University of Insubria

PhD course in Experimental and Translational Medicine
Medicine and Surgery Department
XXXIII cycle

PhD Thesis:

ERG overexpression promotes IFN signalling activation, KRT deregulation and influences secretome composition in an *in vitro* PCa model

Supervisor:

Prof. Ian Marc Bonapace

Candidate:

Dr Maria Giovanna de Marino
ID: 735026

Academic Year: 2019- 2020

Table of content

Abstract 7

Introduction 8

1. Prostate Cancer 8
1.1 PCa biomarkers and TNM Staging 8
1.2 The Heterogeneity of Prostate Cancer 9
1.2.1 Morphological Heterogeneity and Grading 10
1.2.2 Molecular Heterogeneity 11
1.3 PCa Carcinogenesis 12
1.3.1 Prostatic epithelium 14
1.3.2 Molecular Landscape 15
1.3.2 *TMPRSS2:ERG* Fusion gene 18
1.3.3 ETS-Related Gene (*ERG*) 21

2. Epithelial to mesenchymal transition in tumorigenesis 22
2.1 EMT in PCa carcinogenesis 23

3. Epigenetics in PCa 26
3.1 DNA methylation 26
3.2 Histone modification 26

4. The tumour microenvironment 29
4.1 Cells of the tumour microenvironment 30
4.1.1 Cancer-associated fibroblasts 31
4.2 The role of secretome 32

Aim of the thesis 35

Materials and Methods 37

1. Cellular Cultures 37
1.1 Cell lines and culture medium 37
1.1.1 RWPE-1 cells 37
1.1.2 RWPE-1 cells induction 38
1.1.3 Growth curve experiments 39
1.1.4 Wound healing assay 39
1.1.5 Isolation of NAFs and CAFs 39
1.1.5 Fibroblast cultivation 40
1.1.6 Production of RWPE-1 cells conditioned media in K-SFM 41
1.1.7 Fibroblasts exposition to CM from RWPE-1 41
1.2 Subculturing procedures 42
1.2.1 siRNA transfection 42
1.3 Mycoplasma detection 42

2 Analytical Techniques 43
2.1 RNA extraction 43
2.2 RNA Reverse Transcription 43
2.3 Polymerase Chain Reaction 44
2.3.1 Semi-quantitative PCR 44
2.3.2 Quantitative PCR 44
2.4 RNA Sequencing 45
2.4.1 RNA Sequencing analysis 45
2.5 DNA extraction 45
2.5.1 DNA Sequencing 46

2.5.2 DNA Sequencing analysis ... 46
2.6 Bisulphite pyrosequencing ... 47
2.7 Immunofluorence ... 47
2.8 Flow cytometry analysis ... 47
2.9 Protein extraction and quantitation ... 48
2.10 Western Blot ... 48
2.11 ChIP – Chromatin Immunoprecipitation ... 49
2.12 LC-MS analysis ... 49

3 Statistical analysis ... 51

4 Bioinformatics ... 51

Results ... 52

1. RWPE-1 induced cells show a more mesenchymal-like phenotype ... 52
1.1 Immunofluorescence analysis for ERG detection ... 52
1.2 Analysis of RWPE-1 cells growth rate and cell cycle ... 53
1.3 Invasive and migratory properties of RWPE-1 overexpressing ERG ... 54
1.4 Molecular characterization of mesenchymal features in the ERG-induced conditions ... 55
1.5 Timing setting for sequencing analyses ... 57

2. ERG overexpression strongly impacts on the transcriptome profile ... 59
2.1 Explorative analysis ... 59
2.2 Functional analysis ... 64
2.2.1 Enriched network analysis of significant DE genes related to the ERG-induced conditions ... 65
2.2.2 Global network analysis of significant DE genes related to the ERG-induced conditions ... 71
2.2.2.1 Interferon response: type I and II activation (clusters 1 and 2) ... 76
2.2.2.2 Cornification process (cluster 3) ... 77
2.2.2.3 Platelet degranulation (cluster 4) ... 81
2.2.2.4 NABA collagens (cluster 5) ... 82
2.2.3 Comparison of ERG- and AR- dependent global network analysis ... 83

3. The marginal role of DNA methyltransferase in the ERG inducible RWPE-1 cell model 86
3.1 Global DNA methylation evaluation ... 86
3.2 Focused analysis of DNMT3A expression ... 87
3.3 Differentially methylated characterization of ERG induced genes ... 91

4. ERG transcriptional programme is sustained by a huge chromatin remodelling ... 95

5. ERG-dependent secretome alters the phenotype and affects fibroblasts survival ... 97
5.1 Prediction of potential secreted molecules, inferred by RNA-seq and PRAD analyses ... 97
5.2 Proteomic evaluation of CM media with LC-MS ... 98
5.3 Characterization of differentially secreted protein secretion mechanisms ... 103
5.4 Phenotype assessment of fibroblasts exposed to CM ... 104

Discussion ... 107

1. ERG transcriptional programme early affected IFN signalling and estrogenic response 107
1.1 A potential molecular mechanism prompted by ERG overexpression in PCa ... 108
1.1.1 SSPB2 downregulation might affect LMOs' turnover ... 108
1.1.2 IFN signalling activation correlates with a deregulation of estrogenic response ... 110
1.1.3 CAMKIIB-dependent non-canonical STAT1 activation contributes to IFN signalling activity ... 112
1.1.4 TRIM29 and the deregulation of keratin pattern ... 113
1.1.5 ERG-dependent vasculogenesis ... 114
1.2 Comprehensive evaluation of the hypothetical molecular mechanism ... 115

2. An altered chromatin accessibility might sustain ERG transcriptional programme ... 115

3. ERG overexpression strongly impacted on tumour- tumour microenvironment crosstalk 116

Conclusions and future perspectives ***119***

Appendix ***133***

Appendix I – Ethics Committee approval protocol n. 355384628 ..... **133**

Appendix II – Primers **135**

Appendix III –Antibodies **136**

Appendix IV –Network at 120 hours **137**

Appendix V – Signature of Basal and Luminal genes **139**

Appendix VI - 77 Common genes (predicted) secreted **142**

Abstract

The high incidence of *TMPRSS2:ERG* fusion gene and ERG subsequent overexpression is still poorly characterized from a clinicopathologic point of view. This project therefore aimed to investigate if ERG overexpression could be defined as a driver alteration that does the groundwork for tumour initiation, especially through a significant impact on tumour microenvironment. Therefore, we aimed to identify differentially expressed and methylated genes that might be involved in tumour-tumour microenvironment crosstalk, ruling molecule trafficking and affecting stroma cells. The transcriptome analysis showed that ERG overexpression promoted IFN signalling activation, KRT deregulation and estrogenic response downregulation and highlighted the major role of STAT1 and its targets' transcription.

In addition, the DNA-seq analysis revealed that no significant DNA methylation occurred (at least in the selected cell model and time points investigated), while a preliminary analysis on histone modifications highlighted their pivotal role for ERG-dependent transcriptional programme, aimed to cell reprogramming.

Notably, ERG-dependent transcriptional changes reflected on an altered secretome composition, influencing molecule secretion. In particular, the LC-MS analysis of conditioned media (CM) from cells overexpressing ERG stressed the significant enrichment of molecules involved in response to wounding, cell adhesion regulation, chemotaxis and extracellular matrix organization through complex secretory trafficking (both conventional and unconventional). We observed that stroma cells exposed to ERG-induced CM showed an increased survival, suggesting that ERG overexpression played a role on molecule secretion in order to prompt tumour-tumour microenvironment crosstalk.

Introduction

1. Prostate Cancer

Prostate cancer (PCa) is the third leading cause of cancer-related deaths, after lung and colorectal cancer. It generally occurs in men aged 65 years or older. During the last 10 years, the death rate has been reduced by 3.4%, thanks to the early diagnosis and treatment improvements (https://www.cancer.org/cancer/prostate-cancer/about/key-stati stics.html). PCa is due to an abnormal cell proliferation that interests prostate gland tissues, but the related deaths depend on the metastatic growth that affects pelvic and retroperitoneal lymph nodes, spinal cord, bladder, rectum, bone and brain.

1.1 PCa biomarkers and TNM Staging

The PCa treatment is challenging since histologically similar tumours result in different clinical outcomes. Currently, there are no reliable prognostic markers to discriminate indolent tumours from those likely to cause aggressive metastatic disease [1]. Nevertheless, molecular markers are used for the early and non-invasive detection of the disease.

Various diagnostic methods are nowadays used for prostate cancer detection. The prostate-specific antigen (PSA) is a blood test for the early detection but has shown critical issues in the overall recognition of cancer types. Starting from PSA detection, new tests have been proposed as the phi, the 4Kscore or the PCA3 [2]. Particularly, other more recent tests analyse well-known gene changes in cells from urine specimens after a digital rectal exam (DRE), as for the *TMPRSS2: ERG* translocation [3]. Moreover, recent studies have begun to evaluate exosomal biomarkers (*e.g.* survivin and claudin 3) and also tumour microenvironment-associated biomarkers with promising results, that still need a deeper understanding of their diagnostic role [4].

Given the high error rate of listed tests, they are used as an indication, but further efforts are needed to identify more reliable non-invasive tests. If non-invasive screenings in fact highlight abnormalities, further confirmation of PCa can be exert with a transrectal ultrasound, a magnetic

resonance imaging (MRI), or a tissue sample collection with a prostate biopsy. In the diagnostic field, a good knowledge of PCa stage represents an advantage for defining treatments. New technologies are available for an accurate diagnosis, as the MRI that supports the personalized medicine option.

During the last decade, the grading system has been uploaded to finely detect tumour size and localization (T) and its spreading to lymph nodes (N) or metastasis (M). In details, the TNM staging system (*Figure 1*) defines four stages (I to IV) based on tumour extent, where:

- Stage I: a benign tumour,
- Stage II: a not-spread prostate tumour,
- Stage III: a spreading tumour to nearby tissues,
- Stage IV: a metastatic tumour.

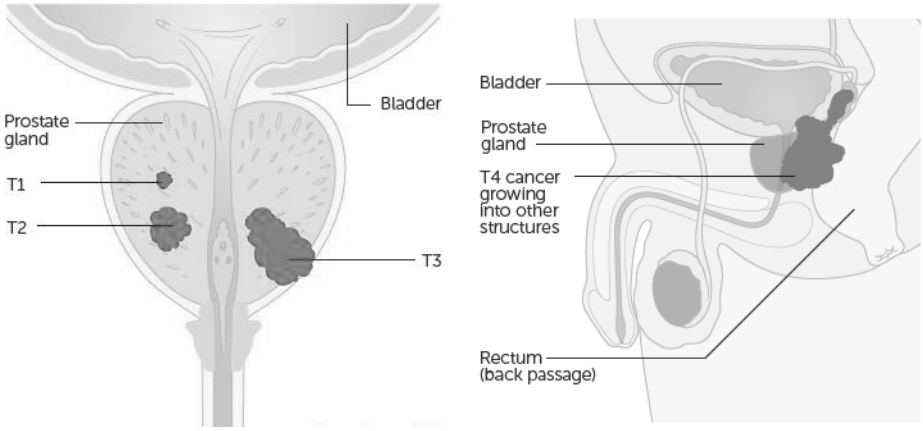


Figure 1| TNM staging scheme of prostate cancer (Cancernz.org).

1.2 The Heterogeneity of Prostate Cancer

The impact of tumour heterogeneity has become pivotal for its clinical drawbacks and this term has been used to define the heterogeneity (i) within a tumour and its microenvironment, (ii) the diversity of its cells and their interactions with each other and also (iii) the miscellaneous genetic profile of tumour cells. Particularly in PCa, heterogeneity has been detected in primary tumours and metastases, and also within separate or individual tumours. Genetic alterations concerning tumour suppressor genes or oncogenes give only a partial view of PCa complexity. Recent studies suggested that heterogeneity correlates with recurrently mutated genes (DDR genes, *RASGRF1*,

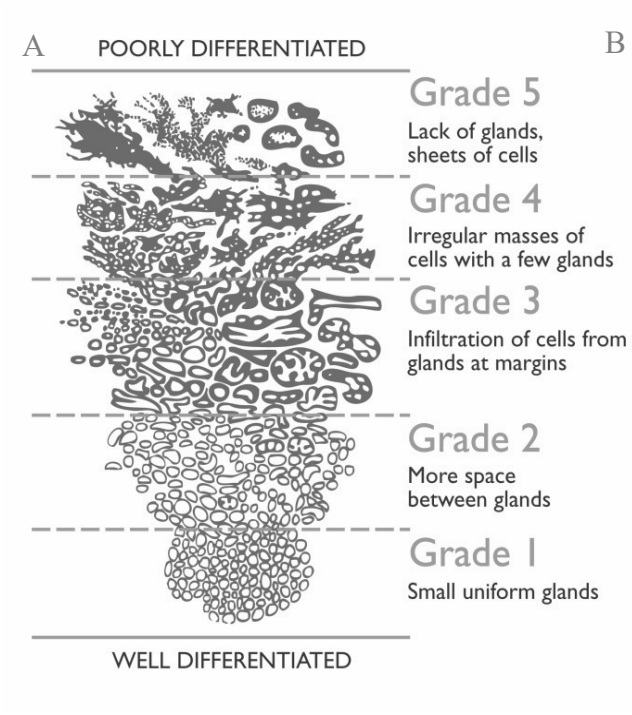
EPG5), somatic variants, SNVs, CNVs, clonal evolutionary processes, variant spectrum and epigenetic differences [4].

1.2.1 Morphological Heterogeneity and Grading

2016 WHO classification of prostate cancer histological categories [5] had highlighted that 95% of tumours are acinar adenocarcinoma with specific morphological patterns that can be associated to meaningful subtypes, in order to define a grading system. The most recent grading score is known as the Gleason score system [6] that focuses on tumour morphology, histology, pathology and defines a comparison between tumour extent and normal tissue.

The Gleason score system (*Figure 2A*) uses a I to V scale to describe the tumour growth cell pattern. In details, each patient is defined by two grades: the primary grade is referred to the crucial area of the tumour, while a second one is assigned to describe cells nearby the relevant area [7][8]. To simplify the Gleason score system, summary groups have been implemented (*Figure 2B*):

- Gleason Group I: Former Gleason 6,
- Gleason Group II: Former Gleason 3+4=7,
- Gleason Group III: Former Gleason 4+3 =7,
- Gleason Group IV: Former Gleason 8,
- Gleason Group V: Former Gleason 9 or 10.



B

TRADITIONAL GLEASON SCORE	NEW GRADING SYSTEM GROUP 1
GLEASON 3+3=6 Only individual discrete well-formed glands	GRADE 1
GLEASON 3+4=7 Predominantly well-formed glands with a lesser component of poorly-formed/fused/cribiform glands.	GRADE 2
GLEASON 4+3=7 Predominantly poorly-formed/ fused/cribriform glands with a lesser component of well-formed glands.	GRADE 3
GLEASON 4+4=8 Only poorly-formed/fused/cribriform glands or -Predominantly well-formed glands with a lesser component lacking or -Predominantly lacking glands with a lesser component of well-formed glands.	GRADE 4
GLEASON 9-10 Lacks gland formation (or with necrosis) with or without poorly-formed/fused/cribriform gland.	GRADE 5

Figure 2| **A)** Gleason score system: I-V scale. **B)** Gleason grading group derived from the traditional Gleason score classification [8].

1.2.2 Molecular Heterogeneity

PCa presents three level of molecular heterogeneity: interpatient, intra- and intertumoral. Concerning interpatient heterogeneity, tumours have shown a unique set of abnormalities that could be imputed to subtypes, with a biological and prognostic meaning. Thanks to the analyses of molecular alterations, some patterns have been identified and classified in seven subgroups related to *ETS* genes and somatic alterations as the *SPOP1*, *FOXA1*, *IDH1* genes [9][10]. Although this, a low rate of tumours could be ascribed to these subgroups and therefore the tumour heterogeneity has become a key point to investigate, in order to focus on more decisive diagnosis, prognosis and treatments [11]. Since multifocality occurs in almost 87% of adenocarcinomas, recent studies placed the intertumoural heterogeneity into the interpatient one, defining multifocal tumours as molecularly peculiar [12].

The intratumoral heterogeneity is instead strongly bound to clonality. Starting from a single focus, the tumour evolves through a branched process acquiring driver and passenger alterations that cooperate for the new clones rise, enhancing tumour spreading.

Overall, the heterogeneity is still understudied, therefore only preliminary approximations can be addressed regarding this topic.

1.3 PCa Carcinogenesis

The prostate gland has been defined in distinct morphologic zones (peripheral, transition and central), according to the localization of prostate diseases. The peripheral zone can undergo acute inflammation that in most cases is also associated with atrophy; some cases reported an inflammatory status also in the transition zone, but this event is likely to be associated with benign prostate hyperplasia (BPH) [13]. Moreover, the peripheral zone also harbours prostate cancer lesions, while the central one is relatively resistant to carcinoma and other diseases. In *Figure 3* zones' localization and prevalence of diseases are reported.

The malignant transformation of prostate cancer is a multistep process, involving many factors. Some inflammatory status as Prostate Inflammatory Atrophy (PIA) and Prostate Intraepithelial Neoplasia (PIN) have been characterized as pre-malignant levels that could potentially turns to invasive carcinomas. PIA is the atrophic lesion that entails a reduction of the gland volume and stroma. In particular, evidence suggested that the focal atrophy can be a precursor of a pre-cancerous status (the high-grade PIN) since expression changes of various hallmark genes were often recognised in both pathological conditions [14][15][16]. The proposed model for the early prostate neoplasia progression is then defined by prior focal atrophy due to many factors (repeated inflammations, diet, onset of autoimmunity) that causes the recruitment of leucocytes and phagocytes. The last ones are the main responsible for the consequent epithelial cell regeneration,

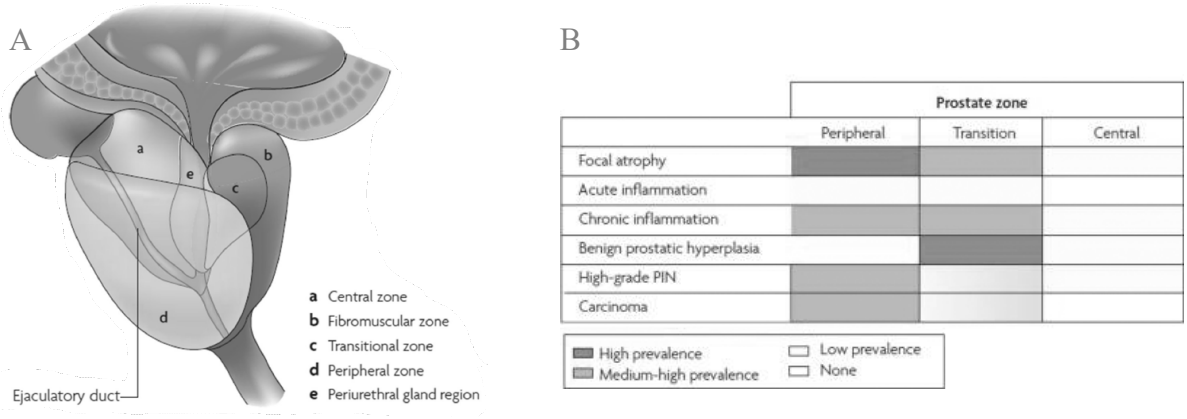


Figure 3|**A)** Organization of prostate gland zones (indicated from a. to e. in legend). **B)** Prevalence of disease for each prostatic zone [14].

since phagocytes locally release reactive oxygen and nitrogen species that cause cell injury and death. Cell proliferation becomes massive, and cells start showing uncertain phenotype between basal and mature luminal cells. Somatic alterations that interest the hypermethylation of the π-class glutathione S-transferase (*GSTP1*) and telomere shortening occur and correlate with an increased genetic instability that might promote the PIN status (low and high-grade) and the early prostate cancer onset. Additional genetic alterations (activation of the ETS family members and *MYC*, as the loss of tumour suppressors as *PTEN*) may then drive to a more malignant status [17][18][19]. Subsequently, in some cases, prostate adenocarcinoma might even turn to castration-resistant prostate cancer (CRPC), even though androgens are no longer physiologically provided. The pivotal features for this resistance have been identified as somatic changes that interest the Androgen Receptor expression (*e.g.* changes in *AR* splice variants), the involvement of growth factors that activate androgen-responsive genes in the absence of androgens, and correlations have been also detected between CRPC and high levels of IL-6 and IL-6R that still sustain AR-transcription [20][21][22]. A brief scheme of the aforementioned process is illustrated in *Figure 4*.

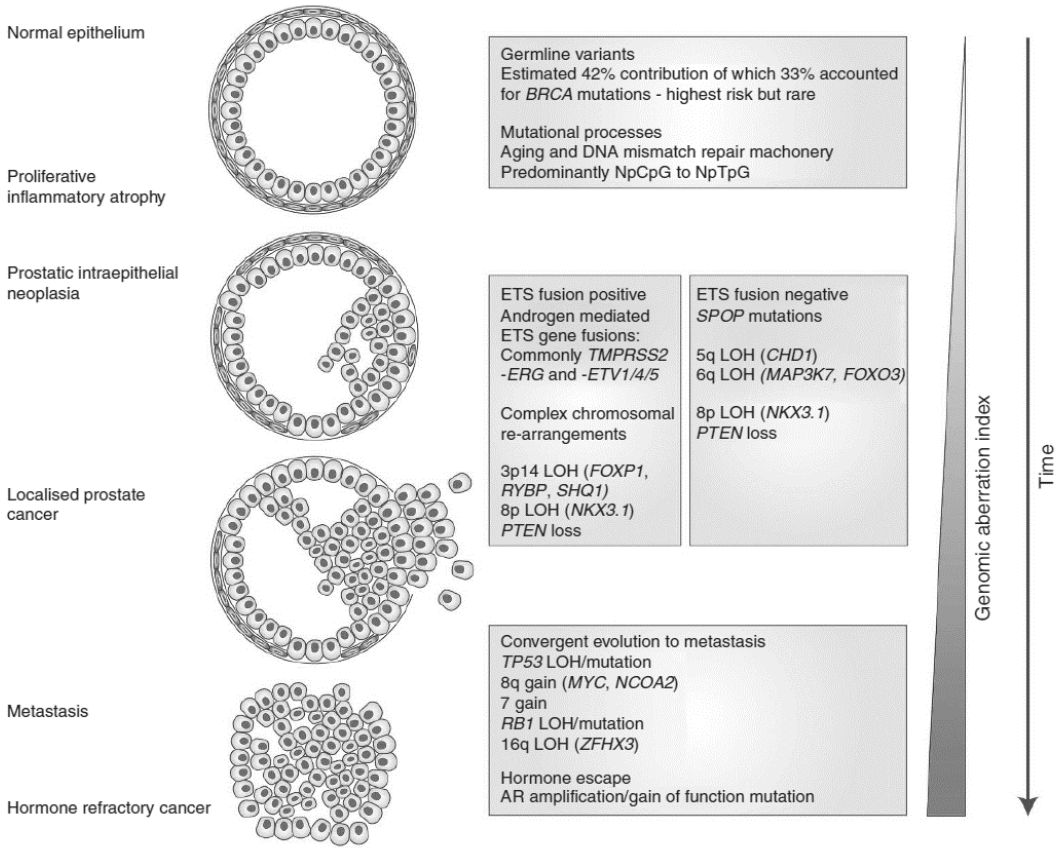


Figure 4| Pathological phases of tumour initiation and progression on the left, relating genomic alterations on the right [19].

1.3.1 Prostatic epithelium

Prostatic epithelium is characterized by three cell types: (i) luminal, (ii) basal, (iii) intermediate cells (*Figure 5*). Luminal cells are androgen-dependent differentiated cells and the lonely with secretory features; they produce and release PSA and express AR, KRT8/KRT18. Basal cells are located between the basement membrane and luminal cells, are not differentiated and typically KRT5/KRT14 positive. Furthermore, intermediate cells are luminally-located and resulted positive for both luminal and basal markers. In addition, neuroendocrine cells are also present in epithelium and generally dispersed among basal and luminal ones. A reduction or a loss of basal features is considered a pre-neoplastic lesion, since adenocarcinomas show luminal phenotype. Tumour initiation thus might be supported by a PCa cell-of-origin, identified as a prostatic stem cell (PSC) that acquire tumour-promoting mutations. The cell-of-origin identification could be useful for better interpreting the tumour heterogeneity; unfortunately, evidence is still insufficient to clearly characterize it [23][24].

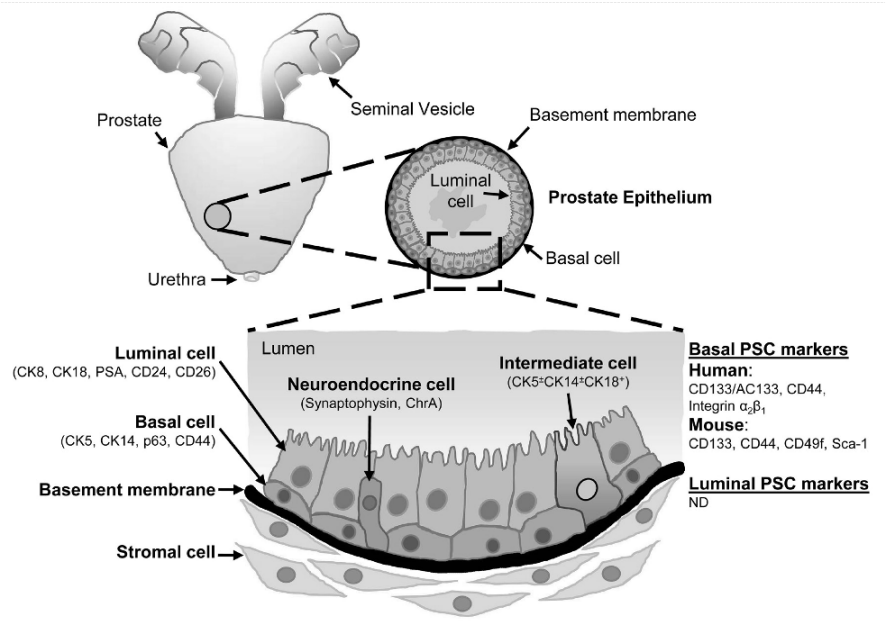


Figure 5| Prostatic epithelium cell types. For each type characteristic markers were reported [22].

1.3.2 Molecular Landscape

As previously highlighted, the molecular heterogeneity defines a low rate of recurrence of genetic abnormalities within tumours. The main alterations that sustain PCa initiation and progression are, of course, genetic predisposition genes and somatic mutations that could exploit oncogenes and result in the loss-of-function of tumour suppressor gene. In details, some pathways and alterations have been recognised to be deregulated in PCa and a variety of tumours, while others are more prostate-specific.

a) Pathways and genetic alterations concerning a spectrum of cancers

• Phosphatinosite 3-kinase pathway

The PI3K pathway is altered in many tumour types and it affects cell proliferation, survival and invasion, thus giving a metastatic potential to PCa. In particular, the most frequent alterations concern the Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha (*PI3KCA*) and the Phosphatased Tensin Homologue (*PTEN*), where the LOH occurs in 60% of patients. PTEN is not properly a PI3K pathway partner in fact it plays as an inhibitor of the PI3K/AKT/mTOR cascade, but its alteration leads to the signalling activation [25][26].

• Ras/Raf/mitogen-activated protein kinase pathway

The MAPK pathway could be activated from growth factors, cytokines, adhesion cell molecules and also by Ras/Raf that might play a role in AR transcription activity. Moreover, evidence suggests that this signal cascade can contribute to tumour progression and independence, but these roles are still not well-clarified [27][28].

• DNA repair genes

A limited number of PCa has shown aberrations in DNA repair genes, including *ATM* and *BRCA2*. Specifically, *ATM* is the cell cycle controller checkpoint, and its aberrant expression causes genomic instability (as resulted in PIN). *BRCA2* is also involved in genomic stability and the homologous recombination for dsDNA break, therefore the biallelic loss that occurs in 3% of cases increases the tumour risk [29].

• TP53

The tumour suppressor gene *p53* resulted mutated in approximately 47% of patients and therefore the genomic stability maintenance is deeply altered [30].

• RB1

Retinoblastoma (*RB1*) is a well-known tumour suppressor gene found mutated in almost 23% of PCa cases. The gene is commonly altered in CRPC, since its impairment seems to correlate with the downstream activation of the androgen receptor pathway, thus conferring resistance [31].

• MYC

MYC encodes for a transcription factor that is highly involved in cycle progression, cell survival and tumorigenesis. *MYC* is often amplified in PCa and it is especially altered in androgen-independent metastatic disease, being a downstream effector of the AR pathway. Moreover, the entire chromosome 8 is often amplified, also altering other oncogenes [31]. From a clinical point of view, evidence suggests that the high nuclear ERG/MYC overexpression correlates with a subset of locally advanced PCa that frequently causes extraprostatic extension, a well-accepted poor prognosis factor [32].

• FOXA1

The Forkhead Box A1 (*FOXA1*) gene is mutated in 3% of prostate cancers. The wild-type levels of *FOXA1* are crucial for the inhibition of cell motility, EMT and AR-independent metastases. In details, FOXA1 influences the AR binding specificity and also its transcriptional programme. This evidence might indicate that *FOXA1* mutations can contribute to abolishing its tumour suppressor activity [33].

• APC

The Adenomatous polyposis coli (*APC*) mutations have an incidence of 3-10% in PCa patients. The protein is a negative regulator of the Wnt pathway, fundamental in cell-cell adhesion, migration, transcriptional activation and apoptosis. The main *APC* alteration concerns its promoter hypermethylation, enhancing EMT and cell migration [34].

b) Prostate cancer-specific abnormalities

• The Hedgehog pathway

The Sonic Hedgehog pathway is involved in prostate growth, patterning and tissue polarity. Various studies showed high levels of its target genes (*GLI* and *PTCH*) in different prostatic cell line models [35], supporting the SHH pathway involvement in the progression to advanced and metastatic PCa (no expression has been detected in normal adult prostate) [36].

• NK3 Transcription Factor Related Locus 1

The *NKX3.1* homeobox is a transcription factor that influences prostate cell fate, inhibiting cancer initiation and it has an androgen-dependent regulation. To this extent, its downregulation is a pivotal event for progression and GWAS identified a mutation of the intron region of a non-coding RNA on chromosome 9 (*PCAT19*) that reduces the affinity for *NKX3.1* binding and therefore sustains the predisposition and aggressiveness in PCa [37].

• LRF

The Leukemia/Lymphoma Related Factor (*LRF*) is a proto-oncogene with tumour suppressive role in PCa and its expression is lost in almost 50% of patients, concomitantly with *PTEN* loss. Physiologically, LRF interacts with SOX9, antagonizing it. In tumour progression then SOX9 is activated and prompts invasion, cytoskeleton alterations as also EMT [38].

• PLZF

The promyelocytic zinc finger transcription factor (PLZF) plays a relevant role in cell proliferation, differentiation, stem cell maintenance and also in innate immune development. PLZF is crucial in AR regulation, inhibiting its functions and activating the mTORC1 inhibitor of the mTOR signal. 15% of PCa cases present a homozygous deletion of *PLZF* with its deregulation that has been associated with a worse prognosis of primary prostate cancer [39].

• CHD1

The chromo-domain helicase DNA binding protein 1 (CHD1) is mutated with an incidence of 26%. *CHD1* is a tumour suppressor gene involved in different mechanisms as the chromatin remodelling or the recruitment of HR for the dsDNA break repair and its deletion causes cell invasion. Terenshchesko *et al.* [40] found that CDH1 loss correlated with the ETS-fusion negative cancer initiation, and it could be responsible for the resulting genomic instability [41].

• SPOP

In 15% of PCa cases, *SPOP* (speckle-type PO2 protein) is affected by nonsynonymous point mutations that modify its tumour suppression function. In normal condition, the SPOP protein degrades the AR activator SRC-3, inhibiting the AR signalling pathway [42].

• AR

The androgen signal is pivotal in PCa. In normal prostate cells, AR binds HSPs and is retained in the cytosol. The DHT availability causes in turn AR-DHT translocation into the nucleus and the binding to regulatory regions of targets genes. In PCa, *AR* commonly undergoes diverse mutations (gene amplifications, point mutations, alterations in splicing that lead to new active variants) that bring to an increased activity. In details, AR regulates different androgen-downstream targets as secreted elements (KLK3, KLK2), stimulators of growth (IGF1R, APP), PI3K members and transcription factors, metabolic enzymes (CAMKK2) and cell-cycle regulators. Moreover, AR is also responsible for the gene fusion between *TMPRSS2* and *ERG* genes, the most frequent alteration in PCa [43].

Interestingly, AR lesions might emerge as a mechanism of resistance against therapies, since they are absent in primary tumours [31] and are more common in metastatic CRPC [44]. Although the AR role in prostate cancer progression has been elucidated, the therapeutic strategies to target it need still to be improved.

1.3.2 *TMPRSS2:ERG* Fusion gene

The most common alterations detected in PCa are the ETS-fusions, with an occurrence of 50% of cases and therefore, tumours have been subdivided into two subclasses as ETS- positive or

negative. The two subclasses define different gene expression profiles [45] and, particularly, ERG-rearrangement is associated with different somatic copy number aberrations and the concurrence of TP53 and PTEN lesions that might define a distinct tumour entity [31]. The main relevance of this classification implies potential therapeutic targets. As an example, it has been detected that the poly(ADP-ribose) polymerase 1 (PARP1) is strictly related to ETS proteins, in fact the use of PARP1 inhibitors reduce ETS-fusion positive PCa xenograft growth [46].

The ETS-related gene (*ERG*) is a member of the ETS transcription factors, physiologically involved in various processes like embryogenesis, vasculogenesis, angiogenesis, haematopoiesis and neuronal development [47]. ETS members share a conserved 85 amino acid sequence, binding to the consensus sequence (5' GGA(A/T) 3'); some members also share a conserved 80 amino acid sequence, the pointed domain (PNT), that has been characterized as a kinase and transcriptional co-regulator site of interaction [48]. Particularly, all members exhibit low binding selectivity, thus protein-DNA interactions and promoter targeting needed to be mediated by TFs [49].

50% of PCa tumours harbour the result of a chromosomal translocation between the Transmembrane Protease Serine 2 (*TMPRSS2*) and *ERG*. These fusions imply an androgen-dependent induction of ETS targets that are certainly involved in tumour progression. Tomlins *et al.* [50] have analysed the role of ERG in benign immortalized prostate epithelial cell lines and demonstrated that ERG overexpression does not affect the cell cycle, and neither can transform cells. Interestingly, the overexpression promotes cell invasion thanks to the downstream activation of a network comprising metalloproteinases and plasminogen activators.

In details, *TMPRSS2* is a prostate-specific androgen-regulated gene that is expressed constitutively through androgen-responsive elements (ARE) in its 5' UTR. The *TMPRSS2: ERG* fusion gene (TE) seemed to be an early event in prostate carcinogenesis and has been detected in High Grade Prostate Intraepithelial Neoplasia (HGPIN), while no alterations were found neither in BPH nor in PIA [51][52]. TE is the results of a chromosomal translocation on chromosome 21 probably due to an interstitial deletion of a 2.8 Mb region (*Figure 6*), since both *TMPRSS2* and *ERG* are on the same chromosome [53]. The resulting fusion gene is constituted by the 5' untranslated sequence of the androgen-regulated *TMPRSS2* and *ERG*, causing ERG androgen-dependent transcription.

Alternative splicing causes the transcription of various TE fusion variants, but the most common is defined by the 4^{th} to 11^{th} exons of *ERG* gene [54]. A synthetic scheme of variants is reported in *Figure 7*. Interestingly, a recent meta-analysis has investigated the clinicopathological aspects of TE occurrence. The fusion gene was associated with stages (no statistical significance

with N), metastasis and Gleason score, while no associations were detected with tumour volume, PCa recurrence or deaths [55][56]. Even though the TE fusion has always been considered as a promising predictive biomarker, contrasting results do not allow a more reliable clinic value.

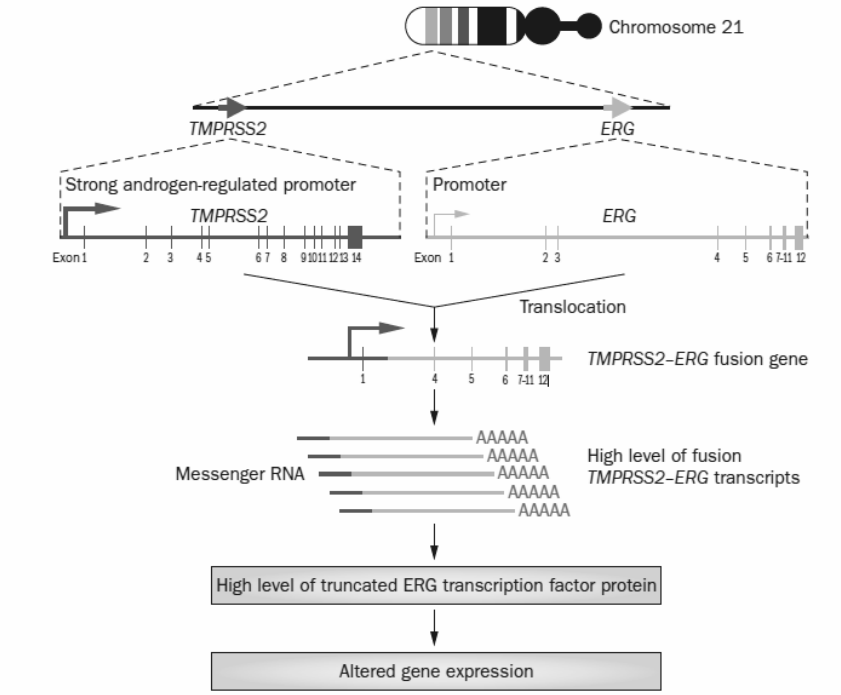


Figure 6| The androgen-dependent promoter of TMPRSS2 fuses with *ERG* gene. Androgen regulation prompts ERG overexpression and defines an altered gene expression profile [51].

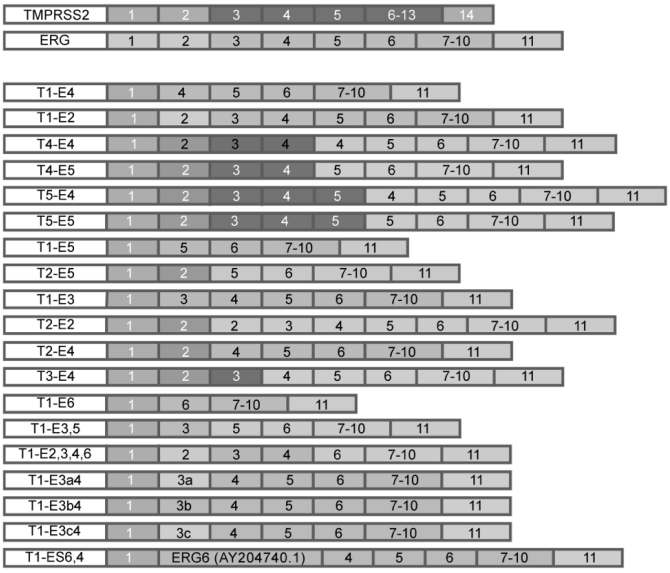


Figure 7| Variants of TE fusion gene. All variants are defined by a peculiar set of exons [53].

1.3.3 ETS-Related Gene (*ERG*)

ERG plays complex roles in physiologic conditions; it is crucial for the embryonic development since it maintains vascular integrity and stability [57] and, hence, it is involved in cell response to vascular inflammation, junction stability and inflammation inhibition acting on IL-8, ICAM-1 and VCAM [58]. Its expression is localized in the nucleus of endothelial cells and increases during vasculature development [48].

In PCa, *ERG* is the most overexpressed oncogene, and it is involved in the malignant progression from PIN to carcinoma (with the concomitant *PTEN* and *TP53* loss). The most relevant ERG-dependent implications relies on cell polarity loss, changes in cell adhesion and invasion, and EMT promotion through the ERG-dependent vimentin expression [59]. The invasive phenotype is supported by the activation of metalloproteases as well of the plasminogen activator and Wnt pathway, but ERG also mediates the upregulation of chemokines, which suggests a relevant role in metastasis [58][60].

