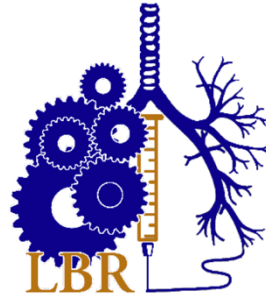




Universidad
Francisco de Vitoria
UFV Madrid

Facultad de Ciencias experimentales
Universidad Francisco de Vitoria



LUNDS
UNIVERSITET

IDENTIFICATION OF YAP-TAZ TRANSCRIPTION FACTOR MOTIF REGIONS IN PRIMARY MOUSE ALVEOLAR EPITHELIAL CELLS

Final Thesis

Irene Urrecho Casla

Biomedicine

Universidad Francisco de Vitoria

Wagner Lab Lung Bioengineering and Regeneration

Supervisors: Hani N. Alsafadi, PhD. Darcy E. Wagner, PhD. Renata Kelly da Palma

July 2022

Model a

Acknowledgments

I would like to thank all the people at the Lung Bioengineering Regeneration center who have helped in this project in any shape and form. Especially to Hani Alsafadi, PhD. student, and Darcy Wagner, PhD. for training and guiding me through the whole process. Inés Mínguez for being my lab partner but most of all for becoming a great friend. To all the students in the lab for supporting me, welcoming me and achieving a great working and training environment.

To my new friends who have become my Swedish family Alejandro Frias, Alejandro Forga, Monica Sartorius and Rick Jobse. Of course, to my family and friends for being my unconditional support, remarkably Vasco Oliveira who has encouraged me during the realization of this project.

Finally, I would like to thank the Universidad Francisco de Vitoria for providing me with the theoretical and practical knowledge necessary for this experience to take place.

ABBREVIATIONS

IPF: Idiopathic Pulmonary Fibrosis

FDA: The Food and Drug Administration

ATI: Alveolar Epithelial Type I

ATII: Alveolar Epithelial Type II

TGF- β : Transforming Growth Factor- β

CTGF: Connective Tissue Growth Factor

FGF: Fibroblast Growth Factor

PDGF: Platelet-Derived Growth Factor

ECM: Extracellular Matrix

Hh: Hedgehog

YAP: YES-associated Protein

TAZ: PDZ-binding Protein

MST1/2: Mammalian Ste20-like Kinases

LATS1/2: Large Tumor Suppressor

TEAD: Transcriptional Enhanced Associate Domain

CUT&RUN: Cleavage Under Targets & Release Using Nuclease

pA-MN: Protein A-nuclease Micrococcal

ChIP: Chromatin Immunoprecipitation

RT: Room temperature

E. coli: *Escherichia Coli*

IGV: Integrative Genomics Viewer

MACS2: Model-based Analysis of Chip-seq

TSS: Transcriptional Start Site

SEACR: The Sparse Enrichment Analysis

DAVID: Database for Annotation, Visualization and Integrated Discovery

AF: Focal Adhesions

GSEA: Gene Set Enrichment Analysis

INDEX

1. ABSTRACT	8
2. INTRODUCTION	9
2.1. Idiopathic pulmonary fibrosis	9
2.2. Lung anatomy and epithelium	9
2.3. Initiation and physiopathology of IPF	9
2.4. Signaling Pathways involved in Idiopathic pulmonary fibrosis	11
2.5. Hippo pathway	12
2.6. Yap/Taz function as regulators of gene expression	13
2.7. The role of Yap Taz in idiopathic pulmonary fibrosis	13
2.9. Hypothesis and objectives	15
3. MATERIALS AND METHODS	16
3.1. Summary of the project action plan	16
3.2. Isolation of mouse alveolar type 2 cells	16
3.3. CUT&RUN experiment	18
3.4. Pipeline for CUT&RUN processing	20
3.4.1. Raw read trimming and alignment	20
3.4.2. Sorting and generation of bw files.....	21
3.4.3. Peak calling and Heat map plots generation	21
3.4.4. IGV analysis of Yap/Taz known target genes	21
3.4.5. Annotation of the gene list	22
3.4.6. Functional annotation analysis.....	22
4. RESULTS	23
4.1. Effective ATII cell isolation and CUT&RUN experiment	23
4.2. DNA alignment results and data normalization	23
4.3. CUT&RUN effectively detects Yap and Taz enrichment near gene transcriptional start sites	24

4.4. Integrative Genomics Viewer confirms the presence in our experimental samples of Yap/Taz well-known target genes.....26

4.5. Gene Annotation and Overlap between Yap and Yap-Taz samples.....27

4.6. Detection of new potential target genes of Yap by STRING28

5. DISCUSSION..... 30

7. ANNEXES..... 36

1. ABSTRACT

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive disease characterized by the formation of fibrotic tissue in the distal part of the lung. In this epithelium-driven disease alveolar epithelial cells type 2 (ATII2) become dysfunctional and promote this fibrotic environment. Yap/Taz are the main mediators in the Hippo signaling; this developmental pathway, among others, is reactivated in IPF and is now being studied as a potential target for drug development. These coactivators bind to specific transcriptional factors (TFs) to regulate the expression of genes associated with proliferation, cell survival, and regeneration.

In the present work, the main objective has been to characterize Yap/Taz transcriptional motifs regions in order to discern the genetic profile regulated by these coactivators.

To achieve this goal, we performed Cleavage Under Targets & Release Using Nuclease (CUT&RUN) assay in primary mouse alveolar type II cells to isolate DNA fragments bound to Yap-Taz+TF complexes followed by a bioinformatic analysis to detect the enrichment of the reads around the Transcription Start Site (TSS) and to carry out the annotation of the detected genes. In addition, we performed functional analysis of our samples using the DAVID and STRING databases to describe the gene's main functions and to find out whether their interaction with Yap was already described in the literature.

We found enrichment of a large number of genes in the TSS and the detection of major target genes in the case of the Yap and Yap/Taz experimental conditions. Moreover, annotation of DNA fragments together with functional analysis identified the presence of extracellular matrix, cell membrane, and cell junction regulatory genes. Finally, by studying the interactions of the genes detected with Yap, new target genes not currently associated with this coactivator were identified. In conclusion, obtaining the Yap/Taz transcriptional motifs in ATII cells validates the use of this CUT&RUN technique for this purpose and will allow in the next steps to predict the transcription factors that bind to these coactivators. Furthermore, the detection of genes with unknown interactions with Yap in healthy ATII cells could support the idea that Yap/Taz are context-dependent coactivators and could be linked to disease-specific transcription factors that can be used as targets.

2. INTRODUCTION

2.1. Idiopathic pulmonary fibrosis

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial disease in which there is excessive accumulation of extracellular matrix (ECM) in the lung parenchyma resulting in scar tissue formation in the distal part of the lung. The progressive thickening of the interstitium and impairment of gas exchange make the disease have significant morbidity and a poor prognosis. IPF is a disorder of unknown cause, but its occurrence has been associated with a combination of genetic and environmental risk factors. Increasing evidence suggests that the pathology involves an aberrant response of the bronchial and bronchiolar epithelium, which contributes significantly to the development and progression of the disease (1). IPF affects 3 million people worldwide and its incidence increases dramatically with age. Although there are two FDA-approved drugs for the treatment of the disease: nintedanib and pirfenidone, after diagnosis, patients have a life expectancy of only 2-5 years. The problem is that these drugs can only slow the progression, so there are currently no treatments that can cure or reverse this disorder (2,3).

2.2. Lung anatomy and epithelium

The lungs are organs made up of a system of tubes with a large number of branches that end in vascularized sacs. Their main function is to conduct air to the alveoli so that gas exchange can take place. Maintenance of the pulmonary epithelium is achieved by the stem and progenitor cells in that region that replace the specialized cells. The airway epithelium is composed primarily of multiciliated cells and secretory cells, such as Club cells and goblet cells. Their function is to produce mucins and glycoproteins that form a thin surface layer of fluid to moisten the air and provide antimicrobial activity. In larger airways such as most of the human lung, the so-called mucociliary epithelium contains basal cells that function as progenitors of these cell populations. By contrast, the air sacs are lined primarily by two distinct cell types: specialized type II alveolar cells (ATII) that are responsible for producing surfactant and type I cells (ATI) that provide a large surface area necessary for gas exchange. In the alveoli, ATII cells are the predominant epithelial progenitor, as they are capable of long-term self-renewal and multipotent differentiation to give rise to alveolar type I (ATI) cells (4,5).

2.3. Initiation and physiopathology of IPF

The cause of IPF is not yet known, but research findings indicate that the alveolar epithelium plays an important role in the physiology of the disease. Genetic risk factors for the occurrence of the disease implicates lung epithelium such as mutations in genes restricted to that area in familiar

forms of the disease or a specific polymorphism in MUC5B gen that is found in 2/3 of fibrosis patients. In addition, environmental exposures such as smoking, respiratory virus infections, or exposure to inhalants such as asbestos have been identified as risk factors for pulmonary fibrosis and possible recurrent sources of damage to the alveolar epithelium (3,6,7). The sum of genetic susceptibility, biological processes underlying aging, and exposure of distal lung tissue to environmental risk factors leads to loss of epithelial stem cells or generation of dysfunctional AT2. These damaged alveolar epithelial cells trigger defective regeneration and pathogenic activation of myfibroblasts by promoting a fibrotic environment. This imbalance of fibrotic mediators is due to the secretion by the surrounding cells of multiple growth factors and cytokines such as transforming growth factor- β (TGF- β), connective tissue growth factor (CTGF), fibroblast growth factor (FGF), platelet-derived growth factor (PDGF) and endothelin. Therefore, this elevated number of myfibroblasts in the tissue secretes exaggerated amounts of extracellular matrix (ECM) which accumulates in the interstitium and remodels the pulmonary architecture. Causing compromise of gas exchange, progressive decline in lung function, and ultimately death (8).

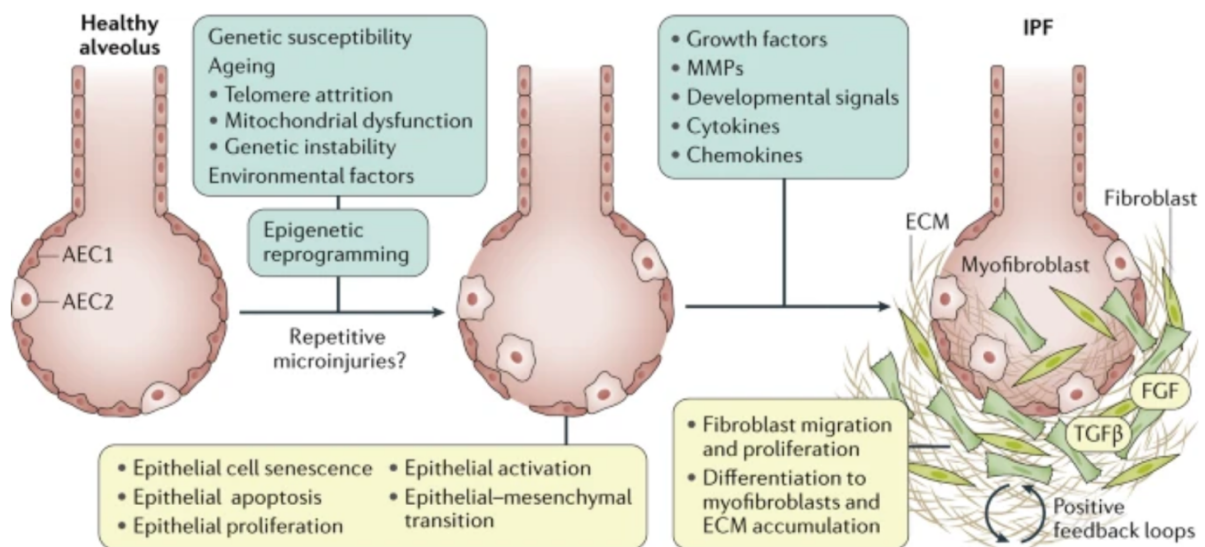


Figure 1. Representation of the pathophysiological mechanism behind the development of Idiopathic Pulmonary Fibrosis. The two main cells that conform the alveolar epithelium of a healthy lung, ATI and ATII, are observed together with the causes of the loss and generation of dysfunctional ATII (genetic susceptibility of the patient as well as the exposure to environmental factors and the biological processes behind aging). The different fibrotic mediators released by this damaged epithelium and the processes they promote such as fibroblast proliferation and differentiation are listed. Finally, it shows an alveolus in IPF situation where the excess of myfibroblasts in the tissue and the accumulation of extracellular matrix are observed. Source:(5)

2.4. Signaling Pathways involved in Idiopathic pulmonary fibrosis

Cell signaling is a multifactorial system that cells use to transfer information from first messengers through different receptors, these signals are then decoded by second messenger signaling intermediates. Signal transduction, as part of cell signaling, describes how cells detect, interpret and react to external and internal stimuli (9). As in many pathological processes, in IPF there is an activation of developmental pathways that should have remained dormant in adults such as TGF- β , Hh, Notch, Wnt, and Hippo (3,5,10). Hh is one of the main cascades responsible for intercellular communication during embryonic development and organogenesis. In IPF Hh signaling dysregulation can promote fibroblast to myofibroblast transition (9). TGF- β modulates several cellular functions, such as proliferation, differentiation, and apoptosis, and is also known to play a role in lung development and repair. TGF- β acts as the main mediator of fibrosis so it is found in excess in many chronic lung diseases including IPF. WNTs are mainly associated with the formation and polarity of the primary body axis, but they act as growth factors in both homeostasis and disease. Aberrant activation of WNT signaling is enough to trigger the expression of a fibrogenic program in fibroblasts (11). Notch signaling is essential for the normal development of multiple organs, including the lung, and its dysregulation has been detected in tissue fibrosis and cancer. The Hippo pathway controls epithelial progenitor cell proliferation, migration, and differentiation in the developing and mature lungs (12). In this project, we will focus on the study of the Hippo signaling since previous studies in Wagner lab show that this pathway seems to be deranged in idiopathic pulmonary fibrosis and as shown in (Figure 2) the main mediators of the pathway Yap/Taz coactivators are enriched in the patient's tissue.

