TNNT2 HABP2 PDLIM7 GADD45A ENTPD2 GJB3 CXCL5 FOXF2 SULT1D1 GADD45B LAMB3 ID2 BNC1 ALDH1A1 CTSW AK1 TBC1D1 ACVR1 RIDA CYSTM1 FGF1 GADD45G SGK1 VCAM1 EPHX1 APBB2 GK ERMP1 MSN RHOB ANGPT2 HSD17B2 RFLNB ABCC3 CSRP1

SPHK1 IFIT3 VIL1 C3 GSTA3 SELENBP1 FGFR2 S100G FKBP7 KLF10 ALB MAN1A CSRP2 TPBG GSTM1 MYO6 AKR1B7 RNF19B BMP1 SLC39A4 MYZAP MMP14 DUSP6 TMPRSS2 ACTN1 PDK1 PLIN2 CP KITL FOXQ1 ACKR3 FOXA2 MT2 S100A13 ACSL4 PRSS23 ADSSL1 CTLA2B ANXA8 BHLHE40 SCX NCAM1 SKIL MT1 HAL CAPG QK SPRR1A ID3 RAB11FIP5 CNN3 ADORA1 CXXC5 FZD2 LCP1 CLRN3 PFKM PDGFB LTBP1 CLDN4 CDH16 SMAGP HS3ST1 HELLS SLCO3A1 SMC2 DDHD1 FHOD3 GGCX RHPN2 RHOQ SHMT1 CXCL1 PFKL TNFRSF1B NEDD9 DUSP1 GM3716 E2F8 CREG1 IQGAP2 SPSB1 GJB4 ENTPD5 PCDH7 PDE7A PITPNC1 SKI PCGF5 LGALS3BP CACNA1G RAB17 CAST AFM DHRS3 SLC46A1 RUNX1 BNIP3 ITGB5 NGEF PKP2 COL7A1 PNKD CTS ABCA1 MYO10 MYO7A WIPF1 SLC26A2 NEDD4L CDH6 DTL SLC2A2 NEU1 HSPA4L CYBB KLF9 CTSF JUP SH3BP2 GSN HHEX TRAF3 PRNP KPNA3 RGL1 FN1 PIP4K2A LRRC8C TNFSF9 TSC22D1 HNF4A HSD3B7 ENO2 ARL4A ABCC2 HSPA2 TGFBR1 CCL7 HIVEP2 JAK2 PLEKHO1 IFI204 MAPT CHRNA1 TIMP1 KLF5 SNAP47 IFI202B EHF EFNA1 UBE2F PELI1 FGD1 SEMA3E CCDC88A DBN1 PKHD1 NREP SORBS1 CSF2 ZAN CTNND1 MAP7D1 PDXK LPGAT1 PACRG LPIN1 SCEL NRIP1 ARL2 MAP3K5 TPGS2 SEMA3C MTMR10 MARCKS FXYD5 SNCA SLC12A2 ELL2 MARCKSL1 KIF1B TRIM59 CASP3 AQP1 SPA17 EPB41L2 C1RA CAMK2D ROBO1 ME1 CREBRF PKIA CDC7 RLIM ALCAM ITPR2 NFIX CELF2

9.3 List of upregulated (log2 fold change +0.6) genes shared in Homo sapiens (human) datasets treated with TGFβ1, TGFβ2, TGFβ3

DHRS2 LTBP2 FN1 CALD1 HECW2 SGPL1 BPGM CRYAB LINC00842 PMEPA1 NEDD9 FBXO32 TCP11L1 KLF12 NLRP1 PDLIM7 ID3 THBS1 TGFBR1 FNBP1L UBASH3B PTPN21 JDP2 BMPR2 COL1A1 ITGAV DNAJB5 PDGFC PDLIM4 ANKLE2 ABHD4 C9orf3 TEAD1 COL4A2 SLC41A2 TPM1 KCNMB4 ST6GALNAC5 C14orf28 SLC1A4 TEPP PTHLH MICAL3 MBOAT2 INHBA SYNC FADS3 GOPC TGM2 TMCC1 MICALCL IGFBP3 MRAP2DCBLD1 WNT5B EDIL3 PRKAB2 COL5A2 VAV2 PHLDB2 FGD6 COL12A1 PXDC1 IL11 IL15 SESN2 RAPH1 STK17B ENC1 SEC14L2 LBH KLF10 PRKCA JUNB AMIGO2 SKIL CTGF SMAD7 ATF3 SPECC1 PAWR GLS SEMA7A ID1 GADD45B HIVEP1 SPHK1 SARS FERMT1 CORO6 CST6 MIR181A2HG COL4A1 PDGFA NUAK1 CEBPG DLC1 LAMC2 TP53I3 ZNF365 OPN3 KCTD11 CASC10 CHST11 INAVA PLCB4 RFLNB SLC46A3 CHAC1 TUFT1 TSPAN2 LOC101928076 FSTL3 C15orf48 MAMDC2 CDKN2B SLCO2A1 MFAP2 CDC42EP3 VMP1 ARHGEF40 FILIP1L GADD45 APBB PHACTR4 TNC LIMA1 ATF4 CSPG4 KIF3C KDM7A BAIAP2L1 ETV6 PLEK2 LOX GTPBP2 BMT2 TGFBI AKT3 BHLHE40 CNN3 CHST15 MN1 CDA JAZF1 RUNX1-IT1 SH3RF1 GLIS2 FOXC1 ZNF281LFN SMURF1 RNF27 SPIRE1 SNX24 ADAM19 TINAGL1 FLNB AGPS PPP1R13L RNF19B PGM3FERMT2 CEMIP2 ITPRID2 GLIPR RBM24 IGF1R MYL12A BLOC1S2 CDH2 DIXDC1 CHML TFEB SERPINE1 ATXN1 HTR1D NT5E