PCa is also characterised by a complex transcriptional crosstalk between AR, ERG and the epigenetic reprogramming. In particular, AR and ERG have a specific affinity to bind distal enhancers and proximal promoters but ERG has also been found to cooperate with histone deacetylase complexes (HDACs) and the Polycomb 2 protein EZH2 to sustain the de-differentiation and to negatively control AR transcriptional programme [61]. Besides, ERG overexpression is associated with a strong induction of MYC expression, promoting cell invasion and survival. Many studies have looked in ERG-dependent EMT and it has been demonstrated that the upregulation of the epithelial-to-mesenchymal transition is based on the epigenetic silencing of the Wnt pathway partners in cooperation with HDAC1 and ERG-associated protein with a SET domain histone methyltransferase (ESET) [62]. The complex transcription programme is also enriched by miRNAs that cooperate or antagonise ERG functions. As examples, miR-200c is well-known to prevent the ERG-related EMT process acting on VIM and ZEB1 expression, while miR-145 and miR-30 bind *ERG* 3' UTR reducing its expression, and were found downregulated in ETS-fusion positive tumours [63][64].

The oncogenic potential of ERG as a pivotal character of a wide network comprising differentiation, motility, growth, invasion and epigenetic control, makes it a fascinating target for diagnosis. The most shared direction for treatments is the use of small inhibitors that compromise ERG's interactions with PARP1 or HDAC1 and also molecules that might block its DNA-binding properties; however, further research is required to explore all the opportunities for treatments.

2. Epithelial to mesenchymal transition in tumorigenesis

Although multiple advances have characterized diagnosis and treatment, the major cancer-related deaths are still caused by metastatic disease. The metastasization involves cells from the primary tumour that invade adjacent tissue or disseminate into the circulation. The circulating tumour cells (CTCs) then extravasate and might colonize distant organs, generally being dormant; in some cases, cells evade dormancy, and the process starts. The molecular mechanisms behind this process have just begun to be unrevealed, but evidence suggests that cancer cells might take advantage of the embryonic epithelial to mesenchymal transition (EMT) programme to make epithelial cells acquiring mesenchymal properties. In details, cancer cells lose their typical apical-basal polarity, concomitantly with cell-cell adhesion and adherence to the basement membrane, and therefore invasion and migration could be exerted [65].

Into this frame, the role of the tumour microenvironment of metastasis (TMEM) seems to be pivotal, in fact macrophages, immune cells, matrix and blood vessels all cooperate and interact each other to sustain the endothelial barrier crossing of cancer cells aimed to reach the circulation (intravasation) and to guarantee their exit to distant organs, in the extravasation step. However, the role of CTCs is unclear, since these cells exhibit a strong heterogeneity with different features: epithelial, mesenchymal or both [66].

EMT is enhanced by paracrine signals and develops at the tumour-stroma interface [67]. The motility that enables the EMT, requires that static structures as actin release to guarantee protrusion from the membranes, like lamellopodia [68]. The first step to allow such a mesenchymal switch implies that the major component of adherens junctions (AJ), or rather E-Cadherin (ECAD), is lost and this causes the leak of cell-cell communications, the disruption of the junction itself and favours the release of proteins. Moreover, ECAD loss entails N-Cadherin (NCAD) and VIM upregulation that confer a mesenchymal-like phenotype, which in turn makes cells more resistant to apoptosis and also enables them to produce ECM compounds that sustain their migrative and invasive properties [65]. The EMT process is characterized by the deregulation of markers that resulted:

- upregulated as NCAD, VIM, Fibronectine (FN1), SNAIL, SLUG, TWIST1, FOXC2, metalloproteinases (MMP2, MMP3, MMP9)
- downregulated as ECAD, BCAD, CYTOKERATIN and DESMOPLAKIN) [69].

A summarizing scheme of the process is reported in *Figure 8*.

The EMT-MET process could be defined as a cycling mechanism that allows the correct embryogenic development of the urogenital organs but could also be prompted in the tumorigenic context. To this extent, SRY-related high-mobility-group box (*SOX*) transcription factors are also recruited to regulate male differentiation. SOX9 is involved in both physiologic and pathologic conditions, since it is a critical partner of the Wnt/β-catenin and the fibroblast growth factor signalling pathways and it upregulates AR expression, thus influencing cancer growth and progression [70].

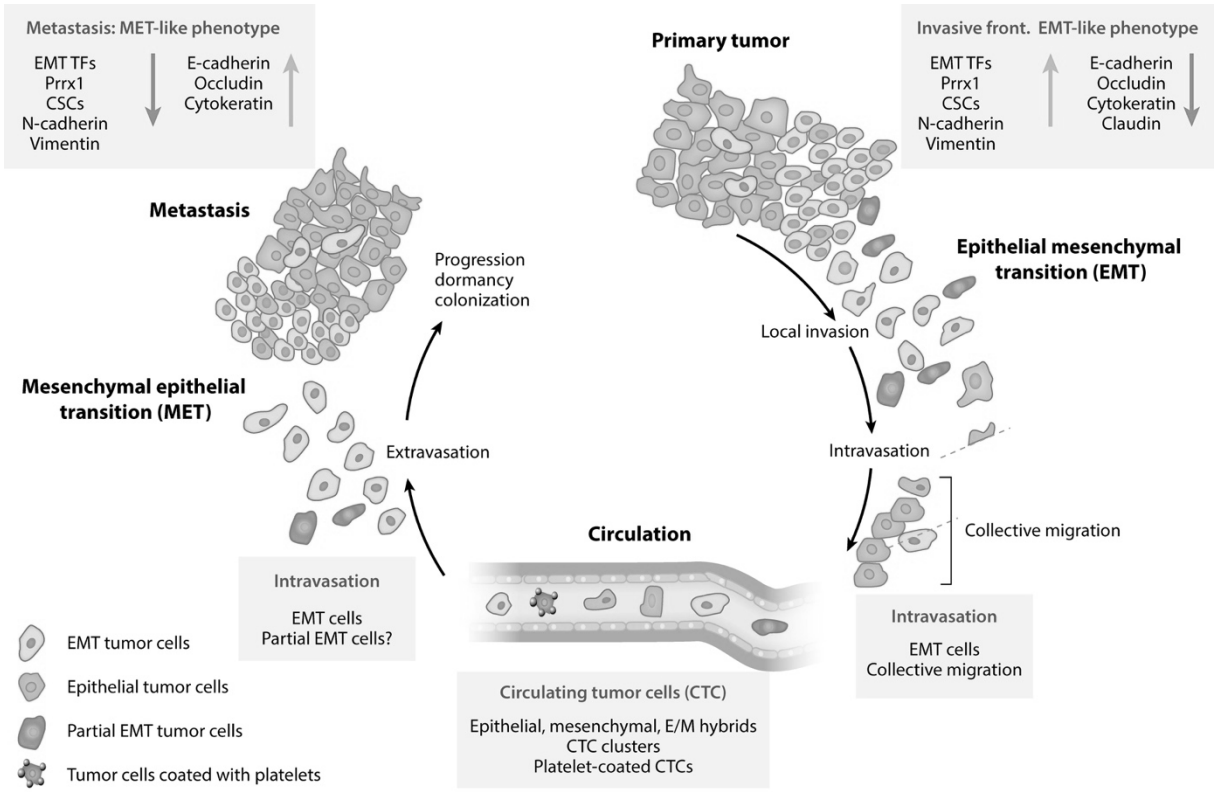


Figure 8| Schematic process of the epithelial-to-mesenchymal transition in metastatic disease. In grey boxes are reported specific markers deregulated in MET and EMT [66].

2.1 EMT in PCa carcinogenesis

Recent evidence assesses the fundamental role of EMT in tumour metastasis, due to changes in the expression of crucial markers that correlate with various pathologic states [71]. The prostate gland is an androgen-dependent tissue, which is deeply regulated by AR signalling that supports epithelial secretory cell survival as cell differentiation and development. However, AR signalling

plays a relevant role also in the tumorigenic context, and its deregulation might promote an AR-dependent EMT induction, even if the exact crosstalk needs still to be elucidated.

To exert EMT, prostate epithelial cells require a cytoskeletal rearrangement and also increased cell plasticity and motility. The crucial actors of this process are mainly TFs and in particular the AR-regulated zinc-finger transcription factors Snail and Slug, members of the TGF- β superfamily [72]. In details, FGF signalling activation promotes SNAI1 expression that acts on E-cadherin promoter in order to silence it [73]. Other EMT regulators in prostate are ZEB1/ZEB2, whose expression is enhanced by AR deprivation and regulated by the miR-200 family, responsible for their targeting. The negative feedback loop involving miR-200/ZEB1/AR plays a central role in pathologic EMT, where it might confer cellular motility, induction of stem-cell properties, apoptotic and senescence prevention [74].

Another key mechanism in prostate development and homeostasis is the estrogenic signalling, involved in cancer progression too. Particularly, the two main ER receptors (ER- α and - β) bind, dimerize and translocate into the nucleus, activating a plethora of targets. In normal prostate, the ER signalling primarily regulates differentiation and proliferation, while in tumours, ER receptors are epigenetically silenced through methylation. In contrast, some recent studies highlighted that in stromal cancer cells, ER-α is upregulated and thus interacts with cancer- associated fibroblasts (CAFs) in order to increase thrombospondin 2 and decrease MMP3 expressions, exerting a protective role and preventing invasion [75]. Furthermore, ER- β role is still controversial even if its deregulation showed a correlation with high Gleason score and poor prognosis, due to the impaired balance of HIF-1α/VEGF. ER- β loss causes VEGF increase that thus drives the EMT (through SNAI1) and also defines a hypoxic state, promoting the mesenchymal switch [76].

TGF- β also is a crucial cytokine regulating apoptosis, immune response and EMT that promotes mesenchymal cell proliferation. In PCa, its suppressive role is altered and, thus, it induces EMT through (i) a Smad-independent mechanism that involves Erk/MAPK, PI3K/AKT and JNK/p38MAPK signalling pathways [77][78], (ii) a Smad-dependent mechanism for the inhibitory functions. In the stromal context, the loss of TGF-β responsiveness of fibroblasts allows an oncogenic regulation of adjacent epithelial cells. In tumours, TGF- β causes the nuclear localization of NF-ϰB, fostering the mesenchymal phenotype [79], and it also acts on the AR signalling crosstalk which activates SNAI1, causing ECAD loss and a cytoskeleton reorganization.

In this complex frame, also EGF signalling defines a relevant crosstalk with prostate cancer. EGF binding to its receptor (EGFR) is fundamental to exert various physiologic and pathologic functions, which include cell growth and vitality as well as migration and invasion [80]. In PCa, their expression is deregulated in an androgen-dependent and independent way and this implies a

well-recognised correlation with an aggressive phenotype, high Gleason score, poor clinic outcome and survival [81]. The EGF/EGFR signalling has been proved to sustain EMT and the relating morphological changes, as the spindle-like cell structures caused by the upregulation of *CDH2* and *FN1* and the downregulation of *CDH1*. In details, EGF activates the protein kinase C (PKC) that in turn stabilizes SNAI1 activation, preventing its deubiquitination by GSK3β [82]. Moreover, the EGF signalling impacts on PCa also through the ROS/STAT3/HIF-1α/TWIST1/NCAD cascade, inducing HIF-1α signalling that in turn upregulates TWIST1, promoting metastatic induction and EMT [83]. Particularly, in TE fusion-positive cancers, EGF interacts with the Src tyrosine kinase pathway for the EMT induction, via the miR-30b silencing that guarantees a higher expression of *ERG* [64][84].

The transcriptional programme that supports EMT is the result of a crosstalk with various signalling specific of the tumour microenvironment that controls the EMT-MET pattern in order to prompt the metastatic growth, making this mechanism potentially useful for clinical insights. To sum up, an overview of signalling involved in the pathologic EMT promotion is presented in *Figure 9*.

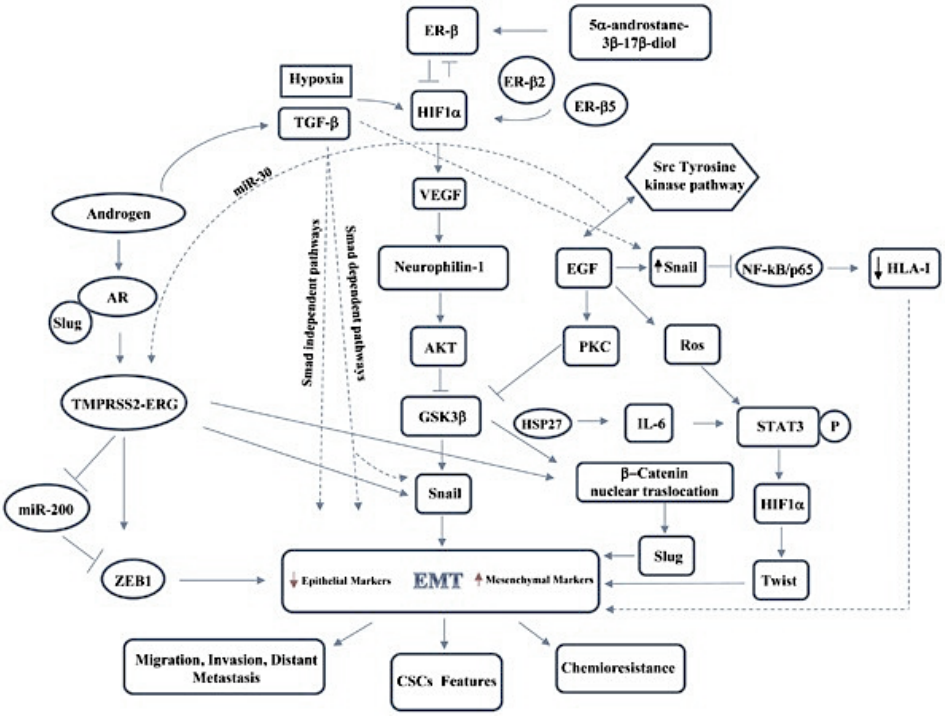


Figure 9| A comprehensive overview of the complex molecular mechanism that regulates EMT in PCa [83].

3. Epigenetics in PCa

3.1 DNA methylation

In recent years the epigenetic dysregulation has been proved to influence PCa progression, as well as genomic alterations. The most known epigenetic alteration in PCa is the global DNA hypermethylation that interests many genes as *GSTP1*, *AR*, *PTEN*, *CDH1* or *p16* that are hypermethylated on their promoter region and thus silenced [85][86][87]. Conversely, hypomethylation causes the upregulation of ERG-dependent genes as the urokinase-type plasminogen activator (*PLAU*) and tissue-type plasminogen activator (*PLAT*) genes that enhance cell invasion and metastatic progression [88]. In details, DNA methylation is a covalent modification that generally occurs in the CpG context, causing the addition of a methyl group to the 5th position of the cytosine carbon ring and forming the 5-methylcytosine (5-mC). The process is ruled by the DNA methyltransferase enzymes (DNMTs) that allow the methyl group transfer. DNMTs are a 4-member family involved in various processes. The DNA methylation maintenance is due to DNMT1 that, through UHRF1 recruitment, can recognise the hemimethylated DNA and fully methylate it, especially in cell division [89]. The *de novo* methylation instead is exerted by two enzymes: DNMT3A and DNMT3B that define new methylation patterns with nonoverlapping actions. Specifically, Pistore *et al.* demonstrated that DNMT3A is upregulated in PCa in order to prompt strong hypermethylation to support EMT [90]. In addition, DNMT3L has no catalytic activity and thus seems to cooperate as an accessory partner of DNMT3A. Furthermore, promoters' methylation generally correlates with a silenced gene expression and recent evidence highlights that instead gene body methylation induces transcription [91].

3.2 Histone modification

The histone proteins are structural components of nucleosomes that characterise chromatin structure and function. Unless highly conserved, histones can undergo posttranslational modifications (PTMs), generally occurring on their N-terminal tails, that affect their properties and thus their functions. PTMs (methylation, acetylation, ubiquitination, SUMOylation, phosphorylation...) cause changes in histone charge and therefore alter the condensed structure, defining active or repressive chromatin status. Many recent studies have identified a crucial role of

histones modifications in PCa onset. Gene expression is enhanced thanks to active marks such as H3K4me3 and H4K9ac, while the repression is driven by repressive marks including H3K27me3, H3K9me2/3 that in both cases interest promoter regions. Histone PTMs as H3K9, H3K18 and H4K12 acetylations or H4R2 and H3K4 demethylations might have a prognostic value in PCa and correlate with Gleason score [92]. Moreover, H3K18ac is associated with an increased risk and tumour relapse, while the mono- and trimethylation of H3K27 is correlated to tumour aggressive growth [93]. Of note, enzymes that ensure these modifications are also crucial in PCa progression. Cell proliferation and migration, for example, have been strongly associated with the deregulation of various methyltransferases as KMT1A, KMT1E, KTM1B, SMYD3 or PRTM5. Moreover, also demethylases play a role in proliferation and malignancy and thus are overexpressed, as LSD1 or the KDM5 family, except KDM5D that instead negatively regulates invasion/progression and is often downregulated in tumour [94].

Other major players in PCa progression are histone deacetylases (HDACs); particularly, HDAC1 is found overexpressed in TE fusion-positive tumours and moreover takes part in a complex regulatory mechanism involving the Polycomb Complex 2 (PRC2) that, through EZH2, allows, as example, the silencing of the cell growth inhibitor DAB2IP, member of Ras signalling [95].

The Nuclear Receptor Binding SET domain protein 2 (NSD2) methyltransferase is associated with an active chromatin status that promotes gene transcription thanks to the demethylation of H3K36. In PCa, it is upregulated by EZH2 and in turn, it modulates TWIST1 expression in order to promote EMT, highlighting the fundamental role of the PRC2 in the tumorigenic context [96].

EZH2 (member of PRC2) plays a role in keeping undifferentiated the embryonic stem cells, catalysing the trimethylation of H3K27 and silencing the expression of diverse targets. It is overexpressed in PCa, revealing its oncogenic role in prompting the epithelial to mesenchymal transition. In 2010 Yu *et al.*, revealed that ERG is a direct regulator of EZH2, promoting its upregulation and, moreover, silencing various targets aiming to perturb cell differentiation [61]. Recently, a PRC2-independent mechanism has been discovered for EZH2 that promotes the AR methylation in an AKT-dependent manner and thus modulates AR recruitment to target sites [97].

Into this frame, also noncoding RNAs revealed their fundamental regulatory roles. These comprise ncRNAs, microRNAs, siRNAs, piRNAs and lnRNASs that can exert oncogenic or tumour suppressive functions. Recent evidence, moreover, suggests that noncoding RNAs might influence proliferation, cell survival, angiogenesis and immune surveillance, playing a role in the tumour microenvironment context.

Cell-derived microvesicles, known as exosomes, are vesicles with a lipid bilayer membrane that could include noncoding RNAs and then allow them to circulate. As examples, miRNA-375 and miRNA-141 have been found included in vesicles in many PCa patients, revealing hidden complex mechanisms of regulation. A deeper understanding of these processes could, in fact, be very meaningful into the diagnostic field, but huge efforts are still required in order to elucidate them [98]. *Figure 10* illustrates the main processes behind the epigenetic regulation in tumours.

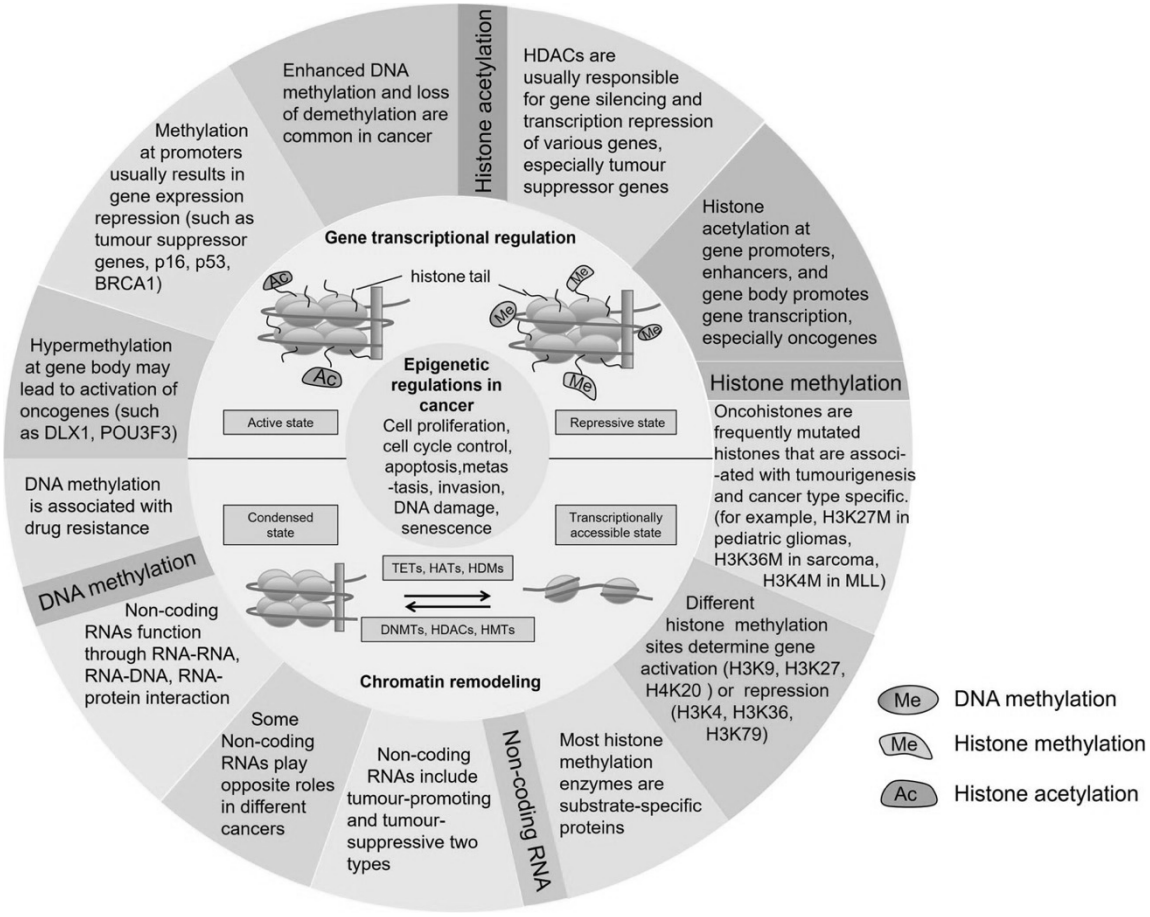


Figure 10| Overview of epigenetic modifications that occur in tumour deregulation, comprising DNA methylation and relevant histone PTMs as methylation and acetylation [91].

4. The tumour microenvironment

The progression towards malignancy that characterises tumours is strongly influenced by the surrounding stroma that mainly affects the tumour cell plasticity, prompting the development of adaptive strategies. This first hint has therefore magnetised the researchers' attention to deepen the composition and the roles of the tumour microenvironment (TME) in progression and metastasis. TME is a complex environment that comprises proliferating tumour cells, tumour stroma, different inflammatory cells as well as blood vessels, and these non-transforming cells generally exert tumour-promoting functions. Reasons behind these processes reside in the tangled intracellular communication that occurs in the microenvironment. The inner communication of TME is due to enzymes involved in inflammation and matrix remodelling processes that alter the physiologic tissue properties, but also growth factors and cytokines/chemokines networks.

To some extent, TME shows similarities with chronic inflammatory states as the wound healing, but also some peculiarities as the presence of macrophages that generally are reprogrammed to release IL-10, prostaglandins or reactive oxygen species aiming to inhibit lymphocytes' recruitment [99][100]. All elements that define TME, synergistically act in order to promote tumour growth.

In details, TME composition (*Figure 11*) comprises various non-malignant cell types as the immune cells (T and B lymphocytes, NK and NKT cells, myeloid-derived suppressor cells, dendritic cells, tumour associated neutrophils), cancer-associated fibroblasts, pericytes,

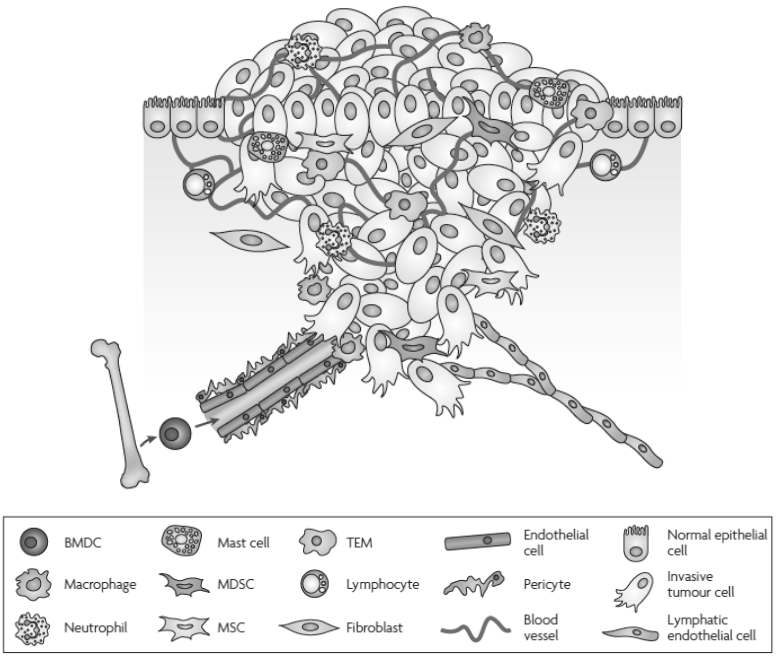


Figure 11| Tumour microenvironment organization and composition [193].

mesenchymal stem cells, vascular endothelial cells, adipocytes and the extracellular matrix compounds.

4.1 Cells of the tumour microenvironment

The exact characterization of the whole TME cell composition is a very complex task, particularly concerning the immune cells and their vast subpopulations. The microenvironment hosts T cells that have different prognostic meanings and between them CD4+ seem to exert pro-tumorigenic activities producing IL-10 and TGF-β with immunosuppressive actions, promoting a worse prognosis [101]. B cells localises at TME adjacent structures and their infiltrative prognostic meaning is still unclear. B lymphocytes play an immunosuppressive role too, especially the subpopulation that produces IL-10 [102].

The major players for migration and invasion are represented by tumour-associated macrophages (TAMs) that are mainly involved in tumour angiogenesis. The crosstalk between TAMs and TME is a very crucial process, in fact macrophages can modify their phenotype according to external conditions and stimuli. In particular, TAM recruitment is promoted by the VEGFs released in the hypoxic context, allowing their accumulation for angiogenesis and cell plasticity promotion [103][104]. TAMs can also derive from the differentiation of myeloid-derived suppressor cells (MDSCs), whose role is still not defined.

The impairment of a correct immune response in the tumour context is also bound to the defective dendritic cells (DCs), which functions are altered due to the inflammatory environment, and in fact, impaired DCs are able to suppress T cell responses [105]. A bivalent role is also played by tumour-associated neutrophils (TANs) which both seemed to enhance ECM degradation to promote angiogenesis, and also inhibit TGF-β response in order to act against disseminated tumour cells. The peculiar role of TANs hence has not clearly been defined [106][107].

Tumour angiogenesis hides complex mechanisms and regulations. The inflammatory signals, as well as the hypoxic conditions, support the activation of quiescent vessels, with the concomitant release of chemokines and factors (VEGFs, PDGFs, FGFs). The derived vascular structures are abnormal with altered shape, branching and functions that facilitate the metastatic process. To this extent, also pericytes are crucial and their depleted coverage in vasculature correlates with poor prognosis in cancer [108][109].

A non-cellular component of TME is the extracellular matrix (ECM) that represents a scaffold for cells and, into this context, plays a role in altering tissue stiffness, elasticity and strength due to the release of angiogenetic growth factors and chemokines. Generally, the typical higher tissue stiffness of tumours is due to cancer-associated fibroblasts, which lay down huge amount of matrix. Moreover, the combined activity of CAFs and TANs that release metalloproteases, contributes to ECM degradation and fosters metastasis [104][110].

4.1.1 Cancer-associated fibroblasts

Recent studies have strongly stressed out that fibroblasts are important promoters of tumour growth and progression. These cells are non-vascular, epithelial or immune cells that embody the crucial component of the connective tissue and are responsible for fibrillar matrix synthesis and deposition. Particularly, they allow the deposition of ECM, the regulation of epithelial differentiation and they also are recruited during inflammation and wound healing. In physiologic conditions, fibroblasts guarantee the ECM homeostasis, ruling the activities of metalloproteinases in the ECM turnover process [111]. Another major role exerted by fibroblasts is the growth factors' secretion, which allows the homeostasis also of adjacent epithelia. In the inflammatory context, these cells improve the ECM secretion and became more proliferating, defining them as "activated" fibroblasts that show some specific markers as the α-smooth muscle actin (α-SMA) and the fibroblast activation protein (FAP), shown in *Figure 12* [112]. CAF role in tumours is controversial, since they initially inhibit the progression during the early stages, and then, after being activated, prompt tumour growth. In details, the crosstalk between the tumour stroma and cancer can follow two pathways: (i) the "efferent" pathway in which cancer cells enhance the reactive response at the stroma level, (ii) the "afferent", where in contrast, the stroma cells prompt malignant features of cancer cells [113][114][115].

Overall, the main roles of CAFs in tumours include the resistance to *anoikis*, a metabolic reprogramming with trophic activities, and in particular, the EMT enhancement. In PCa, Pistore *et al.* have highlighted how CAF-dependent cytokines' release was crucial to activate signalling pathways that caused the deregulation of DNA methyltransferases (except DNMT3A) and UHRF1, in order to define a diffuse hypomethylation that entailed transcriptional activation. Concomitantly, the slight overexpression of the *de novo* DNMT3A might be fundamental to promote cancer stemness and EMT, via GRHL2 and ZEB1 regulation [90] .

In tumours, activated CAFs prompt inflammation through the macrophage recruitment, the neovascularization and the M2 phenotype switch of macrophages that confer immunosuppressive properties to TME promoting tumour growth, migration and invasion through the EMT [116].

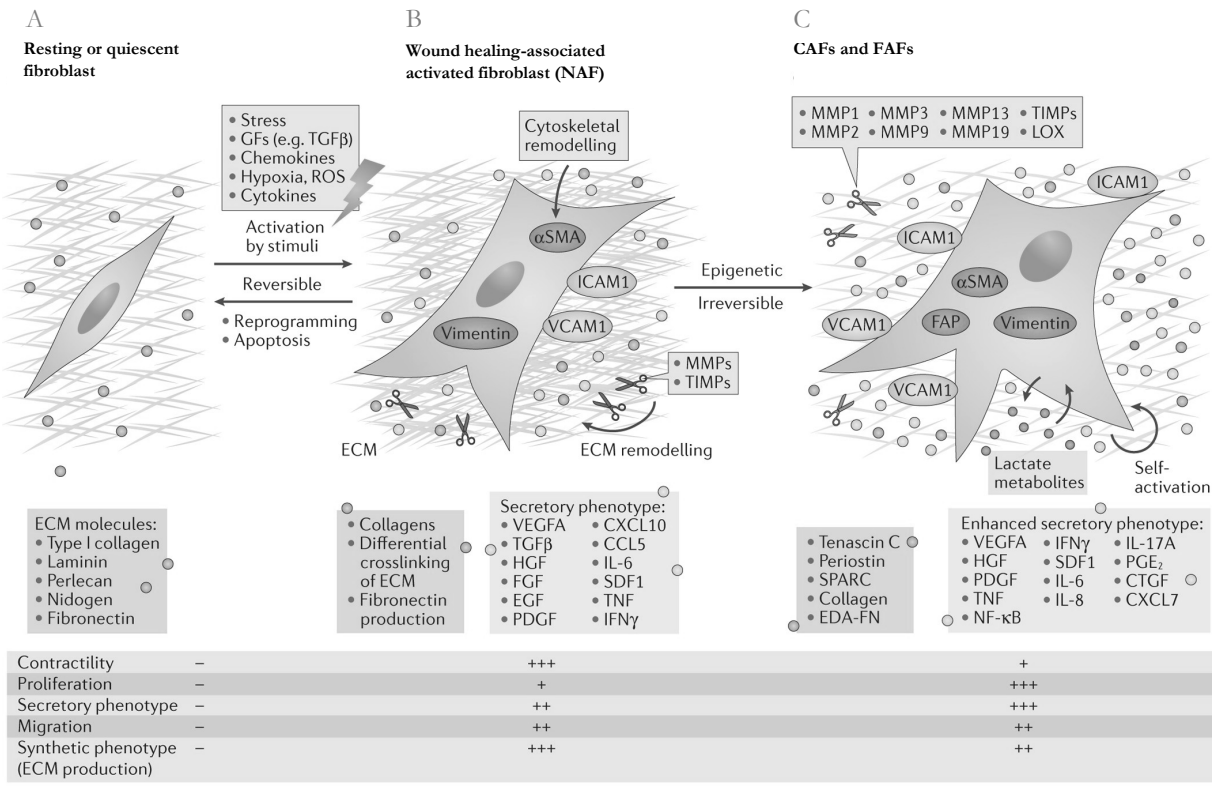


Figure 12|**A)** Inert quiescent fibroblasts embedded in the ECM present a spindle-shape. **B)** NAFs activated by tissue injury, expressing αSMA and vimentin, acquire a stellate shape. **C)** Activated cancer-associated fibroblasts (CAFs) and fibrosis-associated fibroblasts (FAFs) express enhanced secretory phenotype, due to persistent stimuli [115].

4.2 The role of secretome

The secretome comprises a plethora of protein and metabolic compounds that can be released, secreted or detached by cells into the extracellular space and that exert various functions at a cell-cell and cell-extracellular level, affecting proliferation, survival and differentiation [117]. Specifically, two pathways sustain secretion:

(i) the classic pathway in which secreted proteins, after being synthesized by the endoplasmic reticulum, are conveyed to the Golgi apparatus within coat protein

complex II (COPII)-coated vesicles and, after this reorganization are finally released as secretory vesicles [118];

(ii) the non-classic and Golgi-independent pathway which is characterised by a direct translocation to the extracellular space. Proteins are released on the cytoplasm and conveyed through the plasma membrane, as in the case of angiogenetic or inflammatory processes, where fibroblast growth factor 2 (FGF2) and chemokines allow this pattern [119][120].

In particular, the non-classic pathways can be exerted with or without vesicles. In the first case, these intermediate carriers are represented by microvesicular bodies (MVBs) or endosomes that take advantage of autophagosomes or lysosomes, blending with the membrane to directly release their contents. The non-vesicular transport instead is guaranteed through ATP-binding cassette transporters and plasma membranes [121]. Besides proteins or non -protein compounds could be released through extracellular vesicles known as exosomes that have been discovered as pivotal for intracellular signalling regulation, affecting local or distant environment (a schematic view is reported in *Figure 13*).

Due to its involvement in various processes as immune response, coagulation, differentiation and cell signalling, the attention on secretome regulation and function has highly increased in recent years, particularly concerning the tumorigenic context.

In TME, tumour-stroma communication is based on cell-cell interactions as well as on the aforementioned pathways. The physical signals that occur in the microenvironment are thus able to prompt biochemical events that support matrix changes. For example, the *PTEN* loss influences secretome empowers invasion and metastatic activities [122]. Moreover, during the EMT, matrix remodelling is supported also by secretome that increases the expression of MMPs, ECM compounds and factors [123]. Interestingly, secretome activities influence the crosstalk between epithelial and stromal cells. During the early step of cancer progression, in fact, fibroblasts are recruited and activated thanks to the immune and neoplastic cells release of interleukins and growth factors [117]. Due to its promising relevance, deeper analyses on secretome have been done and, in the cancer context, studies on biofluids and conditioned media have been performed in order to identify secreted proteins and biomarkers that are tightly bound to the pathologic conditions. However, in PCa few studies have already characterised the clinical value of secretome and vesicles. Souza *et al.* [124], have demonstrated that extracellular vesicles (EVs) induce alteration in the expression of genes involved in prostate cancer EMT and totipotency (*VIM*, *CDH1*, *CDH2*, *EGFR*, *MDM2*, *MYC*, *ALDH1*, *TMPRSS2*).

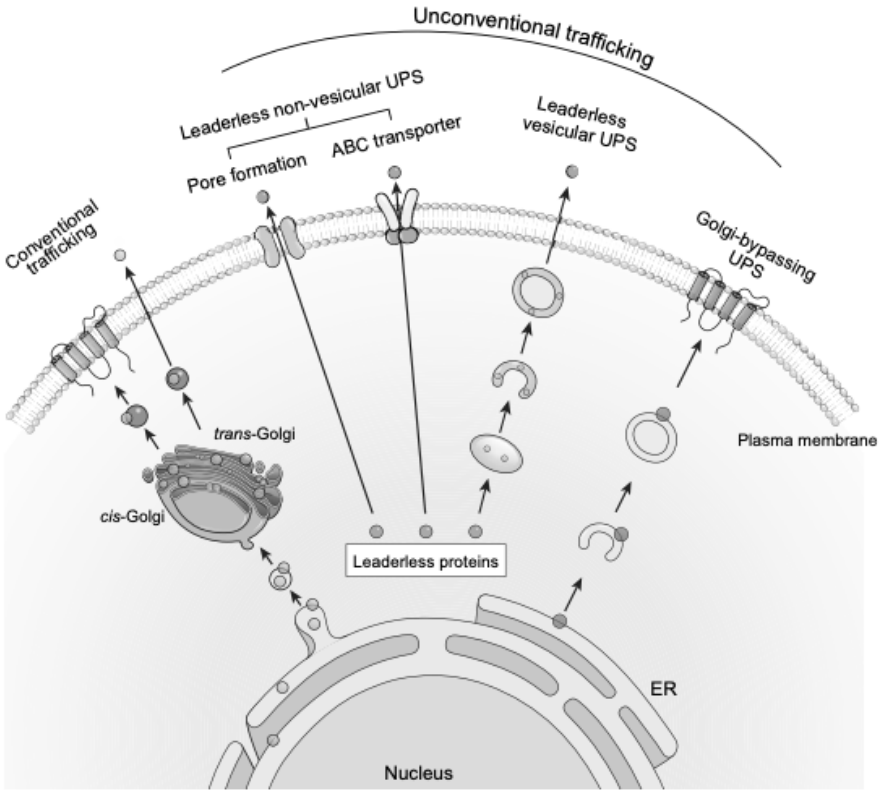


Figure 13| Schematic view of canonical and non-canonical pathway for molecule secretion [120]. The unconventional trafficking comprises (i) leaderless non vesicular UPS, (ii) leaderless vesicular UPS, (iii) Golgi-bypassing UPS.

Aim of the thesis

Prostate cancer is the second leading cause of cancer-related deaths in men and it is characterized by different genomic aberrations and epigenetic deregulations that enhance tumour progression [125]. Particularly, the most frequent alteration concerns the presence of the *TMPRSS2:ERG* (TE) fusion gene that occurs in almost 50%-70% of PCa patients and fosters ERG overexpression in an androgen-dependent manner [126]. Although the high tumour incidence, the clinical outcomes, the prognostic functions as well as the tumour- tumour microenvironment crosstalk still need to be fully elucidated.

Based on these premises, the overall aim of the project was to identify ERG-dependent differentially expressed and methylated genes that might play a role in PCa microenvironment as secreted proteins. In particular, the tumour microenvironment (TME) comprises cellular and noncellular components, where a major role is played by the activated fibroblasts [127]. Furthermore, the tumour-related secretome spans the whole secreted, released or detached proteins of the extracellular (EC) space and, moreover, all the non-protein compounds (DNAs, RNAs, lipids) included in vesicles. Interestingly, these molecules can mediate intracellular signalling pathways promoting invasion, evasion of immune defences and distant organ localization during cancer progression. In this framework, we aimed to clarify if ERG overexpression promotes an altered transcription or enhances the release of factors/molecules that might affect the fibroblast activation and, in general, deregulate tumour microenvironment.

In order to achieve this aim, a benign immortalized epithelial cell line (RWPE-1) engineered with a doxycycline-inducible system to promote ERG overexpression was used. In particular, ERG-dependent differentially expressed genes were firstly intercepted with the RNA-Seq analysis; secondly, the methylome and the chromatin status of ERG-induced RWPE-1 cell were also investigated to contextualise the mechanisms that support ERG-related transcriptional changes, while noncoding RNA role would in future be deepen.