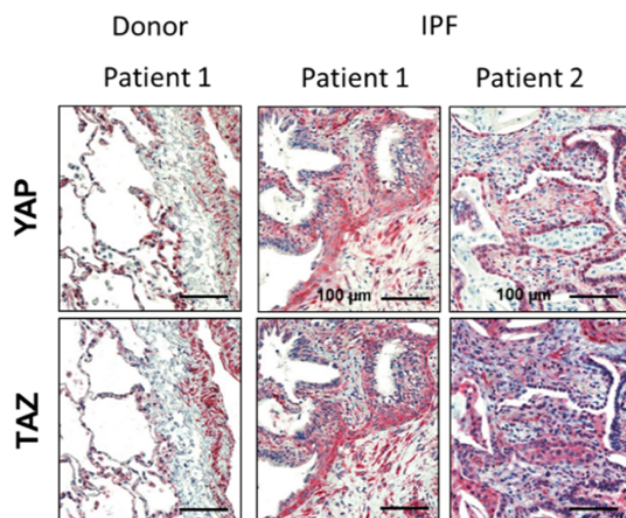


Figure 2. Yap/Taz Immunohistochemistry from healthy and IPF lung tissue. Overexpression of red-labeled Yap/Taz is observed in the lung tissue of patients with IPF compared to control sample. Source: Wagner DE, et al. In preparation

2.5. Hippo pathway

This signaling cascade is essential for the control of organ size, tissue homeostasis, cell proliferation, and tumorigenesis. Its main effectors are YES-associated protein (Yap) and PDZ-binding protein (Taz) whose activation is dependent on their cellular location. When the Hippo pathway is activated a set of kinases (MST1/2 and LATS1/2) phosphorylate Yap/Taz retaining them in the cytoplasm or leading them to degradation. When Hippo is inactivated, phosphorylation does not occur, and these proteins can reach the nucleus and become functional (13). This nuclear translocation of Yap and Taz depends on many extracellular signaling and mechanical stimuli (cellular stress, adhesion, polarity) (14). Yap/Taz are transcriptional coactivators which means that they are not able to bind directly to DNA, but instead, Yap/Taz interact with transcription factors that do have DNA binding sites to regulate gene expression (13). It has been observed that Yap/Taz bind mainly to TEAD-family transcription factors these well-characterized binding partners regulate in physiological conditions cell proliferation, growth, and survival. Nevertheless, they are also known to be able to bind to many other transcriptional factors such as p73, ERBB4, and SMADS (15). Although Yap and Taz are activated in the same manner and share many binding partners, they can also bind different transcription factors. This suggests that while they share some biological functions this overlap is partial.

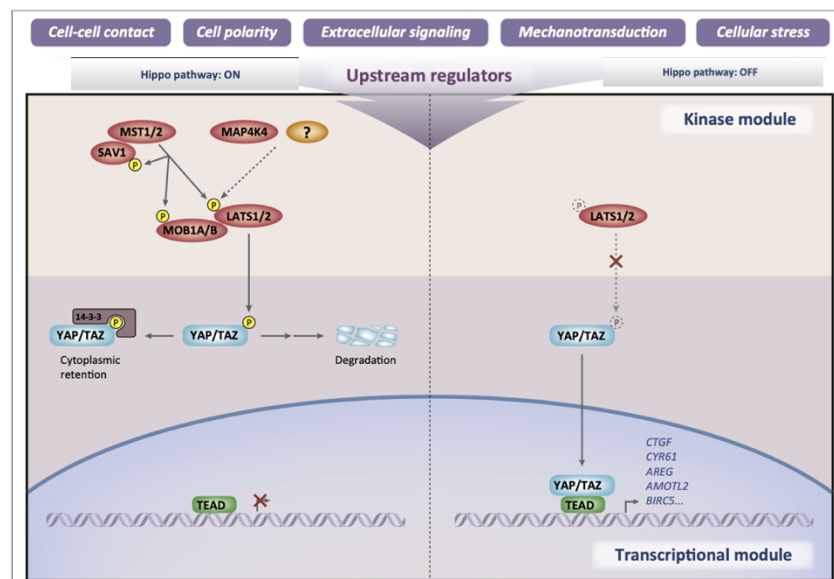


Figure 3. Schematic of the Hippo signaling pathway. The Hippo pathway is composed of a two-kinase module in which MST1 and MST2 with their binding partner Sav1, phosphorylate and activate LATS1 and LATS2. Activated LATS proteins, in cooperation with the co-factor MOB-1, directly phosphorylate Yap and Taz inhibiting their activity through cytoplasmic retention and protein degradation. Only when the pathway is inactivated, hypophosphorylated Yap/Taz can translocate to the nucleus and regulate the expression of their target genes. Source: (13).

2.6. Yap/Taz function as regulators of gene expression

Yap/Taz respond to signals from soluble factors, metabolic pathways, and mechanical signals to control cellular processes such as proliferation, cell migration, cellular plasticity, and regenerative capacity. Accumulating evidence indicates that Yap/Taz promotes proliferation by regulating genes that encode for components involved in cell cycle progression, mitosis, and cell growth. In addition, they stimulate cell survival by up-regulating the expression of several inhibitors of apoptosis. Yap/Taz coactivators also activate transcriptional programs important in the regulation of stem cell potency by promoting target genes such as NANOG, OCT4, MYC, and various components of Polycomb139 group proteins. CYR61 and CTGF, for instance, are classic Yap/Taz target genes and regulate diverse biological processes, such as cell migration, proliferation, and cell adhesion (12,16,17).

2.7. The role of Yap Taz in idiopathic pulmonary fibrosis

Most of the information that is known about the role of these transcriptional coactivators in this disease is about their involvement in fibroblast activation, ECM synthesis, and profibrotic factor expression (18). It has been confirmed that nuclear levels of Yap/Taz are particularly elevated in fibroblasts from IPF patients, and that decreasing Yap/Taz expression could lead to attenuation of matrix synthesis, contraction, proliferation, and differentiation to myofibroblasts (19).

Increasing evidence suggests that the lung epithelium is not only injured in fibrosis, but also have an active role in contributing to disease pathogenesis. More specifically, the group of Dr. Wagner found that the Hippo signaling is deranged in the lung epithelial cells of the mouse model of fibrosis and the human disease. Yap/Taz-controlled target genes are responsible for the proliferation and differentiation of epithelial progenitor cells in the lungs(13). In previous studies, it has been observed that features of other airway cells such as basal and goblet cells have been identified in epithelial cells from idiopathic pulmonary fibrosis patients and that these had aberrant overactivation of Yap/Taz, TGF- β , Wnt, and PI3K signaling cascades (19).

Gene set enrichment analysis (GSEA) of genes upregulated in samples of IPF patients has shown enrichment with genes regulated by Yap/Taz (Figure 4a). This enrichment was not evident for genes targeted by the TEAD transcription factors (Figure 4b). Therefore, Yap/Taz could potentially be binding to transcription factors other than TEAD in this specific pathological situation that could become a therapeutic target. (Unpublished data Wagner lab).

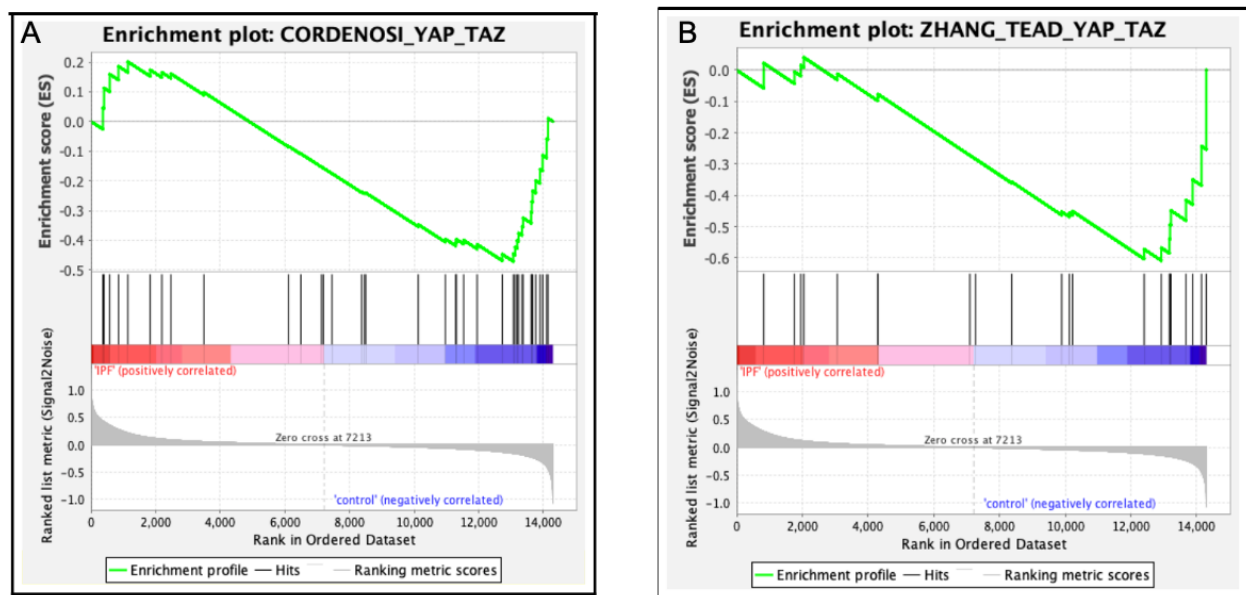


Figure 4. GSEA analysis of Yap/Taz target genes in ATII cells from patients with IPF. (A) Overenrichment of Yap/Taz target genes can be observed in the IPF condition. (B) The genes controlled by the Yap/Taz+Tead factors complex are not found to be over-activated in the pathological condition. Therefore Yap/Taz Hyperactivity is not accompanied by overenrichment of TEAD target genes. source: Wagner DE, et al. In preparation

2.8. Study of protein and DNA interaction

To map these transcription factors in the genome, Cleavage Under Targets & Release Using Nuclease (CUT&RUN) is used as a first approximation. This experimental method like ChIP-seq is one of the techniques used to study the interaction between DNA and proteins that interact with it such as transcriptional factors, activators, inhibitors, etc. In CUT&RUN, an antibody targeting the DNA-binding protein of interest is used to recruit a recombinant protein A micrococcal nuclease (pA-MN). This is an endonuclease capable of cleaving DNA fragments close to the antibody binding sites (20). This epigenomic profiling assay ensures that only the target fragments are solubilized while most of the DNA remains inside the cell. This allows on the one hand to obtain such low background levels and on the other hand, avoids having to make a size exclusion after DNA purification. In addition, this technology requires much less cell input, and costs are lower. CUT&RUN outperforms the most widely-used Chromatin Immunoprecipitation (ChIP) protocols in resolution, signal-to-noise, and depth of sequencing required (21).

2.9. Hypothesis and objectives

Based on the fact that Yap/Taz are context-dependent transcriptional coactivators, and that they are known to be dysregulated in lung epithelium. This study hypothesizes that Yap/Taz are involved in the promotion of fibrosis through transcription factors other than TEAD, the main binding partner in physiological conditions. Therefore, if there are transcription factors involved exclusively in a disease context, the investigation and identification of these binding partners would be very significant as they, for instance, could become therapeutic targets. To get to this point it would be necessary to compare the transcription factors to which Yap/Taz binds in a physiological situation with those that bind in a situation of idiopathic pulmonary fibrosis.

The main objective of this thesis is the identification of Yap/Taz transcriptional motifs in order to discern the genetic profile regulated by these coactivators in healthy lung epithelium. To achieve this the following specific objectives are proposed.

- Effectively isolate type 2 alveolar cells from a healthy mouse lung.
- Perform CUT&RUN in isolated ATII for sequencing DNA bound to complexes containing Yap/Taz.
- Develop and apply an effective bioinformatic pipeline for this and future CUT&RUN analysis.
- Identify Yap/Taz transcription factor motifs.
- Analyze the genetic profile regulated by Yap/Taz.

3. MATERIALS AND METHODS

3.1. Summary of the project action plan

The steps followed to carry out the project are shown in (Figure 5). The cells selected for genetic analysis were primary alveolar type 2 cells. Once isolated from the mouse lung, they were divided into 5 conditions, and CUT&RUN was performed on each of them. The experimental design consists of a control - sample, where the antibody used was a rabbit IgG (no binding expected), control + with an antibody that recognized H3K4me3 (binds to all genes with active transcription), and three experimental samples with antibodies for Yap, another for Taz, and another sample with antibody that recognizes both Yap and Taz simultaneously. Eluted DNA fragments obtained from CUT&RUN were sent for sequencing. Finally, the data received was processed to confirm the presence of genes, and subsequently, each gene in the generated dataset was identified to study its interactions with our target coactivators.