9.4 List of upregulated (log2 fold change +0.6) genes shared in Mus musculus (mouse) datasets treated with TGFβ1, TGFβ2, TGFβ3

SERPINE1 GJA1 FBLN2CTLA2B VEGFC SPSB1 Q GADD45A AK1 ACKR3 LAMB3 APBB2 GJB3 GLDC TNNT2 PDGFA CAPG GADD45B FOXF2 MSN RHOB PDLIM7 RFLNB FHOD3 SGK1 NCAM1 BNC1 ACVR1 PDE7A LTBP1 TBC1D1 RHOQ MYZAP LRRC8C CTSW PKHD1 CTSE FKBP CCL7 SKI CLDN4 NRIP1 ACTN1 CSRP1 CXXC5 KLF10 TPBG CAST PITPNC1 SKIL BMP1 RNF19B CCDC88A ADSSL1 RAB11FIP5 CNN3 SCX PLIN2 CDH6 NEDD9 FZD2 ANXA8 COL7A1 SH3BP2 PDGFB RUNX1 MYO10 FOXQ1 MYO7A PKP2 CACNA1G GJB4 TNFRSF1B DDHD1 CHRNA1 JAK2 SLC46A1 FXYD5 ITGB5 TSC22D1 DBN1 EPB41L2 WIPF1 KPNA3 TRAF3 PIP4K2A PRSS23 BHLHE40 SPHK1 PCDH7 MMP14 TGFBR1 SLCO3A1 CSF2 GADD45G IL11 CTNND1 IFI202B PLEKHO1 KLF9 FN1 GSN HIVEP2 IFI204 SNCA PRNP LPGAT1 ELL2 TIMP1 SNAP47 MARCKS FGD1 UBE2F HSD3B7 KIF1B MAP7D1 MARCKSL1 NFIX NREP ROBO1 TPGS2 TRIM59 PKIA RLIM ARL2

9.5 List of upregulated (log2 fold change +0.6) genes shared in cancer cell line datasets treated with TGFβ1, TGFβ2, TGFβ3

SERPINE1 IL11 LOX COL1A1 PTHLH TAGLN SRPX GLIPR1 THBS1 COL27A1 NPTXR CRLF1 PMEPA1 NAV1 TOX3 CALD1 COL5A2 ITGB3 JUN FN1 FAM43A GXYLT2 PIK3CD SCG2 COL7A1 FLRT3 GLIPR2 FRMD6 ST20-AS1 SPHK1 CTGF CSRNP1 HBEGF PXDC1 SLC2A12 MIR181A2HG AK4 SKIL EPHB1 CCDC50 LMCD1 C3orf80 TPM1 BMPR2 KDM7A STC1 GLS SH3KBP1 CD59 PLEK2 ADAM19 KLHL24 P4HA3 CHST3 PLAUR DOCK4 SERPINE2 OSBPL6 HRH1 LGR4 HMGA2 ATP8B1 AGR2 GRTP1 ARHGAP29 MYO5C PPL CDH1 IFIT3 EMP1 RGCC FAM84B ERMP1 TGM2 PARD6B AQP3 GSE1 THRB RAB27B SYNE2 ITGA2 SLC44A1 NAV2 SERPINB1 PPP1R14C CHDH SLC6A6 GRB10 TNFAIP2 PHLPP1 MXD1 ARHGAP26 SLC20A1 C1orf115 PIK3AP1 PHLDA1 DEPTOR