Concerning the secretome analysis, the (predicted) secreted molecules inferred by transcriptome analyses were further confirmed with a proteomic approach. Notably, proteins of conditioned media obtained from RWPE-1 cells (induced or not) were analysed with LC-MS, in the effort to

underpin secreted compounds that were tightly bonded to ERG overexpression, and that might exacerbate the malignant fibroblasts' phenotype.

Coherently to this aspect, the interdependence between ERG overexpression and the TME, as well as the main features of this crosstalk, were also investigated. To this extent, normal- and cancer-associated fibroblasts (NAFs and CAFs, respectively), obtained from advanced prostate cancer patients subjected to radical prostatectomies (see Appendix I for the Ethics Committee protocol), were exposed to conditioned media obtained from RWPE-1 cells (induced or not) and evaluated for their phenotypic and molecular ERG-related features.

In order to validate all results, the public available database known as PRAD (PRostate ADenocarcinoma, The Cancer Genome Atlas, Cell 2015) was used and patients with a characterized *ERG*$^{+}$ subtype were selected and compared with normal samples. A comprehensive overview of our experimental plan is schematised in the following *Figure 14*.

Given the overall plan of the entire project, the major focus of this thesis was aimed to particularly investigate the transcriptome profile and the secretome interdependence.

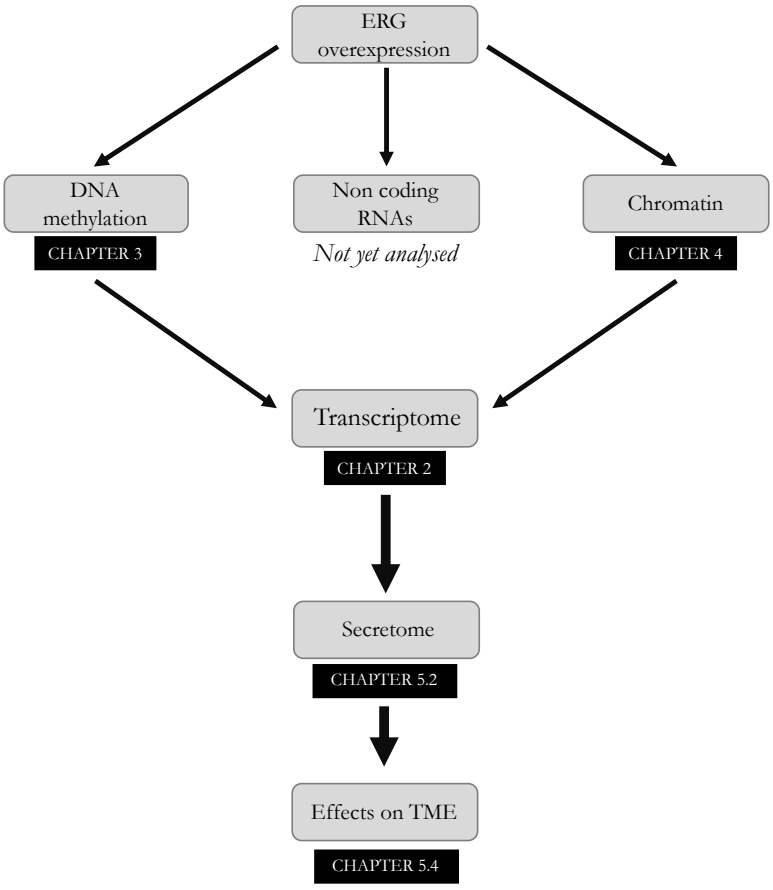


Figure 14| Organization of thesis workflow. In black blocks the reference chapter and paragraphs are indicated.

1. Cellular Cultures

1.1 Cell lines and culture medium

1.1.1 RWPE-1 cells

The cell model used was the RWPE-1 cell line (ATCC® CRL-1160™), an epithelial cell line derived from the peripheral zone of a histologically benign adult human prostate. The cell line was stabilized by a single copy of HPV-18 transfection. Specifically, RWPE-1 cells were engineered with a Tet-On system and thus presented a doxycycline-inducible promoter that allowed ERG overexpression. The inserted vector has been engineered to mimic the TE fusion genes, with the most frequent ERG variant (from exon 4 to 11) and the selection for cells acquiring this vector was obtained through the 1 μM puromycin supplement. The stabilized cell line was kindly donated

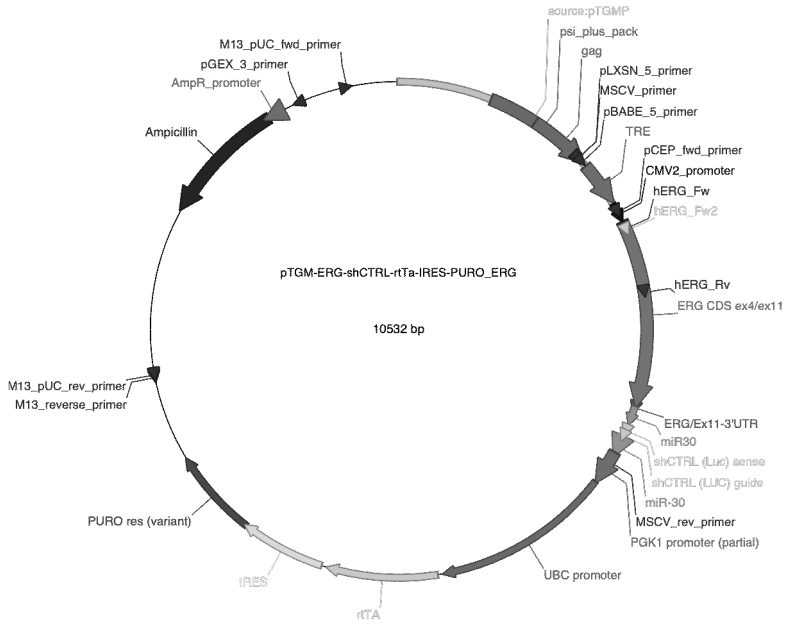


Figure 15| Structure of the vector used for the RWPE-1 engineering.

by Prof. Andrea Lunardi from CIBIO (Centre for Integrative Biology, in Trento) and outlined in *Figure 15*. The optimal doxycycline concentration for ERG induction was estimated as 1μM. Cells were routinely maintained as a monolayer inside an incubator, in a sterile environment with a temperature of 37°C and an atmosphere composed of 5% of CO_2.

The culture medium used is the Keratinocyte Serum-Free Medium (K-SFM) which contains: (i) 1% L-glutamine (an essential amino acid that cannot be synthesized elsewhere) that is an important energy source, (ii) Phenol red (as a pH indicator). K-SFM was complemented with: (i) 0.3% of the total volume Bovine Pituitary Extract (BPE), a mitogen supplement for the serum-free medium that contains a variety of growth factors and hormones with antioxidant activity, and (ii) 5 ng/ml Epidermal Growth Factor (EGF), a potent mitogen factor that triggers mitosis. The Penicillin and Streptomycin antibiotics were also added to prevent bacterial contaminations (1% of total volume) and a 0.1% puromycin of total volume is added to select cells containing the vector.

1.1.2 RWPE-1 cells induction

Depending on the collection time, a different number of cells was seeded in our experiments for each condition. Further details are reported in the table below.

Condition	Number of cells
12 hours	500000 cells/well
24 hours	300000 cells/well
48 hours	200000 cells/well
72 hours	100000 cells/well
120 hours	50000 cells/well

These quantities were experimentally set for a 6-well plate seeding. For inducing ERG overexpression, cells were treated with a 1μM doxycycline and the induction kept following the experimental needs. The culture medium was replaced every 48h.

1.1.3 Growth curve experiments

15000 RWPE-1 cells per condition were double seeded in 24-well plates. The experiment was performed in triplicate and cells were treated with a 1μM doxycycline every 48h to induce ERG overexpression. Experiment was performed up to 168h. For each daily time point cells were harvested, and the cell number estimated as the mean value of two wells per conditions. Cell counting was obtained with the TC20™ Automated Cell Counter, Bio-Rad.

1.1.4 Wound healing assay

800000 RWPE-1 cells were plated in 6 cm dish for each condition (not induced and induced cells) in K-SFM. Cells were left to attach and grow for 48h, in order to reach full confluency. After 48h, a p200 tip was used to make the scratch across the well and the medium replaced. The scratch moment corresponded to the first doxycycline supplement and that time was set as 0h. The induction was performed up to 96h, replacing the media and inducing cells every 48h for the induced conditions. Moreover, every 24h a picture of the wound healing was taken.

In fibroblast wound healing assay, NAFs and CAFs were seeded 1:20 in multiwell p6-wells, starting from a T25 90% confluent flask. After reached the confluence, the scratch was performed up to 92h, taking pictures every 24h. The wound area was evaluated with Adobe Photoshop® and calculated as wound closure percentage.

1.1.5 Isolation of NAFs and CAFs

Samples of normal and cancer tissues were obtained from radical prostatectomies of high-risk prostate cancers (Gleason ≥ 4+4) performed in the surgical unit of The Urology department at San Giovanni Bosco Hospital (Turin), accordingly to the n. 355384628 Protocol approved by the Ethics Committee (see Appendix I).

At the state of the art, the NAF and CAF isolation has been performed in our laboratory, following these steps: (i) surgical samples were fragmented, and the obtained pieces were further plated in a petri dish, with the addition of DMEM supplemented with 20% FBS, 2% Pen/Strep and 1% of Gentamycin, (ii) tissues fragments were then covered with sterilized glass slides, allowing

spontaneous fibroblasts migration from the tissue, (iii) after 2 to 4 weeks, fibroblasts attached to Petri dish were collected by clone selection and expanded in flasks, adding DMEM supplemented with 10% of FBS and 1% of Pen/Strep. A schematic description of the isolation protocol is reported in *Figure 16*.

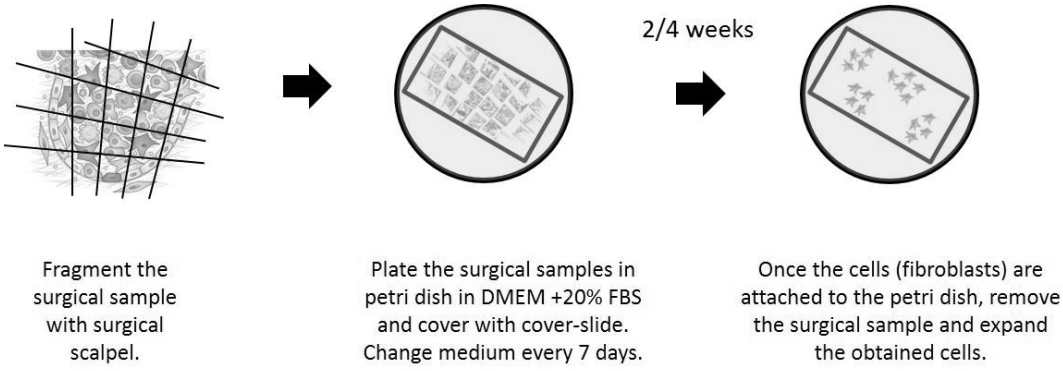


Figure 16| Representation of fibroblasts' isolation protocol, set in our laboratory.

1.1.5 Fibroblast cultivation

NAFs and CAFs were cultured in sterile 75 cm^2 tissue culture flasks filled with DMEM, a basal medium containing amino acids, glucose, vitamins, salts and a pH indicator that needed to be supplemented by: (i) 10% FBS of total volume, (ii) 1% L-Glutamine of total volume, an essential amino acid required in cell medium, since cells cannot synthesize it by themselves, (iii) 1% Pen/Strep of total volume, fundamental to avoid bacterial contaminations. Fibroblasts' culture was maintained in the incubator, in a sterile environment with a temperature of 37°C and an atmosphere composed of 5% of CO_2. Fibroblasts used for this project were previously obtained from radical prostatectomies (according to the isolation protocol previously presented). Normal prostate samples were taken from sections that displayed a normal morphology, whereas cancer-associated ones were taken from neoplastic areas detected by the instrumental analysis, biopsy and imaging/radiological analysis (the histological characterization of the two tissues is reported in *Figure 17*).

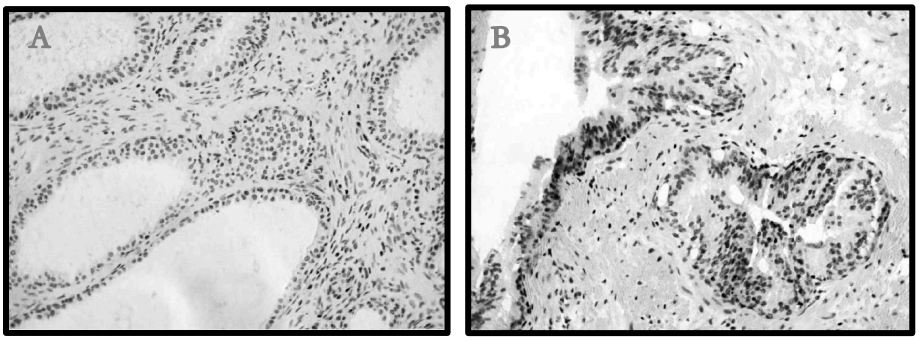


Figure 17|Histological evaluation of prostatic tumour resections: **A)** Normal tumour tissue, **B)** Tumour tissue with Gleason score 3+4.

1.1.6 Production of RWPE-1 cells conditioned media in K-SFM

To produce the conditioned media, 150000 RWPE-1 cells per condition were seeded in a sterile 6 cm Petri dish filled with complete K-SFM and then incubated in a sterile environment with a 37°C temperature and an atmosphere composed by 5% CO_2 for 72h. The culture medium was replaced with a new one every 48h and for the ERG-induced with 0.1% doxycycline 1 μM. At day 6, the lonely ERG-induced cells were supplemented with doxycycline without removing the media, thus the conditioned media would be enriched by proteins and factors released by cells within 120h. The 7th day, media were collected in falcon tubes, clarified by centrifugation at 3000 rpm 4°C for 5 minutes, in order to pellet cells that needed to be removed, the media were then transferred in new tubes and stored at -80°C in 1.5 ml aliquots. A schematic table reports the protocol prospectus:

Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Seed	-	Change media	-	Change media	-	-	Harvest
		No/Induction	-	No/Induction	-	No/Induction	

1.1.7 Fibroblasts exposition to CM from RWPE-1

Fibroblasts were cultured following the culture procedures previously explained. NAFs and CAFs were seeded in multiwell p-6 wells and filled with complete DMEM. After 48 hours from the seeding step, the DMEM was removed, cells rinsed with 4/5 ml of D-PBS and exposed to 600

µl of K-SFM conditioned from RWPE-1 (induced or not). After 120 hours, fibroblasts were collected for different purposes.

1.2 Subculturing procedures

For the subculturing procedure of RWPE-1 cells, the culture medium was discarded, cells were briefly rinsed with D-PBS and then treated with 1 to 2 ml of 0.05% trypsin (for T25 and T75 flasks, respectively); the flask was placed in a 37°C incubator for 6 minutes. After the total cells were detached, the trypsin enzymatic activity was blocked with the same trypsin volume of D-PBS/FBS (1:4). The cell suspension was then moved to a falcon tube, centrifuged at RT and 1000 x *g* for 5 minutes; the supernatant was discarded, cells resuspended in their media and finally quantified (TC20™ Automated Cell Counter, Bio-Rad). After the seeding phase, cells were left 48h to allow a complete adhesion, since RWPE-1 cell grew in a reduced-serum condition.

1.2.1 siRNA transfection

siRNA against DNMT3A was purchased from Invitrogen (s200425). Transfection was performed using Lipofectamine 3000 Reagent (Thermo Fisher Scientific). After 48h from the cell seeding procedure, transfection was performed as suggested in the manufacturer's instructions, using 25 nM final concentration of siRNA and a scramble-siRNA as the internal control. The DNMT3A level was 60% depleted after 96 hours of silencing.

1.3 Mycoplasma detection

Mycoplasma is a serious contaminant of cell cultures and its contamination might cause adverse effect such as changes in metabolism, growth, viability, DNA, RNA, protein synthesis, morphology, virus propagation and can lead to not reliable results and unsafe biological products. The N-GARDE Mycoplasma PCR Reagent set (Euroclone) was therefore used to certify good quality products. The manufacturer's instructions were used, and the products' quality was assessed through electrophoresis.

2 Analytical Techniques

2.1 RNA extraction

RNA was extracted from cells using TRIzol® Reagent (Life Technologies), following the manufacturer's instruction. The media was removed, and cells washed with D-PBS. 1 ml (> $5*10^6$ cells) or 0.8 ml (≅ $1-2*10^6$ cells) of TRIzol® was added and then cells lysate though pipetting. Samples were incubated 5 minutes RT, then 200 µl chloroform 99.5% added and samples shook vigorously for 15 seconds, then incubated 10 minutes RT. Samples were further centrifuged 15 minutes 11500 rpm at 4°C, the aqueous upper phase collected in a fresh tube and 1 µl of GlycoBlue™ (Thermo Fisher Scientific) added. 500 µl 100% isopropanol were added to precipitate RNA, samples incubated 5 minutes RT, centrifuged 10 minutes 11500 rpm at 4°C. Isopropanol was removed, the RNA pellet washed with EtOH 75% (≅ 1 ml) and vortexed, centrifuged 5 minutes 7500 rpm at 4°C. EtOH was carefully removed, the RNA pellet air-dry in a sterile environment, and then resuspend in H_20 DNase/RNase free (from 30 to 60 µl). Quantification was performed incubating samples 10 minutes at 55-60°C to resuspend RNA and then pipetting; the extracted RNA quantified with the QuBit (Qubit™ RNA BR Assay Kit, Thermo Fisher Scientific).

2.2 RNA Reverse Transcription

1-2 µg of RNA were retrotranscribed using the iScript® cDNA Synthesis Kit (Bio-Rad) and following manufacturer's instruction. OligodT were used for the annealing to the 3' end of any polyadenylated RNA molecules. The transcriptase protocol was:

Priming	5 min at 25°C
Reverse Transcription	20 min at 46°C
RT inactivation	1 min at 95°C
Optional step	Hold at 4°C

The obtained cDNA was prepared for the quantitative PCR reaction and diluted 2 ng/µl.

2.3 Polymerase Chain Reaction

Primers were designed by using the open-access tool Primer 3 (http://primer3.ut.ee/), choosing amplicons of approximately 75-250 bp. The selected sequences were *in silico* validated using primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), checking 'Use Thermodynamic Template Alignment' and confirming the specificity for the target of interest. The melting temperature was experimentally defined for each primer pairs.

2.3.1 Semi-quantitative PCR

The semi-quantitative PCR was performed using the iTaq™ DNA Polymerase (Bio-Rad) following the manufacturer's instructions. Amplicons were separated on a 2% agarose gel and visualized with ethidium bromide staining (Sigma Aldrich) at luminometer.

2.3.2 Quantitative PCR

The qPCR was performed in CFX Connect using the 2X iQ™ SYBR Green Supermix (Bio-Rad). Samples were prepared in triplicates and each sample was 20 ul total volume, done by:

iTaq™ Universal Sybr Green	10 µl
H2O dd	3 µl
cDNA (2 ng/µl)	5 µl
Primers (Forward and Reverse)	1 µl (per each)

RNA expression levels of selected genes were normalized with the internal control reference gene (GAPDH). A complete list of used primers is reported in Appendix II. The melting temperature was experimentally determined for each primer pairs.

The PCR protocol:

Polymerase Activation for cDNA	30 sec at 95°C
Denaturation	5 min at 96°C
Annealing/ Extension and Plate Read	30 sec at experimentally determined °C (57-59)
Cycles	39
Melt curve analysis	65-90°C with 0.5°C increment at 2-sec steps

2.4 RNA Sequencing

For RNA-seq 1 μg of total RNA was used for library preparation. Three biological replicates of each library (10 pM) were sequenced on HiSeq1500 (Illumina) for 50 cycles by our collaborators at IIGM (Italian Institute for Genomic Medicine), Turin.

2.4.1 RNA Sequencing analysis

RNA-seq raw data were processed by our collaborator Dr Ivan Molineris at the University of Turin. Sequencing reads were aligned to human reference genome (GRCh38, GENCODE primary_assembly) using STAR v2.7.1a with the following options: (i) outFilterMismatchNmax 999, (ii) outFilterMismatchNoverLmax 0.04, and providing a list of known splice sites in genome index extracted from GENCODE comprehensive annotation v32. Gene expression levels were quantified with featureCounts v1.6.3 (options: -t exon, -g gene_name) using GENCODE basic v32 annotation. Multi-mapped reads were excluded from quantification. Gene expression counts were next analysed using the edgeR package v3.28. Normalization factors were calculated using the trimmed-mean of M-values (TMM) method (implemented in the calcNormFactors function) and RPKM were obtained using normalized library sizes and gene lengths as summarized by featureCounts. Lowly expressed genes (*i.e.*, 1 CPM in less than 3 samples) were filtered out. Differential expression analysis was run by fitting genewise GLMs to the read counts, including all sample groups in the design matrix and performing quasi-likelihood F-test for the interesting pairwise contrasts. To define significant differentially expressed genes we considered multiple testing correction of the P Value, applying the common threshold P value_adj<0.05.

2.5 DNA extraction

DNA was extracted from cells using different methods: Phenol extraction and the QIAamp DNA Mini Kit (Qiagen) for MethylSeq samples.

a) Phenol extraction

The media was removed, and cells washed with D-PBS. Cells were scaped with 500 µl of complete SNET Buffer (for 1 or $2*10^6$ cells). SNET buffer was prepared with 10 mM Tris pH 8, 0.1 M EDTA, 0.5% SDS, 1:200 Proteinase K (20 mg/ml) and 1:400 RNAse (10 mg/ml). Samples incubated 1 to 2 hours at 55°C with a 300-rpm shake. A volume of phenol/chloroform isoamyl alcohol pH 8 was added and samples mixed and centrifuged at 12800 rpm RT for 5 minutes. Aqueous upper phase was collected in new tubes, adding two volumes of EtOH 100%, mixing and centrifuging at 12800 rpm RT. Samples washed with EtOH 70%, centrifuged 5 minutes at 7500 rpm RT and the supernatant removed. Pellet was air-dry in a sterile environment, resuspend in DNase-free water (30 to 60 µl), and quantified with QuBit (Invitrogen).

b) Column extraction

The QIAamp DNA Mini Kit was used following manufacturer's instructions. DNA was collected in SNET buffer and 200-300 µl of starting material used for the extraction.

2.5.1 DNA Sequencing

For DNA-seq, 1 µg of total DNA was bisulfite converted, then amplified and enzymatically fragmented, purified and run with the Illumina Infinium HumanMethylation EPIC BeadChips (Illumina). Three biological replicates were sequenced. The analysis was performed by our collaborator Dr Silvia Polidoro at Fondazione IIGM, Candiolo.

2.5.2 DNA Sequencing analysis

The DNA methylation analysis was performed with the *R* Shiny-based RnBeads 2.0 platform. Single comparisons (Induced vs Not Induced at each time point) were run to intercept differentially methylated residues, starting from raw data and following the RnBeads 2.0 pipeline [128].

2.6 Bisulphite pyrosequencing

Bisulphite conversion was performed using the EpiTect Plus DNA Bisulfite Kit (Qiagen) and following the manufacturer's instructions. Converted DNA was analysed with pyrosequencing in collaboration with Dr Daniela Furlan at the Ospedale di Circolo (Varese). A total of 1 µg of DNA per sample was converted and the time points analysed were from 72-, 120- and 168h.

2.7 Immunofluorence

For IF preparation, 35000 cells were plated in a multiwell p24-wells. After 48 hours, the medium was changed, cells were induced with the doxycycline supplement and the media changed. The following day, medium was removed, and cells washed with D-PBS, 500 µl of paraformaldehyde (PFA) 4% added and samples incubated 15 minutes RT. After PFA removal, samples were rinse three times in D-PBS for 5 min each and then blocked using blocking buffer for 60 min. Primary antibodies were prepared in dilution buffer (1:200 ERG, CST as in Appendix III) and samples incubated overnight 4°C. After three-time wash in D-PBS for 5 min each, samples were incubated in fluorochrome-conjugated secondary antibody (Alexa Fluor™ 488, Goat) diluted in the antibody dilution buffer for 1-2 hr RT in a dark environment. After three-time wash, 300 µl of DAPI staining were added to coverslips slides and then 10 µl mountant used, to fix samples. Blocking buffer: 1X D-PBS, 5% normal serum, 0.3% Triton™ X-100. Antibody Dilution Buffer: 1X D-PBS, 1% BSA, 0.3% Triton™ X-100.

2.8 Flow cytometry analysis

Samples were analysed in collaboration with Prof. Marzia Bruna Gariboldi (DBSV, Biotechnology and Life Science Department, University of Insubria) at the FACScan cytometer (Beckton Dickinson) equipped with a 15 mW, 488 nm and air-cooled argon-ion laser.

Cells were harvested, rinse with D-PBS, suspended with D-PBS/FBS (1:4), counted (at least $5*10^5$), centrifuged 10 minutes 1300 rpm at 4°C. Supernatant was removed and pellet resuspended in 500 µl D-PBS. Pellet was centrifuge 10 minutes 1900 rpm at 4°C, supernatant removed, and pellet resuspended in 500 µl EtOH 70%. Samples were incubated at least 30 minutes at -20°C, then centrifuged 10 minutes 1900 rpm at 4°C. The Propidium Iodide solution was prepared in D-PBS

(PI 50 µg/ml, RNase 20 µg/ml), supernatant removed, and pellet resuspend with 750 µl PI solution, incubating absence of light for 15 minutes.

10000 events were reported and analysed for each sample and all data were processed using the CellQuest software (Beckton Dickinson). The fluorescent PI emission was collected through 575 nm band-pass filter and acquired in *log* mode; the percentage of apoptotic cells was determined based on sub-G1 peaks detected in monoparametric histogram.

2.9 Protein extraction and quantitation

Proteins were collected and extracted using the RIPA Buffer (Sigma Aldrich) complemented with 1:500 volume of Protease Inhibitor Cocktail (PIC 500X, Roche) and of Phenylmethylsulfonyl fluoride (PMSF, Sigma Aldrich). The media was removed and an appropriate volume of RIPA buffer used for scraping cells that were then collected in fresh tubes, incubated 5 min on ice.
Samples sonicated for 12 cycles (30 seconds) + 1-second rest, 80% amplitude on ice and then centrifuged 15 minutes 13000 rpm at 4°C. Proteins were quantified using the BCA assay (Pierce™ BCA Protein Assay Kit, Thermo Fisher Scientific) and read at 575 nm with the iMark™ Microplate Absorbance Reader (Bio-Rad).

2.10 Western Blot

8-10 µg of extracted proteins were added to a 2x Laemmli Buffer (Sigma Aldrich) and samples were then boiled for 5 minutes at 95°C. Proteins separation was performed on gradient gels (4-20% Mini-PROTEAN® TGX™ Precast Protein Gels, Bio-Rad) in reducing conditions and transferred using the Trans-Blot Turbo System (Bio-Rad), with a nitrocellulose membrane. The membrane was incubated in 5% Skim Milk – Tris Buffered Saline – 0.1% Tween20 (TBS-T) blocking solution to bind non-specific sites for 1 hour and then incubated ON with an appropriate primary antibody. Horseradish-peroxidase-conjugated secondary antibodies against rabbit (CST 7074S) or mouse Ig (CST 7076S) 1:5000 was then used. The reaction was detected with an enhanced chemiluminescent kit (ECL, GE Healthcare) and acquired using G: BOX (Syngene). Densitometric analyses were performed using Adobe Photoshop® and normalized using the

housekeeping GAPDH protein as reference. A comprehensive list of used antibodies is reported in Appendix III.

2.11 ChIP – Chromatin Immunoprecipitation

Chromatin immunoprecipitation assay was performed according to the Farnham Lab Protocol (https://gozani-lab-website.github.io/website/docs/Chromatin-Immunoprecipitation-ChIPs-Protocol-from-Farnham-Lab.pdf), adapted to cell suspension. Briefly, cells were treated with 1% formaldehyde and collected in 1X PBS. Cell pellets were re-suspended in lysis buffer (5mM PIPES pH8, 85mMKCl, 0,5% NP40, 1X Proteinase inhibitors) and sonicated for 10 seconds 18 times on ice (BRANSON S250 digital sonicator, Branson, Danbury, CT, USA). Sonicated chromatin was pre-cleared for 1 hour at 4°C using 40µl of pre-washed G-agarose beads (KPL). Pre-cleared chromatin was quantified using Qubit® dsDNA BR Assay Kit (Life Technologies). 10µg of chromatin was incubated overnight at 4°C in dilution buffer (1% SDS, 10mM EDTA pH8, 50mM Tris-HCl pH8, 1X proteinase inhibitors) with 10µg of specific antibodies. Five percent of the total non-immunoprecipitated lysate was used for input control. Antibody coupled chromatin was incubated using 40µl of pre-washed protein G agarose beads (KPL) for 2 hours at 4°C. The beads were extensively washed, and the DNA was extracted with the Chromatin IP DNA Purification kit (58002, Active Motif) and resuspended in water. Chromatin immunoprecipitation products were amplified using iTaq™ DNA Polymerase (Bio-Rad) and specific primers (Appendix II). To better interpret the ChIP results, the densitometric analysis was performed using Adobe Photoshop®. First, the obtained values were normalized on input signals as follow: IP / Densitometric value (INPUT), for NI and I and then the two conditions compared. Primers for semiQ PCR of ChIP targets were obtained as previously described (see Results 2.3) and using the genomic sequence as reference. Primer sequences were listed in Appendix II.

2.12 LC-MS analysis

The conditioned media obtained from RWPE- 1 cells were processed with Liquid Chromatography-Mass Spectrometry in collaboration with Dr Marco Gaspari at the University of Catanzaro.

An aliquot of 1 ml of conditioned media was diluted by adding 150µl of 10% SDS, 150µl of 500 mM dithiothreitol, 75µl of 1M tris buffer (pH 8.0), 125µl of deionized water. Proteins were denatured for 10 min at 95 °C and digested by filter-aided sample preparation for proteome analysis (FASP) as originally described by Wisniewski *et al.* [129] with minor modifications. Briefly, filter units were YM-10 (Cat No. MRCF0R010, Millipore) and tris buffer was at pH = 8.0; after loading 600µl of the reduced protein solution onto the filter and concentrating the solution by centrifugation, all washing steps (8M urea, 100 mM tris buffer) were performed using aliquots of 200µl. After alkylation with iodoacetamide as described in the original FASP protocol and additional 2 x 200µl washes in urea solution and 2 x 200µl washes of 50mM triethylammonium bicarbonate buffer (TEAB) solution, overnight digestion was carried out at 37 °C by adding 60µl of digest buffer containing 200ng of trypsin proteomics grade (Sigma) in 50mM TEAB (pH 8.5). Peptides were recovered by adding 140µl of HPLC-grade water to the filter, and by centrifuging the filters for 25 min at 14,000 g (final volume: 200µl of tryptic digest). An aliquot corresponding to 40% of the digest volume was purified by strong cation exchange StageTips [130] using Empore extraction disks (Sigma-Aldrich); the eluate (10µl of 500mM ammonium acetate in 20% acetonitrile) was evaporated and resuspended in 20µl of mobile phase A. Tryptic peptides were analysed by nanoscale LC-MS/MS by a Top-12 data-dependent analysis method run on a Q-Exactive "classic" instrument from Thermo Fisher Scientific, essentially as described in Laria *et al* [131]. Briefly, after a preliminary injection using a short chromatographic gradient (30 min) aimed at estimating peptide content, approximately 200ng of protein digest were analysed by reversed phase nanoscale liquid chromatography coupled to a quadrupole Orbitrap mass spectrometer (Q-Exactive "classic" Thermo Fisher Scientific) as described. A nanoLC capillary column (75µm i.d., length: 10cm), packed with 3µm C18 silica particles (Dr Maisch GmbH, Ammerbuch, Germany) was used to separate the peptides. Mobile phase A was 0.1% formic acid, 2% acetonitrile, whereas mobile phase B was 0.1% formic acid, 80% acetonitrile. The flow rate was 300nl/min, and peptides were separated using a linear gradient of phase B from 6% to 42% in 60 min, and from 42% to 100% in 8 min. MS acquisition was done using a top-12 data-dependent acquisition (DDA) mode. Full scan m/z range was 350-1800 and resolution was 70000; MS/MS scans were at 35000 resolution; AGC target was 106 for full MS and 105 for MS/MS; maximum injection time (ms) was 50 ms for full MS, and 120 ms for MS/MS. Mass window for precursor ion isolation was 1.6 m/z. Normalized Collision energy was 25. Dynamic exclusion was 20s. Raw data were processed for protein identification and quantitation using the software MaxQuant version 1.6.2.6a [132], using default parameters, with the following exceptions: (i) minimum peak length 4, (ii) minimum ratio count 1, (iii) protein quantification was based on unique peptides, (iv) the match-between runs

option was activated (match time window: 0.3 min, alignment window: 10 min). Tandem mass spectrometry data were searched by the Andromeda search engine by querying the Human Reference Proteome Database (74788 sequences, accessed in December 2019). The contaminants list in MaxQuant was updated by adding all protein identifications obtained by an LC-MS/MS analysis of the medium performed in triplicates and processed by Proteome Discoverer 1.4 (Thermo Fisher Scientific) using default parameters, peptide-spectrum matches (PSM) validation by Percolator, and by including all hits with a minimum of 2 high-confidence peptides. The "protein groups" txt file was uploaded in Perseus software (v1.6.2.1) for statistical data analysis. After removing entries belonging to the reverse database and to the contaminants database, LFQ values were log_2-transformed. Only proteins quantified in at least three replicates and in at least one group were kept, corresponding to 1681 protein groups. After missing value imputation using default parameters in Perseus, differentially abundant proteins were assigned by performing a paired two-sided Student's T-test with S0 = 0.2 as correction factor and permutation-based estimate of FDR (FDR < 0.05).

3 Statistical analysis

Control and induced conditions were tested with One-way ANOVA, multiple comparisons. A p-value <0.05 was considered statistically significant. Analyses were performed using Microsoft Excel 365, GraphPad Prism V8, SPSS v25 IBM software.

4 Bioinformatics

The Gene Set Enrichment Analysis (GSEA) for each condition was obtained with the GSEA software (Broad Institute). The Enrichment analysis was run with the web based Metascape tool (https://metascape.org) and PPI networks were built using STRING v10.5 (https://string-db.org). The parameters were: (i) molecular action for edges meaning, and (ii) experiments and database as active interaction source. The network analysis was obtained on Cytoscape 3.8.0 (https://cytoscape.org/) and the Network Analyzer tool and clusters calculated with the MCODE algorithm. The Cancer Genome Atlas (PRAD database) was directly downloaded by the public portal GDC Data Portal (Cell 2015 version). Clinic, genomic, proteomic, transcription and expression data were downloaded to further confirm experimental data.

Results

1. RWPE-1 induced cells show a more mesenchymal-like phenotype

The phenotypic features of RWPE-1 cells engineered By Prof. Andrea Lunardi's Lab (CIBIO, Trento) were firstly evaluated at different levels.

1.1 Immunofluorescence analysis for ERG detection

To assess the correct localization and the magnitude of cells overexpressing ERG, we performed an IF analysis evaluating the induction after 24h, using as the conjugated antibody the Alexa Fluor™ 488 for ERG detection (*Figure 18*).

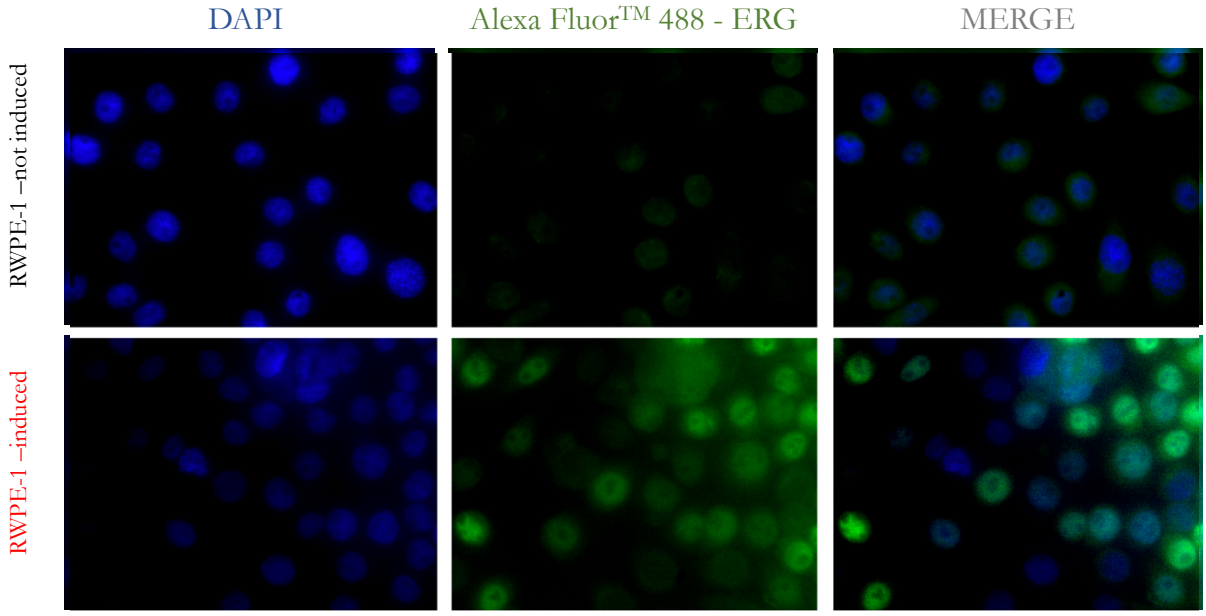


Figure 18| Confocal IF analysis on RWPE-1 induced or not at 24h. The Alexa Fluor 488 was used for ERG detection. Magnification 60x.

ERG signal perfectly lined up with DAPI staining, showing a nuclear localization. In details, the green signal in the not induced cells might be due to an autofluorescent phenomenon or to an endogenous ERG expression that would be further analysed. This result showed that the cell model

is effectively able to induce ERG expression after the doxycycline supplement (even in a short term) and coherently defined ERG subcellular localization in the nucleus, as previously seen by Shah *et al.* [48].

1.2 Analysis of RWPE-1 cells growth rate and cell cycle

In order to evaluate the effect of ERG induction on RWPE-1 cells concerning their growth, we performed a growth rate analysis collecting cells every 24h and inducing cells every 48h. In *Figure 19*, the growth curve showed a lower trend of proliferation for ERG-induced cells in a time-dependent manner; the difference between Not Induced (NI) and Induced (I) conditions increased in the last time points analysed. Coherently, also Tomlins *et al.* [50] demonstrated that cells overexpressing ERG both permanently or in transient models did not increase cell proliferation.

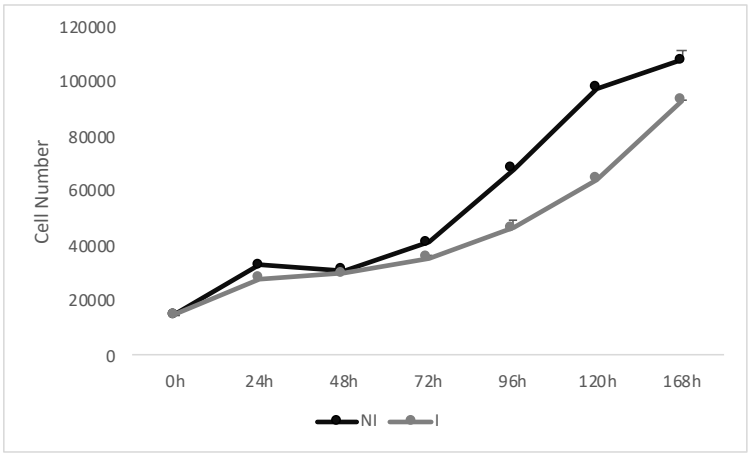


Figure 19| Growth curve of RWPE-1 cells induced I (in grey) or not induce NI (black).

Therefore, to better elucidate this different trend, we further investigated the cell cycle phase patterns and the apoptotic rate in the two conditions. To focus the analysis on the most variable conditions, the 72-, 120- and 168-hour time points were considered, as reported in *Figure 20*.