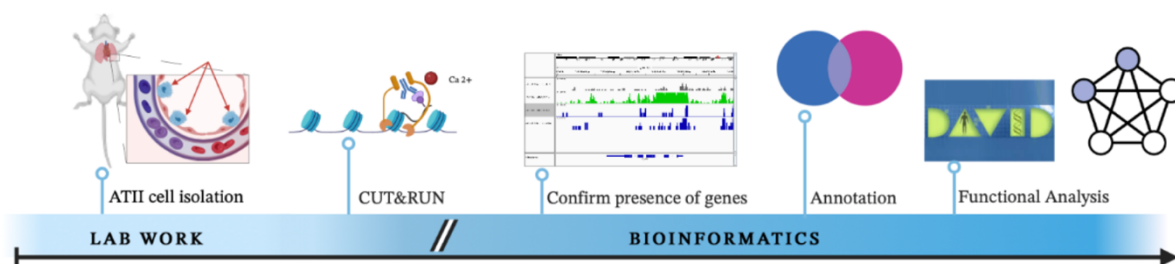


Figure 5. Overview project plan. Each experimental and analytical step is shown in the timeline. Source: manually created by BioRender

3.2. Isolation of mouse alveolar type 2 cells

Preparation of mouse lungs and initiation of epithelial cells isolation: After the dissection of the mouse, the right ventricle of the heart was perfused with PBS to remove blood from the lungs. The lung was then explanted and embedded into a device called 3DLD (22) that allowed the introduction of the dispase (Fisher Scientific, 11553550) into the lobes through the trachea with the use of a syringe. This device separated the lung lobes from the trachea and then the separated lobes were incubated in dispase solution for 45 minutes at room temperature. Up to this point the experimental protocol was carried out in the animal facility and I could only be an observer of the process. Next, the lungs were minced with the help of forceps in a Plus medium that contains DNase I (Table 1), and the resulting cell suspension was passed through a series of nylon filters gradually (100 μ m, 40 μ m, and 10 μ m) to generate a final filtrate. This filtrate was centrifuged, and the cell pellet was resuspended in Minus medium (Table 1) (Figure 6a).

Removal of fibroblast and macrophages: the cell suspension was seeded into 10 cm Petri dishes and incubated at 37°C for 30 minutes. After that, the fraction not adhering to the plate was collected and centrifuged at 800g for 10 minutes, and the final cell pellet was resuspended in MACs buffer (Table 1) up to a total volume of 15 ml (Figure 6b).

Removal of CD31+ CD45+ cells: Cells were counted using the Neubauer chamber and centrifuged at 800g for 10 minutes. For the preparation of Microbeads solution per 10 million cells (90ul of MACs buffer + 10ul of CD31 Microbeads + 10ul of CD45 Microbeads) were needed. The cell pellet obtained was resuspended in this Microbeads solution and incubated at 4°C for 15 minutes. The cell suspension was diluted in MACs buffer (15 mL) and then centrifuged at 800g for 10 minutes resuspending the pellet in MACs buffer (2 ml per lung). MACs separator was used to place the columns after being washed 3 times with MACs buffer. The cell suspension was then carefully pipetted into each column and washed again with MACs buffer (10 ml). ATII cells were collected.

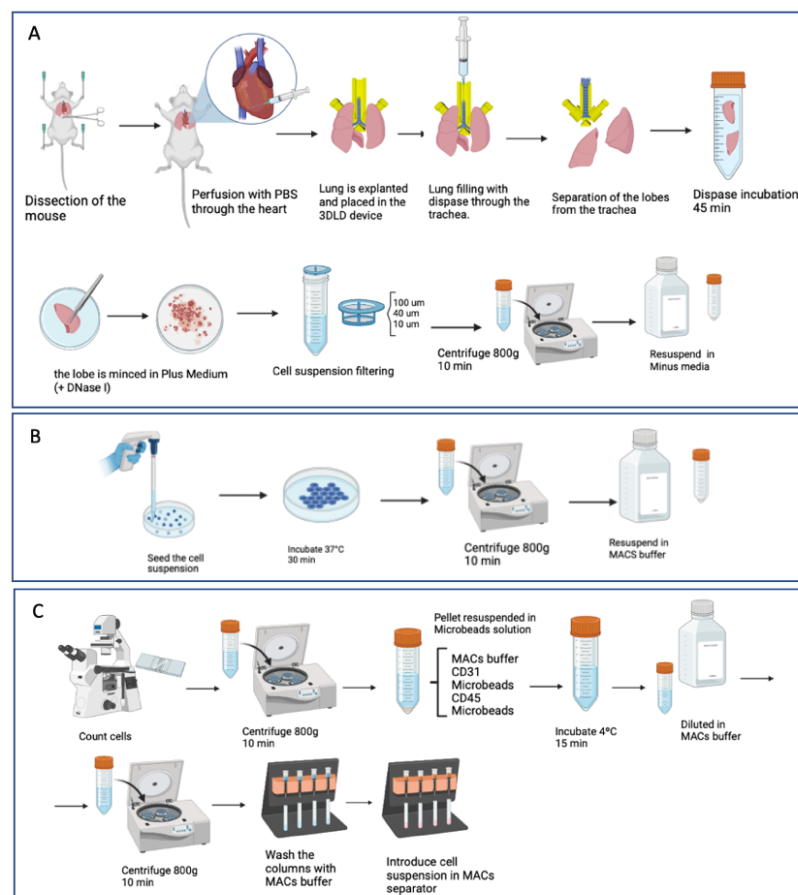


Figure 6. Graphical representation of the ATII cell isolation protocol. A) Tissue extraction and first steps of the cell isolation, The 3DL D device used has been developed by Hani Alsafadi. B) Removal of the macrophages and fibroblast. C) Removal of the CD31+ y CD45+ cells. Source: manually created by BioRender

Table 1. Recipe of the buffers used in the ATII cell isolation

Solution	Components	Stock	Final Concentration	Volume/mass to use
Minus Medium	DMEM (11554546, Gibco)	NA	NA	500 ml
	Glucose (1413411.1211, Saveen & Werner)	NA	3.6 mg/ml	1.8 g
	GlutaMAX (13462629, Gibco)	1X	2%	10 ml
	Pen/Strep(15140122, Thermo Fisher)	100U/ml	1U/ml	5 ml
	HEPES (11464805, Fisher Scientific)	1M	10 mM	5 ml
Plus Medium	Minus Medium	NA	NA	250 ml
	DNASE I (A3778.0050, VWR international AB)	NA	0.04 mg/ml	10 mg
MACs buffer	MACS rising solution (130-091-222, Miltenyi Biotec)	NA		237.5 ml
	MACS BSA (130-091-376, Miltenyi Biotec)	100%	5%	12.6 ml

3.3. CUT&RUN experiment

CUT&RUN experiments were performed following the EpiCypher CUTANA protocol (annex). Briefly, the Concanavalin A-coated magnetic beads (11 ul) were activated with 100 ul of bead activation buffer (Table 2), and the ATII cells were rinsed three times each with 100 ul /sample of wash buffer (Table 2) followed by 3 min 600x g spin at RT. The cells bound to the ConA beads were incubated with the primary antibody overnight at 4°C. The antibodies used were the following; negative control IgG antibody (EpiCypher 13-0042), positive control antibody (H3K4me3: EpiCypher 13-0041, Yap antibody (Yap antibody cell signaling D8H1X), Taz antibody (Anti-TAZ antibody Abcam ab84927), Yap-Taz antibody (Yap/Taz antibody cell signaling D24E4). The next day permeabilized cells retained on the beads were washed three times with 200 ul of cell permeabilization buffer (Table 2) and then incubated with 2,5 ul of pAG-MNase for 10 min at RT. The slurry was washed again with cell permeabilization buffer and the Antibody-associated protein-DNA complex was released from chromatin by adding 1 µL 100 mM CaCl₂ for 2 hours at 4 degrees. To stop MNase enzymatic activity by chelating Ca²⁺ ions 33 ul of Stop buffer (Table 2) was added and to perform the DNA normalization in the afterward analysis 0.5 µL *E coli* Spike-in DNA to each sample. The DNA fragments were released by 10 min incubation at 37°C in a thermocycler and the supernatant containing the CUT&RUN-enriched DNA was collected and transferred to new 1.5 ml tubes to perform DNA purification.

DNA purification: First, 420 μ l of binding buffer (CUTANA DNA purification kit) was added to each sample and loaded onto a DNA column in a collecting tube. A centrifugation step was performed at 16000x g, 30 sec, RT, and the flow-through was discarded. It was then necessary to wash the column 3 times by adding 200 μ l of DNA washer buffer and performing centrifugations at 16,000x g, 30 sec, RT, the supernatant was discarded each time and additional centrifugation was performed to dry the column. Later, the column was transferred to new collecting tubes and for DNA elution, 12 μ l of DNA elution buffer (CUTANA DNA purification kit) was added and then let stand for 5 minutes. Finally, we centrifuged at 16,000x g, 1 min, RT and kept the flowthrough.

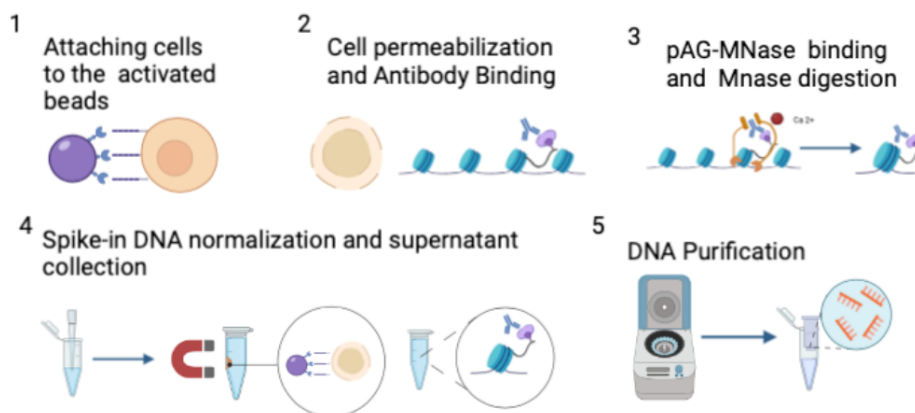


Figure 7. CUT&RUN protocol overview. Cells are attached to Concanavalin A-magnetic beads. Subsequent cell permeabilization, antibody binding, and targeted pAG-MNase digestion are performed on these cells immobilized on the beads. The magnetic beads are captured each time the buffers are changed. Source: manually created by BioRender

Table 2. Buffer recipes for CUT&RUN

Bead Activation Buffer	20 mM HEPES, pH 7.9
	10 mM KCl
	1 mM CaCl ₂
	1mM MnCl ₂
Pre-Wash Buffer	20 mM HEPES, pH 7.5
	150 mM NaCl
Wash Buffer	Pre-Wash Buffer
	25X Protease inhibitor
	1M Spermedine

Cell Permeabilization Buffer	Wash Buffer
	5% digitonin
Antibody Buffer	Cell permeabilization Buffer
	0.5 M EDTA
Stop Buffer	340 mM NaCl
	20 mM EDTA
	4 mM EGTA
	50 μ g/mL RNase A
	50 μ g/mL Glycogen

3.4. Pipeline for CUT&RUN processing

The summary workflow of the analysis of the CUT&RUN experiments that were carried out consists of taking as input FASTQ files of pairwise sequencing reads and performing a set of analytical steps: trimming of adapter sequences, alignment with the reference genome, peak calling, estimation of the cut matrix at single-nucleotide resolution, IGV analysis, and data visualization.

3.4.1. Raw read trimming and alignment

Two rounds of trimming were carried out to remove the adapter sequences needed in sequencing. The Trimmomatic tool was used for the first round of adapter contamination removal. Next, the Kseq tool developed by Qian Zhu *et al* was applied as it is shown to be capable of processing adapter sequences that were not detected by Trimmomatic such as those up to 6 bp from the 3' end of each read. Bowtie2 was used to align our sequences with the reference genomes. In our case, each mate of the pair was aligned with the *mus musculus* genome and with the *E. coli* genome (Script S.1). The resulting alignment data were imported into a table showing the total number of reads, the number of aligned reads detected with each genome and the percentage of the sample aligned to each genome (Script S.2).

For the normalization of the data, the ratio of *E. coli* DNA to *mus musculus* DNA was calculated for each sample and the necessary scale factors for a correct normalization were obtained. For this purpose, the following calculations were applied:

To obtain DNA ratio, the aligned concordantly exactly 1-time *E. coli* data was divided by the total number of reads in the sample.

$$\text{DNA ratio} = \frac{\text{aligned concordantly exactly 1 time } E \text{ coli}}{\text{total reads in sample}}$$

To calculate the scale factor, the DNA ratio of each sample is divided by the ratio of the one that has the lowest value. This sample, in our case the positive control, will have a scale factor of 1.

$$\text{Scale factors} = \frac{\text{DNA ratio each sample}}{\text{lowest ratio among samples}}$$

3.4.2. Sorting and generation of bw files

For the implementation of the Sorting steps, the `filter_below` tool was applied, where the sequenced read pairs that appeared duplicated were eliminated and fragments were filtered between those with more than 120 bp and those with less than 120bp. This is because fractions with ≤ 120 -bp are more likely to have TF binding sites. SAMtools was used to create an index file for each of our BAM files the purpose of this is to have the ability to retrieve alignments that overlap with a specific region without having to read the entire alignment file (23). We operate with `bamCoverage` to transform BAM files to bigWig files. The bigWig format is an indexed binary format useful for this dense, continuous data and served as input for some of the visualization commands that it will run in deepTools (Script S.3) (24).

3.4.3. Peak calling and Heat map plots generation

To carry out this computational method, the Model-base Analysis of ChIP-seq (MACS2) tool was implemented, which allowed us to identify the areas of the genome that have been enriched with aligned reads (Script S4). To visualize the data of the files in bigwig format Heat map plots that show the enrichment around the transcription start site (TSS) of the different genes were created. This region was selected since many DNA-binding proteins are known to be close to the TSS of their Target genes. To generate these plots the creation of an intermediate file is necessary. In this Count Matrix is where you set the reference window to generate the figure in our case ± 500 bp around the TSS of the genes (Script S5) (24).

```
Mac2 callpeak -t $bam_file -g mm -f BAMPE -n $bam_file_name  
--outdir $outdir -q 0.01 -B --SPMR --keep-dup all
```

3.4.4. IGV analysis of Yap/Taz known target genes

For easy visual exploration of the genomic data, The Integrative Genomics Viewer (IGV) was used. This analysis aimed to verify that the main Yap/Taz target genes were present in Yap/Taz samples in the ATII cells compared to the negative control (IgG) and the positive control (H3K4me3).

3.4.5. Annotation of the gene list

The Sparse Enrichment Analysis (SEACR) it's a peak calling algorithm designed specifically for processing paired-end CUT&RUN data, achieving superior selectivity than common ChIP-seq peak callers such as MACS2. It was used to identify the regions of the genome enriched by the DNA fragments and to eliminate erroneous peaks caused by multiple mapping at repeated regions or false-positive sites (25). The Homer software was applied to associate the peaks with nearby genes and as a result we obtained two lists of annotated Yap/Taz-enriched genes, the relaxed and the stringent. In the first one, since it has a lower threshold, a greater number of signals are included than in the stringent one (Script S6).

To filter the result list, tidyverse and ggVennDiagram packages of Rstudio were used to compare which genes were activated in the sample with the Yap antibody and in the sample with Yap and Taz antibodies and a Venn diagram was created.

3.4.6. Functional annotation analysis

DAVID bioinformatics database was used to perform a functional analysis of the identified genes from the CUT&RUN experiment. Due to the number of genes, the Yap and Yap/Taz overlap list was used as an input. Genes were divided into groups according to their biological functions. Each subgroup was introduced into the STRING database and a plot was generated where each gene was represented by a sphere and the interactions between them by lines of different colors. The goal was to compare whether there was a theoretically or experimentally known relationship with Yap.

4. RESULTS

4.1. Effective ATII cell isolation and CUT&RUN experiment

A total of 912 000 cells/ml were obtained, therefore in a final volume of 18 ml, the total number of cells would be 16 million. Based on the EpiCypher protocol it was decided to use 1 million cells per condition to perform the CUT&RUN. As a result of this experimental procedure, 10 ul of supernatant of 5 samples (IgG, H3K4me3, Yap, Taz, Yap/Taz) that contains the purified DNA fragments were obtained and sent for sequencing.

4.2. DNA alignment results and data normalization

To carry out the different analysis on the genes isolated in CUT&RUN, it was necessary to trim the isolated DNA fragments which means removing the adaptor sequences introduced for the sequencing. Later, their alignment with the reference genomes of *mus musculus* and *E.coli* was carried out using the Bowtie2 tool. It was necessary to take into account that from around 3-5 million total reads onwards the sample was considered to be good and therefore the alignment had worked (26). The data resulting from both alignments are shown in Table 3. The total reads in the negative control, the positive control, and Yap/Taz sample were above 5 million, and in the case of Yap was slightly below but inside the range (4159949 total reads). However, it was detected that the Taz sample only had a value of around 5 thousand total reads, which was well below the expected value. Another quality control of the CUT&RUN experiment was the percentage of total alignment. The aim was to have in the experimental samples more than 50% of the isolated fragments aligned with the *mus musculus* genome. The results obtained in the positive, negative, Yap, Taz, and Yap/Taz samples had the following values respectively 51.65%, 71.93%, 50.05%, 58.13%, 46,23%. As a final control, the sum of the alignments of *E.coli* and *mus musculus* should not exceed 100%. It can be seen that this percentage was not reached in any of them. To continue with the bioinformatics analysis of the data obtained, it was imperative to normalize the content of the different samples. For this purpose, it was necessary to calculate the DNA ratio between the DNA of *E.coli* and the number of total reads of each one of them and, based on this, the scale factor needed to complete the normalization. The sample with the lowest DNA ratio, in this case, H3K4me3+, was used to calculate the scale factor to be applied to the remaining conditions. Therefore, when applying the formula explained in materials and methods the scale factor of the positive control was 1. The negative control since we were not interested in

amplifying the signal, was also considered to have a scale factor of 1. The values used to normalize the rest of the conditions are shown in Table 4.

Table 3. Alignment results from IgG, H3K4Me3+, Yap, Taz, Yap and Taz samples obtain with Bowtie2

Sample name	total reads	aligned concordantly exactly 1 time	overall alignment rate	Sum of the alignment rates
Mouse13.IgG(C-).mus musculus	13896449	4987675	51,65%	82,61%
Mouse13.IgG(C-).E.coli	13896449	4203467	30,96%	
Mouse13.H3K4me3(C+).mus musc	22070840	13710324	71,93%	86,96%
Mouse13.H3K4me3(C+).E.coli	22070840	3235932	15,03%	
Mouse13.Yap.mus musculus	4154949	1548003	50,05%	83,51%
Mouse13.Yap.E.coli	4154949	1356559	33,46%	
Mouse13.Taz.mus musculus	55416	24535	58,13%	78,12%
Mouse13.Taz.E.coli	55416	10798	19,99%	
Mouse13.Yap/Taz.mus musculus	9606112	3261142	46,23%	73,31%
Mouse13.Yap/Taz.E.coli	9606112	2539556	27,08%	

Table 4 Normalization of the data

Sample	DNA ratio	Scale factor
Mouse13.IgG(C-)	0,302484973	1
Mouse13.H3K4me3(C+)	0,146615716	1
Mouse13.Yap	0,326492335	2,226857697
Mouse13.Taz	0,194853472	1,329008089
Mouse13.Yap/Taz	0,264368769	1,803140731

4.3. CUT&RUN effectively detects Yap and Taz enrichment near gene transcriptional start sites

To evaluate the correct isolation of DNA fragments carried out by CUT&RUN, heat maps have been generated showing the gene enrichment close to the TSS in the whole *mus musculus* genome in the positive control, negative control, Yap, Taz, and Yap/TAZ samples. The window chosen to generate the figure is +/- 500 bp around TSS as it has been determined in previous analyses performed at Wagner lab as the optimal distance. As expected, the negative control has almost no gene activation around TSS (Figure 8a). Indeed, the presence of aligned reads near the TSS of a large number of genes has been demonstrated in the positive control sample and in the Yap and Yap/Taz experimental conditions (Figure 8b, d, e). However, we confirmed again that the Taz sample is defective as the genes in the map are not well detected, and therefore no signals from the alignment reads can be identified (Figure 8c).

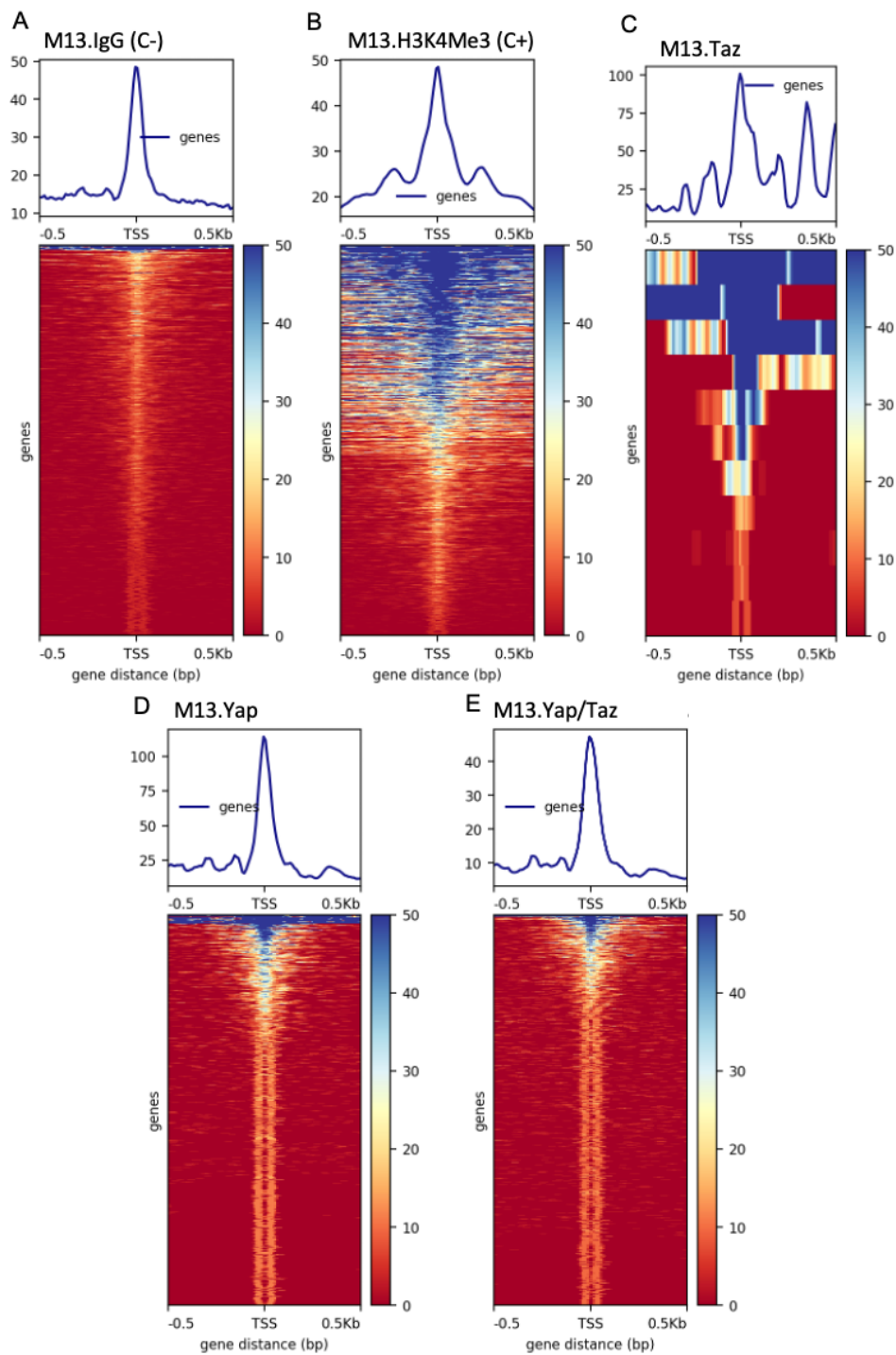


Figure 8. Coverage heatmap of aligned sequences of CUT&RUN experiments results. A heat map plot of the signals from total aligned reads to *mus musculus* genome was obtained by the MACS2 tool. The region within 0.5 kb around the annotated transcription start site (TSS) is presented in controls (IgG and H3K4me3) and experimental (Yap, Taz, and Yap/Taz) samples. The detection of TSS enrichment in the experimental conditions indicates that Yap /Taz antibodies capture TF motif sequences.

4.4. Integrative Genomics Viewer confirms the presence in our experimental samples of Yap/Taz well-known target genes

To check whether the main Yap/Taz target genes were recollected in our sample, IGV was used to visualize and explore the dataset. The presence of fragment sequences at the chromosomal location of *Ctgf* is observed (Figure 9a). This signal is detected in the positive control sample and in the Yap and Yap/Taz samples. This result under the experimental conditions indicates that among all isolated DNA fragments that interacted with the Yap/Taz+Transcription Factor complex this gene was present. Alignment reads are also found in the chromosomal location of *Cyr61* and *Birc5* as shown in (Figures 9b,c), this confirms the presence of these genes in the Yap and Yap/Taz conditions. In addition, *Nuack2*, a gene involved in biological processes such as the cellular response to glucose deprivation or the negative regulation of apoptosis, was also detected in the experimental samples, as can be seen by the accumulation of fragments in the *Nuack2* localization (Figure 9d). The fact of detecting new genes that are not currently considered targets of these coactivators is an interesting result to investigate further. The presence of more Yap/Taz Targets (Figure S1) was verified.