Curiously, for all the time points analysed, a slight reduction of cells in G0/G1 phase from 72- to 168h was observed, while, conversely, the G2/M phase increase peaked at 168h in both conditions, suggesting a time-dependent and not ERG-dependent effect. No difference, instead, could be detected for cells in S phase as was also previously observed by Tomlins *et al.* [50]. At the

same time, the percentage of apoptotic cells was less than 1.5% for each condition, highlighting that not even apoptosis could explain the different trend in cell growth.

These preliminary validating analyses, thus, might suggest that ERG overexpression is not crucial to increase cell proliferation but might be pivotal for a cell reprogramming that supports invasion.

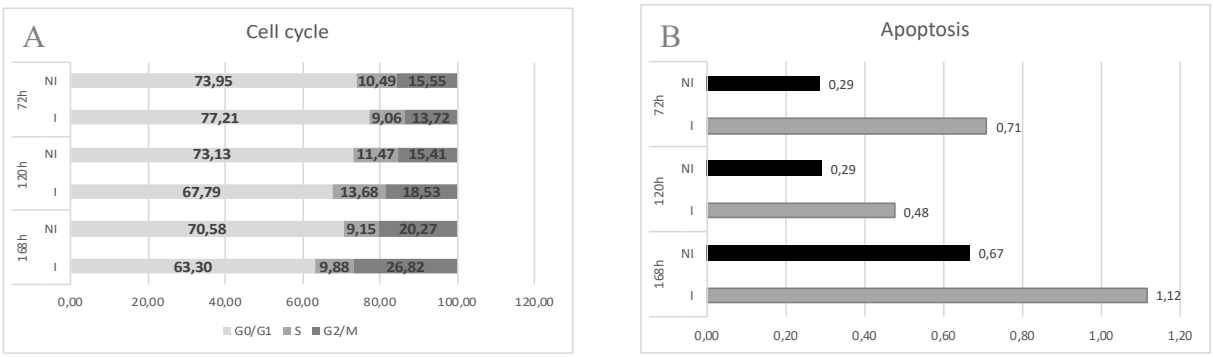
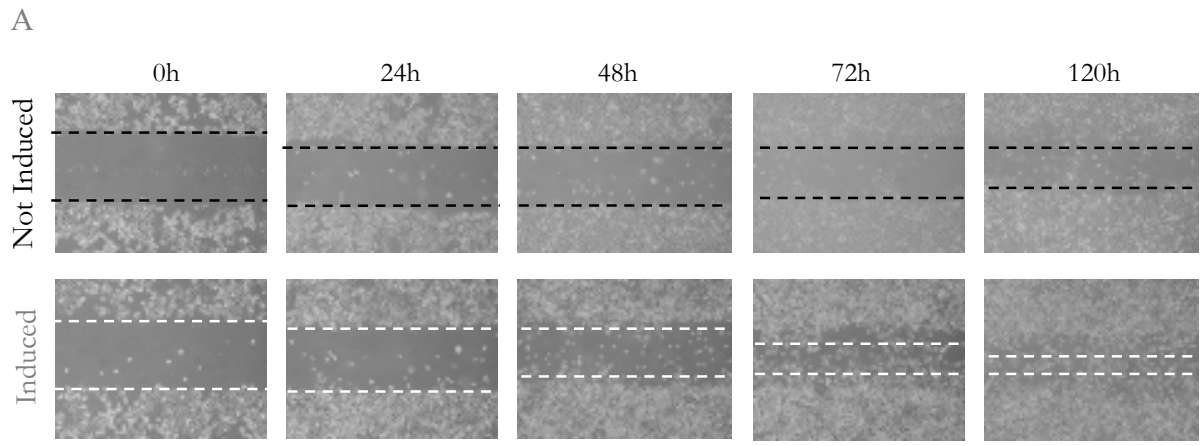


Figure 20| **A)** Cell cycle phase analysis for the three selected time points. **B)** Percentage of apoptotic cells in the three time points for the two conditions NI (in black) and I (in grey).

1.3 Invasive and migratory properties of RWPE-1 overexpressing ERG

To test the invasive features of ERG-induced RWPE-1 cells, a wound healing assay was performed (*Figure 21*). Since RWPE-1 cells are a benign cell model, the assay was performed up to 120h, where a significant difference in terms of wound closure fold change could be assessed, although, as previously shown, induced cells had a slower growth rate. Both these aspects confirmed that induced RWPE-1 cells might acquire mesenchymal features to prompt invasive and migratory behaviours.



B

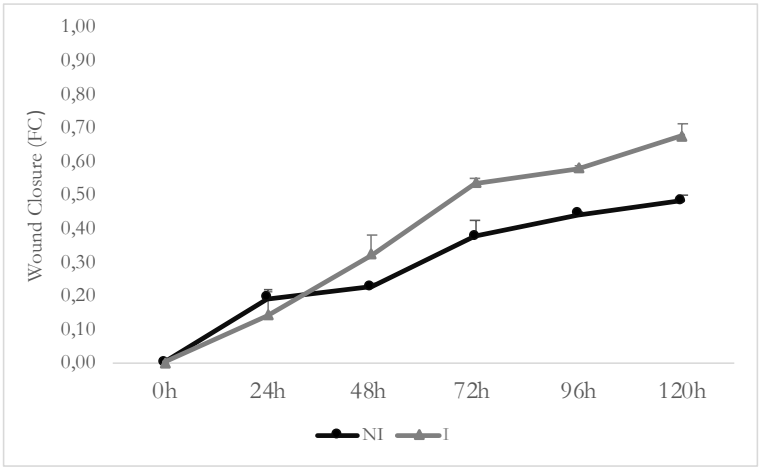


Figure 21 | **A)** Wound healing assay of the NI (Not Induced) and I (Induced) conditions performed up to 120h. Magnification 40X. **B)** Fold change (FC) of wound closure, normalized on the 0h time point within each condition.

1.4 Molecular characterization of mesenchymal features in the ERG-induced conditions

In order to confirm the mesenchymal behaviour previously observed, a molecular evaluation of factors and markers of the epithelial-to-mesenchymal transition (EMT) as well as of its crucial regulators was performed, analysing transcript and protein levels (*Figure 22*).

The qPCR and WB analyses revealed that the maintenance DNA methyltransferase 1 (DNMT1) and its recruiting enzyme UHRF1 were significantly downregulated in ERG overexpressing cells. On the other hand, concerning the two *de novo* DNMTs, DNMT3B resulted significant downregulated at a transcript and protein level, even if significance was only assessed in WB analysis. DNMT3A showed, instead, conflicting trends between qPCR and WB results that will be better deepened in the following paragraphs, since it has a crucial role in EMT regulation in prostate cancer [90].

EZH2 (Enhancer of zeste homolog 2) is an histone-lysine methyltransferase that catalyses the addition of the methyl group to H3K27me2 as part of the Polycomb Repressive Complex 2 (PRC2), and resulted highly expressed in PCa, especially in *ERG*$^{+}$ tumours, where it was found to be directly regulated by ERG, correlating with an epigenetic reprogramming [61]. Despite this evidence, EZH2 was found downregulated in our cell model, suggesting that an early effect of ERG overexpression negatively affects EZH2 expression. Coherently, *ZEB1* (a direct repressor of *CDH1*) was significantly upregulated at 72- and 120h, confirming previous evidence that highlighted a possible ERG direct regulation of ZEB1 in PCa harbouring the TE fusion gene, in

fact ERG binding motifs were detected in *ZEB1* promoter [133]. In addition, the EMT suppressor GRHL2 and the EMT positive marker CDH2 did not showed any statistically significant change, highlighting that the molecular mechanisms in which there are involved were not really affected by ERG overexpression, and therefore that other mechanisms might occur. Furthermore, CDH1 downregulation and the concomitant VIM upregulation strongly suggested an EMT promotion of RWPE-1 cells when ERG is overexpressed. Particularly, VIM expression was already shown to be ERG-dependent [59]. Future analyses would also try to assess the potential reversibility of ERG-induced target deregulation, removing the induction after 120h and evaluating the resulting molecular profile.

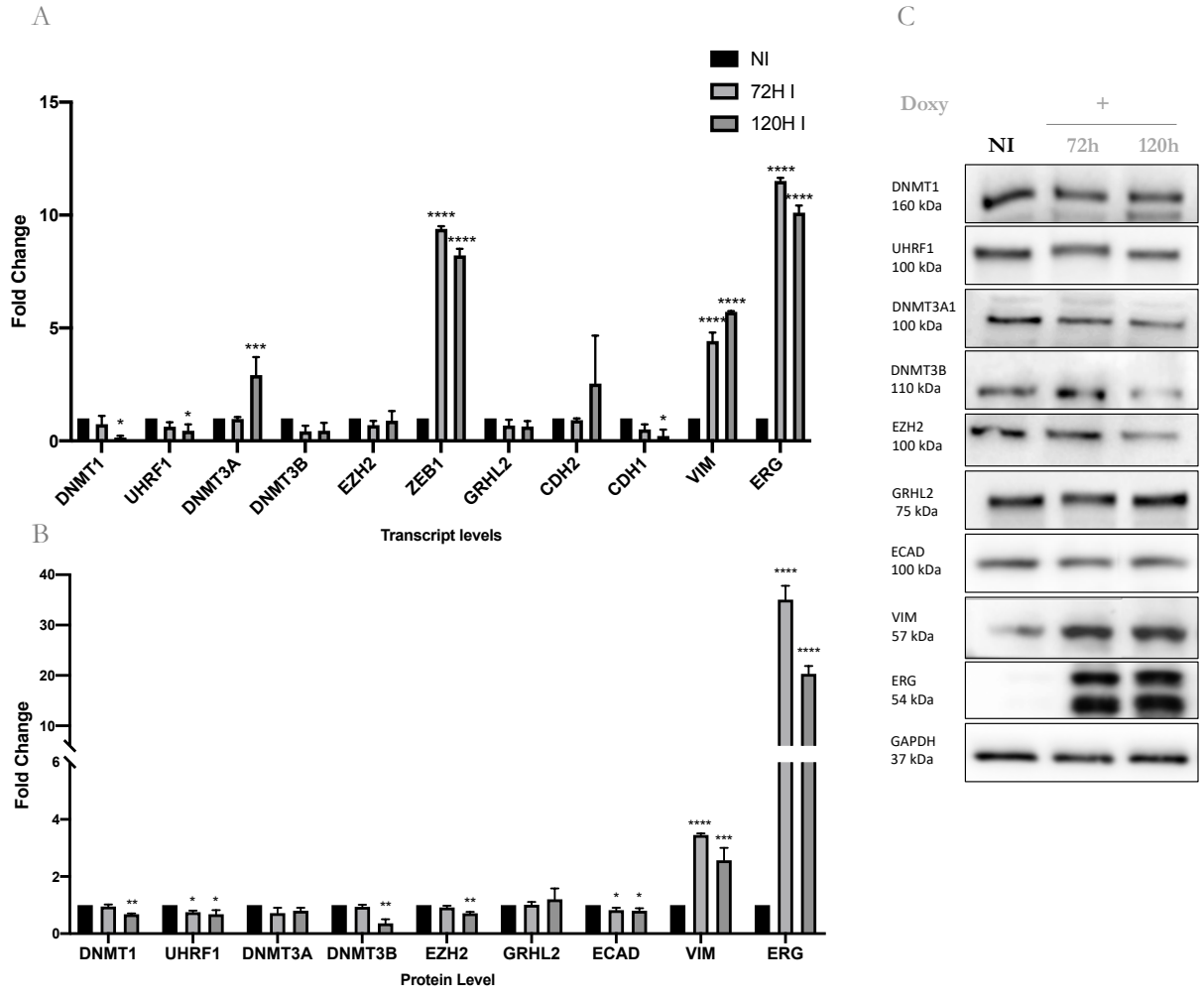


Figure 22|**A)** Transcript levels of mesenchymal markers, factors and DNMTs obtained by qPCR analysis. **B)** Densitometric analysis of the relating WB. Transcript and protein analyses were normalized using GAPDH as reference. **C)** Protein level evaluation for mesenchymal markers/factors and DNMTs with the Western blot assay. Statistical significance was assessed with an Anova One-way test using the multiple comparison option and comparing each I vs NI conditions. *P value≤ 0.05, ** P value≤ 0.01, *** P value≤ 0.001, ****P value ≤ 0.0001.

The analysed EMT markers and factors supported the hypothesis that although ERG overexpression alone is insufficient to promote an increase in cell proliferation, it is relevant for fostering the epithelial-to-mesenchymal transition, in fact a switch from epithelial to mesenchymal features was observed in our experimental conditions.

1.5 Timing setting for sequencing analyses

Whether ERG overexpression alone is sufficient to define a specific clinicopathologic status, needed still to be clarify. Therefore, we aimed to better elucidate if ERG induces early effects that do the groundwork for the malignant onset, based on its transcriptional (activating and repressing) activity. In particular we selected different time points to distinguish "early" and "late" effects. As reported in previous results, ERG overexpression determined a moderate promotion from an epithelial to a mesenchymal switch in a "late" time, at 120h. Therefore, we investigated if some changes were still detectable in a short window time, at 12h and 24h from the first induction procedure.

Firstly, the molecular pattern of factors and targets involved in EMT and epigenetic regulation were analysed at the aforementioned time points (*Figure 23*). The transcript and protein analyses revealed that, as expected, EMT-related factors and markers acquired statistically significance in a late time, excluding vimentin that was significantly upregulated also at 12h. In addition, some interesting changes in DNMTs expression were observed already at 12h (even if not significant), where the peak of ERG transcript and protein level was already assessed. Therefore, these "early" (12- and 24h) and "late" (120h) time points were further used for sequencing analyses.

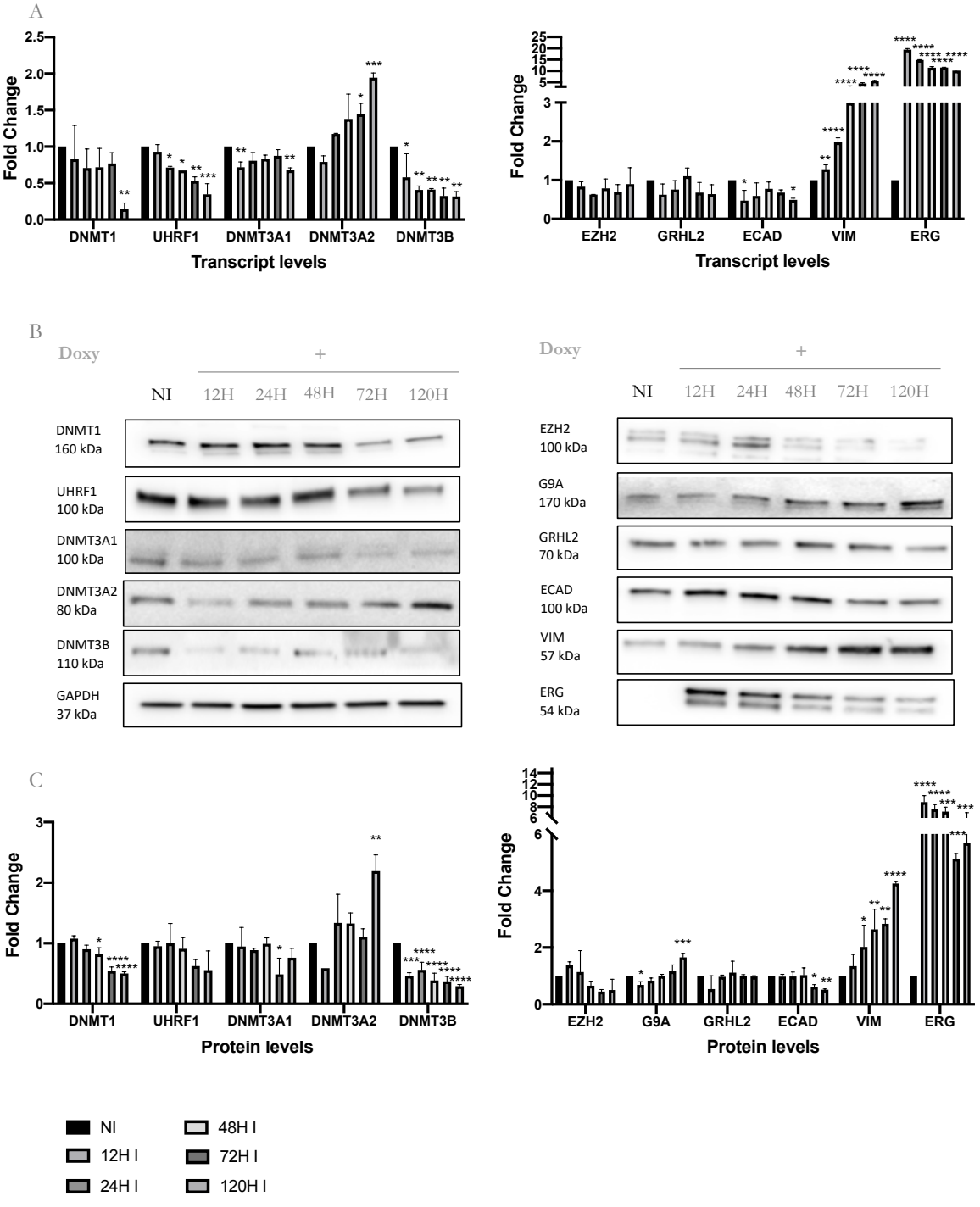


Figure 23| Molecular evaluation of RWPE-1 cells induced (I) or not (NI). **A)** qPCR analysis. **B)** WB analysis. **C)** Densitometric analysis. In both analyses GAPDH was used as reference. Statistical significance was evaluated with an Anova One-way using the multiple comparison option, evaluating I vs NI. *P value≤ 0.05, ** P value≤ 0.01, *** P value≤ 0.001, ****P value ≤ 0.0001.

2. ERG overexpression strongly impacts on the transcriptome profile

2.1 Explorative analysis

The transcriptome profile of RWPE-1 cells overexpressing ERG with the related counterpart was performed using the same time points previously defined, in order to intercept differentially expressed (DE) genes. Three independent replicates at 12-, 24- and 120h were then run using the HiSeq system by our collaborators at IIGM, in Turin.

The quality control step detected a number of reads per sample that in all cases exceeded 15M, therefore optimal conditions and no contamination were found in our samples. This premise is fundamental to try to better interpret the following aspect of the explorative analysis performed. In particular, a Principal Component Analysis (PCA) of the 500 most variable genes was done to evaluate the overall distribution of samples between different conditions and time points (*Figure 24*). The PCA showed that the first two components accounted for almost 75.2% of variance, where PC1 clearly separated samples for the time points (12-, 24 and 120h) while PC2 strikingly separated them for the treatment (NI and I). Moreover, the three replicates showed high similarity and we concluded that no batch effect was detectable.

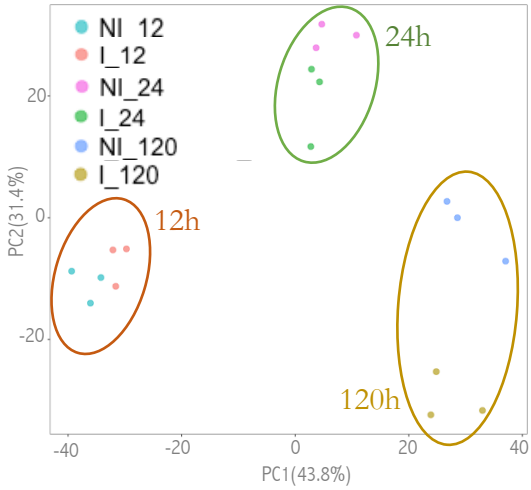


Figure 24| PCA analysis of the 500 most variable DE genes. The PCA was obtained with the iGEAK [139] tool and each dot of the same colour represents the batch of a single sample. The colour code is reported in the top left legend.

Unexpectedly, the NI conditions of the time points analysed did not cluster together, even though the only experimental difference relayed on the culture timing. Particularly, we speculated that this

difference reflected a biological meaning that we would try to better identify with further analyses and that could be partially due to a cell-cell contact effect.

On one hand, the differentially expressed genes (DEG) analysis comparing I vs NI conditions at each time point was firstly performed by our collaborator Dr Ivan Molineris (IIGM, Turin). For each gene a log_2FC, a P value and a P value-adjusted with the multiple correction were estimated. The criterion for the significance was set at P value_adj<0.05.

Firstly, aiming to describe the biological global differences between I versus NI, a Gene Set Enrichment Analysis (GSEA) was performed using all genes obtained from single comparisons and the related log_2FC values; particularly the gene lists included:

- 14518 genes for the 12h condition,
- 14518 genes for the 24h condition,
- 10445 genes for the 120h condition.

Analyses were run using the GSEA software (4.0.3 version) and the "Run GSEA Preranked" method. As setting parameters, we used the Hallmark 7.2 Gene Set database, with 1000 permutations and the "No_Collapse" options. For each condition then, only Normalised Enrichment Scores (NESs) with a P value_adjusted< 0.05 were considered (*Figure 25*).

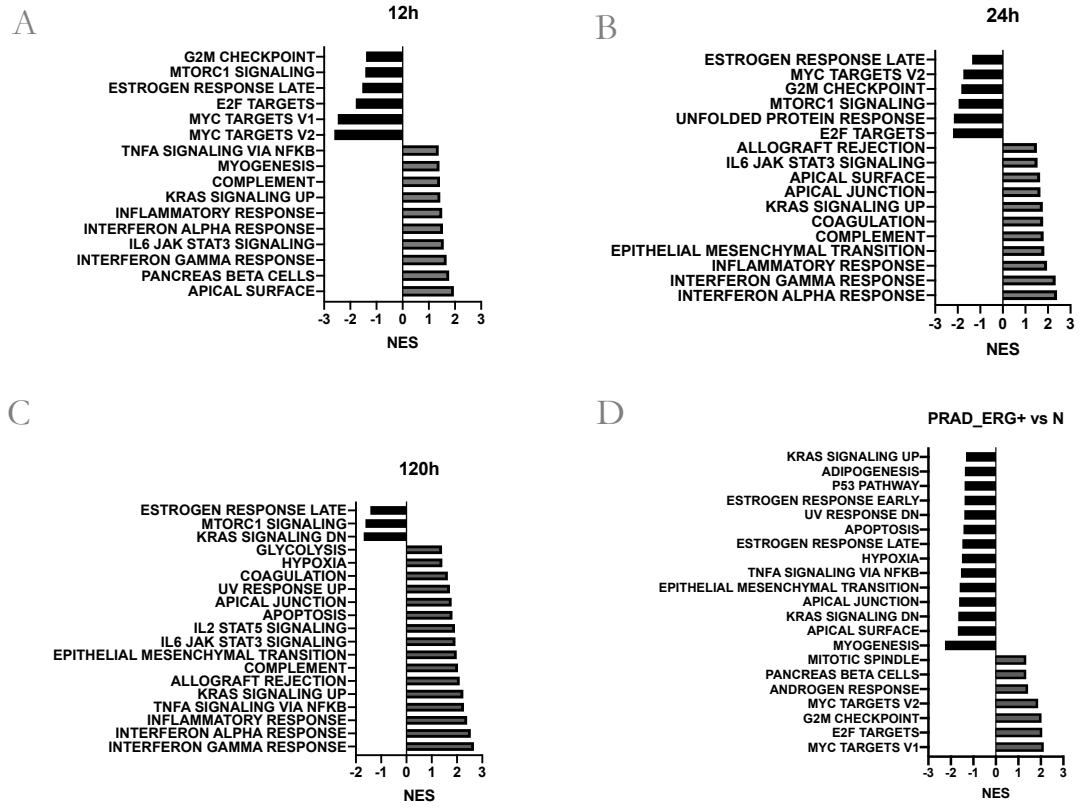


Figure 25| GSEA of each experimental condition and of the PRAD subset. The bar plots show Normalised Enrichment Scores significant for P value-adjusted<0.05. **A)** 12h, **B)** 24h, **C)** 120h and **D)** PRAD subset.

Particularly, to confirm our results using a real tumour cohort, the PRAD database from the "The Cancer Genome Atlas" (TCGA) was downloaded. Using the clinicopathologic subtype classification, we further selected *ERG*$^+$ patients (131) and normal (52) samples, defining a focused subset. For the aforementioned analysis, the PRAD subset was processed with the edgeR Galaxy tool to obtain the DEG analysis of the *ERG*$^+$ vs N comparison. The resulting 20531 genes were then analysed for the GSEA, using the same setting parameters previously defined.

The GSEA highlighted that from 12- to 120h some processes resulted particularly prompted in the ERG overexpressing conditions. Of note, in all conditions the Interferon Response (Alpha and Gamma), the inflammatory response, the complement, the KRAS signalling upregulation and STAT3 signalling resulted highly represented, while the estrogenic late response showed a negative NES. Particularly, EMT emerged at 24- and 120h, coherently with previous assays performed. Furthermore, the PRAD subset comparison highlighted the complexity of *ERG*$^+$ tumours, where the estrogenic late response recapitulated previous results while the EMT resulted curiously negatively enriched.

The role of the estrogenic signalling in prostate development and in tumour progression has largely been investigated. The estrogenic activity is mediated by two nuclear receptors: ERα and ERβ encoded by *ESR1* and *ESR2*, respectively. Although the loss of ERβ expression is clearly associated with PCa growth and progression, the role of ERα is still unclear [134]. Particularly, a recent study highlighted that the TE occurrence correlates with *ESR1* higher expression levels and PCa progression [135]. Curiously, in our experimental conditions *ESR1* showed a significant negative fold change peaking at 120h of ERG induction (log_2FC=-1.21), while an opposite trend was observed in PRAD dataset. This evidence seemed to suggest that ERG overexpression affects ERs regulation, inducing its upregulation in adenocarcinomas to support cancer progression. No data instead were detected for *ESR2* levels.

Overall, this preliminary analysis might indicate that some mechanisms could be considered as early events of ERG overexpression that do not find correspondence in developed tumours, while the ERG-dependent estrogenic late response might instead be a crucial mechanism from the pre-malignant to the malignant status.

On the second hand, in order to avoid any kind of bias and therefore minimize the time-effect, the DEG list was further filtered. The major problem in reducing the time-effect of NI conditions implied the possibility to remove DE genes relevant for a time-dependent induction. Therefore, triplicates for each time point and their NI counterparts were evaluated using the log_2FC instead of the trimmed Count Per Million (CPM) values. Genes showing a P value_adjusted< 0.05 and a

$log_2FC>|1|$ in at least one of the three conditions were then selected. Through this approach, 1237 genes were found DE in the three time points and we chose to further filter them, removing low trimmed count genes (that showed a difference of 0- 1.9 CPM reads, in the NI vs I and vice versa comparisons). This filtering procedure helped to determine a DEG list of 507 genes that were selected as the candidate genes for further evaluation.

Focusing on 507 genes, we performed a new PCA (*Figure 26*), where the quality of the DE gene list obtained, and a reduced time-dependent effect was shown. In this case, PC1 accounted for 41.3% of the explained variance that correlated with the induction, while PC2 represented the time-effect variance (35%).

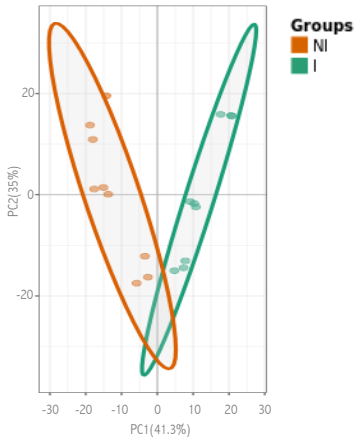


Figure 26| PCA plot for the 507 gene selection. Group colors are indicated in the top right legend. The first two components are reported.

Moreover, we firstly explored our DEG list to evaluate how genes were up- or downregulated due to ERG overexpression, therefore volcano plots were obtained (*Figure 27*).
Particularly, setting threshold at $log_2FC>|1.5|$ we obtained:

- 31 up- and 4 downregulated DE genes at 12h,
- 90 up- and 29 downregulated DE genes at 24h,
- 149 up- and 60 downregulated DE genes at 120h.

The same gene list was also used to totally evaluate differences within conditions, as in heatmap in *Figure 28*. The heatmap highlighted that genes could be clustered accordingly to different aspects and overall, we determined that, unless values in the NIs were not properly corresponding, the trend and therefore the log_2FC values between the different I time points and the relating NI counterparts, could be compared (the 507 gene selection in fact was obtained using log_2FC values and not the CPM ones).

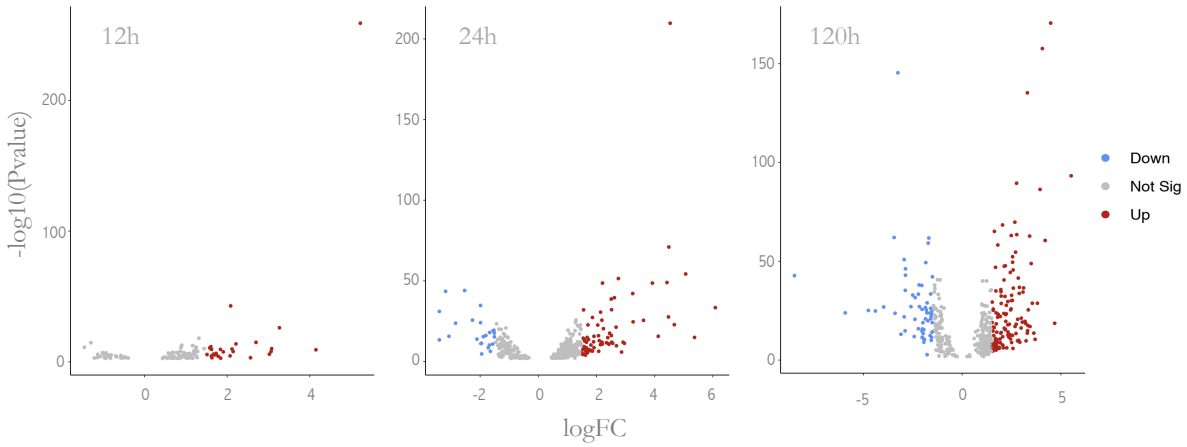


Figure 27| Volcano plots for each condition. Significance was set for P value_adj <0.05 and log_2FC>|1.5|. Blue and red dots represent significantly downregulated and upregulated genes, respectively. In grey, not significant ones.

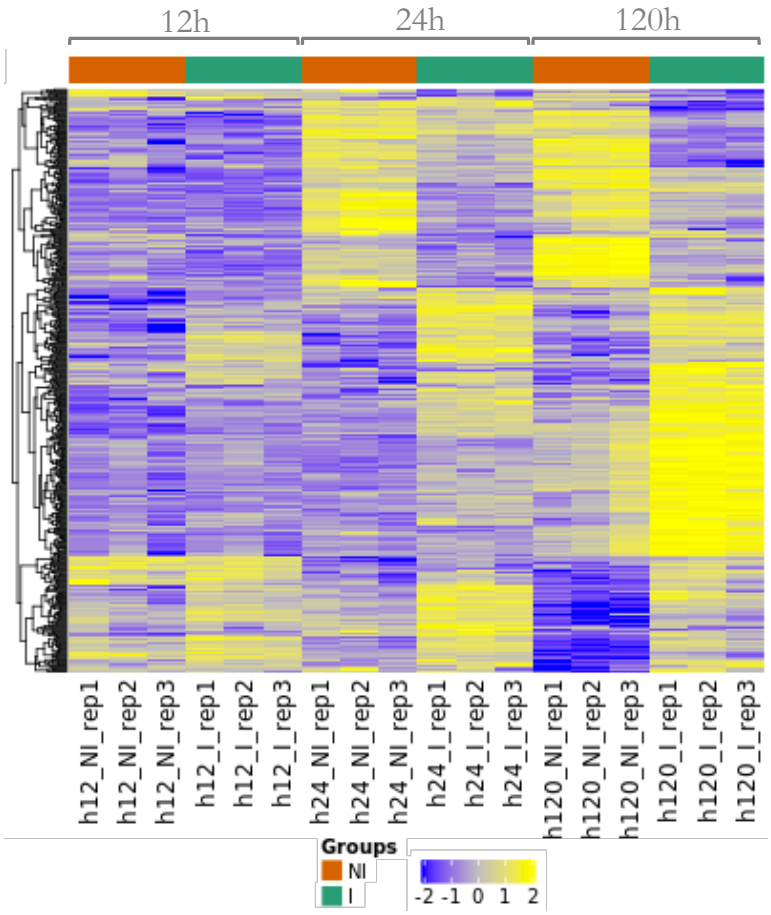


Figure 28| Heatmap of the 507 DE genes obtained by filtering procedures. Samples are presented with their replicates and Groups (NI or I) are identified in green or orange, respectively. Values for each gene were log_2 scaled from the CPM values and the colour code is reported in the bottom legend. The analysis was performed using the web tool FaDA.

2.2 Functional analysis

From a biological point of view, we better characterized the 507 DEG list to define an overview of processes and mechanisms that resulted altered due to ERG-overexpression in a time-dependent manner. Therefore, we performed a multiple enrichment analysis, selecting genes significant for the $log_2FC>|1.5|$ at each time point. Particularly, the number of selected genes per condition were the same resulted from the volcano plots, previously reported. Starting from the gene name list, the multiple enrichment was performed using the custom analysis ("Express Analysis" option) of Metascape, selecting as the reference databases: the Biological Process of the Gene Ontology (GO:BP) and the Ontologies of Canonical Pathways and Hallmark Gene Set. Results are reported in *Figure 29*.

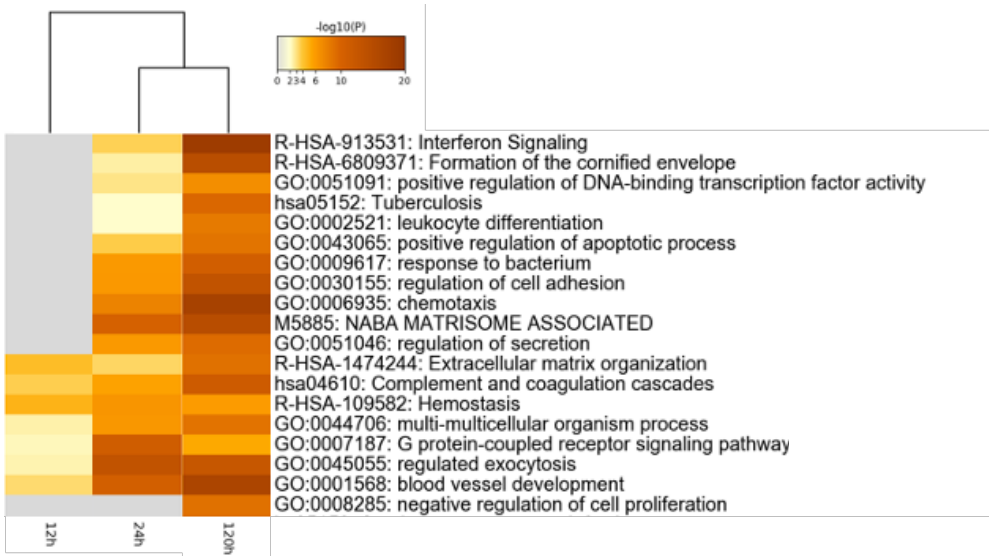


Figure 29| Enriched heatmap of the three experimental conditions obtained with genes significant for $log_2FC>|1.5|$. The heatmap was built in Metascape and the BP, canonical pathways and hallmark gene set ontologies were used as reference. The colour code in the top right legend is referred to the -log(P) calculated for the single process enrichment.

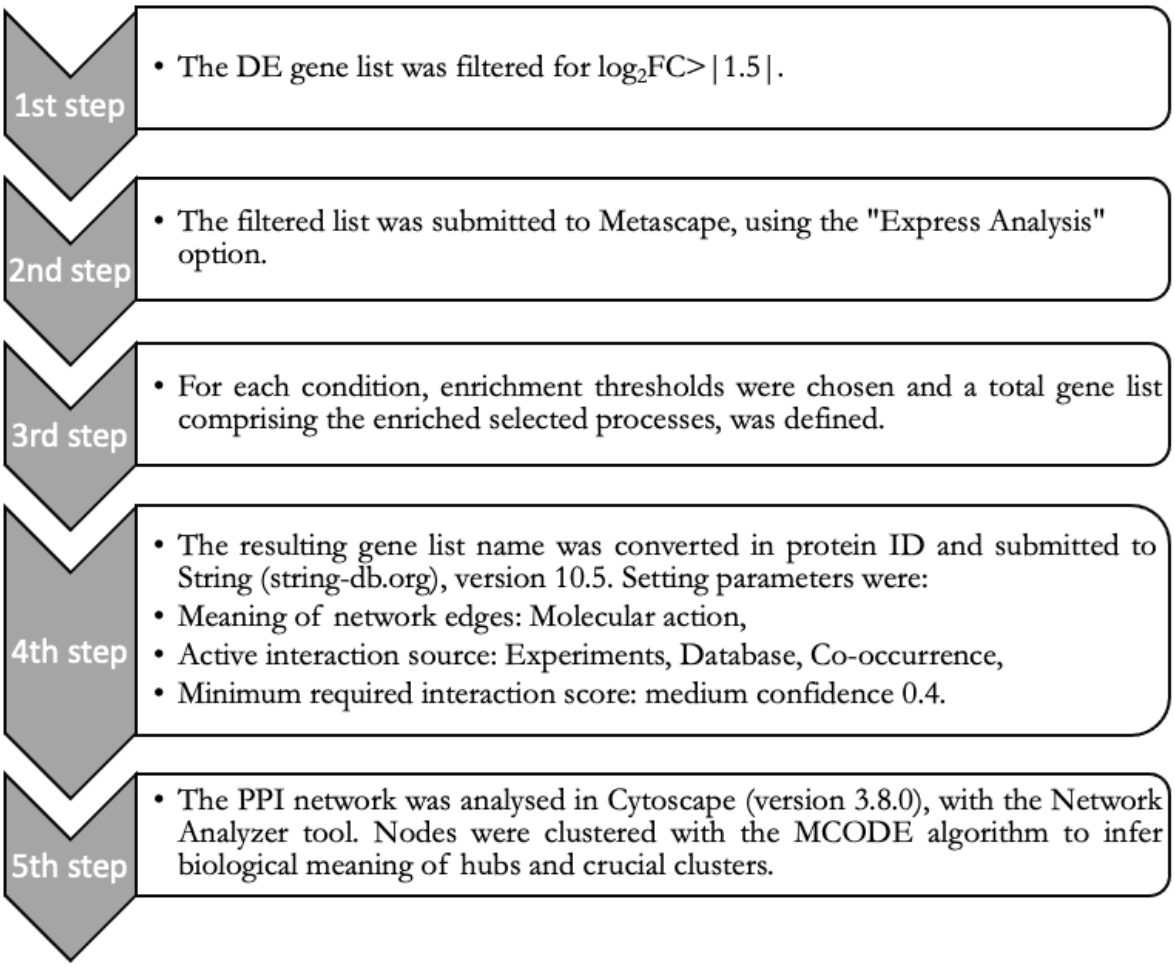
Interestingly, we observed an increasing trend from 12h to 120h for significant terms as the blood vessel development, the multi-multicellular organism process, the complement and coagulation cascade; a trend that peaked at 24h for the exocytosis, the G-protein coupled receptor signalling pathway, or that resulted reduced at 24h as in the extracellular matrix organization case. The major significance was instead assessed in the 24- and 120h time points, generally peaking at 120h as for the Interferon signalling, regulation to adhesion and so forth.

The multiple enrichment analysis thus highlighted that in an "early" phase ERG overexpression is already setting bases for a more complex pattern, chiefly associated to blood vessel development and cell response and organization, as also previously emerged in GSEA.

2.2.1 Enriched network analysis of significant DE genes related to the ERG-induced conditions

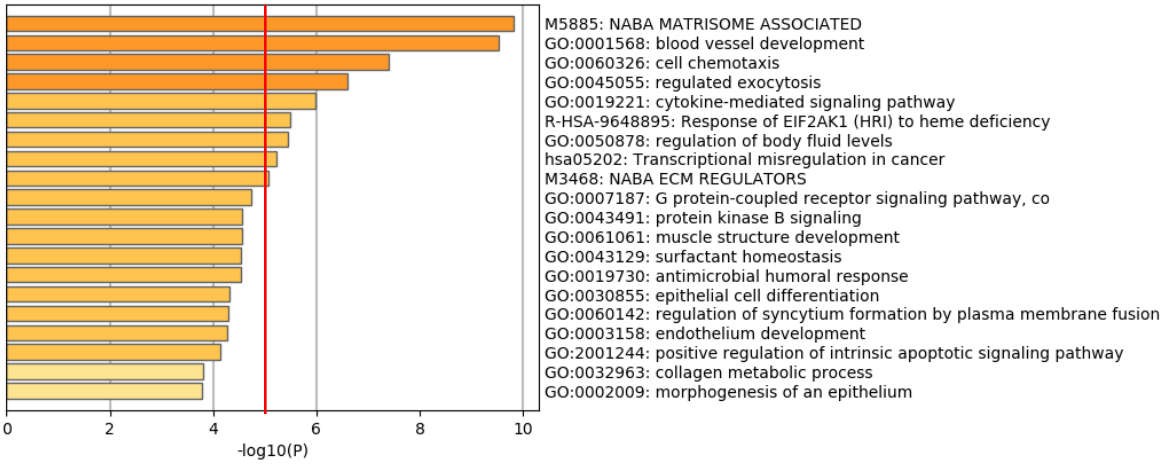
Previous analyses had highlighted that different biological processes resulted perturbed when ERG is overexpressed in a benign prostate epithelial model, supporting the idea that the high occurrence of the TE fusion gene in PCa has a driving role, even if it does not reflect a specific clinical status or meaning. Starting from these premises, we aimed to characterized which could be defined as exclusive mechanisms due to ERG overexpression, in a "early" or "late" timing, mimicking the steps of prostate cancer initiation and transformation. Moreover, this analysis was also aimed to investigate if we could intercept some ERG-direct processes to try to determine which could be the molecular regulation promoted by ERG overexpression. To answer these questions, we chose a systems biology approach combined with a preliminary enrichment step and therefore network analyses were defined.

Firstly, the approach was used to evaluate which were the DE genes that behaved as hubs in relevant biological processes and that were directly or indirectly perturbed by ERG overexpression. We started from the DEG lists of each condition and then proceeded with the following workflow:



Due to the limited number of genes, we did not obtain any enrichment for the 12h condition (data not shown). Inversely, the 24h condition revealed the following enriched terms: (i) the Naba matrisome associated, (ii) blood vessel development, (iii) cell chemotaxis, (iv) regulated exocytosis, (v) cytokines-mediated signalling pathway, (vi) the response of EIF2AK1, (vii) regulation of body fluid levels, (viii) transcriptional misregulation in cancer, and (ix) Naba ECM regulators (*Figure 30*).

A



B

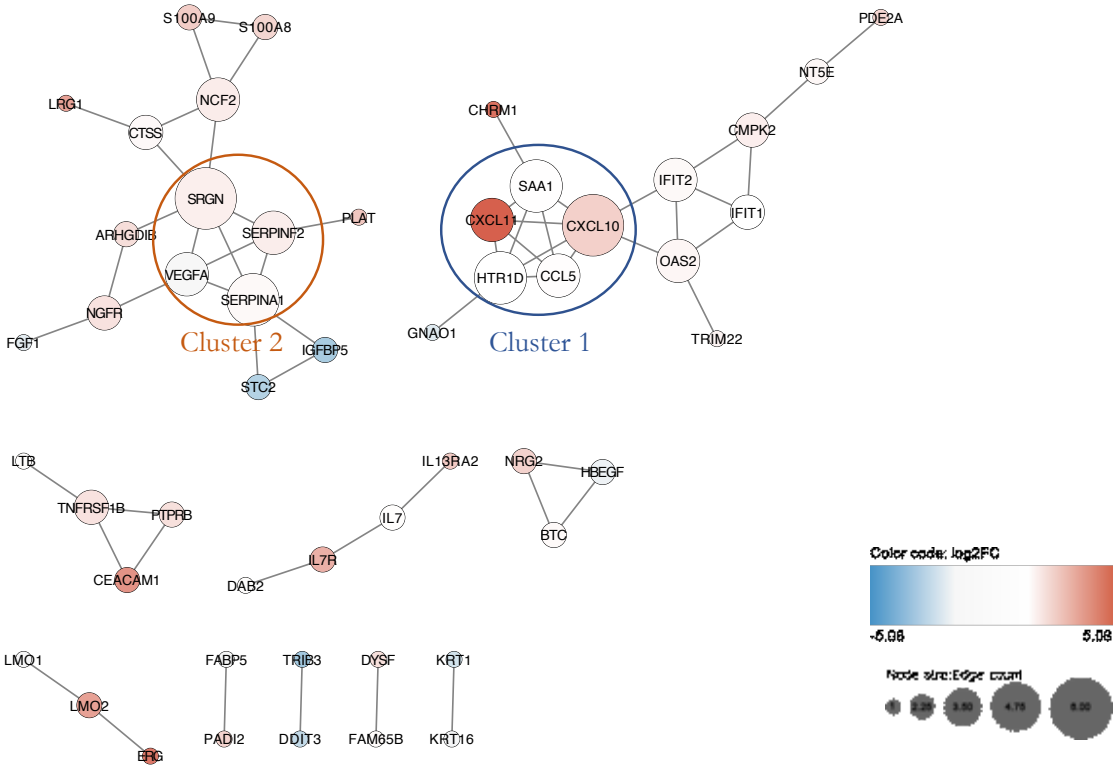


Figure 30| **A)** Enrichment analysis obtained with Metascape and relating significance expressed as -log10(P). The red line indicates the threshold selected. **B)** Network derived from the enrichment analysis. Protein colour code reflected the log_2FC of DEG analysis, while nodes size depends on the degree centrality score of the network, as shown in the bottom right legend.

The threshold was set at P value= 0,00001 that corresponded to -log10(P)=5, as indicated by the red line. Genes enriching these terms were then used for the network building, as previously described, and clustered using the MCODE algorithm [136]. The algorithm revealed two different clusters, that were further characterized detecting the hub proteins, defined as nodes with the highest Degree Centrality (sum of in- and outdegree) in the network. Particularly:

- Cluster 1 represented the lymphocyte chemotaxis term, and the hub protein was CXCL10 (the IFN- γ induced protein 10).
- Cluster 2 represented the platelet degranulation term, and the hub protein was SRGN (Secretory Granulate Proteoglycan core protein).

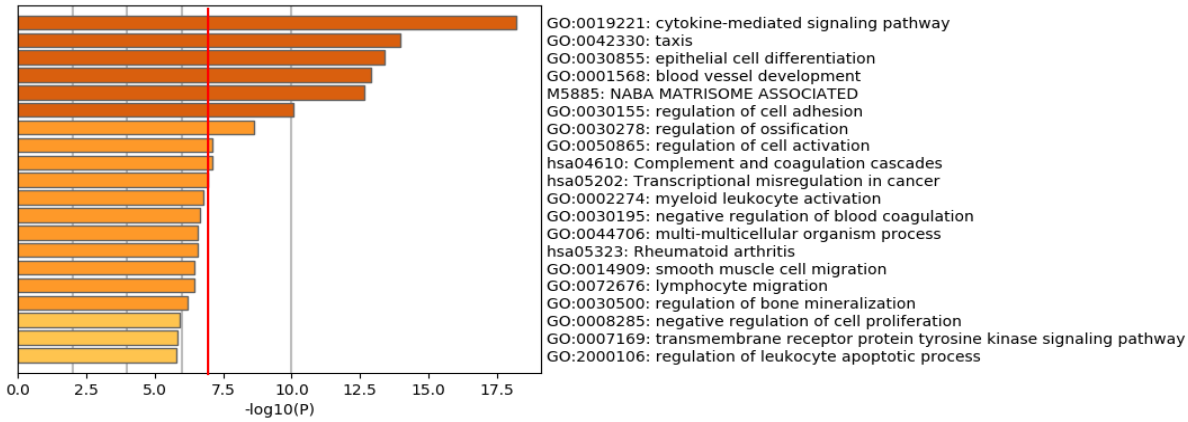
The same workflow was applied for the 120h time point analysis and results illustrated in *Figure 31*. In this case, due to the large number of significant processes, the enrichment threshold was set at P value= 0,0000001 (or rather -log10(P)=7). The gene lists selected for the network analysis resulted: (i) cytokine-mediated signalling pathway, (ii) taxis, (iii) epithelial cell differentiation, (iv) blood vessel development, (v) Naba matrisome associated, (vi) regulation of cell adhesion, (vii) regulation of ossification, (viii) regulation of cell activation, (ix) complement and coagulation cascade, (x) transcriptional misregulation in cancer. Interestingly, some terms were also previously obtained in the 24h enriched network analysis, highlighting that relevant biological processes resulted in the "early" time point were still relevant at 120h, acquiring complexity. The network analysis defined the following clusters and hub genes:

- Cluster 1 represented the Interferon α/β response, the hub protein was OAS2 (2'-5'-oligoadenylate synthetase 2).
- Cluster 2 represented the lymphocyte migration, the hub protein was C3 or rather the complement component 3.
- Cluster 3 represented the cornification process, the hub protein was KRT6C (Keratin 6C).

The enriched network analysis revealed that ERG might really influence the microenvironment fate, since different hints emerged concerning lymphocytes chemotaxis and migration, even at 24h. The Interferon response was strongly significant also at 120h, while the cornification process was curiously found downregulated in our data. Further observations of these two processes will be later discuss.

In addition, we tried to intercept common mechanisms between the pre-malignant condition represented by our cell model when ERG was induced, and adenocarcinomas presented in

A



B

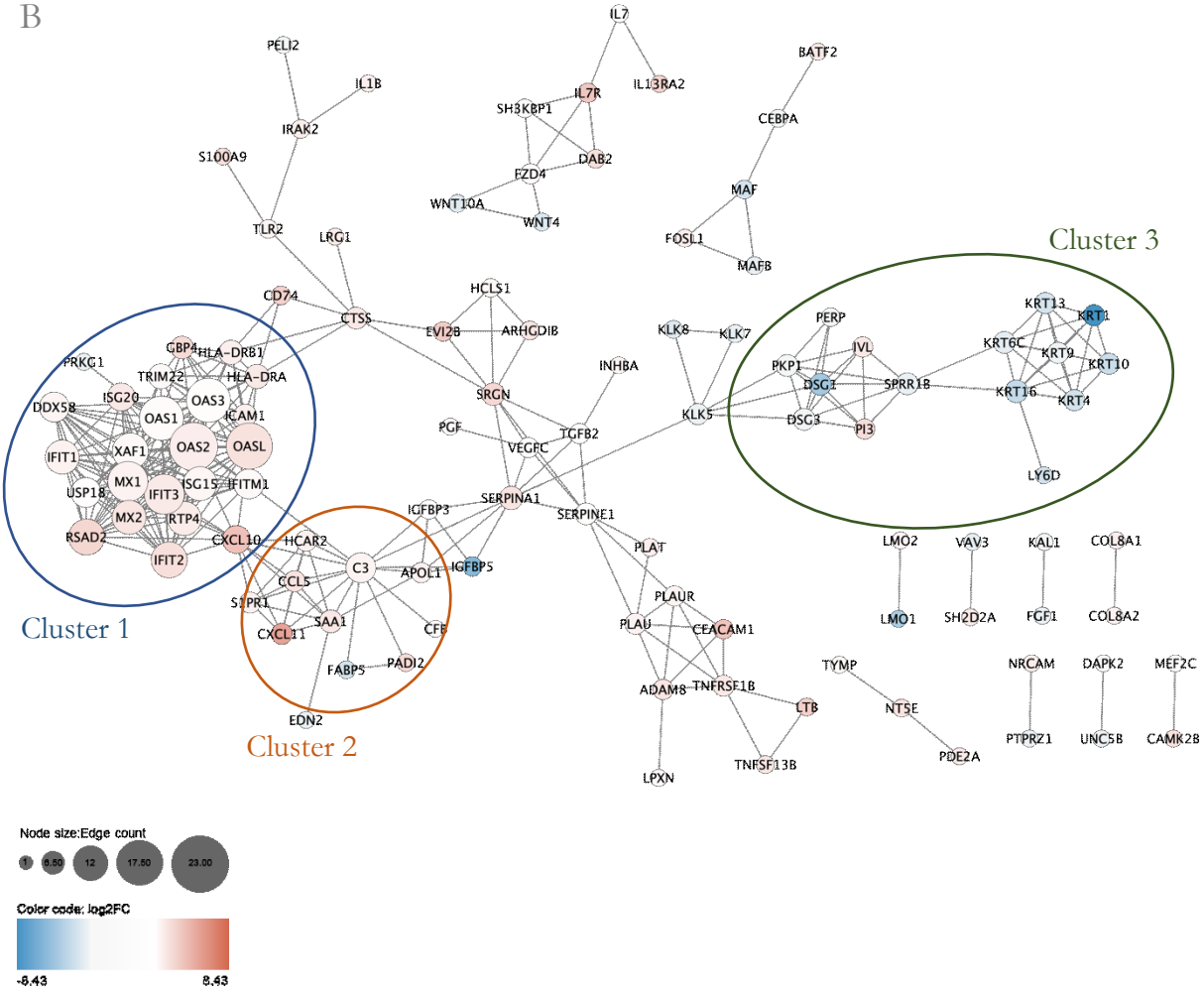


Figure 31 | **A)** Enrichment analysis obtained with Metascape and relating significance expressed as -log10(P). The red line indicates the threshold selected. **B)** Network obtained from the enrichment analysis. Protein colour code reflected the log_2FC of DEG analysis, while nodes size depends on the degree centrality score of the network, as shown in the bottom left legend.

the PRAD database. Of course, hypothetical common mechanisms could also share different gene lists, taking into account that other pathways could be crucial in tumours in order to sustain specific biological functions. Therefore, we performed an enrichment analysis in Metascape for the PRAD subset gene list (1857 DE genes significant for P value-adjusted <0.05 and selected for $log_2FC>|1.5|$), using the same parameters (*Figure 32*).

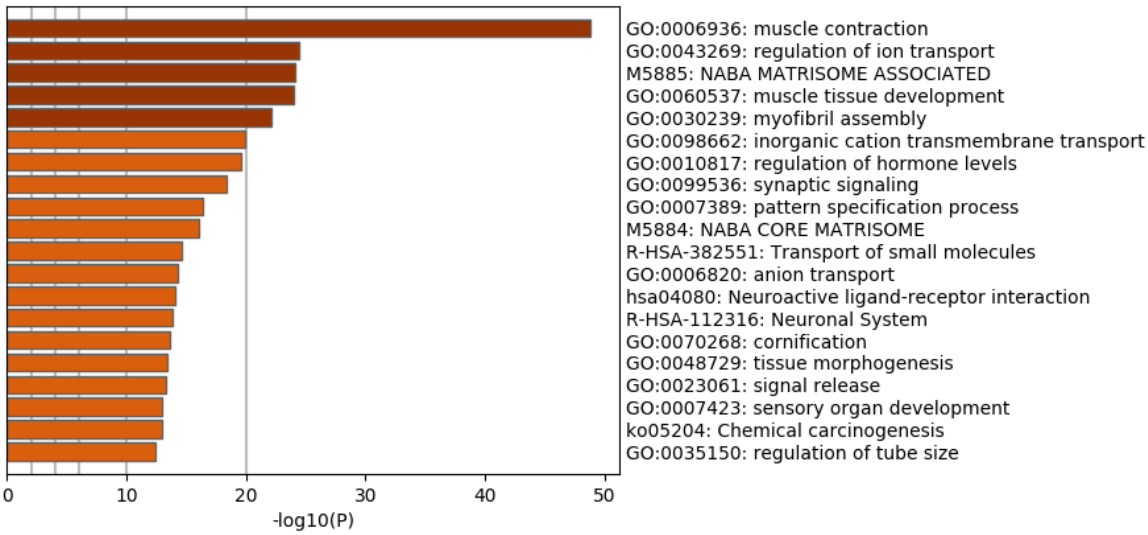


Figure 32| Enrichment analysis of DE genes from the PRAD subset, selected for the P value_adj<0.05 and the $log_2FC>|1.5|$, using Metascape.

The Naba matrisome associated term was found in common between 120h and PRAD data. This term defines the group of genes encoding extracellular matrix and extracellular matrix-associated compounds that comprise:

- ECM-affiliated protein,
- ECM regulators,
- ECM-remodelling enzymes, crosslinkers, proteases, regulators,
- Secreted factors,

obtained from a protein specific domain study [137]. Particularly, ECM represents a complex scaffold of proteins that are crucial in the TME context, orchestrating cellular processes deregulated during cancer progression (proliferation, invasion, migration, angiogenesis..); yet its peculiar composition remained to be explored [138].

Our results suggested that even on the long run, the alteration of the ECM composition could be defined as an ERG-dependent effect. The biological rationale could be related to the promotion

of cell migration, inhibition of apoptosis and proliferation as well as the chaotic vascular formation that does the groundwork to favour tumour progression and metastasis [139][140].

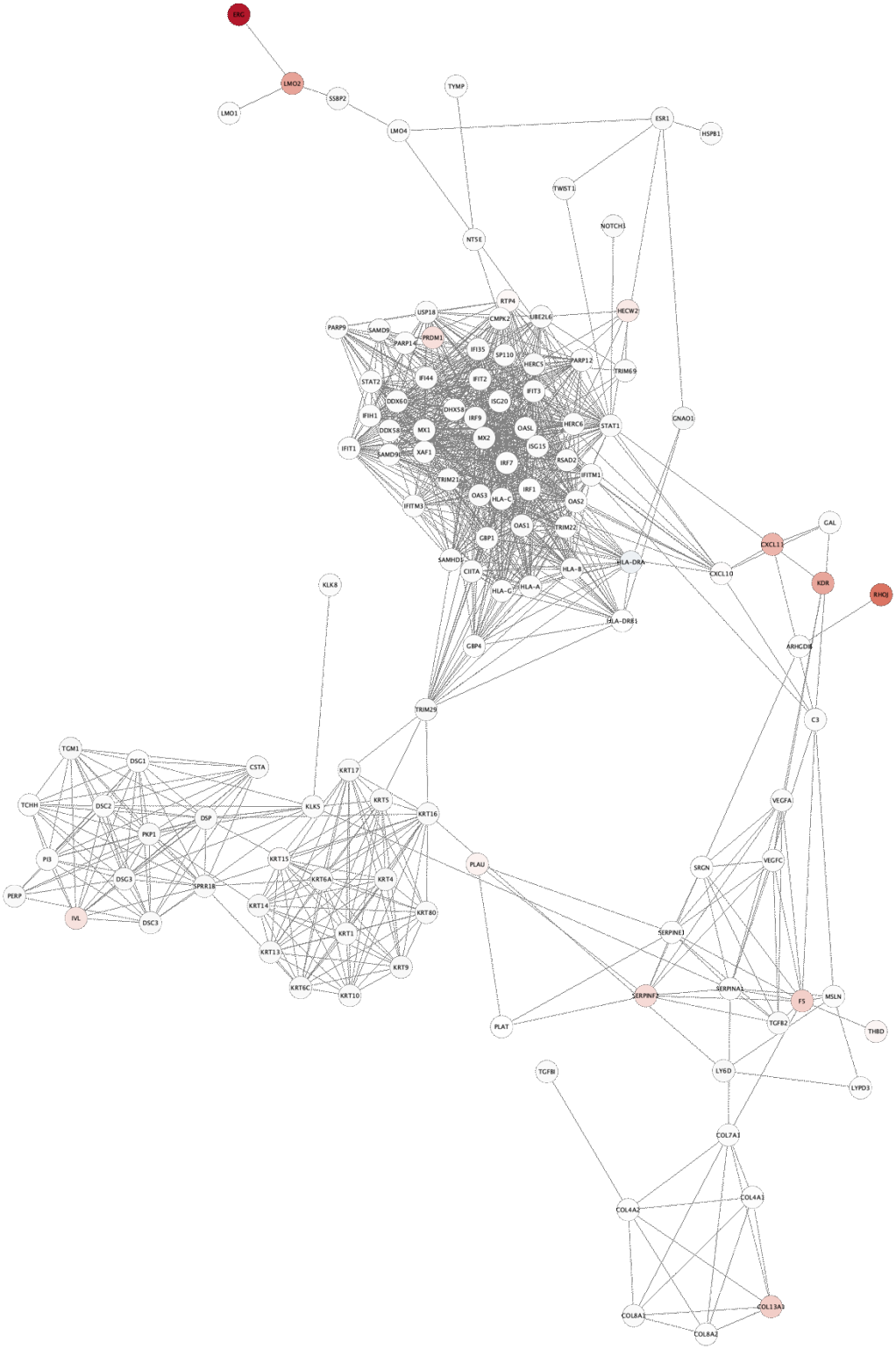
2.2.2 Global network analysis of significant DE genes related to the ERG-induced conditions

Beyond the enriched network analysis aimed to intercept strongly represented biological terms and hubs, we performed also a global network analysis of the 507 DE genes for each time point. The overall aim was to define ERG interactions to intercept crucial hubs that could or not be directly transcriptionally regulated therefore all genes were considered, without filtering for the log_2FC. Firstly, we converted Gene Symbol into Protein ID and submitted the entire gene list in String-db (Version 10.5), with the same setting parameters previously defined (see Results 2.2.1), and then used Cytoscape to perform the network analysis. The entire network at 120h is reported in Appendix IV, while the most relevant nodes for each condition were shown in *Figure 33*. The global network analysis was obtained with the "Network Analyzer" tool in Cytoscape and hub protein was defined as the highest Degree Centrality node one.

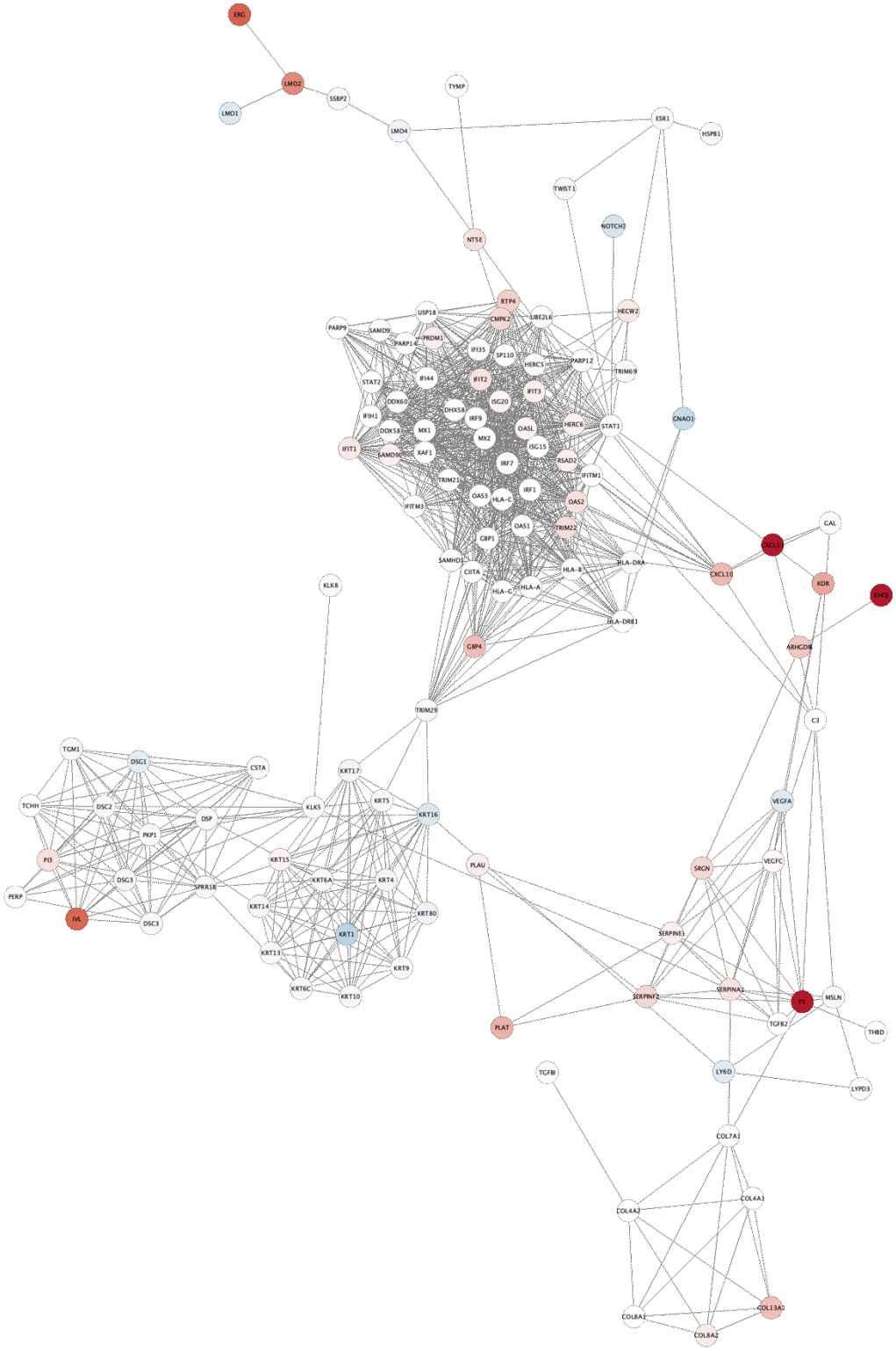
Direct ERG edges were properly determined by deregulated Lim-only proteins (LMOs), crucial in hematopoietic stem cell maintenance and involved in PCa migration and invasion, further deepen in the discussion section [141][142].

The most relevant hub protein (Degree 57) resulted the Signal transducer and activator of transcription STAT1, a transcription factor activated by the Interferon signalling. Although it is merely transcriptionally upregulated at 120h (log_2FC= 1.22), STAT1 could also be regulated through signal transduction. Of note, due to the complexity of the global network analysis, we chose to focus on nodes that could be considered ERG neighbours, or rather nodes with higher Closeness Centrality and that were found to be potentially connected to ERG in different studies (in physiologic as in pathologic conditions). In details, we tried to infer the biological meaning of edges in order to depict a potential molecular mechanism deregulated by ERG overexpression, that will be later discuss.

A


ERG
LMO2
LMO1
SSBP2
TYMP
LMO4
ESR1
HSPB1
TWIST1
NOTCH3
NT5E
RTP4
USP18
CMPK2
UBE2L6
HECW2
PARP9
SAMD9
PRDM1
PARP14
IFI35
SP110
HERC5
PARP12
TRIM69
STAT2
IFI44
IFIT2
IFIT3
DDX60
ISG20
OHXS8
IRF9
IFIH1
DDX58
MX1
OASL
HERC6
STAT1
GNAO1
IFIT1
SAMD9L
XAF1
MX2
ISG15
IRF7
RSAD2
IFITM1
TRIM21
OAS3
HLA-C
IRF1
OAS2
IFITM3
OAS1
TRIM22
GBP1
SAMHD1
CIITA
HLA-B
HLA-DRA
CXCL10
CXCL11
GAL
KDR
RHOJ
KLK8
HLA-A
HLA-C
HLA-DRB1
GBP4
ARHGDIB
TRIM29
C3
TCM1
DSG1
CSTA
KRT17
TCHH
DSC2
KLK5
KRT5
KRT16
DSP
PKP1
VEGFA
PI3
KRT15
PLAU
SRGN
VEGFC
DSG3
KRT6A
KRT4
SPRR1B
PERP
KRT14
KRT80
IVL
DSC3
KRT1
SERPINE1
KRT13
KRT9
KRT6C
KRT10
SERPINF2
SERPINA1
F5
MSLN
TGFB2
THBD
PLAT
LY6D
LYPD3
TGFBI
COL7A1
COL4A2
COL4A1
COL8A1
COL13A1
COL8A2

B


ERG
LMO2
LMO1
SSBP2
TYMP
LMO4
ESR1
HSPB1
TWIST1
NOTCH3
NT5E
RTP4
USP18
CMPK2
UBE2L6
HECW2
PARP9
SAMD9
PARP14
PRDM1
IFI35
SP110
HERC5
PARP12
TRIM69
STAT2
IFI44
IFIT2
IFIT3
ISG20
ODX60
DHX58
IRF9
IFIH1
ODX58
MX1
OASL
HERC6
STAT1
GNAO1
IFIT1
SAMD9L
XAF1
MX2
ISG15
IRF7
RSAD2
IFITM1
TRIM21
OAS3
HLA-C
IRF1
OAS2
IFITM3
OAS1
TRIM22
GBP1
SAMHD1
CIITA
HLA-B
HLA-DRA
CXCL10
CXCL1
GAL
KDR
RHOJ
KLK8
HLA-G
HLA-A
HLA-DRB1
GBP4
ARHGDIB
TRIM29
C3
TGM1
DSG1
CSTA
KRT17
TCHH
DSC2
KLK5
KRT5
KRT16
DSP
PKP1
PI3
KRT15
KRT6A
KRT4
DSG3
SPRR1B
PERP
IVL
DSC3
KRT14
KRT80
KRT1
KRT13
KRT9
KRT6C
KRT10
PLAU
VEGFA
SRGN
VEGFC
SERPINE1
SERPINF2
SERPINA1
F3
MSLN
TGFB2
THBD
PLAT
TGFBI
LY6D
LYPD3
COL7A1
COL4A1
COL4A2
COL8A1
COL13A1
COL8A2

C

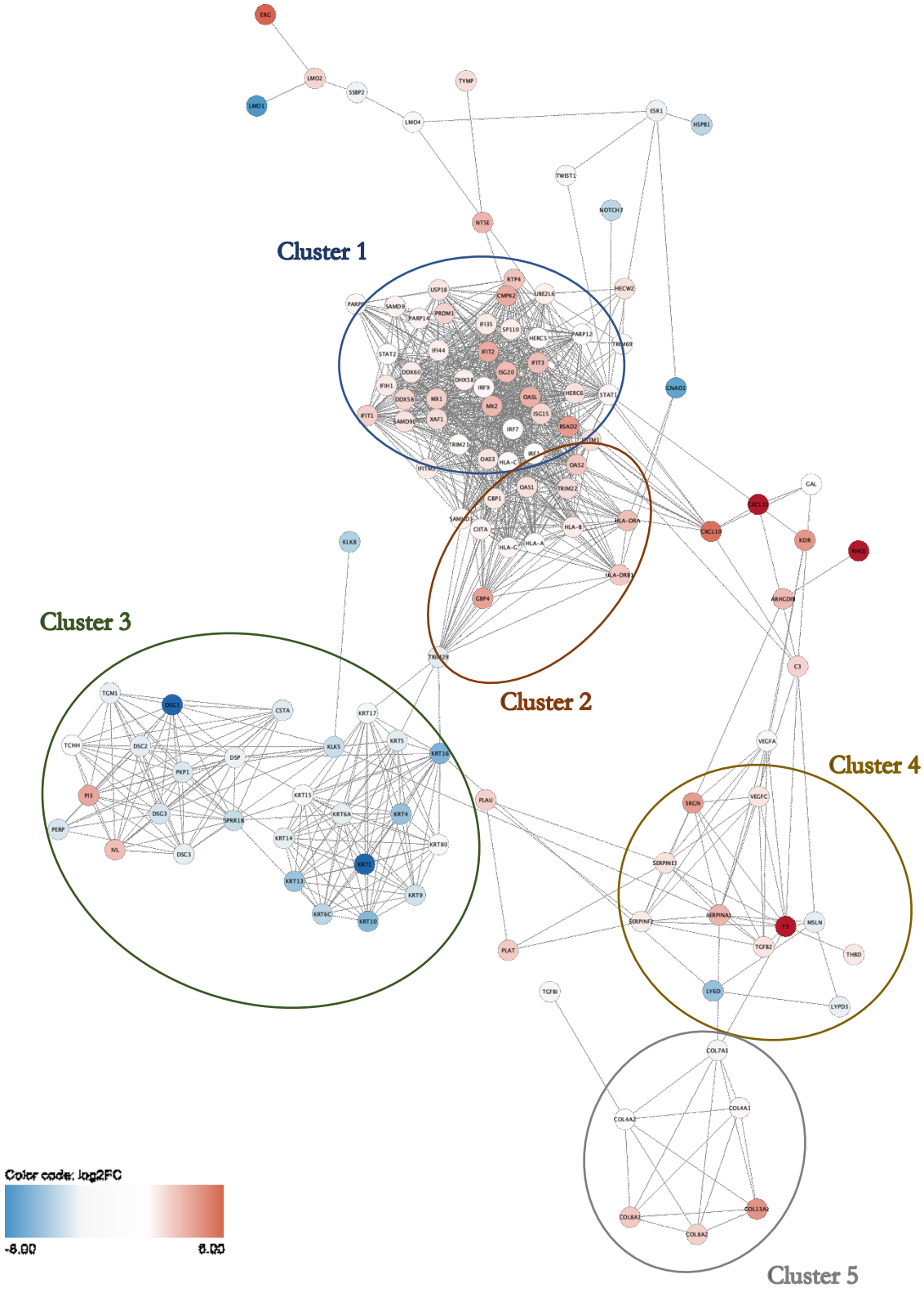


Figure 33| Global network analysis, selection of the 507 DE genes. **A)** 12h, **B)** 24h, **C)** 120h. Colour code is indicated in the bottom left legend. Relevant clusters were circled and further characterized.

Global analysis revealed that ERG overexpression was strongly associated to type I and II IFN signalling activation, with the upregulation of various targets. Particularly, some of the most relevant targets were also confirmed with qPCR analyses (*Figure 34*).

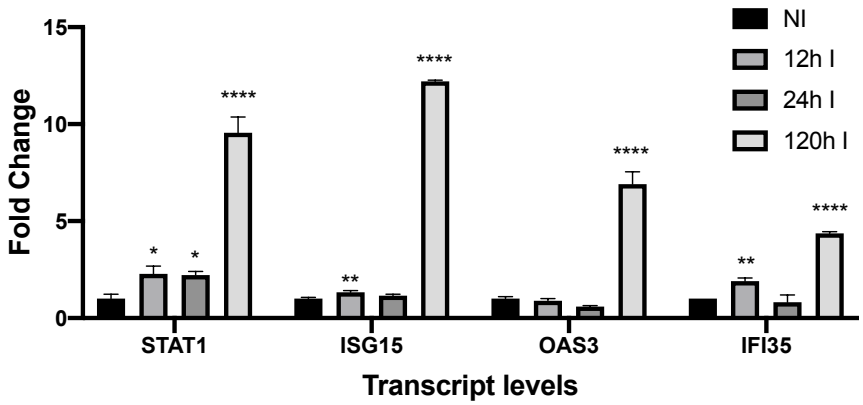


Figure 34| qPCR analysis of IFN signalling selected targets. Statistical significance was evaluated with an Anova One-way test using the multiple comparison option and comparing each I versus its NI. *P value≤ 0.05, ** P value≤ 0.01, *** P value≤ 0.001, ****P value ≤ 0.0001.

Overall, the global network analysis showed five major clusters obtained with the MCODE approach (reported in *Figure 33 (C))* and for their biological meaning was further evaluated with Metascape. The resulting clusters were associated to:

- Interferon α/ β signalling (cluster 1),
- Interferon γ signalling (cluster 2),
- Cornification process (cluster 3),
- Platelet degranulation (cluster 4),
- NABA collagens (cluster 5).

Both type I and II IFN signalling, the platelet degranulation and the NABA collagens GO terms resulted upregulated, while the cornification process was found downregulated. Expression values of relevant nodes for each cluster were showed in the following paragraphs. Moreover, the potential molecular mechanisms that bound ERG overexpression and the deregulation of specific biological terms, would be later discussed.

2.2.2.1 Interferon response: type I and II activation (clusters 1 and 2)

The IFN signalling is characterized by cytokines with antiproliferative, antiviral and immunomodulatory effects that mediate the transcriptional regulation of a plethora of genes involved in various biological processes. In our cell model, type I and II IFN signalling defined the cluster 1 that was strongly upregulated at 120h of ERG overexpression. The expression values of nodes that particularly characterized both signalling was therefore showed in *Figure 35*.

ID	12h	24h	120h	PRAD subset
LMO1	0,13	-1,52	-4,39	-0,97
SSBP2	-0,42	-0,81	-1,16	-0,49
LMO4	0,03	-1,13	-0,72	-0,92
STAT2	0,18	0,54	1,04	0,31
IRF7	0,32	-0,02	1,07	0,68
STAT1	-0,22	0,17	1,22	0,26
IFITM3	0,26	0,81	1,45	-0,41
IFIH1	0,18	0,8	1,48	0,42
ISG15	0,06	0,94	1,88	1,29
LMO2	3,06	3,62	1,97	-0,69
HLA-DRB1	0,82	0,38	2,2	-0,37
OAS2	0,21	1,7	2,27	0,84
IFIT3	0,01	1,36	2,45	-0,56
CAMKIIB	1,58	3,25	2,77	0,63
IFIT2	0,12	1,64	2,91	-0,61
ESR1	-0,53	-0,23	-1,22	
HLA-A	-0,01	0,03	1,02	
IRF1	0,49	0,71	1,02	
HLA-C	0,16	0,4	1,14	
IRF9	-0,17	0,35	1,21	
CIITA	0,03	0,58	1,27	
HLA-G	0,6	1,03	1,27	
IFI44	-0,02	0,91	1,38	
HLA-B	0,1	0,4	1,45	
IFI35	0,12	0,4	1,47	
OAS1	-0,14	0,8	1,68	
IFITM1	-0,1	0,53	1,76	
IFIT1	0,15	1,51	2,08	
HLA-DRA	-1,2	0,01	2,44	
ISG20	0,57	1,25	2,49	
OASL	0,01	1,25	2,78	

Figure 35| Fold changes of selected nodes significant for FDR<0.05 obtained by RNA sequencing analyses in experimental and PRAD data, defining clusters 1 and 2. In red, genes that didn't show correspondence in the PRAD subset. Missing values referred to not significant DE genes in PRAD.

2.2.2.2 Cornification process (cluster 3)

The cornification GO term describes a type of programmed cell death that occurs in epidermis, causing the accumulation of specific proteins (keratin, involucrin, loricrin) and lipids (fatty acids and ceramides). Global network analysis revealed that the majority of nodes defining cluster 3 were downregulated when ERG was induced, peaking at 120h (expression values were showed in *Figure 36*).

Particularly, keratins conserved a typical epithelial cell type-specific expression and were extensively used as epithelial markers in physiologic as in pathologic conditions [143].

ID	12h	24h	120h	PRAD subset
IVL	1,66	4,43	2,46	-3,74
KRT13	-0,58	-0,71	-2,83	-3,18
KRT16	-0,32	-1,56	-3,43	-2,86
KRT4	-0,77	-1,10	-2,91	-2,66
KRT9	-0,71	-0,64	-1,79	-2,21
DSG3	-0,34	-1,02	-1,70	-1,88
SPRR1B	-0,55	-0,91	-1,95	-1,84
KRT14	-0,69	-0,91	-1,32	-1,79
KRT6A	-0,59	-0,74	-1,32	-1,79
KRT15	1,08	1,28	-0,74	-1,67
KRT5	-0,43	-0,56	-1,41	-1,63
KRT1	-0,09	-2,23	-8,43	-1,58
TRIM29	-0,56	-0,4	-1,39	-1,52
CSTA	-0,15	-0,27	-1,49	-1,31
KRT6C	-0,44	-0,64	-2,20	-1,25
KRT17	-0,59	-1,06	-1,05	-1,14
DSC3	-0,49	-0,71	-1,32	-1,13
KRT80	-0,24	-1,13	-0,53	-0,66
PKP1	-0,39	-0,44	-1,58	-0,59
KRT10	-0,36	-0,70	-3,24	-0,22
DSC2	-0,10	-0,52	-1,37	1,15
DSG1	0,40	-1,4	-4,73	
TGM1	-0,90	-0,95	-1,28	
DSP	-0,47	-0,45	-1,20	
TCHH	0,52	1,03	0,37	
PI3	0,17	1,7	2,99	

Figure 36| Fold changes of selected nodes significant for FDR<0.05 obtained by RNA sequencing analyses in experimental and PRAD data, defining clusters 3. In red, genes that didn't show correspondence in the PRAD subset. Missing values referred to not significant DE genes in PRAD.