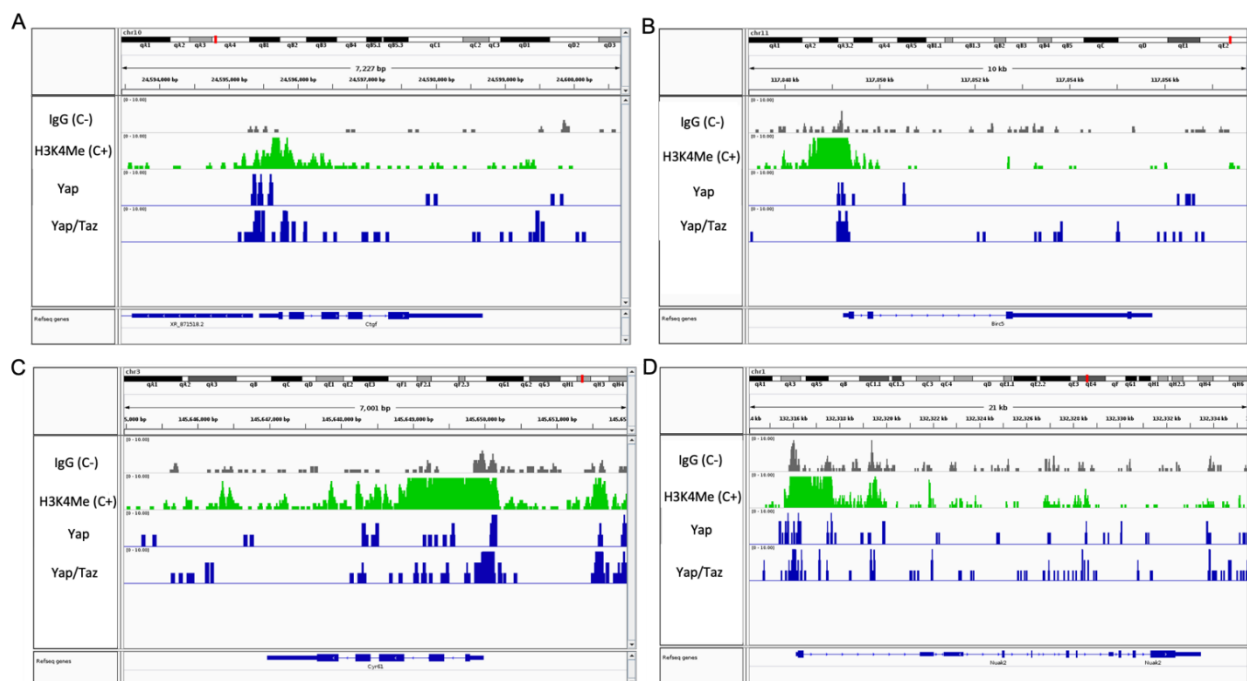


Figure 9. Study of gene presence in control positive, negative, Yap, Yap/Taz conditions. The sequencing files in bam format were used as input to IGV. The graph indicates the presence of the gene in the sample by placing each sequence fragment in the genome. The x-axis represents the location on the chromosome and the y-axis represents a scaling of the number of reads per bp. Yap/Taz motif sequences overlap with H3K4me3 marker on the promoter regions of well-known Yap/Taz target genes.

4.5. Gene Annotation and Overlap between Yap and Yap-Taz samples

Homer software was applied to generate the list of genes that have been detected in our sample as Yap/Taz+TF complexes target genes taking SEACR results as input. The tool searched the chromosomal location of the DNA fragments and identify them based on the genes that are located at the same site. For each annotated gene there is a peak score value that corresponds to the reliability of the peak identified by the tool. Table 5 shows the peak score obtained by the genes previously analyzed with IGV in the relaxed and stringent list. As can be seen in the table we do not have data on the peak score of Birc5 in the Yap and Yap-Taz stringent lists as this gene does not appear in those datasets.

Table 5. Peak score of example genes from the generated datasets

Gene Name	Peak score stringent-Yap	Peak score stringent-Yap/Taz	Peak score relaxed-Yap	Peak score relaxed-Yap/Taz
Cgtf	4	6	4	6
Birc5			2	3
Cyr61	3	4	4	9
Nuak2	3	4	3	7

In order to reduce the list of identified genes to have a manageable volume of them to perform functional analysis, the annotation captured in the Yap sample and the Yap/Taz sample were compared using the R packages tidyverse and ggVennDiagram. The results obtained are shown in (Figures 10a,b). With a peak score threshold greater than one, it is found that about 84% of the annotated genes were in both lists. On the other hand, if a peak score of more than 10 is established, the percentage is reduced to 18%.

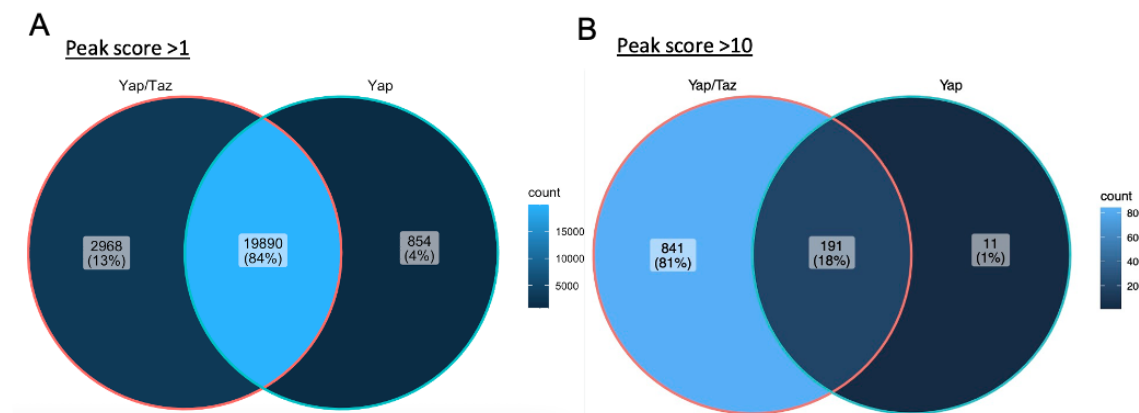


Figure 10. Overlap genes in Yap and Yap/Taz samples. A Venn diagram illustrating the number of genes that match in both datasets. To obtain the graph, a peak score filter of more than 1 (A) and of more than 10 in (B) was established.

These 191 genes identified in the stringent list and with peak scores greater than 10 were introduced into the DAVID database. As a result, the classification of genes according to biological function can be seen in (Figure 11). Functions clearly related to Yap/Taz such as cell or tight junction extracellular matrix and cell membrane can be observed but others such as synapse, secreted, cytoplasmic vesicles have also been detected.










Term	RT	Genes	Count
Cell junction	RT		18
Cytoplasm	RT		51
Extracellular matrix	RT		7
Tight junction	RT		5
Cytoplasmic vesicle	RT		10
Synapse	RT		8
Cell membrane	RT		32
Secreted	RT		18
Cell projection	RT		13

Figure 11. Functional annotation chart of Yap/Taz regulated genes.

4.6. Detection of new potential target genes of Yap by STRING

In pursuit of determining whether the identified genes were currently related in the literature to Yap, the sublists generated for each biological function were introduced in the STRING database together with the Yap protein. As a result, the database generates a plot per sublist where each gene is represented by a sphere and the interactions between them by lines of different colors.

When studying the genes isolated from ATII cells that have been associated with cell junction we can see that only 5 out of 18 (Ajuba, Limd1, Tjp1, Akp5, Pard6b, Cldn34d) are currently known to have a direct relationship with Yap (Figure 12a). While for the even more specific function of tight junction it can be observed that in all but 2 of the genes in the list, their association with Yap was already known (Figure 12b). The cytoplasm category is the one that contained the greatest number of genes since it is a function that encompasses many more processes, nevertheless, there are many genes with no connection with our target coactivator (Figure 12c). None of the genes involved in extracellular matrix processing showed prior association with Yap. However, after searching for more indirect associations through intermediary proteins in SRING, the WNT7A a Wnt ligand could be related (Figure 12d). The results of the interactions found in the cell membrane and secreted functions can be seen in (Figures 12e, f). The rest of the sublists with less relevant functions for the IPF study are shown in annexes (Figure S2).

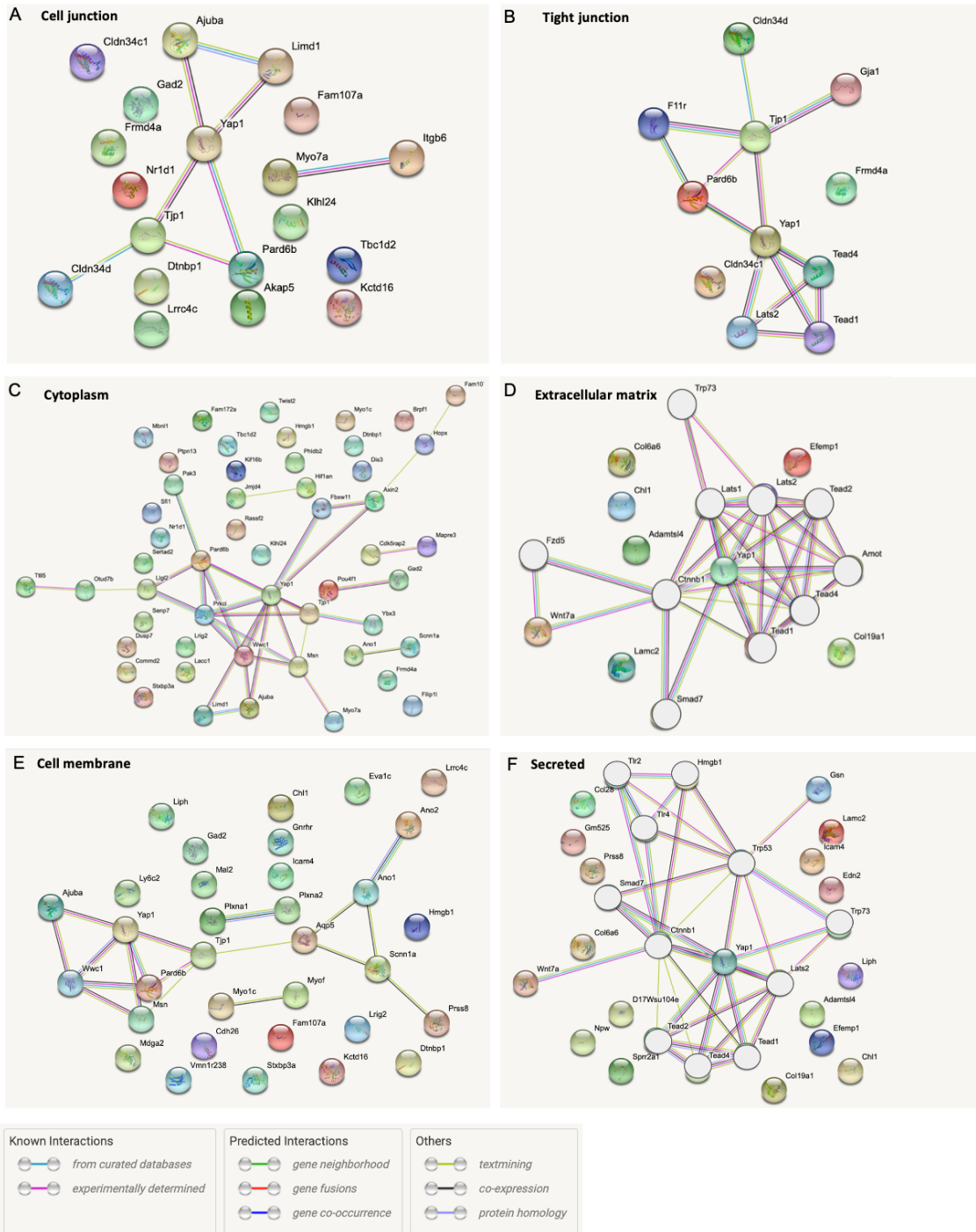


Figure 12. Interactions study between the gene dataset and Yap protein using STRING database. Genes are represented as spheres and interactions are visualized as lines. The colors of these junctions as can be seen in the legend show either the type of association found or how this interaction was discovered. In the extracellular matrix and secreted functions, the MORE option of STRING was used so that intermediate genes not belonging to the list are covered with a gray sphere.

5. DISCUSSION

In IPF, the cells of the alveolar epithelium become dysfunctional losing their regenerative capacity and contributes to the formation of the characteristic scar tissue in the lung. These disease driver cells undergo a reactivation of many developmental signaling cascades including Hippo pathway. Their main mediators Yap/Taz are context-dependent coactivators and therefore, depending on the stimulus of the microenvironment around the cells, they will bind to specific TFs and promote the expression of different genes. If we hypothesize that Yap/Taz bind to different transcription factors in IPF conditions and we identify them, they could serve as targets for drug development. The cell line chosen to carry out this study on the genes regulated by Yap/Taz is primary alveolar type 2 cells. The main disadvantage of primary cultures, which are cells isolated directly from, in this case, mouse lung tissue, is that they have a limited number of passages as opposed to immortalized alveolar epithelial cells such as MLE12. However, by using these cells we ensure that the activation pattern and the genes regulated by Yap/Taz are not altered by the immortalization process and resemble as closely as possible physiological conditions. It is necessary to take into account that Yap/Taz are context-dependent coactivators and therefore their activation is clearly regulated by mechanical signals such as cell density, extracellular matrix, type of support, etc. For this reason, as soon as the cells were isolated from the tissue they had been subjected to CUT&RUN, which means they were not seeded at any time except for the 30 min used to eliminate the fibroblasts within the cell isolation protocol. However, 30 minutes should not alter the behavior of the ATII cells since they most likely will continue in a shock period where their cell metabolism is paused.

As previously mentioned, one of the advantages of using CUT&RUN is that it has a low sample requirement, namely, the minimum number of cells needed to carry out the process is 5 thousand cells (26). In this case, due to the use of the cell isolation protocol developed by *Alsafadi et al.* a total of 16 million ATII have been successfully isolated. However, it has been decided to use 5 million cells for the experiment, which means 1 million cells per condition. *Peter J Skene et al* by performing the same CUT&RUN experiment but changing the initial cell number from 600,000 to 10 million cells showed that after normalization with *E.coli* the number of excision events is proportional to the initial cell number. This suggested that high data quality is maintained with low input material and therefore we had a sufficient quantity of cells in the sample to obtain an optimal result in the CUT&RUN.