Of note, prostatic epithelium lineage can be classified in three distinct types with a typical keratin pattern [144]:

- Basal ($KRT5^+$, $KRT14^+$),
- Luminal ($KRT8^+$, $KRT18^+$),
- Intermediate ($KRT5^+$, $KRT14^+$, $KRT8^+$, $KRT18^+$).

In our cell model, *KRT8* and *KRT18* resulted highly expressed also in NI conditions and no significant changes were detected after induction (at any time point). Inversely, *KRT5* and *KRT14* were significantly downregulated, peaking at 120h and we further confirmed this trend with a qPCR analysis (*Figure 37*).

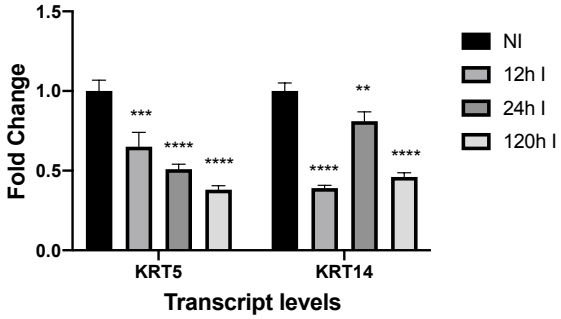


Figure 37| qPCR of KRT5 and KRT14. Statistical significance was evaluated with an Anova One-way test using the multiple comparison option and comparing I vs NI. ** P value≤ 0.01, *** P value≤ 0.001, ****P value ≤ 0.0001.

Some recent hints highlighted that ERG overexpression enhanced a luminal-like phenotype, increasing KRT8/KRT18 expression and concomitantly reducing KRT5/KRT14 one. This lineage plasticity switch was supported by an ERG-dependent epigenetic reprogramming, aiming to attenuate the basal phenotype in prostate cancers. In details, Li *et al.* [145] demonstrated that promoters of KRT8/KRT18 showed an increased chromatin accessibility and higher H3K27ac levels; conversely KRT5/KRT14 accessibility was considerably reduced, as well as H3K27ac levels (in *Figure 38*, ChIP-Seq of specific genes were reported).

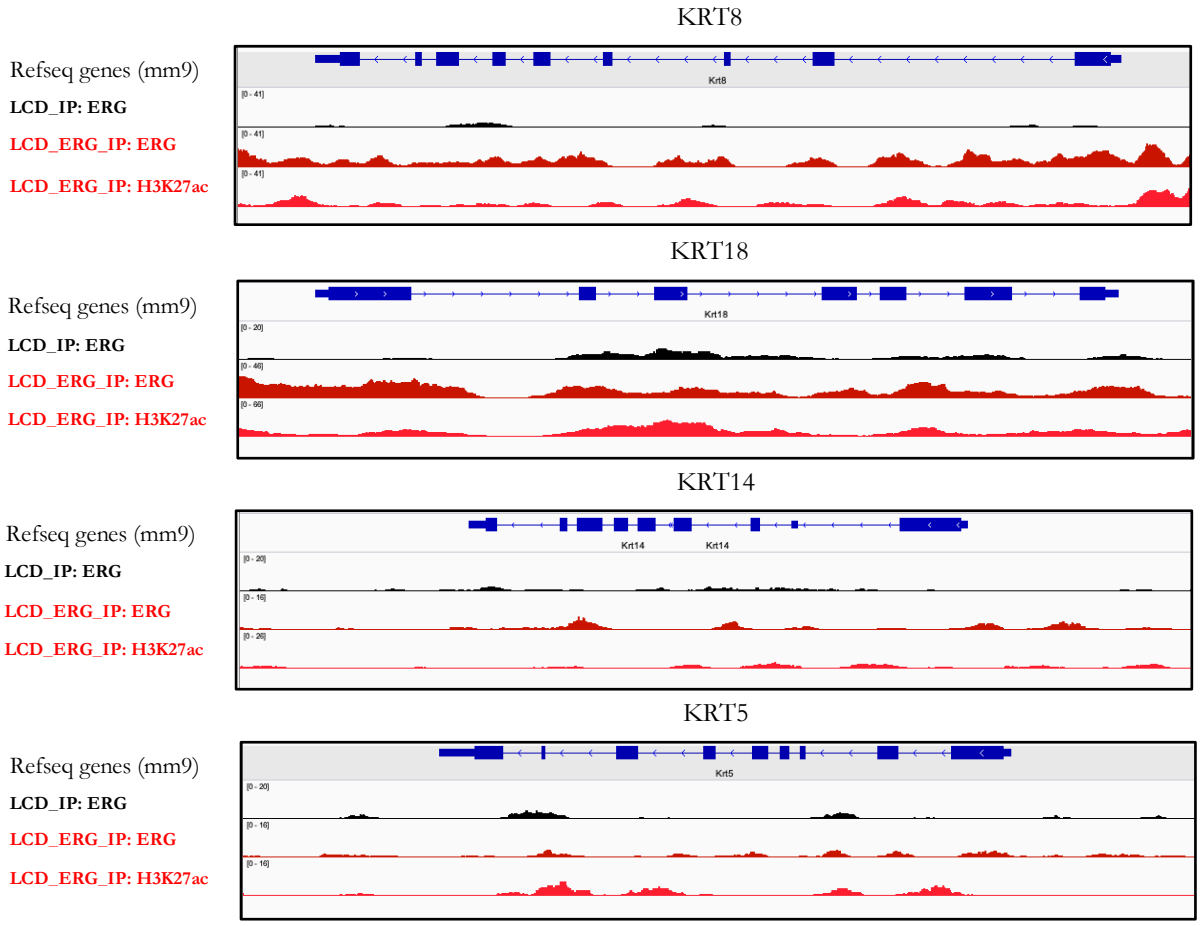
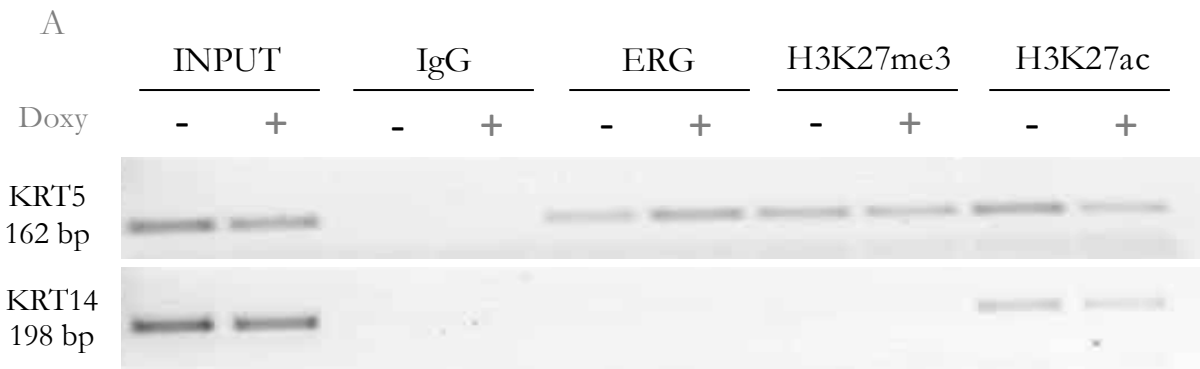


Figure 38|ChIP-Seq data obtained from biosino.org/node/OEX002110. Sequencing analysis was performed on luminal cells derived from murine organoids (LCD) and LCD overexpressing ERG (LCD_ERG). Samples were immunoprecipitated for ERG and H3K27ac. KRT8, KRT18, KRT14, KRT5 were reported as targets of interest. Sequences were annotated with the mm9 reference genome, accordingly to indication reported in Li *et al.* work [145].

Furthermore, a very preliminary ChIP analysis on RWPE-1 cells (*Figure 39*), seemed to suggest a similar regulation also in our model, showing a reduced level of H3K27ac for KRT5 and KRT14 when ERG was induced at 120h, while no differences were shown for H3K27me3 levels.



B

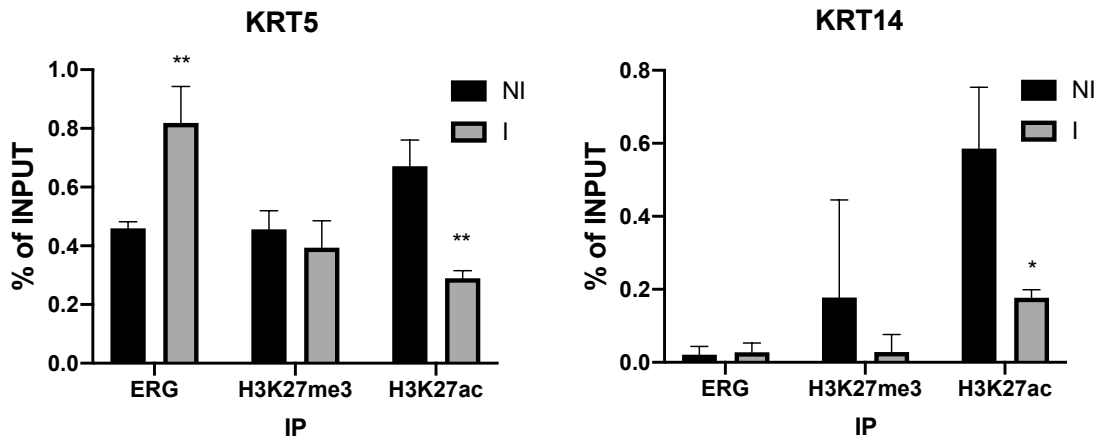
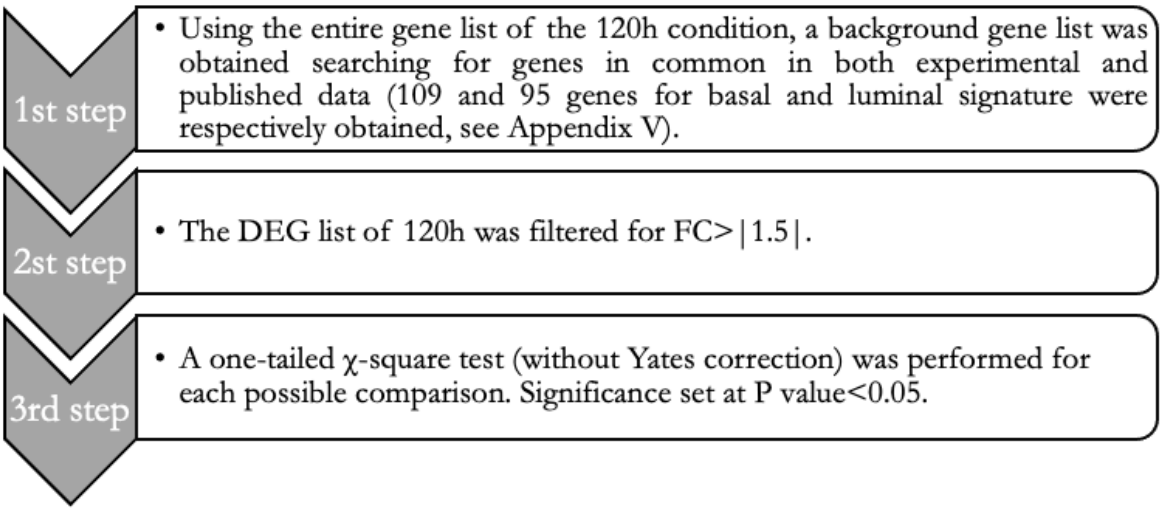


Figure 39| **A)** Single ChIP analysis of RWPE-1 cells induced or not for ERG overexpression at 120h. IP analysed: ERG, K3K27me3 and H3K27ac. KRT5 and KRT14 were selected as targets. ChIP result was evaluated with a semiQ PCR analysis. **B)** Densitometric analysis were estimated for each condition (NI and I) and expressed as % of INPUT used.

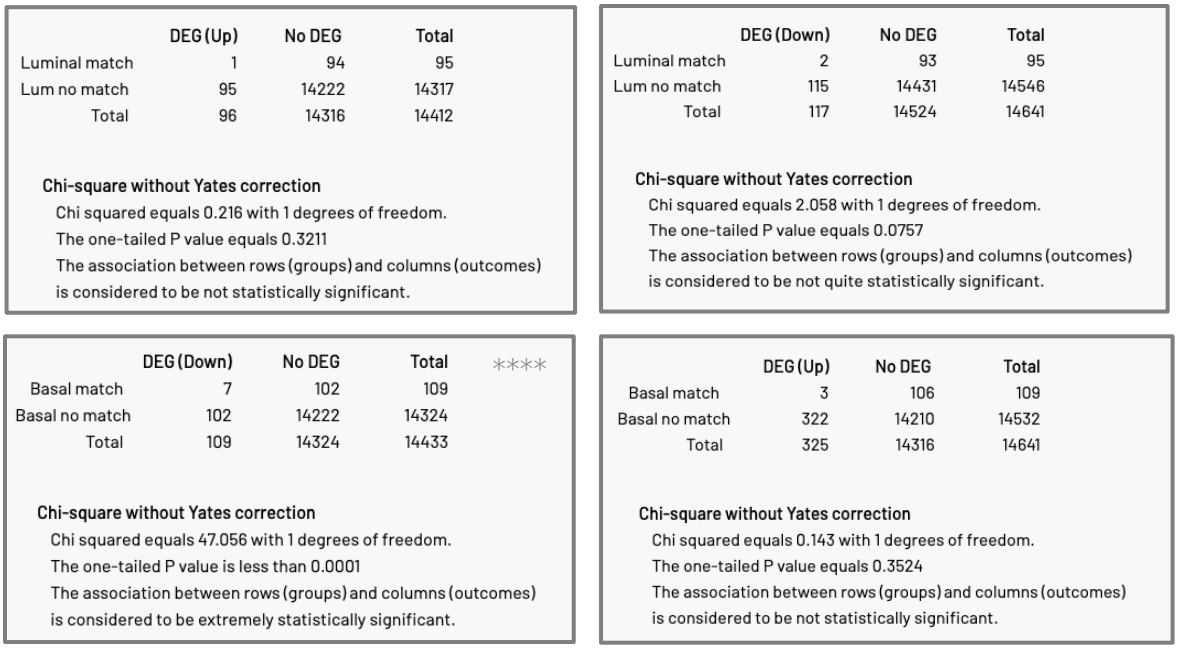
In details, this single analysis also highlighted that ERG did not immunoprecitated with KRT14, although seemed to be involved in its epigenetic regulation and also suggested that ERG might play a central and direct role in the acetylation regulation.

Of course, this was just a preliminary analysis that needed further replicates and also the characterization of other PTMs, in order to identify how keratin expression is regulated, even though showed consistence with previously published data.

Given these premises, we speculated that our model mimicked an "intermediate" step where basal markers resulted already downregulated while luminal ones were not yet prompted. A preliminary confirmation was obtained comparing our DEG list at 120h to basal and luminal metagene lists, recently published by Mapelli *et al.* [146]. The detailed workflow was showed below:



Interestingly, the only basal vs downregulated DEG comparison was statistically significant (all comparisons were showed in *Figure 40*), supporting the hypothesis that at 120h ERG overexpression started to induce a lineage switch, with the downregulation of basal-specific markers, while luminal ones were not yet altered. Nevertheless, further and more extensive analyses needed to be done to deepen this aspect and better characterize also the not induced profile of RWPE-1.



	DEG (Up)	No DEG	Total
Luminal match	1	94	95
Lum no match	95	14222	14317
Total	96	14316	14412

Chi-square without Yates correction
Chi squared equals 0.216 with 1 degrees of freedom.
The one-tailed P value equals 0.3211
The association between rows (groups) and columns (outcomes) is considered to be not statistically significant.

	DEG (Down)	No DEG	Total
Luminal match	2	93	95
Lum no match	115	14431	14546
Total	117	14524	14641

Chi-square without Yates correction
Chi squared equals 2.058 with 1 degrees of freedom.
The one-tailed P value equals 0.0757
The association between rows (groups) and columns (outcomes) is considered to be not quite statistically significant.

	DEG (Down)	No DEG	Total	****
Basal match	7	102	109	
Basal no match	102	14222	14324	
Total	109	14324	14433	

Chi-square without Yates correction
Chi squared equals 47.056 with 1 degrees of freedom.
The one-tailed P value is less than 0.0001
The association between rows (groups) and columns (outcomes) is considered to be extremely statistically significant.

	DEG (Up)	No DEG	Total
Basal match	3	106	109
Basal no match	322	14210	14532
Total	325	14316	14641

Chi-square without Yates correction
Chi squared equals 0.143 with 1 degrees of freedom.
The one-tailed P value equals 0.3524
The association between rows (groups) and columns (outcomes) is considered to be not statistically significant.

Figure 40| One-tailed χ-square test without Yates's correction for all comparisons between luminal /basal signatures and 120h DEG down – or upregulated. The entire list of genes used for the test were reported in Appendix V.

2.2.2.3 Platelet degranulation (cluster 4)

The GO term "platelet degranulation" defines the process through which platelets release secretory granules containing mediators (histamine, serotonin, proteins), through a regulated exocytosis. Our results highlighted that ERG overexpression caused a strong upregulation of some genes in an early time window (12h). In PRAD subset, almost all nodes showed curiously an opposite regulation (as indicated in *Figure 41*). This result suggested that ERG played a role in platelet degranulation in the early phases of tumour initiation, increasing the expression of specific markers.

ID	12h	24h	120h	PRAD subset
VEGFA	0,05	-1,50	-0,82	-0,52
SERPINF2	1,82	1,91	1,44	-1,63
VEGFC	0,54	1,15	1,58	0,52
TGFB2	-0,16	0,41	1,62	-0,81
SERPINA1	0,55	1,63	2,69	-1,28
SRGN	0,77	1,87	3,20	-0,65
F5	2,08	5,00	4,70	2,34
RHOJ	4,13	5,37	5,81	-0,93
SERPINE1	0,50	1,34	1,65	

Figure 41| Fold changes of selected nodes significant for FDR<0.05 obtained by RNA sequencing analyses in experimental and PRAD data, defining cluster 4. In red genes that didn't show correspondence in the PRAD subset. Missing value referred to not significant DE genes in PRAD.

2.2.2.4 NABA collagens (cluster 5)

Collagen is a major constituent of basement membrane and ECM and, during carcinogenesis, cancer cells adhere to these compounds through integrins or laminins, exerting migratory activity. In particular, Burns *et al.* [147] showed an increase collagen content in biopsies from prostate cancer tissues that unfortunately did not reflect a clinical meaning. Although the association to clinicopathologic features is still controversial, the role of collagen in tumour progression resulted strongly correlated with higher invasion, promoted through the RhoC GTPase activation [148]. A recent work highlighted that the TE fusion and the subsequent ERG overexpression enhance the expression of bone marrow markers as COLA1A, that was not found significant in our DE gene list [149]. Although the uncertain biological role of collagens and ERG overexpression, we particularly found *COL13A1* upregulated since 12h of induction, while curiously *COL7A1* was downregulated. A panel of fold changes for all cluster 5 nodes is reported in *Figure 42*. In addition, the same genes in the PRAD subset showed an opposite expression, suggesting that collagen organization could be an early effect of ERG-dependent activity.

ID	12h	24h	120h	PRAD subset
COL7A1	0,10	0,03	-1,11	-0,54
COL4A2	0,42	0,55	1,02	-0,67
COL4A1	0,58	0,99	1,19	-0,50
COL8A2	0,48	1,44	2,06	-0,64
COL8A1	-0,24	0,89	2,22	0,51
COL13A1	2,13	2,44	3,44	-1,38

Figure 42| Fold changes of selected nodes significant for FDR<0.05 obtained by RNA sequencing analyses in experimental and PRAD data, defining cluster 5. In red genes that didn't show correspondence in the PRAD subset.

2.2.3 Comparison of ERG- and AR- dependent global network analysis

Although the *TE* fusion alteration depends on AR activity, a quite recent study [61] highlighted that ERG overexpression caused the disruption of the AR signalling itself, in PCa cell lines. The *TE* fusion and the consequent ERG overexpression attenuated in fact AR signalling, since AR and ERG shared various transcription targets. Therefore, using the PRAD selection of ERG^{+} patients (131), relevant nodes of the global network were further analysed to infer a possible exclusive role of ERG in terms of targets regulation.

In details, a subsequent classification of the 131 ERG^{+} patients following the combination of *AR* and *ERG* expression was done, and we obtained 32 patients per group:

- the AR^{High}/ERG^{Low} group, patients with *AR* expression value higher that the median (calculated on all the 131 ERG^{+} patients), combined with a lower than the median *ERG* expression value, obtained with the same criteria;
- the AR^{Low}/ERG^{High} group, patients with *AR* expression value lower than the median (calculated on all the 131 ERG^{+} patients), combined with a higher than the median *ERG* expression value.

Then a differentially expressed genes analysis between AR^{High}/ERG^{Low} vs AR^{Low}/ERG^{High} patients was performed with the edgeR method on the Galaxy platform and data filtered for FDR<0.05. Genes corresponding to relevant proteins previously defined in our global network analysis were investigated. Particularly, we compared proteins of the five relevant clusters between all the

available conditions: Induced vs Not Induced (at 12-, 24- and 120h), the PRAD subset *ERG*$^{+}$ vs Normal samples and the PRAD *ERG*$^{+}$ subset AR^{High}/ERG^{Low} vs AR^{Low}/ERG^{High}.

This preliminary analysis revealed that at least some nodes of the cornification cluster (DSG3, KRT1, KRT5, KRT6A, KRT6C, TRIM29) showed an opposite trend in the AR^{High}/ERG^{Low} vs AR^{Low}/ERG^{High} condition, revealing that the deregulation previously assessed might be due to an exclusive ERG-dependent effect. Particularly, the estrogenic response also was inversely regulated in patients with high levels of *AR* expression (data are reported in *Figure 43*).

Given the high overlap of AR and ERG binding sites shown by Massie *et al.* [150], our result might indicate that ERG overexpression enhances an aberrant expression of AR transcriptional programme acting on its targets and perturbing its canonical network .

Of note, in order to expand the magnitude of samples and go deeper in the antagonizing role of ERG and AR transcription programmes, we would aim to find other databases to perform similar analyses. Unfortunately, no databases showed comparable clinicopathologic indications, therefore we used this preliminary result just as a hint of the potential exclusive role of ERG overexpression in ruling some processes as the cornification and the estrogenic response. More detailed analyses would be necessary to determine the contrasting activities of AR- and ERG-dependent transcriptional programme and the pivotal ERG role in the aforementioned biological terms.

	EXPERIMENTAL DATA			PRAD	
ID	**12h**	**24h**	**120h**	**ERG+ vs N**	**AR+/ERG- vs AR-/ERG+**
CLUSTER 1 - CLUSTER 2					
ESR1	-0,53	-0,23	-1,22		0,72
IRF7	0,32	-0,02	1,07	0,68	-0,62
STAT1	-0,22	0,17	1,22	0,26	0,78
CIITA	0,03	0,58	1,27		0,92
IFIT1	0,15	1,51	2,08		0,71
HLA-DRA	-1,2	0,01	2,44		0,66
IFIT3	0,01	1,36	2,45	-0,56	0,94
IFIT2	0,12	1,64	2,91	-0,61	0,76
CLUSTER 3					
KRT1	-0,09	-2,23	-8,43	-1,58	1,65
KRT10	-0,36	-0,70	-3,24	-0,22	-0,61
KRT6C	-0,44	-0,64	-2,20	-1,25	1,64
DSG3	-0,34	-1,02	-1,70	-1,88	1,84
PKP1	-0,39	-0,44	-1,58	-0,59	1,01
KRT5	-0,43	-0,56	-1,41	-1,63	1,29
TRIM29	-0,57	-0,41	-1,39	-1,52	1,01
KRT6A	-0,59	-0,74	-1,32	-1,79	1,77
DSP	-0,47	-0,45	-1,20		0,46
CLUSTER 4					
VEGFA	0,05	-1,50	-0,82	-0,52	0,93
SERPINA1	0,55	1,63	2,69	-1,28	0,75
SRGN	0,77	1,87	3,20	-0,65	0,53
CLUSTER 5					
COL4A1	0,58	0,99	1,19	-0,50	0,70
COL8A1	-0,24	0,89	2,22	0,51	1,02

Figure 43| Selection of significant DE genes in common between different available datasets. Missing values in the ERG^{+} vs Normal condition highlighted genes that were not statistically significant (the threshold was FDR<0.05). In red, genes with a different trend between Induced vs Not Induced at 120h and the AR^{High}/ERG^{Low} vs AR^{Low}/ERG^{High} conditions were reported. Missing values referred to not significant DE genes.

3. The marginal role of DNA methyltransferase in the ERG inducible RWPE-1 cell model

3.1 Global DNA methylation evaluation

To elucidate the significant deregulation of DNMTs and UHRF1 seen in previous analyses, together with a slight but statistically significant *CDH1* downregulation, a focused pyrosequencing analysis was performed. Particularly, we analysed *CDH1* promoter cytosines to investigate if ERG overexpression directly affected CDH1 methylation on regulatory regions. Moreover, also long interspersed nucleotide element-1 (LINE-1) were analysed, aiming to understand if the global DNA methylation level was altered when ERG overexpression was induced. In details, the LINE-1 methylation was found to correlate with the global DNA methylation [151], therefore was used as bait for the evaluation of global DNA methylation changes in this study.
Unexpectedly, all cytosines of the *CDH1* promoter investigated were found at methylation 0% (data not shown) in both NI and I conditions, and no differences were reported not even in the LINE-1 levels that all peaked at about 50% of methylation, as reported in *Figure 44*.

These negative results shed light to the complexity of ERG-dependent regulation and might suggested that the DNA methylation could be a limited regulatory mechanism in this model. In fact, *CDH1* downregulation may be due to an ERG-indirect activity that instead might be mediated by epigenetic modifications or non-coding RNAs regulation. Concerning LINE-1, moreover, in all cases the methylation percentage was assessed at almost 50%, revealing that globally the DNA methylation did not show strong differences and therefore that might not be a pivotal ERG-dependent regulation, or at least not in the time window analysed.

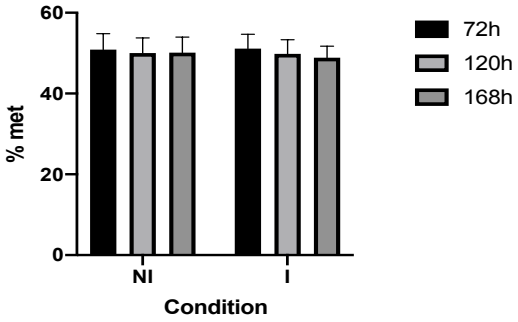


Figure 44| Methylation percentage of LINE-1 in both NI and I samples for 72-, 120- and 168h. No statistically significant differences were detected.

3.2 Focused analysis of DNMT3A expression

As previously observed, transcript and protein levels of DNMT3A from previous results seemed to be conflicting. A preliminarily explanation for this apparent paradox could be related to the various *DNMT3A* transcript variants and protein levels experimentally detected. We found that primers used for *DNMT3A* expression amplify a C-terminal region that is shared by all known transcript variants, while the primary antibody (see Appendix III) detects only full-length DNMT3A protein, recognising the N-terminal epitope.

In particular, *DNMT3A* encodes for a full-length variant (*DNMT3A1* or *DNMT3Aa*) and for a smaller one (*DNMT3A2* or *DNMT3Ab*), which transcription starts from an alternative downstream intronic promoter. The resulting DNMT3A2 protein lacks 233 aa of the N-terminal region in human (*Figure 45*). Further analyses, moreover, revealed that the two transcript variants shared the same DNA methyltransferase activities, while showed a different subcellular localization; DNMT3A1 is found concentrated in heterochromatin, while DNMT3A2 is associated to euchromatin. In addition to this, DNMT3A1 is expressed ubiquitously at low levels, while DNMT3A2 expression is predominant in embryonic stem and carcinoma cells, suggesting that the two variants have distinct targets and functions [152][153].

Therefore, DNMT3As and histone PTMs were evaluated in ERG-induced model. As showed in *Figure 46,* the DNMT3A2 protein was significantly upregulated at 120h of induction and a downregulating trend could be seen for DNMT3A1. Curiously, the upregulation of DNMT3A2 correlated with the global increase of the repressive histone mark H3K9me2 and the bivalent histone mark H3K27me3/ H3K4me3, while the transcriptional activating histone mark H4K12ac resulted significantly decreased.

In order to further investigate if histone PTM changes were affected by DNMT3A2, we used not induced RWPE-1 cells and further silenced the full-length DNMT3A1 variant, analysing thus the same markers previously seen from 72- to 168h. In *Figure 47*, silencing results are showed.

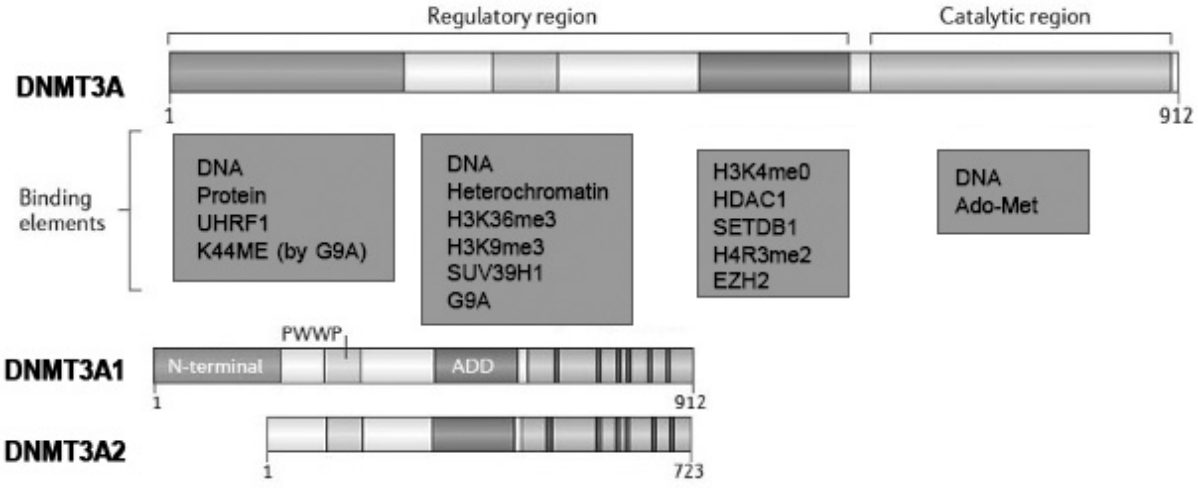
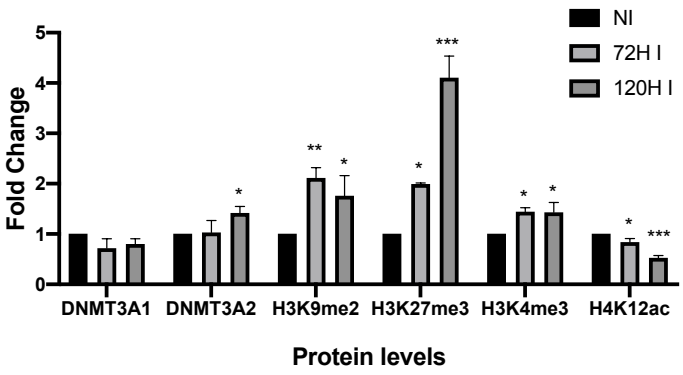


Figure 45| Schematic structure of DNMT3A transcript variants. On the top of DNMT3A domains with their relative recognized binding elements are listed. On the bottom, the two DNMT3A1 and DNMT3A2 are showed with their domains and the total amino acid number. Particularly, DNMT3A2 lacks the N-terminal domain compare to DNMT3A1 [153].

A



B

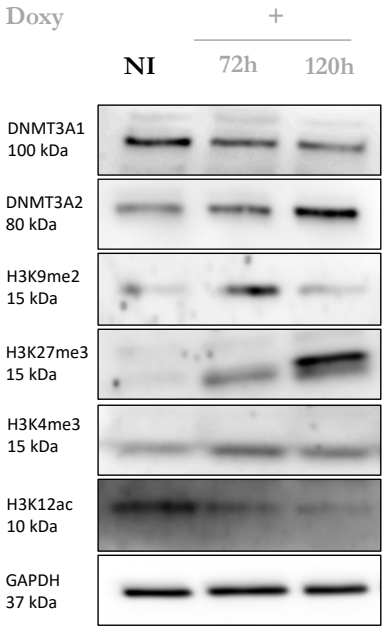


Figure 46| Protein levels of DNMT3A variants and histone marks. **A)** Densitometric analysis of WB normalized using GAPDH as reference. **B)** Protein levels evaluation with the WB assay. Statistical significance was evaluated with an Anova One-way test using the multiple comparison option and comparing I versus NI. *P value≤ 0.05, ** P value≤ 0.01, *** P value≤ 0.001.

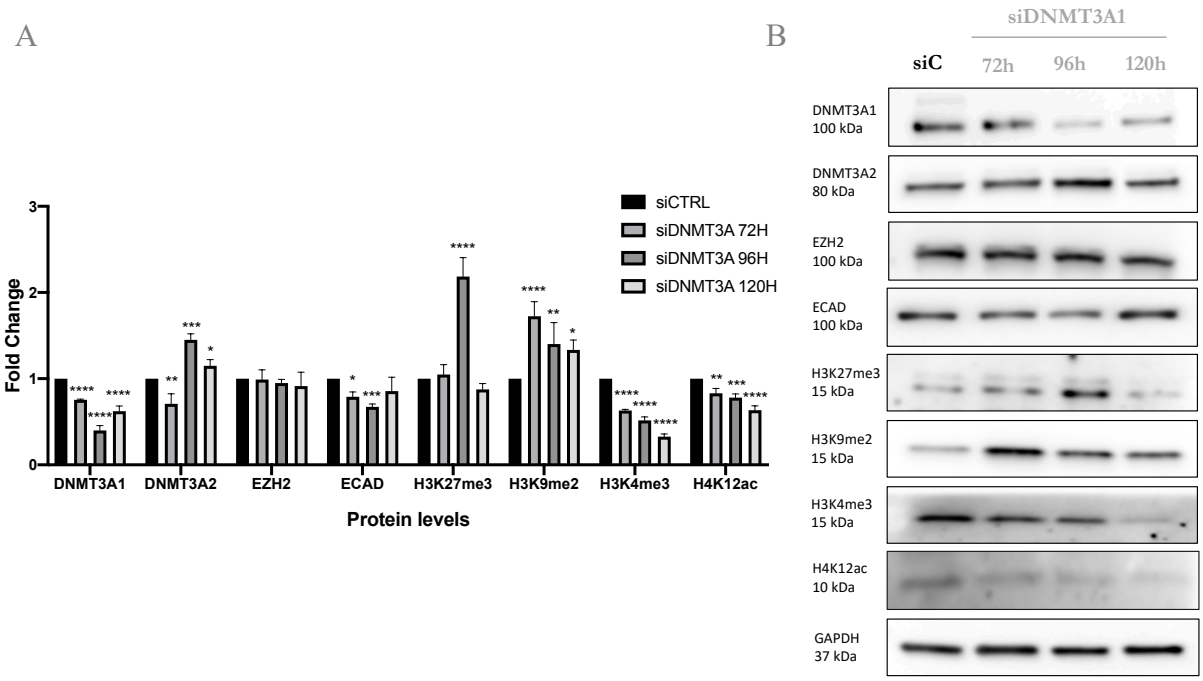


Figure 47| siDNMT3A1 results. **A)** Densitometric analysis of WB normalized using GAPDH as reference. **B)** Protein level evaluation with the WB assay. Statistical significance was evaluated with Anova One-way test using the multiple comparison option and comparing I versus NI. *P value≤ 0.05, ** P value≤ 0.01, *** P value≤ 0.001, **** P value≤ 0.0001.

The DNMT3A1 silencing effect peaked at 96h (almost 50% of reduction) with a concomitant significant increase of DNMT3A2. The reduction of the DNMT3A1 seemed to correspond to a compensatory mechanism that promoted DNMT3A2 upregulation, with a downstream effect on histone modifications. Particularly, H3K9me2 and H3K27me3 levels were found to be significantly increased at 72- and 96h respectively, suggesting that the combination of DNMT3A1 downregulation and DNMT3A2 upregulation might exert an effect on the global histone methylation pattern, in a repressive manner. In fact, both histone marks associated to active transcription (H3K4me3 and H4K12ac) were found significantly decreased.

Of course, the pattern of histone modifications between the ERG-induced experiments and the DNMT3A1-silenced ones in not induced cells showed differences, but what really stood out from these analyses was the crucial relevance of DNMT3A2 and the potential interdependence of histone PTMs and DNMT3A2 regulation.

ECAD and EZH2 proteins were also evaluated, since ECAD deregulation correlates with the promotion of epithelial-to-mesenchymal transition, and EZH2 is one of the most common histone-lysine writers. While EZH2 didn't show any significant change, ECAD level decreased in accordance with DNMT3A1 downregulation. This result could be interpreted as a direct effect of

DNMT3A1 downregulation, due to its ability of epigenetically silence *CDH1* promoter [90], or as a DNMT3A2 upregulation effect. We speculated that the second hypothesis better answers to this uncertainty, since also in the pyrosequencing assay of induced RWPE-1 cells, *CDH1* cytosines were found to be 0% methylated. In addition, in gastric cancer DNMT3A2 is able to promote a Snail-dependent EMT, repressing *CDH1* expression through the epigenetic silencing mediated by H3K27me3 and H3K9me2 on its promoter [154].

Furthermore, we also investigated the transcript levels in our experimental induced conditions, as shown in *Figure 48*, revealing that instead both variants resulted significantly upregulated at 120h, in contrast with previously shown protein levels.

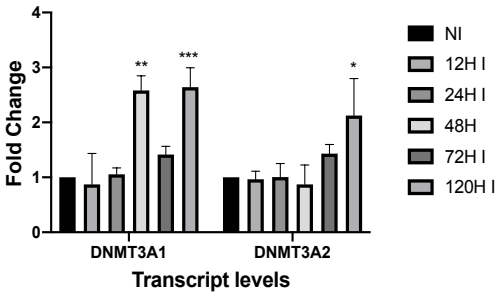


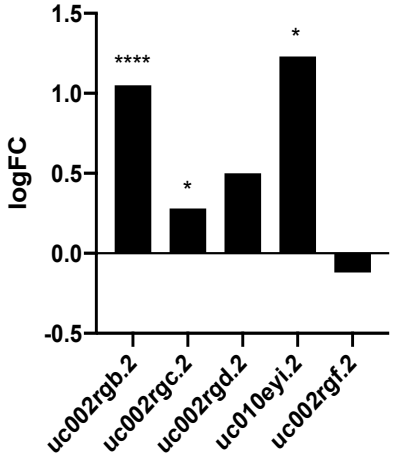
Figure 48| Transcript levels of the two DNMT3A variants evaluated for each experimental condition. The *GAPDH* gene was used as reference. Statistical significance was evaluated with an Anova One-way test using the multiple comparison option and comparing each I (Induced) with NI (Not Induced). *P value≤ 0.05, ** P value≤ 0.01, *** P value≤ 0.001.

These results might suggest that the different regulation on DNMT3As is exerted in a post-transductional fashion, and therefore might not be ERG-directly dependent.

Of course, it still remains unclear if ERG overexpression acts somehow in this potential transcript variant modulation, therefore further and more focused analysis would be required in order to characterize the exact molecular mechanism. In addition to this, a preliminary evaluation in PCa samples was analysed using the PRAD data. The transcript variants' database was downloaded, then all samples were filtered for their subtype and only normal (52 samples) and *ERG*$^{+}$ tumours (131 patients) were selected. *ERG*$^{+}$ vs N differentially expressed (DE) transcript variant analysis was performed, and data reported in *Figure 49*.

The analysis revealed that variant 2 (*DNMT3A2*) was the most significantly different in *ERG*$^{+}$ tumours, but also highlighted a coherent trend compare to qPCR analyses, where the two full-length variants (variant 1 or rather *DNMT3A1* and variant 3) resulted upregulated. Furthermore, variant 4 (a variant that lacks the C-terminal domains and that is associated to the full-length variants) showed no significance.

To conclude, this preliminary data started to shed light to the complexity of DNA methyltransferases 3A role in tumour deregulation, but of course further and more focused analyses needed to be done.



USCS Name	Gene name
uc002rgb.2	DNMT3A variant 2 (NM_153759)
uc002rgc.2	DNMT3A variant 3 (NM_022552) FL
uc002rgd.2	DNMT3A variant 1 (NM_175629) FL
uc010eyi.2	DNMT3A variant 1 (NM_175629) FL
uc002rgf.2	DNMT3A variant 4 (NM_175630)

Figure 49| DE transcript variants of DNMT3A obtained by PRAD analysis, *ERG*+ tumours versus Normal samples. The reference name followed the GRCh37 and USCS nomenclature. FL indicates transcript variants annotated as Full Length. *FDR≤ 0.05, ****FDR ≤ 0.0001.

3.3 Differentially methylated characterization of ERG induced genes

Due to previous hints concerning changes in DNMT expression and protein levels, we further analysed hypothetical variations of single cytosines with the Illumina Infinium HumanMethylation EPIC BeadChips and three independent replicates for each time point (12-, 24- and 120h) were evaluated, in collaboration with Dr Silvia Polidoro at Fondazione IIGM (Candiolo). In details, raw data were processed with the RnBeads 2.0 pipeline [128], comparing induced versus not induced samples at each selected time point.

The threshold set to assess significance was the βvalue>|0.2| and the statistical significance between induced and control samples was determined with the FDR< 0.05. At 12- and 24h, few cytosines were found significant for P value and not for FDR, while at 120h some cytosines passed the threshold and were significant for the FDR, therefore were further investigated. Particularly, as shown in *Figure 50*, the highest percentage of significant cytosines belonged to open sea areas (87%), while the left over 13% accounted for the N- and S-shore cytosines, respectively 3 and 10%.

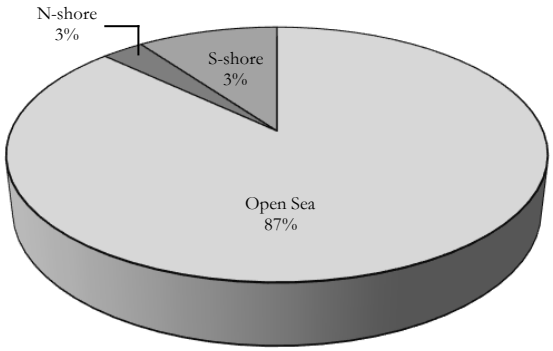


Figure 50| Pie chart of DM cytosines at 120h categorise in three groups: Open sea, N- shore and S-shore.

Unexpectedly, although premises seemed to suggest a relevant regulatory role of DNMTs in our cell model, we couldn't observe any strong ERG-dependent DNA methylation effect, especially for the "early" time points, where almost zero cytosines resulted significant.

Global DNA methylation changes and also the cytosines belonging to specific regions (genes, promoter, CpG Island) were analysed for each time point, as reported in *Figure 51*.

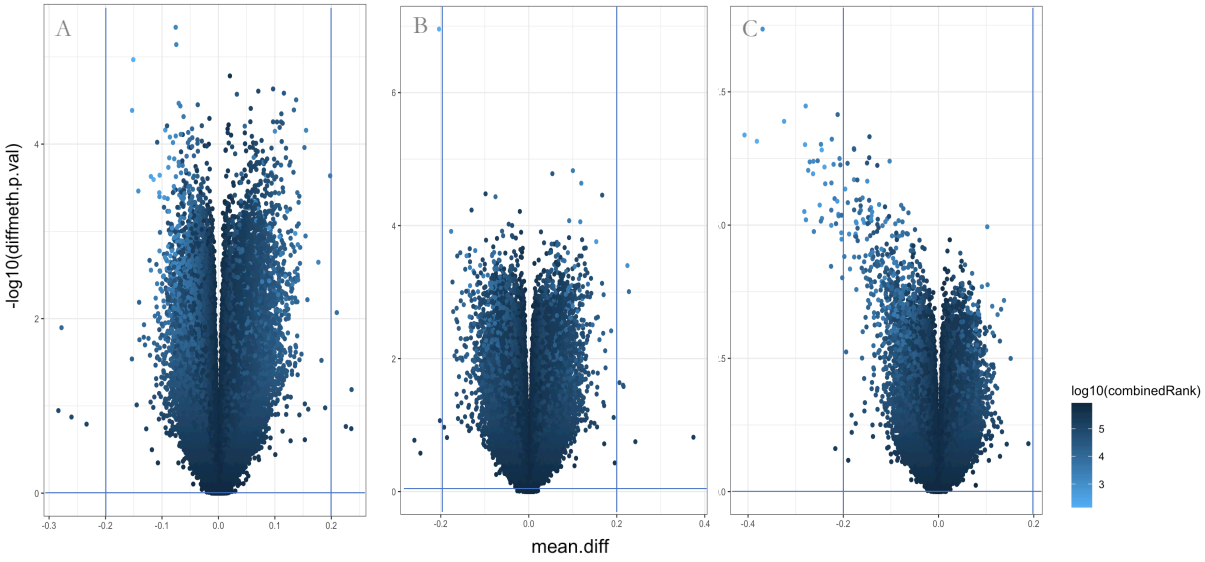


Figure 51| Volcano plot of differentially methylated cytosines quantified by difference metric. Colour scale according to combined ranking is reported in the bottom right legend. Blue lines indicate the minimum threshold set. **A)** 12h, **B)** 24h and **C)** 120h.

In general, CpG Islands (CGIs) are defined as region enriched with more than 50% of CpG nucleotide clusters that span almost 200 bases. The majority of CGIs in human genome is located in the promoter regions and generally unmethylated, while their methylation is associated to gene silencing during development or in pathologic conditions. Nevertheless, methylation also occur in

region with lower CpG content that are 0-2 kb close to CGIs and named shores, particularly divided in north and south shores, respectively upstream or downstream of the CGIs. Beyond the gene body methylation that correlates with transcriptional activation, 70% of differentially methylated CpG island shores are associated with gene reprogramming [155][156]. On the other hand, "open sea" indicates all other genomic regions that cannot be classified as shores, shelves (2 - 4 kb from CGIs) or islands.

Even if some cytosines resulted statistically significant, their role could not be better investigated from these results and might suggest that DNA methylation was poorly involved in ERG-dependent transcriptional programme.

Moreover, a deeper analysis of cytosine variations was also performed, selecting for CGs that belonged to common regulatory regions, or rather: promoter, genes and islands, separately (*Figure 52*). Focused analyses revealed that almost all CGs evaluated were not DM in our experimental conditions, since did not pass the βvalue>|0.2| threshold. In details, at 12h some CGs were significant for the combined P value although they were not significant for FDR and were not further analysed.

Despite preliminary premises, the DNA methylation analysis didn't show any statistically significant DM gene, revealing that the DNA methylation was not strongly involved in our cell model regulation, while other epigenetic processes might be key players.

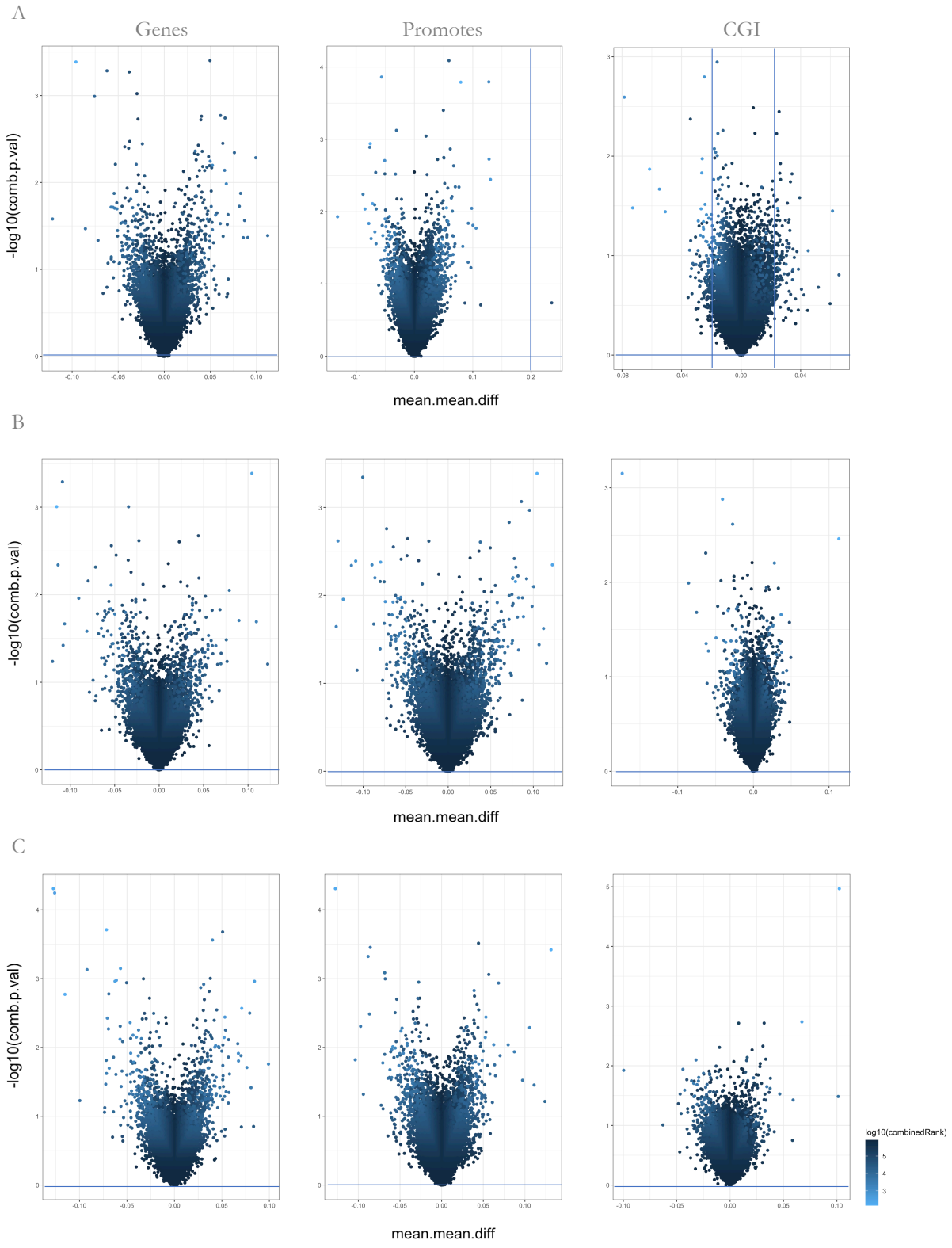


Figure 52| Volcano plot for differential methylation quantified with the difference metric. The significance metric used is the combined P value. Different regions (genes, promoters and CGI) are reported for each experimental condition. Blue lines define the threshold set. **A)** 12h, **B)** 24h, **C)**120h. The colour code is reported in the bottom right legend.

4. ERG transcriptional programme is sustained by a huge chromatin remodelling

Previous analyses highlighted a strong transcriptional effect when ERG expression was induced with a concomitant weak methylation involvement. Therefore, we speculated that chromatin accessibility and modifications might instead be pivotal for sustaining ERG transcriptional programme. To preliminarily investigate this aspect, the global level of different histone modifications was analysed (*Figure 53*) from 12- to 120h.

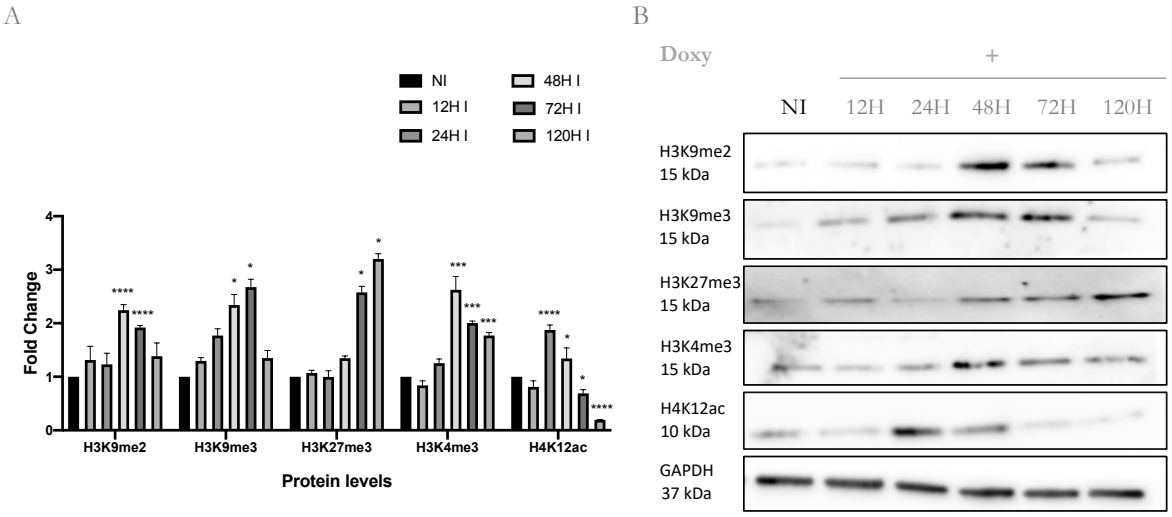


Figure 53 | **A)** Densitometric analysis of histones modifications obtained from WB normalization using GAPDH as reference. **B)** Protein level evaluation of WB assay. Statistical significance was evaluated with Anova One-way test using the multiple comparison option and comparing I versus NI. *P value≤ 0.05, ** P value≤ 0.01, *** P value≤ 0.001, **** P value≤ 0.0001.

Interestingly we observed that histone PTMs showed different trends in a time- and induction-dependent fashion. In details, at 48- and 72h both H3K9me2 and H3K9me3 resulted significantly increased. H3K9 methylation is a heterochromatin mark and the me2/me3 methylation determined a more condensed and transcriptionally inactive chromatin status [157] that seemed to be replaced at 120h with an increase of H3K27me3 levels, a repressive mark too. Moreover, H3K4me3 levels resulted inverted compare to H3K27me3 ones, considering the 48-, 72- and 120h time points. This trend seemed to perfectly line up with the bivalent mark' role of H3K4me3/H3K27me3 that generally locate on the same promoter domains, exerting opposite activities especially for development genes [158]. H4K12ac has a promoter localization and active transcriptional role, as seen for H3K4me3 [159]. Overall, the preliminary western blot analyses revealed that during the early phase, ERG overexpression promoted a global transcriptional activity with a more open and active chromatin status that was then reversed at 120h.

Of course, further and focused analyses are necessary in order to disentangle the relevance of these global modifications. Nevertheless, we observed consistent histone PTM changes linked to ERG overexpression that found coherence with recent papers, which showed the pivotal role of ERG in driving the prostate cell fate reprogramming through chromatin re-organization [61][145].

In addition to this, we further analysed if PTM changes were supported by ERG-transcriptional regulation of writer and eraser enzymes, as well as cofactors involved in the epigenetic modifications. Therefore, using the EpiFactors database (https://epifactors.autosome.ru/), we matched significant DE genes with cofactors or enzymes relevant in epigenetics regulation, and highlighted the potential role of resulting proteins (*Figure 54*).

GeneID	FC h12	FC h24	FC h120	Function	Complex name	Target	Specific_target	Product
ARRB1	0,39	1,24	0,99	Histone modification	#	histone	#	#
CBX4	0,00	-1,50	-0,85	Histone modification read	PRC1	histone	H3K9me3	#
IKZF3	0,06	0,71	1,96	TF	#	DNA	DNA motif	#
JDP2	0,25	-0,94	-1,06	Chromatin remodeling, Histone modification erase cofactor	#	DNA	#	#
KDM7A	0,20	-1,24	-0,73	Histone modification erase	#	histone	H3K9me2, H3K27me2, H4K20me1	H3K9, H3K27, H4K20
PADI2	0,43	2,46	3,06	Histone modification	#	histone	H2AR, H3R, H4R	H2ARci, H3Rci, H4Rci
PADI3	-0,03	3,18	1,42	Histone modification	#	histone	H2AR, H3R, H4R	H2ARci, H3Rci, H4Rci
PRDM1	1,72	1,41	1,99	Histone modification write cofactor	#	histone	H3K9	H3K9me

Figure 54|DE significant genes matched with the EpiFactors database to infer the epigenetic role of factors and cofactors found in our experimental data. Fold change of expression, function, target and resulting products are indicated.

Of note, KDM7A, an eraser of H3K9 and H3K27 methylation, resulted slightly downregulated in a time- and induction-dependent manner and therefore could correlate with an increased methylation of the aforementioned PTMs. The same coherence was also seen for the PRDM1 cofactor that promotes H3K9 methylation and resulted transcriptionally activated by ERG overexpression. Particularly, PRDM1 has no intrinsic histone methyltransferase activity, even though it interacts with the H3 lysine methyltransferase G9A [160], the histone lysine deacetylases HDAC1 and HDAC2 [161] and H3 lysine demethylase LSD1 [162], not found as significant DE genes in our conditions.

Nonetheless, the transcriptional regulation of PTM enzymes seemed to be a minor effect and we speculated that the driving correlation between ERG overexpression and chromatin status was due to changes in chromatin accessibility, coherently with the results of Li *et al.* [145], that were previously discussed.

5. ERG-dependent secretome alters the phenotype and affects fibroblasts survival

The following purposes of this project were aimed to characterize the possible crosstalk between the transcriptional deregulation due to ERG overexpression and a potential effect of ERG-target on the tumour microenvironment. Particularly, we dealt with these aspects with diverse and complementary approaches.

Firstly, a prediction of molecules that might play a role as secreted factors was inferred from the transcriptome analysis of differentially expressed genes, using experimental and PRAD data.

Secondly, also *in vitro* data were collected, using RWPE-1 cells (induced or not) to produce conditioned media that were further analysed with LC-MS, in order to isolate differentially secreted molecules between induced versus not induced conditioned media (CM).

Thirdly, we tried to explore the phenotypic effects of the RWPE-1 (induced or not) -derived CM on a model of TME cells, the normal and cancer -associated fibroblasts.

5.1 Prediction of potential secreted molecules, inferred by RNA-seq and PRAD analyses

The enriched network analysis helped to define that the overexpression of ERG could be crucial for different processes connected to the crosstalk between the tumour and the relating microenvironment (*i.e.* matrisome deregulation, chemotaxis, exocytosis...). Therefore, ERG transcriptional programme might determine a differential expression of genes that play a role in secretion context. Our first analysis aimed to predict which of the DE genes could be potentially secreted as protein and could affect TME, both in our cell model and PRAD data. Since the most relevant effect of induced cells was assessed at 120h, all analyses concerning the secretome focused on this specific time point.

Firstly, the 507 significant DE genes were matched to the protein names of the (predicted) secreted protein database from the Human Protein Atlas (https://www.proteinatlas.org/), and the same procedure was also used for the PRAD subset, in order to identify common terms between the experimental conditions and the tumour data (even if genes enriching the specific terms could also be different). In the first case, 105 DE genes were found to be (predicted) secreted at 120h, while 996 out of 12654 DE genes resulted to be (predicted) secreted in the PRAD subset. The resulting gene lists were further analysed in Metascape ("Express analysis"), to depict the biological relevance of (predicted) secreted proteins in both cases. Particularly, some terms resulted in

common as: (i) extracellular matrix organization, (ii) regulation of IGF transport and uptake, (iii) cell chemotaxis, (iv) humoral immune response, (v) response to wounding, (vi) regulation of cell adhesion; some instead showed similarities as: (i) matrisome components (associated, core, ECM regulators and glycoproteins, secreted factors) and (ii) the regulation of blood vessel (morphogenesis and development), illustrated in *Figure 55*.

Strikingly, the common biological terms shared the same gene lists; in fact, 77 DE genes encoding for (predicted) secreted molecules were found in common between the experimental and the PRAD sets (see Appendix VI for the entire list).

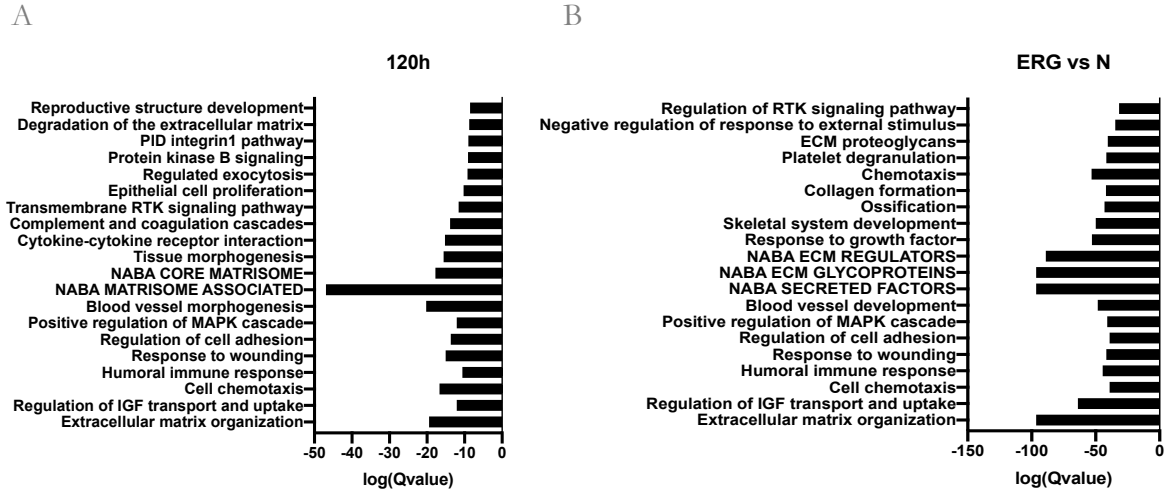


Figure 55| Bar plots of the log (Q values) obtained from the Metascape enrichment analysis of DE genes matched with the (predicted) secreted proteins of the Human Protein Atlas database, in the 120h (**A**) and *ERG*+ vs N PRAD, (**B**) sets.

5.2 Proteomic evaluation of CM media with LC-MS

The production of CM was obtained cultivating RWPE-1 cells for 7 days, corresponding to 120h of induction. Particularly, the last induction procedure was performed with a direct doxycycline supplement, in order to concentrate the CM protein content. Collected media were further processed and analysed with LC-MS by our collaborator Dr Marco Gaspari at the University of Catanzaro. 196 molecules resulted differentially secreted when ERG expression was induced and were evaluated for their biological meaning. To this purpose, the list of differentially secreted proteins was enriched for the GO terms: Biological Process (BP) and Molecular Function (MF), using GOrilla (http://cbl-gorilla.cs.technion.ac.il/) and the entire list of peptides recognised by the LC-MS analysis were used as background list (196 molecules of the target list vs 1681

molecules of the background one). The enriched terms reported in *Figure 56* were filtered for FDR<0.05. The most relevant BP terms were: (i) angiogenesis, (ii) cell adhesion, (iii) biological adhesion; while concerning MF, all significant terms showed similar NESs (Molecular transducer, transmembrane signalling receptor and signalling receptor activities).

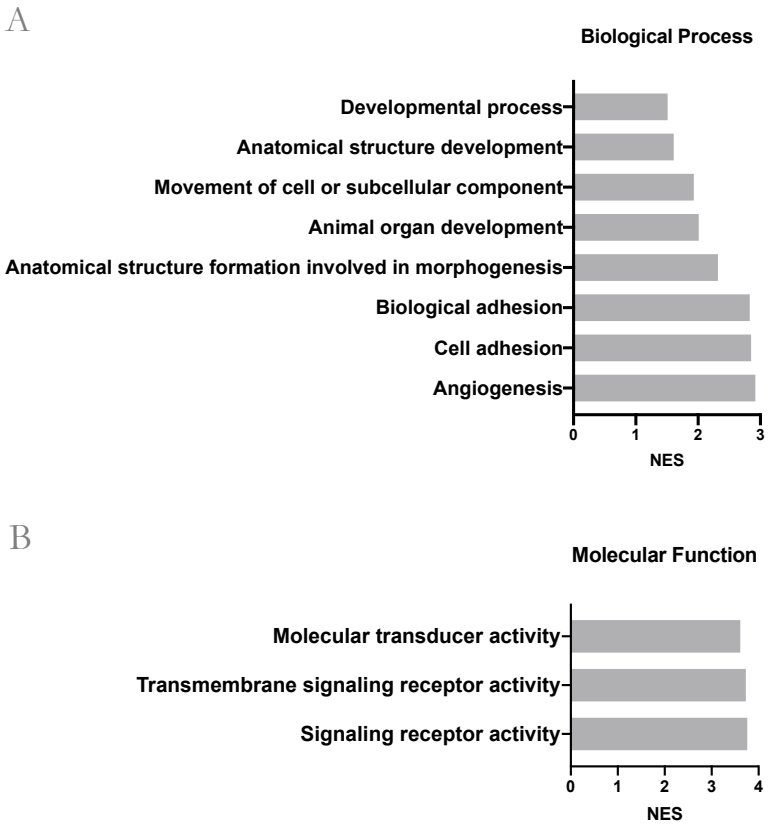


Figure 56| Normalised Enrichment Score of Biological Process (**A**) and Molecular Function (**B**) of the enrichment analysis performed with GOrilla. All reported terms were significant for FDR<0.05.

Furthermore, we aimed to investigate if some secretome-related processes could be deregulated by ERG transcriptional programme, therefore the differentially secreted list was also evaluated with the "Express Analysis" of Metascape *(Figure 57)* and the obtained enriched terms were further compared with 120h and PRAD significant processes previously emerged (*Figure 58A*), searching for overlapping terms. Particularly, we obtained that (i) the response to wounding, (ii) the regulation of cell adhesion, (iii) chemotaxis and (iv) extracellular matrix organization showed correspondence between all sets analysed.

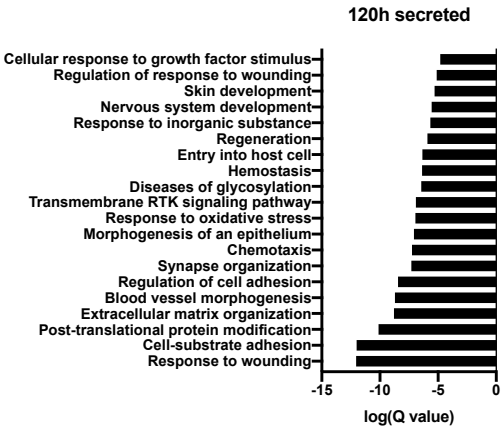
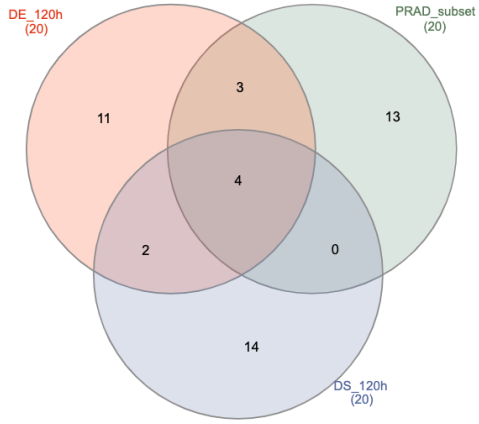


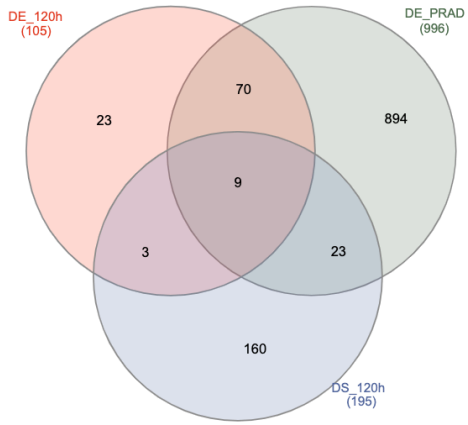
Figure 57| Bar plots of the log (Q values) derived from the Metascape enrichment analysis of differentially secreted proteins obtained by LC-MS analysis.

A



[DE_120h] and [PRAD_subset]:
Regulation of Insulin-like Growth Factor (IGF) transport and uptake
Positive regulation of MAPK cascade
Humoral immune response
[120h DE] and [120h DS protein]:
Blood vessel morphogenesis
Transmembrane receptor protein tyrosine kinase signaling pathway
[DE PRAD] and [120h DS protein]:
[120h DE] and [DE PRAD] and [120h DS protein]:
Response to wounding
Regulation of cell adhesion
Chemotaxis
Extracellular matrix organization

B



[120h DE] and [DE PRAD] and [120h DS protein]:
C1R
C1S
COL7A1
FST
GAL
KLK8
PLAU
TGFBI
VEGFA

Figure 58| **A)** Venn diagram of terms enriched in the following sets: DEG at 120h in Induced vs Not Induced condition, DEG of PRAD in *ERG*+ patients vs Normal samples, DS proteins at 120h in Induced vs Not Induced conditions. Terms in common are reported in table. DE (Differentially Expressed), DS (Differentially Secreted). **B)** Venn diagram of corresponding genes/proteins.

The same comparison was also performed using the lists of significant genes/proteins (*Figure 58B*) for each condition and 9 molecules (C1R, C1S, COL7A1, FST, GAL, KLK8, PLAU, TGFBI, VEGFA) were found in common, highlighting that ERG transcriptional programme impacted on the differential expression of some molecules that were also differentially secreted in experimental data as in tumours. In details, C1R, C1S and PLAU were transcriptionally upregulated and also highly secreted, while COL7A1, KLK8, TGFBI and VEGFA resulted downregulated and less secreted. Concerning the left-over ones, we could not assess a correspondence between transcription levels and magnitude of secretion (data not shown), therefore were not further considered.

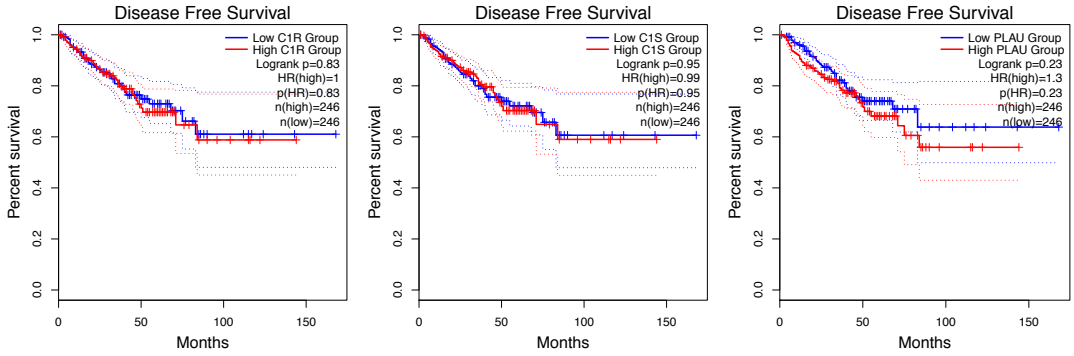
A preliminarily role of these factors in terms of survival was evaluated in GEPIA2 (http://gepia2.cancer-pku.cn/#survival). The survival analysis was estimated in the entire PRAD dataset (550 cases: 52 normal and 498 tumour samples), using single or cumulative signatures and the disease-free survival rate (with the Relapse-Free Survival, RFS) investigated. Parameters used were:

- Methods: RFS,
- Group cutoff: Median,
- 95% Confidence Interval,
- Dataset: PRAD.

Single gene analyses were reported in *Figure 59*, while cumulative gene signature in *Figure 60*.

Single survival analyses revealed that highly secreted factors did not showed any effect on survival, while the lower secreted ones highlighted significant Hazard Risk and LogRank P values, except for VEGFA and TGFBI. Interestingly, RFS of downregulated factors seemed in contrast with what we could expected, in fact a lower expression of these factors correlated with a better prognosis. Cumulative analyses (*Figure 60*) fostered these trends and we found that no significance was assessed for the upregulated signature, while the downregulated one showed strong significance (LongRank P value= 4.3e-05), excluding VEGFA.

A



B

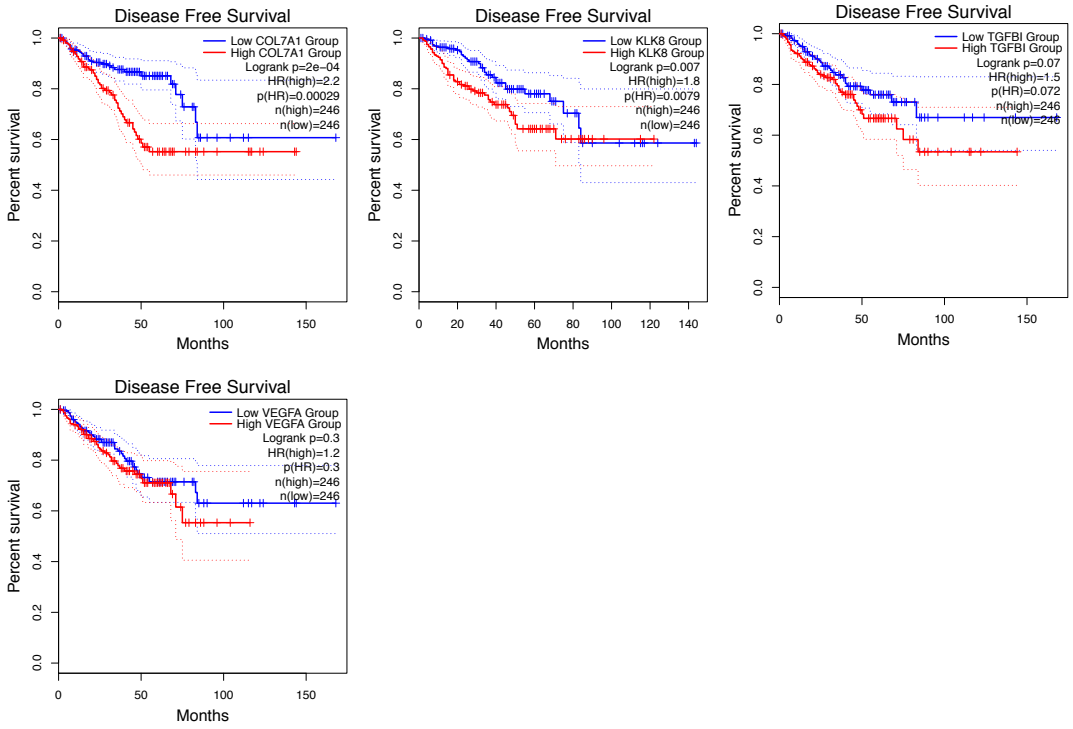


Figure 59 | RFS analyses obtained in GEPIA2, "Survival Analysis" of PRAD. **A)** Genes found as higher secreted factors. **B)** Genes found as lower secreted factors. Significance was assessed for Logrank P value<0.05.

A B C

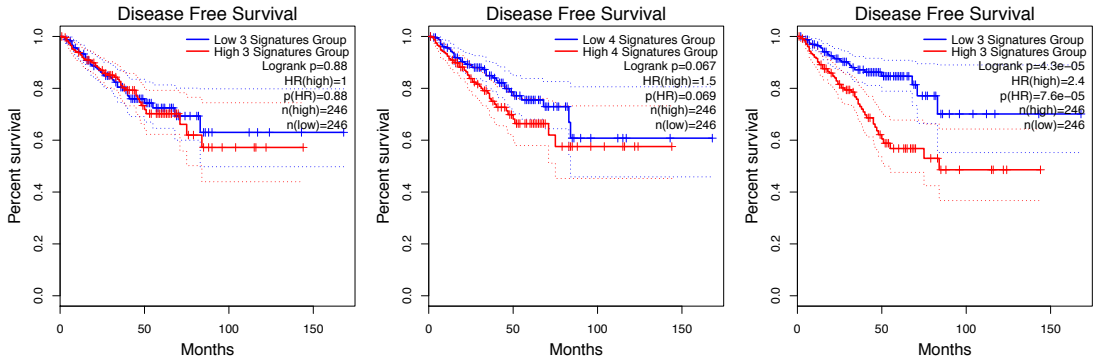


Figure 60| **A)** Signature of C1S, C1R, PLAU did not show significance. **B)** Signature of COL7A1, KLK8, TGFBI, VEGFA. **C)** Signature of COL7A1, KLK8, TGFBI. Significance was assessed for Logrank P value<0.05.

5.3 Characterization of differentially secreted protein secretion mechanisms

The canonical protein secretion is defined by a process involving cargo proteins that move from the endoplasmic reticulum (ER) to the Golgi complex and finally to the plasma membrane. This process implicates the recognition of a "leader sequence", or rather a signal peptide at the N-terminus, that allowed the vesicular fusion. Last discoveries, however, highlighted the existence of other secretion pathways as: the extracellular secretion of leaderless proteins (vesicular and non-vesicular) and a Golgi-bypassing route that both defined the Unconventional Protein Secretion (UPS) [120][121].

Aiming to characterize the complexity of the comprehensive secreted molecules in our experimental conditions, we used different tools to retrospectively discriminate secretion mechanisms. Proteins secreted through conventional ways are characterized by a leader sequence, whose location can be predicted by the analysis of the FASTA sequence using SignalP-5.0 (http://www.cbs.dtu.dk/services/SignalP-5.0/); particularly this tool is able to predict the location of the cleavage sites of proteins and is based on a neural network architecture. All proteins showing a cleavage site were defined "secreted through the conventional pathway". FASTA files of not matching proteins were instead submitted to SecretomeP 2.0 (http://www.cbs.dtu.dk/services/SecretomeP/) that offered a combination of parameters to infer a non-classically secreted protein, when the sequence showed a NN-score/SecP>0.6 (in mammals, as in our case); proteins that fitted the NN-score were therefore labelled as "secreted through non canonical pathways". Furthermore, proteins negative in both analyses were further investigated to be potentially secreted through vesiculation in exosomes, using the ExoCarta database (http://www.exocarta.org/). Results of the analysis were summarised in the pie chart in *Figure 61*.

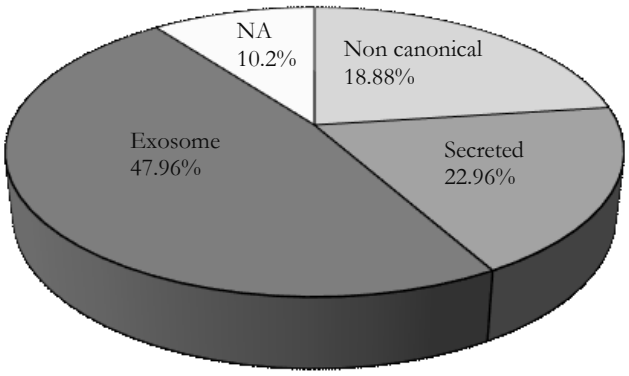


Figure 61 | The pie chart illustrates the magnitude of the differentially secreted proteins accordingly to possible mechanisms of trafficking, inferred through a bioinformatic approach. NA represented all Not Assessed proteins.

We obtained that the majority of molecules found in CM could be potentially secreted through exosome vesiculation (47.96%), while non canonical and canonical pathways accounted for 18.88% and 22.96%, respectively. Unfortunately, 10.2% of significant secreted molecules could not be classified with this approach (were indicated as not assessed, NA).

5.4 Phenotype assessment of fibroblasts exposed to CM

To underpin the crosstalk between epithelial and TME cells mediated by secreted proteins, we also looked at potential phenotype changes affecting NAFs and CAFs exposed to K-SFM conditioned from RWPE-1 cells (induced or not). Particularly, we analysed the transcriptional and protein profile of NAFs and CAFs, following well- characterized molecular markers [163] (*Figure 62*).