After carrying out the assay, the DNA in the supernatant recollect for each condition was sent for sequencing to Wallenberg Centre For Molecular Medicine in Linköping (27). The results received

were the isolated fragment sequences. As explained in detail in materials and methods these sequences were first trimmed to remove the adaptor sequences and then aligned with the *mus musculus* and *E.coli* genomes. In order to analyze the results obtained from the alignment (Table 1), it is necessary to take into account that based on EpiCypher CUTANA protocol from 3-5 million total reads onwards the sample is considered to be good and therefore a successful alignment. The total reads in the controls, Yap, and Yap/Taz samples are above 3 million. However, it was detected that the Taz sample only had a value of around 5 thousand total reads, which was well below the minimum value. This indicates that some experimental error occurred in that sample that prevented the cut fragments from being collected in the supernatant. Perhaps problems with cell permeabilization or loss of sample during DNA purification. As additional quality controls, we use the percentage of alignment of the total reads with *mus musculus* genome and the sum of the overall alignment of each genome (Table 1). As commented in the results all samples have a percentage of alignment with *mus musculus* above 50%. Consequently, this implies that more than half of the contents of all samples were of sufficient quality or length to be aligned by the tool. As for the overall alignment sum, in no case does the result exceed 100%. Therefore, it is concluded that in the conditions of positive and negative controls, Yap, and Yap/Taz the CUT&RUN seems to be carried out properly, however in the Taz sample errors seem to have occurred in the development of the experiment. In future CUT&RUN experiments I could measure the amount of DNA in the sample after purification to detect if any of the samples are defective in order to repeat that condition before sequencing and bioinformatics analysis.

As a result of TSS enrichment analysis (Figure 8) where the purpose is to ensure that we detect the presence of genes in our samples. It can be concluded that the negative control has almost no gene activation around TSS. The low detections that can be observed are probably since rabbit IgG can sometimes bind randomly to certain proteins and transcription factors from the mouse (Figure 8a) (28). The presence of aligned reads near the TSS of a large number of genes in the Yap and Yap/Taz experimental conditions indicates that Yap/Taz antibodies used in CUT&RUN successfully captured TF motif sequences (Figure 8b,d,e). No signal around TSS can be identified in the Taz sample as it was expected from the alignment results (Figure 8c).

As can be seen in the graphs obtained from IGV (figure 9) Yap/Taz motifs overlap with H3K4me3 marker in the promotion regions of *Ctgf*, *Birc5*, *Cyr61*. It has been demonstrated then that with the CUT&RUN experimental methodology the presence of important Yap/Taz target genes can be detected. The appearance of these Yap-associated genes corroborates that the pAG-MNase enzyme cleavage has been specific and therefore the genes obtained have been regulated by

Yap/Taz coactivators. In addition, as genes not associated with Yap (Figure 9d) were also found in our sample, it was decided to continue the analysis of the gene expression profile.

After the annotation of the genes with Homer software, it was decided to study the peak score of the isolated genes. A Peak score is a measure that establishes the quality of the signal detected by SEACR. Peak scores of all the genes shown in (Table 5) are around 2-9 in both the relaxed and stringent lists. In addition, there is no data for Birc5 in the stringent list since it was not found in that dataset. This implies that although we know that Birc5 is Yap/Taz Target gene (29). SEACR did not consider the peak identified in that gene as high enough to enter it in the stringent list. This drawback could be solved with the implementation of LIANA, another computational tool in development that seems to establish an even better threshold for peak calling (30).

In order to follow up on the analysis of the genes obtained, the Yap and Yap/Taz sample datasets were compared to identify the repeated genes. The unfiltered comparison contained an 84% match rate however when the threshold was set to more than 10 in the peak score only 18% matched. This means that there were many identifications with low peak score values that matched in both lists but in higher peak scores, fewer genes matched. This overlap list generated for further analysis includes genes that are on both the Yap and the Yap/Taz samples and also have a peak score greater than 10 (Figure 10). Many genes including some important targets (Table 5) are left out of this analysis, but it was decided to use this parameter as the resulting number of 191 genes was sufficient for the results to have biological significance.

For a better biological understanding of this overlap list, these genes were introduced in DAVID database and as result 9 different functional categories showed up (Figure 11).

It is widely described in the literature that the cell membrane is responsible for receiving a large number of stimuli from the extra- and intracellular environment and acts as an interface to transduce these signals through signaling cascades such as Hippo. In addition, cell membrane has many domains such as adherens junctions, tight junctions, or focal adhesions that transmit stimuli from the extracellular matrix and from other cells (31). It has been shown that many components of cell-cell junction- and cell polarity-related protein complexes regulate the Hippo pathway and in turn, Yap/Taz activation controls the extracellular matrix and cell-cell interactions generating positive feedback (32). For example, activation of Yap/Taz regulates RhoGAPs, which consequently remodels the cytoskeleton, and this, in turn, affects the initiation and maintenance of AFs (31). For this reason, functions such as cell or tight junction, ECM, and cell membrane are extremely associated with Yap and the fact that we have detected genes in our sample that fall into those categories is another confirmation of CUT&RUN's success.

To continue with the analysis of the Yap interactome. DAVID-generated sublists were introduced into STRING. Although this is a database of known and predicted protein interactions it obtains information from genomic context predictions, high-throughput lab experiments, co-expression, automated textmining, and previous knowledge in databases (33).

The fact that we have detected Yap/Taz unknown interactions supports the idea that Yap/Taz are context-dependent and therefore the genes that they regulate are changeable. In this study, we have obtained the gene expression profile of ATII in a physiological situation and although it would be necessary to repeat this experiment in ATII cells from mice with the disease, these new targets detected support the hypothesis that the transcription factors that Yap/Taz binds to, could be different in IPF.

The presence of genes that were only found in the Yap sample and not in the Yap/Taz sample may be due to nonspecific binding of either the antibody or the pAG-MNase. Most genes present only in the Yap/Taz sample should be genes regulated by Taz exclusively although they may also be fragments obtained from nonspecific binding. Therefore an interesting analysis that would have been useful for this project would be to compare the experimental gene list that was generated with a KO Taz/Yap gene dataset to confirm that the information obtained is specific and correct. For future research in order to identify transcriptional factors that bind to Yap and Taz in a physiological condition, it will be necessary to apply motif search and footprinting analysis to associate transcriptional motifs in our gene sample with a list of possible TFs. Once we have this, we can repeat the whole analysis with mouse ATII modeling idiopathic pulmonary fibrosis. In the end, the final output will be a theoretical list of Yap/Taz Binding partners in the disease. Nevertheless, this bioinformatic data will need to be confirmed by experimental work that allows the isolation of these transcription factors and their identification.

The main conclusions of this project are as follows:

- Effective isolation of primary alveolar type 2 cells from a healthy mouse has been achieved.
- Technique validation of the CUT&RUN for detecting Yap and Taz transcriptional motifs.
- The bioinformatic pipeline developed for CUT&RUN is applicable to samples obtained from both mice and humans in either healthy or IPF tissue.
- Obtaining Yap/Taz transcriptional factor motif regions and gene expression profile in healthy ATII.
- Identification of new Yap/Taz target genes by STRING.

6. BIBLIOGRAPHY

1. Chakraborty A, Mastalerz M, Ansari M, Schiller HB, Staab-Weijnitz CA. Emerging Roles of Airway Epithelial Cells in Idiopathic Pulmonary Fibrosis. *Cells* [Internet]. 2022 Mar 1 [cited 2022 Jun 28];11(6). Available from: [/pmc/articles/PMC8947093/](#)
2. Sgalla G, Iovene B, Calvello M, Ori M, Varone F, Richeldi L. Idiopathic pulmonary fibrosis: pathogenesis and management. *Respir Res* [Internet]. 2018 Feb 22 [cited 2022 May 30];19(1). Available from: [/pmc/articles/PMC5824456/](#)
3. Hewlett JC, Kropski JA, Blackwell TS. Idiopathic Pulmonary Fibrosis: Epithelial-mesenchymal interactions and emerging therapeutic targets. *Matrix Biol* [Internet]. 2018 Oct 1 [cited 2022 May 30];71–72:112. Available from: [/pmc/articles/PMC6146058/](#)
4. Barkauskas CE, Chung MI, Fioret B, Gao X, Katsura H, Hogan BLM. Lung organoids: current uses and future promise. *Development* [Internet]. 2017 Mar 3 [cited 2022 Jun 28];144(6):986. Available from: [/pmc/articles/PMC5358104/](#)
5. Sun T, Huang Z, Zhang H, Posner C, Jia G, Ramalingam TR, et al. TAZ is required for lung alveolar epithelial cell differentiation after injury. *JCI Insight* [Internet]. 2019 Jul 7 [cited 2022 May 30];4(14). Available from: [/pmc/articles/PMC6675554/](#)
6. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet*. 2017 May 13;389(10082):1941–52.
7. Parimon T, Yao C, Stripp BR, Noble PW, Chen P. Molecular Sciences Alveolar Epithelial Type II Cells as Drivers of Lung Fibrosis in Idiopathic Pulmonary Fibrosis. [cited 2022 May 30]; Available from: [www.mdpi.com/journal/ijms](#)
8. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Prim* 2017 31 [Internet]. 2017 Oct 20 [cited 2022 Jun 18];3(1):1–19. Available from: [https://www.nature.com/articles/nrdp201774](#)
9. Effendi WI, Nagano T. The Hedgehog Signaling Pathway in Idiopathic Pulmonary Fibrosis: Resurrection Time. *Int J Mol Sci* 2022, Vol 23, Page 171 [Internet]. 2021 Dec 24 [cited 2022 Jun 13];23(1):171. Available from: [https://www.mdpi.com/1422-0067/23/1/171/htm](#)
10. Sun M, Sun Y, Feng Z, Kang X, Yang W, Wang Y, et al. New insights into the Hippo/YAP pathway in idiopathic pulmonary fibrosis. *Pharmacol Res*. 2021 Jul 1;169:105635.
11. Chanda D, Otoupalova E, Smith SR, Volckaert T, De Langhe SP, Thannickal VJ. Developmental Pathways in the Pathogenesis of Lung Fibrosis. *Mol Aspects Med* [Internet]. 2019 Feb 1 [cited 2022 Jun 13];65:56. Available from: [/pmc/articles/PMC6374163/](#)
12. Kim CL, Choi SH, Mo JS. Role of the Hippo Pathway in Fibrosis and Cancer. *Cells* 2019, Vol 8, Page 468 [Internet]. 2019 May 16 [cited 2022 Jun 14];8(5):468. Available from: [https://www.mdpi.com/2073-4409/8/5/468/htm](#)
13. Hansen CG, Moroishi T, Guan KL. YAP and TAZ: a nexus for Hippo signaling and beyond. *Trends Cell Biol*. 2015 Sep 1;25(9):499–513.
14. Panciera T, Azzolin L, Cordenonsi M, Piccolo S. Mechanobiology of YAP and TAZ in physiology and disease. *Nat Rev Mol Cell Biol* 2017 1812 [Internet]. 2017 Sep 27 [cited 2022 May 30];18(12):758–70. Available from: [https://www.nature.com/articles/nrm.2017.87](#)
15. Kim MK, Jang JW, Bae SC. DNA binding partners of YAP/TAZ. *BMB Rep* [Internet]. 2018 [cited 2022 May 30];51(3):126–33. Available from: [https://pubmed.ncbi.nlm.nih.gov/29366442/](#)
16. Totaro A, Panciera T, Piccolo S. YAP/TAZ upstream signals and downstream responses. *Nat Cell Biol* [Internet]. 2018 Aug 1 [cited 2022 Jun 14];20(8):888. Available from:

- /pmc/articles/PMC6186418/
17. Moya IM, Halder G. Hippo–YAP/TAZ signalling in organ regeneration and regenerative medicine. *Nat Rev Mol Cell Biol* 2018 204 [Internet]. 2018 Dec 13 [cited 2022 Jun 14];20(4):211–26. Available from: <https://www.nature.com/articles/s41580-018-0086-y>
 18. Tang W, Li M, Yangzhong X, Zhang X, Zu A, Hou Y, et al. Hippo signaling pathway and respiratory diseases. *Cell Death Discov* [Internet]. 2022 Dec 1 [cited 2022 Jun 18];8(1). Available from: /pmc/articles/PMC9021242/
 19. Noguchi S, Saito A, Nagase T. YAP/TAZ Signaling as a Molecular Link between Fibrosis and Cancer. *Int J Mol Sci* [Internet]. 2018 Nov 20 [cited 2022 Jun 18];19(11). Available from: /pmc/articles/PMC6274979/
 20. Zhu Q, Liu N, Orkin SH, Yuan GC. CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. *Genome Biol* [Internet]. 2019 Sep 9 [cited 2022 Jun 18];20(1). Available from: /pmc/articles/PMC6734249/
 21. Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* 2018 135 [Internet]. 2018 Apr 12 [cited 2022 Jun 13];13(5):1006–19. Available from: <https://www.nature.com/articles/nprot.2018.015>
 22. Alsafadi HN, Stegmayr J, Ptasinski V, Silva I, Mittendorfer M, Murray L, et al. Simultaneous isolation of proximal and distal lung progenitor cells from individual mice using a 3D printed guide reduces proximal cell contamination of distal lung epithelial cell isolations. *bioRxiv* [Internet]. 2022 May 10 [cited 2022 Jun 28];2022.05.10.491312. Available from: <https://www.biorxiv.org/content/10.1101/2022.05.10.491312v1>
 23. samtools(1) manual page [Internet]. [cited 2022 Jun 19]. Available from: <http://www.htslib.org/doc/samtools.html>
 24. Visualization of peaks | In-depth-NGS-Data-Analysis-Course [Internet]. [cited 2022 Jun 19]. Available from: https://hbctraining.github.io/In-depth-NGS-Data-Analysis-Course/sessionV/lessons/10_data_visualization.html
 25. Meers MP, Tenenbaum D, Henikoff S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics and Chromatin* [Internet]. 2019 Jul 12 [cited 2022 Jun 29];12(1):1–11. Available from: <https://link.springer.com/articles/10.1186/s13072-019-0287-4>
 26. CUTANA™ CUT&RUN Technology | Learn about CUT&RUN [Internet]. [cited 2022 Jun 19]. Available from: <https://www.epicypher.com/technologies/cutana/cut-and-run>
 27. Claudio Cantù - Linköping University [Internet]. [cited 2022 Jun 20]. Available from: <https://liu.se/en/employee/claca29>
 28. Technical Data Sheet: Rabbit IgG Antibody, CUTANA™CUT&RUN_Negative Control. [cited 2022 Jun 20]; Available from: www.epicypher.com
 29. Hong W, Guan KL. The YAP and TAZ transcription coactivators: key downstream effectors of the mammalian Hippo pathway. *Semin Cell Dev Biol* [Internet]. 2012 [cited 2022 Jun 20];23(7):785. Available from: /pmc/articles/PMC3459069/
 30. GitHub - Tingvall/LIANA [Internet]. [cited 2022 Jun 19]. Available from: <https://github.com/tingvall/LIANA>
 31. Rausch V, Hansen CG. The Hippo Pathway, YAP/TAZ, and the Plasma Membrane. 2020 [cited 2022 Jun 19]; Available from: <https://doi.org/10.1016/j.tcb.2019.10.005>
 32. Cai X, Wang KC, Meng Z. Mechanoregulation of YAP and TAZ in Cellular Homeostasis and Disease Progression. *Front Cell Dev Biol*. 2021 May 24;9:1333.
 33. About - STRING functional protein association networks [Internet]. [cited 2022 Jun 17]. Available from: <https://string-db.org/cgi/about>

7. ANNEXES

CUT&RUN complete protocol: <https://www.epicypher.com/content/documents/protocols/cutana-cut&run-protocol.pdf>

Table S1: CUT&RUN buffer components

Components	Source	Cat #
HEPES	Sigma-Aldrich	H3375
KCl	Sigma-Aldrich	P3911
CaCl ₂	Sigma-Aldrich	C1016
MnCl ₂	Sigma-Aldrich	203734
Molecular biology grade H ₂ O (RNase, DNase free)	VWR	VWRV02-0201-0500
NaCl	Sigma-Aldrich	S5150-1L
EDTA (prepare 0.5 M stock at pH 8.0)	Sigma-Aldrich	E5134
EGTA (prepare 0.5 M stock at pH 8.0)	Sigma-Aldrich	E3889
RNase A	Thermo Fisher Scientific	EN0531
Glycogen	Sigma-Aldrich (Roche)	10930193001
Spermidine trihydrochloride*	Sigma-Aldrich	S2501
Digitonin (store aliquots of 5% stock in DMSO at -20°C)	Millipore Sigma	300410
DMSO	Sigma-Aldrich	D8418-100ml
cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail	Roche	11836170001
Trypan blue	Thermo Fisher Scientific	T10282

Table S2: CUT&RUN reagents

Item	Vendor	Catalog No.
Concanavalin A (ConA) Conjugated Paramagnetic Beads	EpiCypher	21-1401
CUTANA pAG-Mnase	EpiCypher	15-1016 or 15-1116
Rabbit IgG Negative Control Antibody	EpiCypher	13-0042
<i>E. coli</i> Spike-in DNA	EpiCypher	18-1401
CUTANA DNA Purification Kit	EpiCypher	14-0050

Script S1: Step1_Allignment&Trimming

```

#!/bin/bash
#SBATCH -A lu2022-7-7
#####
## Load modules
module load Trimmomatic/0.38-Java-1.8.0_162
module load GCC/7.3.0-2.30 OpenMPI/3.1.1
module load Bowtie2/2.3.4.2
module load SAMtools/1.9
#####
# Define functions & paths
workdir=/lunarc/nobackup/projects/snic2020-12-34/Irene_project
adapterpath=$workdir/tools/adapters
proj=$workdir/projects/mouseATII
len=42
# path for Ecoli bowtie2 index
bt2ecoli=$workdir/genomes/Escherichia_coli_K_12_MG1655_NCBI_2001-10-15/Escherichia_coli_K_12_MG1655_NCBI/2001-10-15/Sequence/Bowtie2Index
trimdir=$proj/trimmed
trimdir2=$proj/trimmed3
logdir=$proj/logs
aligndir=$proj/aligned
alignspike=$proj/aligned_to_ecoli
if [ $3 == 'hs' ]; then
bt2idx=$workdir/genomes/Homo_sapiens_UCSC_hg38/Homo_sapiens/UCSC/hg38/Sequence/Bowtie2Index
elif [ $3 == 'mm' ]; then
bt2idx=$workdir/genomes/Mus_musculus_UCSC_mm10/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index
else
    echo "species argument not inputted"
    exit
fi
mkdir $trimdir $trimdir2 $logdir $aligndir $alignspike
>&2 echo "##### STARTING Trimming"
>&2 echo "file $1 trimming started"
infile=$1
base=$(echo $infile | awk '{gsub("_R1.fastq.gz","")}1')
basec=`basename $infile _R1.fastq.gz`
>&2 echo "Input file is $infile"
>&2 date
#trimming paired-end
#good version

```

```

>&2 echo "Trimming file $base ..."
>&2 date
java -jar $EBROOTTRIMMOMATIC/trimmomatic-0.38.jar PE -threads 1 -phred33
"$base"_R1.fastq.gz "$base"_R2.fastq.gz $trimdir/"$basec"_1.paired.fastq.gz
$trimdir/"$basec"_1.unpaired.fastq.gz $trimdir/"$basec"_2.paired.fastq.gz
$trimdir/"$basec"_2.unpaired.fastq.gz
ILLUMINACLIP:$adapterpath/Truseq3.PE.fa:2:15:4:4:true LEADING:20 TRAILING:20
SLIDINGWINDOW:4:15 MINLEN:25
>&2 echo "Second stage trimming $base ..."
>&2 date
$workdir/tools/kseq_test $trimdir/"$basec"_1.paired.fastq.gz $len
$trimdir2/"$basec"_1.paired.fastq.gz
$workdir/tools/kseq_test $trimdir/"$basec"_2.paired.fastq.gz $len
$trimdir2/"$basec"_2.paired.fastq.gz
>&2 echo "Aligning file $base ... with bowtie"
>&2 date
(bowtie2 -p 2 --dovetail --phred33 -x $bt2idx/genome -1
$trimdir2/"$basec"_1.paired.fastq.gz -2 $trimdir2/"$basec"_2.paired.fastq.gz)
2> $logdir/"$basec".bowtie2 | samtools view -bS - >
$aligndir/"$basec"_aligned_reads.bam
>&2 echo "Finished BowTie Alignment"
>&2 date
>&2 echo "Aligning file $base to E. coli"
>&2 date
(bowtie2 -p 2 --dovetail --phred33 -x $bt2ecoli/genome -1
$trimdir2/"$basec"_1.paired.fastq.gz -2 $trimdir2/"$basec"_2.paired.fastq.gz)
2> $logdir/"$basec"ecoli.bowtie2 | samtools view -bS - >
$alignspike/"$basec"_aligned_reads.bam
>&2 echo "Finished BowTie Alignment against E. coli"
>&2 date

```

Script S2: Step2_botwie2data_table

```
#!/bin/bash
# syntax
# 1.3.generate_table_of_bowtie_data.sh project_name
workdir=/lunarc/nobackup/projects/snrc2020-12-34/Irene_project
proj=$workdir/projects/mouseATII
fqdir=$proj/fastq
logdir=$proj/logs
result=$logdir/bowtie_summary.txt
echo "File"> $result;
echo "total reads" >> $result;
echo "paired total" >> $result;
echo "aligned concordantly 0 times" >> $result;
echo "aligned concordantly exactly 1 time" >> $result;
echo "aligned concordantly >1 times" >> $result;
echo "----" >> $result;
echo "aligned concordantly 0 times" >> $result;
echo "of the above aligned discordantly 1 time" >> $result;
echo "----" >> $result;
echo "pairs aligned 0 times concordantly or discordantly" >> $result;
echo "mates make up the pairs" >> $result;
echo "of the above aligned 0 times" >> $result;
echo "aligned exactly 1 time" >> $result;
echo "aligned >1 times" >> $result;
echo "overall alignment rate" >> $result;
btfiles=$(find $logdir -type f -name "*.bowtie2")
for file in $btfiles
do
base=`basename $file .bowtie2`
echo $base > tmp2.txt
cat $file | awk '{print $1;}' >> tmp2.txt
paste -d'\t' $result tmp2.txt > tmp3.txt
cat tmp3.txt > $result
rm tmp2.txt
rm tmp3.txt
done
```


Script S3: Step3_Sorting

```

#!/bin/bash
#SBATCH -A lu2022-7-7
module load GCC/7.3.0-2.30 OpenMPI/3.1.1
module load deepTools/2.5.4-Python-3.6.6
module load SAMtools/1.9
workdir=/lunarc/nobackup/projects/unic2020-12-34/Irene_project
adapterpath=$workdir/tools/adapters
tools=$workdir/tools
proj=$workdir/projects/mouseATII
aligndir=$proj/aligned
sortdir=$proj/positionsort
scaleData=$proj/scale_factors_irene.txt
bamCov=$proj/bamCov_unscaled
comMat=$proj/comMat
mkdir $bamCov $comMat
#####
module load GCC/7.3.0-2.30
module load SAMtools/1.9
files=$(find $aligndir -type f -name "*.bam")
for file in $files
do
basec=`basename $file .bam`
base=$(echo $file | awk '{gsub(".bam","")}1')
>&2 echo "base: $base"
>&2 echo "basec: $basec"
#requires the filter_below.awk file for filtering
##### STEP2a. sort bam files
cd $proj
mkdir sorted dup.marked dedup
>&2 echo "Sorting bam... ""$base".bam
>&2 date
mkdir namesorted
samtools sort -n -o namesorted/"$basec".bam "$base".bam
>&2 echo "If this text appears without errors: sorting works ok."
##### STEP2b. Mark duplicates
>&2 echo "Marking duplicates... ""$base".bam
>&2 date
mkdir fixmate positionsort
samtools fixmate -m $proj/namesorted/$basec.bam $proj/fixmate/$basec.bam
echo "fixmate works"
samtools sort $proj/fixmate/$basec.bam -o $sortdir/$basec.bam
echo " sort from fixmate ends here. "

```



```
samtools markdup $sortdir/$basec.bam $proj/dup.marked/$basec.bam
echo "marking duplicates ends here"
>&2 echo "Removing duplicates... "$base".bam"
>&2 date
samtools markdup -r positionsort/$basec.bam dedup/$basec.bam
mkdir sorted.120bp dup.marked.120bp dedup.120bp
>&2 echo "Filtering to <120bp..."$base".bam"
>&2 date
samtools view -h namesorted/$basec.bam |awk -f $tools/filter_below.awk
|samtools view -Sb - > sorted.120bp/$basec.bam
samtools view -h dup.marked/$basec.bam |awk -f $tools/filter_below.awk
|samtools view -Sb - > dup.marked.120bp/$basec.bam
samtools view -h dedup/$basec.bam |awk -f $tools/filter_below.awk |samtools
view -Sb - > dedup.120bp/$basec.bam
done
```

Script S4: Step4_Peakcalling

```

#!/bin/bash
#SBATCH -A lu2022-7-7
# Load Modules
module load GCC/7.3.0-2.30 OpenMPI/3.1.1
module load deepTools/2.5.4-Python-3.6.6
module load SAMtools/1.9
workdir=/lunarc/nobackup/projects/unic2020-12-34/Irene_project
adapterpath=$workdir/tools/adapters
tools=$workdir/tools
proj=$workdir/projects/mouseATII
aligndir=$proj/aligned
sortdir=$proj/positionsort
scaleData=$proj/scale_factors_irene.txt
bamCov=$proj/bamCov_unscaled
comMat=$proj/comMat
logdir=$proj/macs2.logs
outdir=$proj/macs2
mkdir $bamCov $comMat $logdir $outdir
#####
module load GCC/7.3.0-2.30
module load SAMtools/1.9
files=$(find $aligndir -type f -name "*.bam")
for file in $files
do
basec=`basename $file .bam`
base=$(echo $file | awk '{gsub(".bam","")}1')
>&2 echo "base: $base"
>&2 echo "basec: $basec"
#### PEAK CALLING WITH MACS2
module purge
module load GCC/6.4.0-2.28 OpenMPI/2.1.1
module load MACS2/2.1.2.1-Python-2.7.14
bam_file=$proj/dup.marked.120bp/"$basec".bam
dir=`dirname $bam_file`
base_file=`basename $bam_file .bam`
cd $outdir
macs2 callpeak -t $bam_file -g mm -f BAMPE -n $base_file --outdir
$outdir -q 0.01 -B --SPMR --keep-dup all 2>
$logdir/"$base_file".macs2
>&2 echo "Converting bedgraph to bigwig... ""$basec".bam
sort -k1,1 -k2,2n "$base_file"_treat_pileup.bdg >
"$base_file".sort.bdg
$tools/bedGraphToBigWig "$base_file".sort.bdg
$tools/mm10.chrom.sizes "$base_file".sorted.bw
rm -rf "$base_file".sort.bdg
>&2 echo "$1 Finished"
done

```

Script S5: Step5_Heatmapgeneration

```

#!/bin/bash
# Load Modules

module load GCC/7.3.0-2.30 OpenMPI/3.1.1
module load deepTools/2.5.4-Python-3.6.6
module load SAMtools/1.9
workdir=/lunarc/nobackup/projects/unic2020-12-34/Irene_project
adapterpath=$workdir/tools/adapters
tools=$workdir/tools
proj=$workdir/projects/mouseATII
aligndir=$proj/aligned
sortdir=$proj/positionsort
scaleData=$proj/scale_factors_irene.txt
bamCov=$proj/bamCov_scaled
comMat=$proj/comMat
mkdir $bamCov $comMat
#####
#module purge
#module load GCC/5.4.0-2.26 OpenMPI/1.10.3
#module load icc/2017.1.132-GCC-6.3.0-2.27 impi/2017.1.132
#module load ifort/2017.1.132-GCC-6.3.0-2.27 impi/2017.1.132
#module load BEDTools/2.26.0
#mkdir $proj/bed
module load GCC/7.3.0-2.30 OpenMPI/3.1.1
module load deepTools/2.5.4-Python-3.6.6
module load SAMtools/1.9
mkdir $proj/comMat1
awk '{ system("computeMatrix reference-point --referencePoint TSS -b 500 -a 500 -
R /lunarc/nobackup/projects/unic2020-12-
34/Irene_project/projects/mouseATII/macs2/"$1"_aligned_reads_summits.bed -S
/lunarc/nobackup/projects/unic2020-12-
34/Irene_project/projects/mouseATII/bamCov_scaled/"$1".bw --skipZeros -o
/lunarc/nobackup/projects/unic2020-12-
34/Irene_project/projects/mouseATII/comMat1/"$1"_TSS.gz --outFileSortedRegions
/lunarc/nobackup/projects/unic2020-12-
34/Irene_project/projects/mouseATII/comMat1/"$1"_genes.bed") }' $scaleData
echo "done comuting matrixes"
awk '{ system("plotHeatmap -m /lunarc/nobackup/projects/unic2020-12-
34/Irene_project/projects/mouseATII/comMat1/"$1"_TSS.gz -out
/lunarc/nobackup/projects/unic2020-12-
34/Irene_project/projects/mouseATII/comMat1/"$1".png --heatmapHeight 10 --zMax
50") }' $scaleData
echo "done making Heatmaps"

```

Script S6: Step6_Annoation

```

#!/bin/bash
#SBATCH -A lu2022-7-7
workdir=/lunarc/nobackup/projects/snic2020-12-34/Irene_project
proj=$workdir/projects/mouseATII
tools=$workdir/tools
mm10=$tools/mm10.chrom.sizes
seacr=$tools/SEACR-1.3/SEACR_1.3.sh
igg=L11-C12-M-mATII-N_S11
sample1=L12-D7-M-mATII-P_S12
sample2=L15-D10-M-mATII-B_S15
module load GCC/10.2.0
module load BEDTools/2.30.0
mkdir $proj/bedgraphs $proj/seacrOut
for x in $igg $sample1 $sample2; do
bedtools bamtobed -bedpe -i $proj/namesorted/$x"_aligned_reads.bam" >
$proj/bedgraphs/$x.bed
awk ' $1==$4 && $6-$2 < 1000 {print $0}' $proj/bedgraphs/$x.bed >
$proj/bedgraphs/$x.clean.bed
cut -f 1,2,6 $proj/bedgraphs/$x.clean.bed | sort -k1,1 -k2,2n -k3,3n |
grep -v "\." > $proj/bedgraphs/$x.fragments.bed
bedtools genomcov -bg -i $proj/bedgraphs/$x.fragments.bed -g $mm10 >
$proj/bedgraphs/$x.fragments.bedgraph
done
module load GCC/9.3.0 OpenMPI/4.0.3
module load R/4.0.0
for y in $sample1 $sample2; do
$proj/bedgraphs/$igg.fragments.bedgraph non relaxed
$proj/seacrOut/$y"_relaxed.output"
$proj/bedgraphs/$igg.fragments.bedgraph non stringent
$proj/seacrOut/$y"_stringent.output"
$proj/bedgraphs/$igg.fragments.bedgraph norm relaxed
$proj/seacrOut/$y"_norm_relaxed.output"
$proj/bedgraphs/$igg.fragments.bedgraph norm stringent
$proj/seacrOut/$y"_norm_stringent.output"
sh $seacr $proj/bedgraphs/$y.fragments.bedgraph 0.1 non relaxed
$proj/seacrOut/$y"_noIgg_relaxed.output"
sh $seacr $proj/bedgraphs/$y.fragments.bedgraph 0.1 non stringent
$proj/seacrOut/$y"_noIgg_stringent.output"
done
module purge
module load GCC/7.3.0-2.30
module load homer/4.10
mkdir $proj/seacrAnnotated
peakfiles=$(find $proj/seacrOut -type f -name '*.bed')
for file in $peakfiles
do
base=`basename $file .bed`
annotatePeaks.pl $file mm10 >
$proj/seacrAnnotated/$base."annotated.bed"
echo "$base is done!!"
done

```

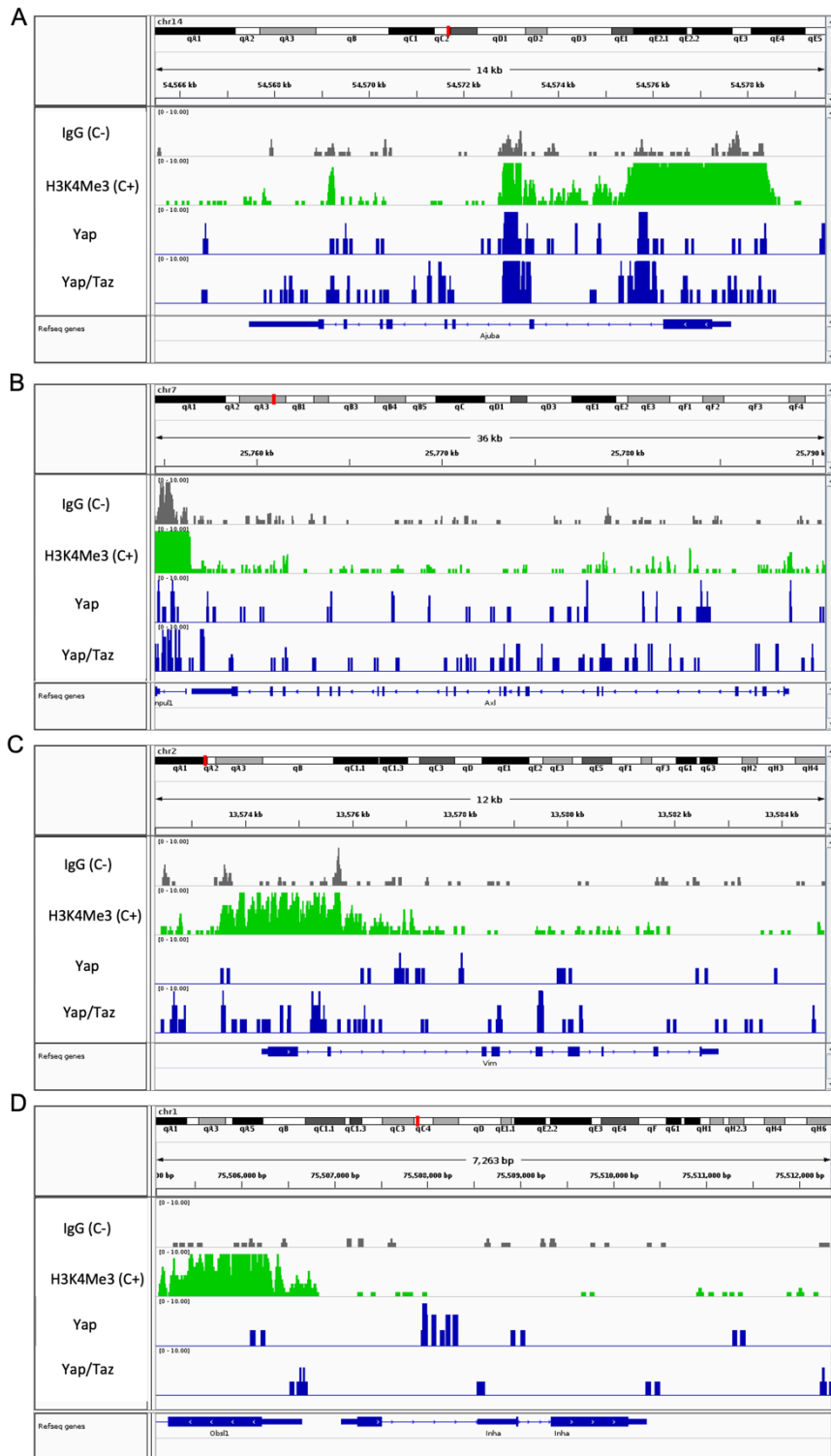


Figure S1: IGV Study of gene presence in control positive, negative, Yap, Yap/Taz conditions. (A) Confirmation of Ajuba presence in the sample. (B) Confirmation of Axl presence in the sample. (C) Confirmation of Vim presence in the sample. (D) Confirmation of Inha presence in the sample.

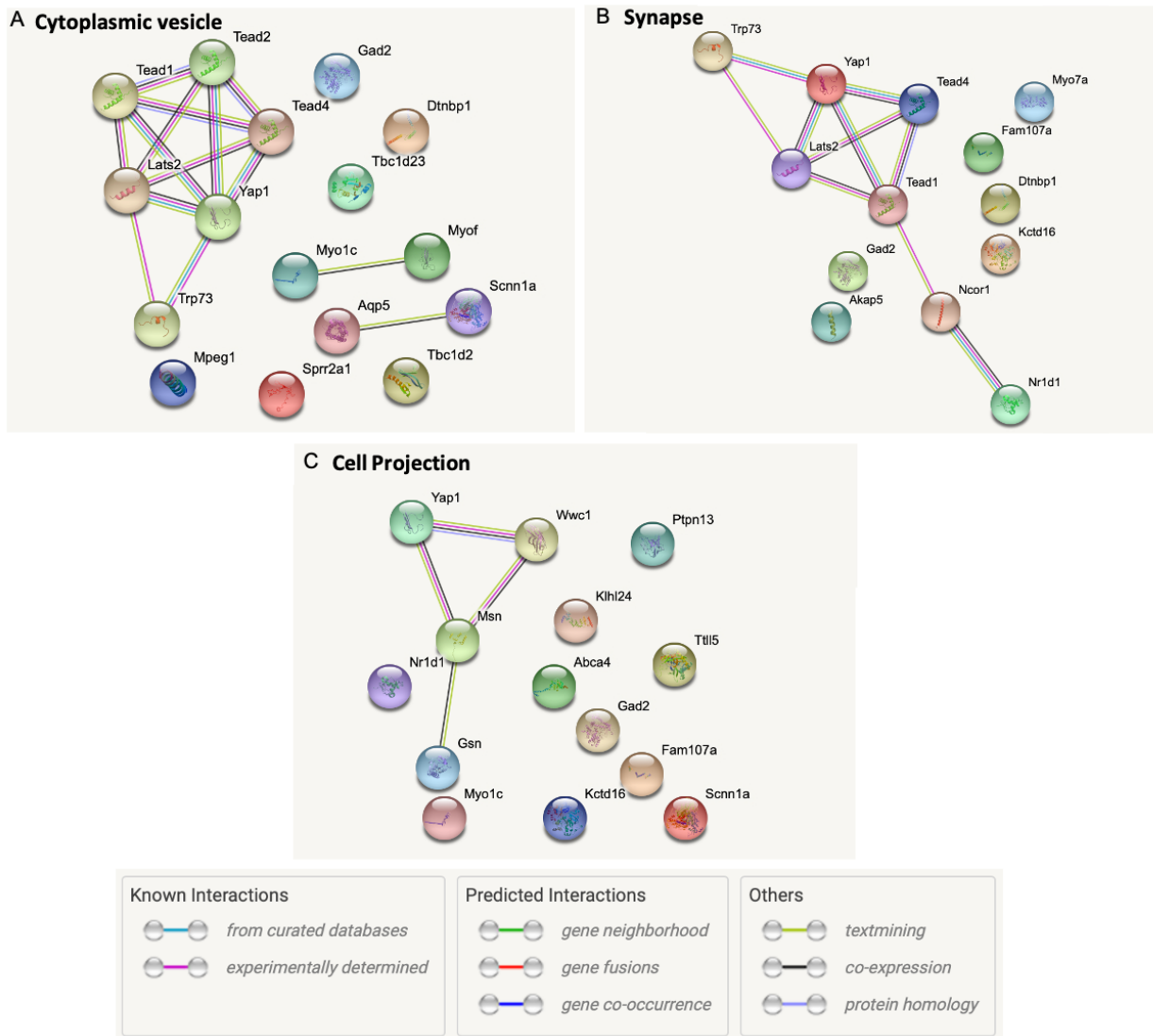


Figure S2. Interactions study between the gene dataset and Yap protein using STRING database. (A) Interaction between Yap and genes associated with Cytoplasmic vesicles. **(B)** Interaction between Yap and genes associated with Synapse. **(C)** Interaction between Yap and genes associated with cell projection.