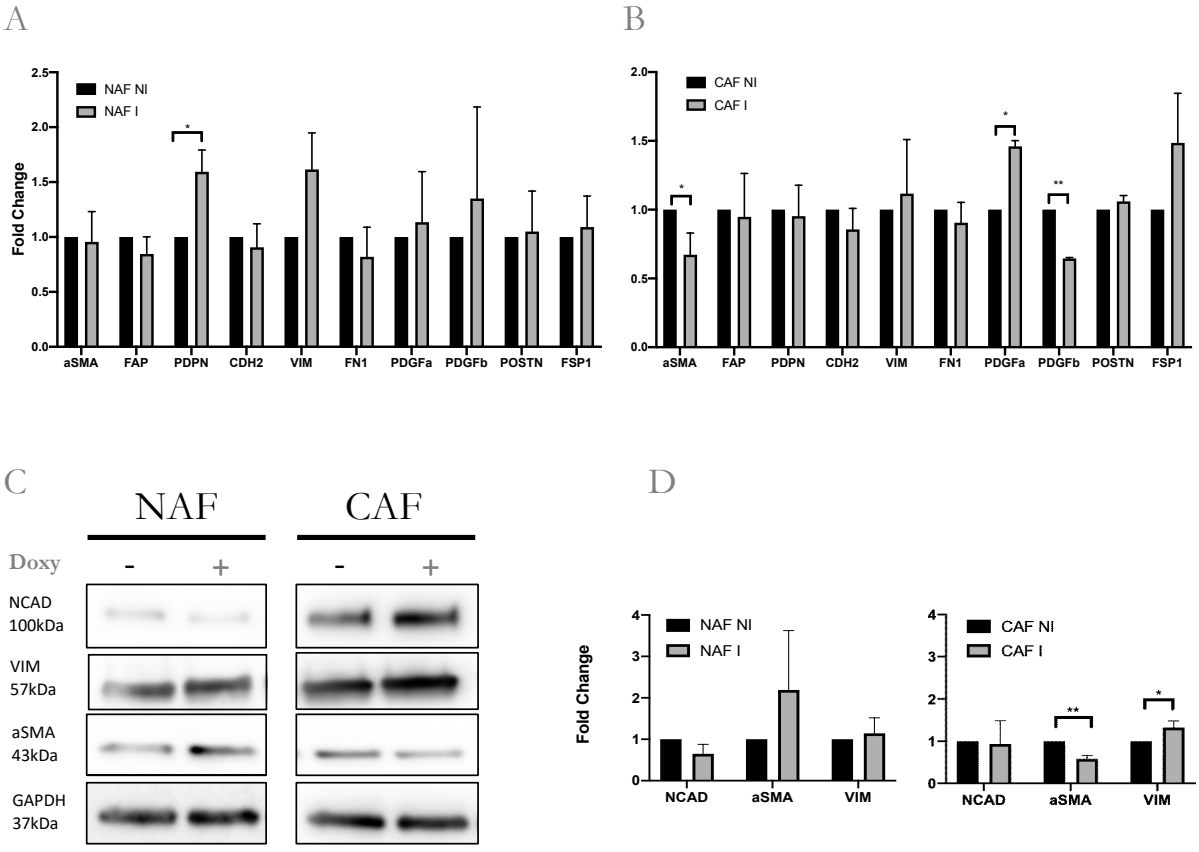
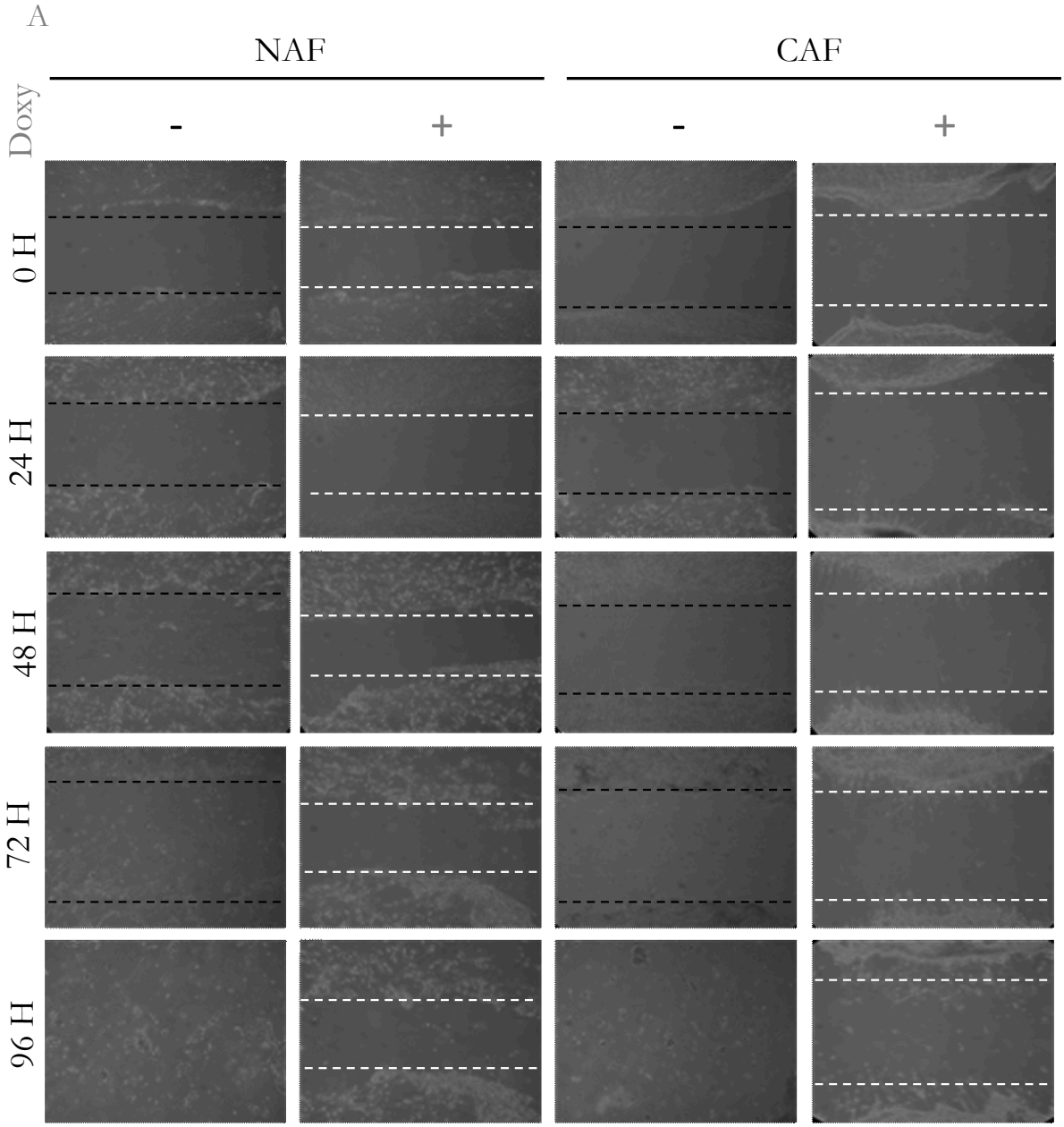


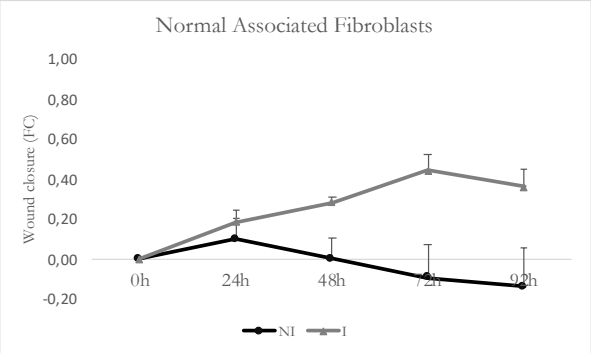
Figure 62| qPCR analysis of NAFs (**A**) and CAFs (**B**) exposed to induced or not CM. NAFs showed not significant changes except for PDPN; CAFs instead showed significance for αSMA, PDGFα and PDGFβ. **C)** Protein level evaluation of Western blot assay. **D)** Densitometric analysis. Statistical significance was evaluated with t-Student test using GAPDH as reference. *P value≤ 0.05, ** P value≤ 0.01.

Concerning NAF profile, qPCR analyses revealed that the induced-derived CM promoted an increase of "activated fibroblasts" typical markers as VIM and PDPN (that resulted statistically significant); on the other hand, αSMA and PDGFβ resulted reduced in CAFs exposed to induced CM, while upregulated were instead VIM and PDGFα levels. Due to the limited markers analysed and significance, it was difficult to define a clear-cut meaning of induced CM effects in terms of molecular changes, therefore deeper analyses would be necessary to clarify these aspects.

In addition to this, we further evaluated if the exposure to CM might affect fibroblast invasive and migratory properties, therefore wound healing assays were performed in triplicate (*Figure 63*).



B



C

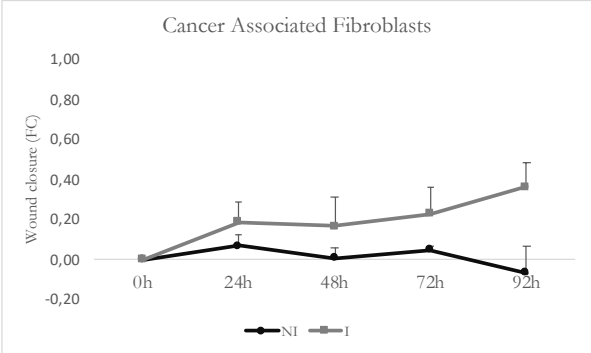


Figure 63 | **A)** Wound healing assay of NAFs and CAFs exposed to NI and I conditioned media. Fold change of wound closure for NAFs (**B**) and CAFs (**C**).

Particularly, these experiments revealed that fibroblasts did not acquire more invasive and migratory properties, since the different trend between NI and I observed was primarily due to an increased mortality in NI samples. At 96h from the wound healing, NI exposed NAFs and CAFs resulted almost detached. Therefore, the effect seen was mainly related to an increased mortality of fibroblasts exposed to not induced conditioned media, that peaked at 96h.

The analysis suggested that the secretome produced by RWPE-1 cells induced for ERG overexpression was enriched by molecules that might positively affect the fibroblast fate, in both NAFs and CAFs. Further analyses would be required in order to exclude a doxycycline-dependent effect. Of course, more extensive analyses are required in order to evaluate the magnitude and the effectiveness of this consequence, and also to intercept the crucial players of the increased survival.

Discussion

1. ERG transcriptional programme early affected IFN signalling and estrogenic response

The correlation between ERG and prostate cancer progression has emerged over the last decade, when Tomlins *et al.* [51], highlighted that the majority of prostate cancers harbour a chromosomal translocation promoting its aberrant expression. Although frequent, the resulting fusion gene showed poor clinicopathologic correlations and moreover inability in neoplastic transformation initiation [164]. This cryptic aspect magnetised our attention, therefore we tried to better elucidate if ERG overexpression could be considered a passenger or a driver alteration for prostate carcinogenesis, particularly ruling the tumour microenvironment.

Using then a benign prostatic epithelial cell model, we aimed to identify possible direct/indirect processes that not only could preliminarily do the groundwork for tumour progression, but also might have a relevant impact on the tumour-tumour microenvironment crosstalk.

In this scenario, different time points of ERG overexpression were followed, in order to describe how ERG-related effects could change in a time-dependent fashion. Of note, the "late" time point selected (120h of ERG induction), was sufficient to resemble phenotypic features already showed in previous studies. Particularly, we confirmed that ERG overexpression did not correlate with cell proliferation [50], while significantly promoted motility, invasion and a more mesenchymal-like phenotype acquisition [165].

Furthermore, to deepen early and late effects of ERG overexpression, the transcriptome profile was investigated though the RNA-Seq analysis. Explorative and functional analyses helped to reveal that some biological processes were already impaired at 12h of ERG overexpression, acquiring complexity in a time-dependent manner. The type I and II IFN responses, the complement and the STAT3 signalling resulted upregulated from 12h, while the estrogenic response was early downregulated. The EMT promotion instead, emerged as a middle-late effect (from 24h) and was mainly prompted at 120h, also confirmed by the wound healing assay.

In addition to this, the global network analysis highlighted a crucial role of STAT1 in ERG transcriptional programme, since it was the highest degree node into the network. STAT1 has been for a long time considered a tumour suppressor gene [166], while recent studies had revealed that an altered or prolonged STAT1 activation is associated to tumour development [167][168]; of note, different evidences also reported that the bivalent role might be correlated to different STAT1 isoforms [169]. The network analysis helped thus to characterize a hypothetical molecular mechanism that correlated ERG overexpression to the pivotal role of STAT1. In particular, ERG indirect transcriptional downregulation of the estrogenic response was associated to the strong activation of the type I and II IFN signalling, and targets of both signalling were found deregulated when ERG was induced. Deepening the biological role of crucial nodes, we defined a hypothetical molecular mechanism that might describe ERG-dependent changes.

1.1 A potential molecular mechanism prompted by ERG overexpression in PCa

The network analysis revealed interesting correlation between ERG and various nodes involved in crucial biological processes and that were recapitulated by clusters previously defined, even though the exact molecular nature of these connections needed still to be fully elucidated (see *Figure 32*, Appendix IV). Particularly, we tried to infer a molecular meaning of relevant edges, in order to underpin ERG direct/indirect involvement in their deregulation. This analysis resulted in a potential molecular mechanism that might be totally associated to ERG overexpression.

1.1.1 SSPB2 downregulation might affect LMOs' turnover

Global network analysis showed that ERG primary edges resulted with Lim-only proteins (LMO2, LMO4), physiologically involved in hematopoietic stem cell maintenance and whose deregulation instead correlated with migration and invasion in PCa [170]. Although the exact PPIs are still poorly characterized, ERG-dependent LMOs' role showed interesting cues. In details, DEG analysis showed that LMO2 and LMO4 resulted respectively up- and downregulated up to 12h, suggesting a direct ERG-dependent effect whose biological relevance was further addressed.

In T-ALL leukemia ERG regulates LMO2, a transcriptional co-factor that does not directly bind to the DNA, thus is complexed with other transcription factors as LDB1, SCL, GATA1 or GATA2. Particularly, LMO2 seemed to foster tumour progression and its altered expression in

immature hematopoietic cells, defined a first hit for the lymphoid precursor accumulation [141][171]. Coherently, LMO2 is overexpressed in PCa and it correlates with a higher motility, invasiveness and tumour aggressiveness. Consistent with this, Ma *et al.* [170] showed an indirect correlation between LMO2 upregulation and E-cadherin downregulation in PCa, suggesting a crucial role in cell-cell adhesion deregulation and invasion. The hematopoietic stem cell maintenance depends on the LMO2 multi-protein complex stoichiometry, a tightly controlled process ruled by the single stranded DNA-binding protein (SSBP2). Impairment of this process in fact negatively affects the developmental process [142]. In details, SSBP2 controls the turnover of LMOs and thus inhibits the E3 ubiquitin-protein ligase RLIM, preventing the proteasomal degradation of LDB1 and LMO2. A transient increase of LMOs also prevents RLIM-mediated turnover and it blocks the terminal differentiation of hematopoietic cells [172][173]. SSBP2 suppresses cell proliferation in prostate normal tissue, while in cancer its promoter hypermethylation and subsequent gene downregulation correlates with tumour staging and advanced prostate cancer. Our experiments showed a similar trend of SSBP2 expression level (downregulated at 12 and 120h), supporting emerging evidence. In head and neck carcinoma, SSBP2 also controls the transcriptional accumulation of LDB1 and LMO4 through poly-ubiquitination, since their overexpression particularly correlate with invasion, migration and angiogenesis. Various studies demonstrated a tissue specific oncogenic or tumour suppressor activity of SSBP2, while LMO4 role in solid tumours is still controversial. Overall, this analysis suggested a potential direct and early effect of ERG overexpression in the SSBP2-mediated LMOs' turnover, aimed to impair the physiologic hematopoietic development, supporting tumour progression. A schematization of the aforementioned process is reported in *Figure 64*.

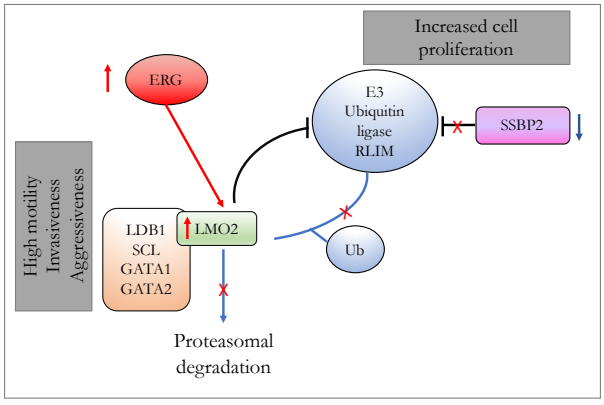


Figure 64| Schematic view the hypothetical mechanism of the SSBP2-dependent LMOs' turnover, promoted by ERG overexpression.

1.1.2 IFN signalling activation correlates with a deregulation of estrogenic response

The different approaches used to deepen ERG exclusive transcriptional programme, highlighted a central role of type I and II IFN signalling in our experimental conditions. In details, peculiar biological processes that emerged through the global network analysis seemed to be the consequence of an altered IFN response activation, with STAT1 playing a pivotal role. The strong deregulation of cluster 1 and 2 nodes referred to the concomitant activation of type I and II IFN signalling, and in particular almost all genes were upregulated at 24h and 120h, with HLA-A, HLA-DRA, STAT1 downregulated only at 12h. The biological meaning of IFN strong upregulation is difficult to assess, mainly because still vague is the characterization of its role in cancers. However, published evidence was useful to infer a possible mechanism and biologic impact of this deregulation.

In breast cancer, a reduced expression of estrogenic receptors correlates with HLA class II protein upregulation (in details HLA-DM and Ii). HLA-II or MHC-II are the major histocompatibility complex molecules involved in the adaptive immune response, including the anti-tumour immunity. The higher HLA-II expression is associated with IFN- γ increase, even if other factors are implicated in their expression. HLA-II expression is controlled by the class II transactivator (CIITA) that does not directly bind the DNA but allows the recruitment of transcription factors to HLA-II promoter. In turn, CIITA expression is due to the IFN-γ inducible promoter pIV activation; therefore, when IFN-γ binds to its receptor (IFNGR), there is an autophosphorylation of JAK1/JAK2 and the receptor itself on Tyr^{440}, which results in the phosphorylation on Tyr^{701}, dimerization and nuclear translocation of STAT1 [174]. Into the nucleus, pSTAT1 binds the IFN-activated sequences (GAS) and promotes IRF1 expression that, in turn, binds the CIITA pIV promoter, allowing its transcription. Recent studies underlined the role of oestradiol (E2) in HLA-II inhibition, through the CIITA downregulation in breast cancer, properly interfering with IFN-γ signalling [175]. This mechanism suggests that E2-dependent interference of the CIITA and STAT1 signalling, causes a reduced patient survival. A more recent study in breast cancer, revealed that type I and II IFN signalling are both able to activate the ISGF3 complex (formed by STAT1-STAT2-IRF9), that is highly expressed in breast cancers compared to normal tissues, since it is probably involved in ERα-dependent gene regulation in $ER\alpha^+$ cancer cells [176]. The ISGF3 complex translocates into the nucleus, where it binds IFN stimulated response elements (ISREs) of ISGs, allowing their transcriptions [177]. In our model, actually, it might be that both type I and II IFN signalling are concomitantly activated, since we found STAT2, IRF9 and various ISGs overexpressed.

The main hypothesis is that the IFN signalling is relevant for the anti-tumour immune response that cancer cells are able to evade in order to prompt a malignant progression. This complex mechanism in obviously based on the immune cells activity and cytokine actions [178]. Although this, some evidence highlighted that in other circumstances, when prolonged or generally altered, this signalling could also trigger tumour promotion and progression. Of note, Snell *et al.* [179] point out that cancers might promote a chronic deregulation of inflammation and immunosuppression that turns into a dysfunctional immune state, difficult to contrast also through therapies. Therefore, a chronic inflammation could be a primary mechanism for cancer progression. Coherently, the canonical characterization of IFN role in cancers as beneficial thanks to T cell response activation, has been recently questioned [179]. Our apparently contradictory levels of ISGs, IRFs and all the other DE genes playing a role into the IFN signalling, could therefore be in line with the last updates concerning the IFN response in tumours, aimed to promote a chronic inflammatory status. In details, the major role seems to be played properly by STAT1. The IFN-γ activation entails a prolonged activation of STAT1 and its downstream targets, while on the other hand, IFN-α /-β and also -γ prompt the ISGF3 complex formation and allow ISGs transcription. In tumours, *STAT1* is often considered a tumour suppressor gene, in contrast with *STAT3* that, as oncogene, fosters tumour growth and survival. Such clear-cut interpretation has been recently doubted and the different modulation of STAT1 (transcriptionally or post-transductionally) has been associated with non-canonical STAT1 activity [167]. Some evidence, in fact, revealed how the ambiguous role of STAT1 activation inhibits the anti-tumour immune response of tumour-associated macrophages (TAMs), causing T-cells apoptosis and therefore a poor outcome. Moreover, a recent study highlighted that the immune surveillance in cancers can be reverted through a plethora of mechanisms. Using some PCa cell models, Ayub *et al.* [167] showed that miR-2909 increased STAT1 transcription and phosphorylation and in turn promotes ISGs expression, without affecting IFN production. In details, miR-2909 expression entails the upregulation of ISGs (as ISG15), OAS1, HERC5 that are well-characterized to promote cell proliferation in PCa. Strikingly, all the aforementioned targets of STAT1 activation were also found upregulated in our model. The driving hypothesis, therefore, is that type I and II IFN responses are altered due to a chronic stimulation of STAT1 that, through its target upregulation, supports an odd inflammatory status that can promote tumour initiation/progression. *Figure 65* reports the hypothetical molecular mechanism discussed.

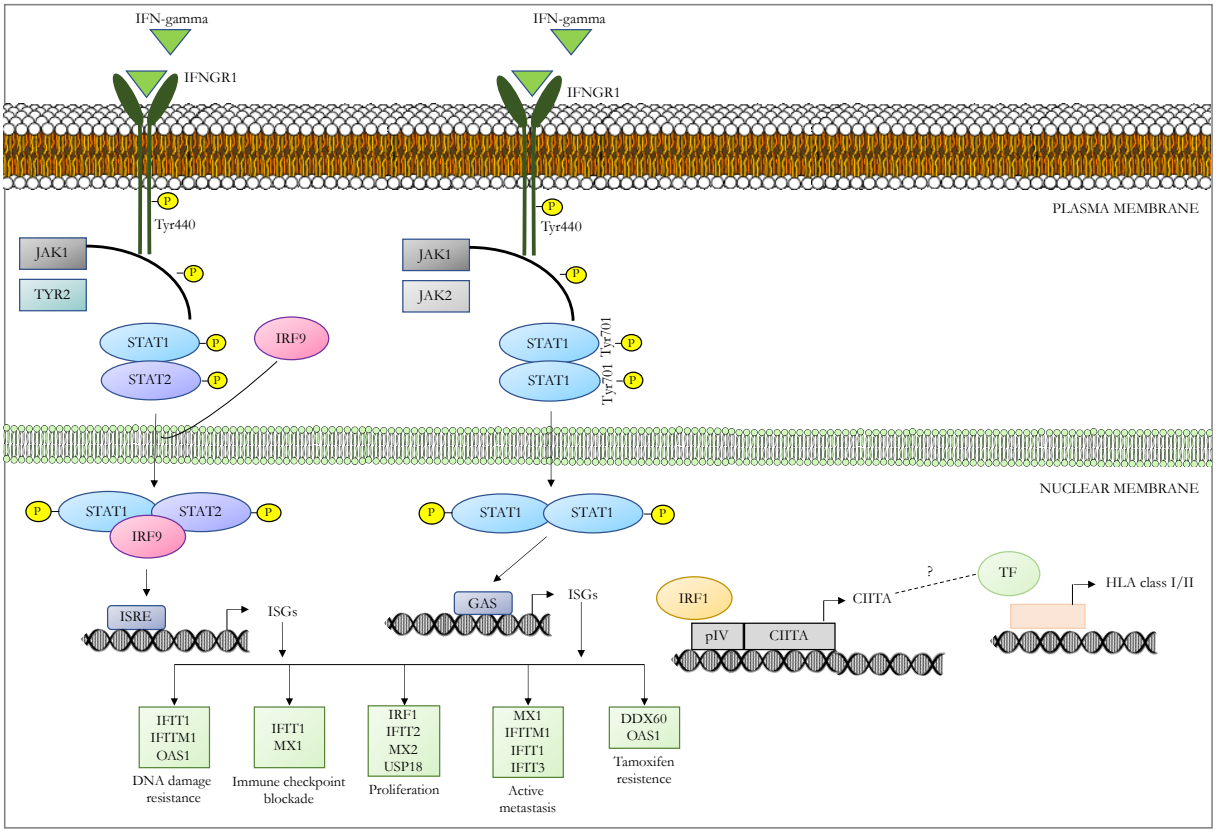


Figure 65| Hypothetical overview of IFN Type I/II signalling activation in our *in vitro* model. The scheme indicates two different phosphorylation processes that influence STAT1 activation, and a list of STAT1-targets that resulted differentially expressed in our RNA-Seq data, involved in various biological functions supportive for PCa progression.

1.1.3 CAMKIIB-dependent non-canonical STAT1 activation contributes to IFN signalling activity

Another relevant node belonging to cluster 1 and 2 of global analysis resulted the Ca^{2+}/calmodulin-dependent kinase II B (CAMK2B), upregulated up to 12h of ERG induction and mildly upregulated also in PRAD data. *CAMKIIB* is upregulated in different tumours and it phosphorylates a plethora of targets as enzymes, ion channels, kinases and transcription factors, all involved in altered proliferation, differentiation and survival. Recent evidence suggests that IFN-γ signalling can determine an increase of Ca^{2+} release, fostering *CAMKIIB* expression [180][181]. In CRC, *CAMKIIB* upregulation activates in turn the MEK/Erk signalling, promoting the phosphorylation and the subsequent proteasomal degradation of cyclin-dependent kinase inhibitor 1B ($p27^{Kip1}$), causing cell cycle progression and inducing the S-G2/M switch that sustains cancer invasion and metastasis. The most interesting role of this kinase is therefore related to STAT1

phosphorylation. Nair *et al.* [180], demonstrate that the IFN-γ -dependent *CAMKIIB* activation promotes STAT1 phosphorylation of Ser^{727} in a JAK1/JAK2-independent fashion. This mechanism suggests that STAT1 activation can be transcriptionally or post-transductionally regulated, through different processes connected to the IFN signalling stimulation.

1.1.4 TRIM29 and the deregulation of keratin pattern

Another important IFN-dependent mechanism can be keratin deregulation, confirmed also in PRAD data. The cornification process might in fact be mediated by the activation of CIITA that is known to react with TRIM29, even if the exact nature of this reaction has not been yet fully understood (it might be mediated by ICAM-1, transcribed after CIITA activation).

TRIM29 is a member of the tripartite motif (TRIM) family proteins involved in various processes as intracellular signalling, cell development, apoptosis, protein quality control and carcinogenesis and it also takes part in the assembly of the DNA repair proteins. The exact role of TRIM29 in tumours is uncertain, since it seems to have a tissue-specific value. In details, TRIM29 has been characterized as a selective marker of basal cells within the prostatic epithelium. One of the most frequent alteration in PCa is the loss of a basal cell layer in the prostatic acini and coherently with this, TRIM29 is strongly downregulated in tumour samples compared to the normal counterpart [182]. The same downregulation is also observed in squamous cell carcinoma, where a lower expression of TRIM29 is associated with cell migration and invasion [183]. Strikingly, TRIM29 co-localized with keratins into the cytoplasm and its downregulation is also associated with an altered keratins pattern; in particular, TRIM29 forms a complex with keratins through its C-terminal domain and causes their alteration. To this extent, the chronic inflammation stimulated by the aberrant activation of the IFN signalling, correlates with the loss of epithelial features, through the downregulation of a plethora of targets. Particularly, keratins have been characterized in normal and benign prostate samples, as specific markers of cell subpopulations: Pan, luminal, basal, intermediate and stem, with a peculiar biological meaning [184]. Nonetheless, their deregulation in PCa correlates with a higher invasiveness, revealing a central role with what we previously assessed. In squamous cell carcinoma, TRIM29 localizes and forms complexes with keratins and keratin-interacting proteins and its downregulation causes a "nondiffuse" keratin pattern associated to migration and invasion [182][183]. We hypothesised that in PCa the same mechanism can occur, mainly because one of the most relevant alteration associated to tumour progression is the loss of basal cells, that in turn correlates with the downregulation of some basal

markers as KRT5 and KRT14 (downregulated in our cell model). Moreover, the downregulation of TRIM29, as in our experiments, is also associated to an increased expression of EMT markers as vimentin [185] and therefore strongly associated to the mesenchymal transition.

These insights thus might indicate that ERG overexpression significantly impacts on KRTs' pattern enhancing invasiveness, through an epigenetic reprogramming that impairs cell adhesion and polarity [145].

1.1.5 ERG-dependent vasculogenesis

Furthermore, as largely documented, in physiologic conditions ERG is crucial for endothelial differentiation and it acts as a gatekeeper of the endothelial maintenance in an anti-inflammatory status, through the repression of pro-inflammatory molecules (ICAM-1, VCAM, PAI1, IL8), while its altered expression impairs vascular development. During tumour angiogenesis, one of the early steps is represented by the "angiogenetic switch", consisting in the upregulation of pro-angiogenetic factors and neoangiogenesis initiation. The new vascular structures result altered with an irregular aspect that impacts on vascular network and function [186]. Moreover, ERG also regulates the endothelial cells lumen formation, acting on direct targets as the RHO GTPases RHOJ [187] that in our data resulted strongly upregulated, while in tumours (PRAD data) it showed a slight downregulation, suggesting an initial role in PCa onset.

Interestingly, ERG and other ETS factors show their binding motif [5' GGA(A/T) 3'] into the promoter region of genes as *KDR* (encoding for VEGFR2), a pivotal regulator of the endothelial biology [188]. The vascular endothelial growth factor receptor (VEGFR2) mediates cell proliferation and survival and is activated by a strong ligand-dependent tyrosine phosphorylation. In aortic endothelial, VEGF stimulation (through VEGFR2) is able to promote a transient STAT1 tyrosine phosphorylation presumably on Ser^{727}. It might be that in tumours, VEGF stimulation enhances a constitutive activation of STAT1 signalling with an uncontrolled growth [189]. The upregulation of *KDR* was observed in experimental data and not in PRAD. This evidence thus suggests that ERG plays a role in vasculature development in the early phases of cell transformation.

1.2 Comprehensive evaluation of the hypothetical molecular mechanism

Overall, we hypothesized a strong ERG-dependent IFN signalling activation aimed to do the groundwork for an altered scenario that correlates with estrogenic response downregulation, the promotion of an altered vasculogenesis and moreover the reprogramming of epithelial cells to a more differentiated phenotype, with basal-like features loss.

The transcript levels of the aforementioned targets were also investigated in real tumours, from the TCGA database. This comparison revealed a minor relevance of type I and II IFN signalling, while still strongly downregulated was the cornification process, characterized by basal-specific keratin downregulation. This confirmation was then useful to isolate a specific ERG-dependent process in line with the epithelium alteration known to be relevant in PCa carcinogenesis. To further consolidate the exclusive role of ERG in this regulation, the transcriptional programmes of *AR* and *ERG* were compared using the TCGA data. Coherently, all relevant genes showed an opposite trend between *AR* higher expression versus *ERG* overexpressing samples. This additional hint thus better stressed the unique role of ERG on these targets and also supported the idea that AR and ERG have different transcriptional programmes, sharing many binding sites [61].

Surely, more focused analyses are needed in order to better characterize the molecular mechanisms that correlate ERG overexpression to the type I/ II IFN signalling activation and the cornification deregulation, even if some turning points already emerged from our results.

2. An altered chromatin accessibility might sustain ERG transcriptional programme

To define a comprehensive evaluation of these hypotheses, we also investigated the epigenetic modifications that could sustain ERG transcriptional programme, therefore a Methyl-Seq analysis was performed. Unexpectedly, the time points considered did not show any significant differentially methylated cytosine, therefore we exclude a crucial role of DNA methylation at least in the cell model and time windows analyses.

Even though preliminary, some evidence instead supported the huge involvement of chromatin regulation in the ERG overexpression context. The global histone PTM evaluation stressed a complex histone regulation, occurring in an ERG- and time- dependent manner. Curiously, the global effect on histone modifications did not reflect a consistent transcriptional deregulation of

chromatin remodelling writer and eraser enzymes. Recent studies had in fact highlighted the central role of EZH2 and the Polycomb Complex 2 in the ERG-dependent transcription [61][190]; Yu *et al.* demonstrated that ERG transcriptional activity is EZH2-mediated therefore it is upregulated in PCa. Unexpectedly, our data showed an opposite regulation of EZH2 expression, apparently in contrast with the H3K27me3 increase. The marginal role of chromatin remodelling enzymes and the concomitant huge effect on histone PTMs might therefore be interpreted as an altered accessibility to chromatin, causing the impairment of enzyme binding and activity. Supportive results illustrated that ERG overexpression induces global chromatin interaction changes [191] particularly affecting cell lineage identity [145]. Experimental and PRAD data suggested that the enzymatic involvement could be mainly detectable in transformed tissues, while ERG overexpression represented a primary turning point also in the early phases of tumour initiation.

ERG is nowadays considered a master transcription factor that exerts its regulation through direct interaction with chromatin, instead of a canonical direct gene body binding [145]. The complexity of the epigenetic regulation would further be investigated with more focused approaches.

3. ERG overexpression strongly impacted on tumour- tumour microenvironment crosstalk

Malignancy and progression of tumour is largely mediated by the surrounding stroma that affects tumour plasticity. The tumour-stroma crosstalk is properly enhanced by different kinds of intracellular communication that emerged in the tumour microenvironment. In this scenario we investigated the hypothetical role of ERG overexpression in affecting the TME and its cellular component phenotype. The prediction of secreted molecules inferred from gene expression (experimental and PRAD data), already revealed a relevant impact of ERG overexpression in processes as: (i) extracellular matrix organization, (ii) regulation of IGF transport and uptake, (iii) chemotaxis, (iv) humoral immune response, (v) regulation of cell adhesion and (vi) ECM compounds, (vii) response to wounding, (viii) blood vessel regulation/development. The LC-MS analysis of conditioned media obtained from induced or not RWPE-1 cells, highlighted consistence with predictions' results; in fact, matching all datasets, the main enriched processes resulted: (i) response to wounding, (ii) regulation of cell adhesion, (iii) chemotaxis and extracellular matrix

organization. This correspondence suggested a relevant role of ERG-dependent transcription in secretome composition and activity, and we found that:

- C1S, C1R, PLAU resulted upregulated and higher secreted,
- COL7A1, KLK8, TGFBI, VEGFA were instead downregulated and lower secreted.

Preliminarily evaluation in terms of survival, highlighted that downregulated and less secreted molecules (VEGFA excluded) seemed relevant for the higher relapse-free survival in cumulative analyses, evaluated on all PRAD cases. The prognostic value of ERG has been largely studied, although contrasting evidence emerged. As found by Taris *et al.* [192], ERG overexpression represents an advantage for the overall survival and progression free survival in clinically localized as in CRPC samples; other studies instead highlight null or negative correlations between ERG expression and the prognostic value. Features responsible for this apparent paradox can in fact be correlated to *AR* expression and the tumour heterogeneity, therefore further analyses would be required. Moreover, it would be necessary to assess a specific role of significant molecules obtained from our data, taking into account the complexity of secretion mechanisms that we defined through a bioinformatic approach to be largely represented by vesicled molecules.

Preliminarily, the effect of ERG overexpression on TME cell was investigated using normal and cancer fibroblasts as cell model, therefore fibroblasts were exposed to ERG induced or not CM and their phenotype analysed. Although PDPN resulted upregulated in NAFs exposed to induced CM, suggesting a more "activated" phenotype, αSMA, PDGFα and PDGFβ levels in induced-exposed CAFs didn't allowed a clear phenotype identification. Therefore, the molecular evaluation failed to determine a possible stroma ERG-dependent phenotype change, at least for the markers investigated.

In addition to this, exposed fibroblasts were also evaluated for the migratory behaviour though the wound healing assay. In details, the analysis did not strongly reflect a proper "wound response", but instead showed a strong impact on cells survival for both induced-exposed NAFs and CAFs compared to the not induced counterpart. These interesting findings thus might suggest that the transcriptional programme prompted with ERG overexpression, not only acted through a chromatin remodelling to define the bases for epithelial changes towards a luminal lineage, but also affected the stroma fate, through the release of compounds that influence tumour microenvironment. The ERG-dependent inflammatory status with type I/II IFN response activation might set the scene also for the reactive stroma, providing secreted molecules that enhance fibroblasts' survival. Future analyses would try to elucidate if NAFs and CAFs exposed to not-/induced CM alter their typical secretome for a reverse effect on epithelial cells.

Further and more extensive analyses are required for each of the hints emerged from our work, though different aspects already found correspondence with published evidence. Of note, another relevant issue that deserves a more focus analysis relies on the significance of DNMT3A transcript variants, their roles in PCa progression and chromatin changes, marginally dealt with in the present work.

Conclusions and future perspectives

The overall aim of the project was to identify regulatory processes (DNA methylation, chromatin rearrangements, ncRNAs) that determined a specific ERG-dependent transcriptional programme aimed to alter the tumour-tumour microenvironment crosstalk and activate TME cells, ruling secretion and trafficking. This thesis particularly focused on the interdependence between ERG-dependent transcription and the differential molecule secretion (schematised in *Figure 66*). In details, we aimed to identify DE genes that could be translated into differentially secreted molecules and/or act on the secretion process and their regulation, in order to better characterized the crosstalk between tumour and tumour microenvironment.

Firstly, we confirmed that ERG overexpression defined a mesenchymal switch of epithelial cells that also acquired invasive and migratory features, without affecting their proliferation. The resulting EMT was defined as a long-term effect, while ERG transcriptional programme started to do the groundwork for tumour initiation in an early phase. Although the cell model used highlighted some critical issues, the Gene-Set Enrichment Analysis primarily confirmed that at 12h of induction, ERG overexpression was already responsible for the transcriptional upregulation of type I/II IFN signalling, inflammatory response and a concomitant downregulation of the estrogenic signalling, while the EMT activation was observed up to 24h. Further validating analyses with PRAD data, revealed a limited role of IFN and inflammatory responses in tumours, while totally confirmed the estrogenic response downregulation. These results suggested that ERG overexpression set bases for tumour initiation through an odd inflammatory status, that could no more be assessed in developed tumours.

Supportive hints emerged also with a systems biology approach, since the global network analysis underlined the central role of STAT1 activation (though canonical and non-canonical mechanisms). Its activation in turn supported the upregulation of IFN signalling and, moreover, seemed to promote the downregulation of the estrogenic response and the cornification process, through the downregulation of TRIM29 and the consequent definition of a "nondiffuse" keratin pattern associated to migration and invasion [182][183]. Coherently, Li *et al.* [145] showed that keratin deregulation is the result of an epigenetic ERG-dependent reprogramming aimed to altered the epithelium lineage plasticity, attenuating basal-like features. Furthermore, preliminary analyses

of epigenetic regulation, underlined a limited DNA methylation involvement, while chromatin status and accessibility were found pivotal and in line with emerging highlights.

Secondly, we tried to depict secretome role and composition through a combined approach based on a transcriptional prediction of differentially secreted genes/molecules (from experimental and PRAD data) and a proteomic analysis of conditioned media obtained from ERG induced or not RWPE-1 cells. Interestingly, we demonstrated that some biological processes characterizing molecule secretion were prompted by ERG transcriptional programme; in details, the two approaches highlighted overlap for: (i) response to wounding, (ii) regulation of cell adhesion, (iii) chemotaxis, (iv) ECM organization. Of note, the secretome biological effect was investigated with a preliminary analysis, using NAFs and CAFs as a TME cell model. We showed that fibroblasts exposed to ERG induced conditioned media were able to survive compared to the not induced-exposed counterpart. This first evidence therefore suggested that ERG transcriptional programme determined a secretome composition that positively correlated with fibroblast survival.

Finally, future perspectives would focus on relevant hints emerged throughout this work and are below reported. Particularly we aimed to:

(i) Validate interesting ERG induced targets in an ERG-silenced PCa cell model,
(ii) Deepen the molecular mechanism hypothesized performing biochemical studies to define ERG- dependent interactome,
(iii) Analyse STAT1 activation and relating phosphorylation to characterize its role and its activity on the ERG-dependent IFN signalling activation,
(iv) Examine the chromatin status of our cell line in order to better explain the strong histone modifications that occur during ERG induction on promising targets as keratins,
(v) Cross-validate crucial molecules emerged from transcriptional prediction and secretome analyses, using databases and *in vitro* experiments on NAFs and CAFs,
(vi) Phenotypically and morphologically characterize NAFs and CAFs exposed to ERG induced conditioned,
(vii) Underpin the exclusive role of DNMT3A transcript variants in PCa, especially concerning histone PTMs.

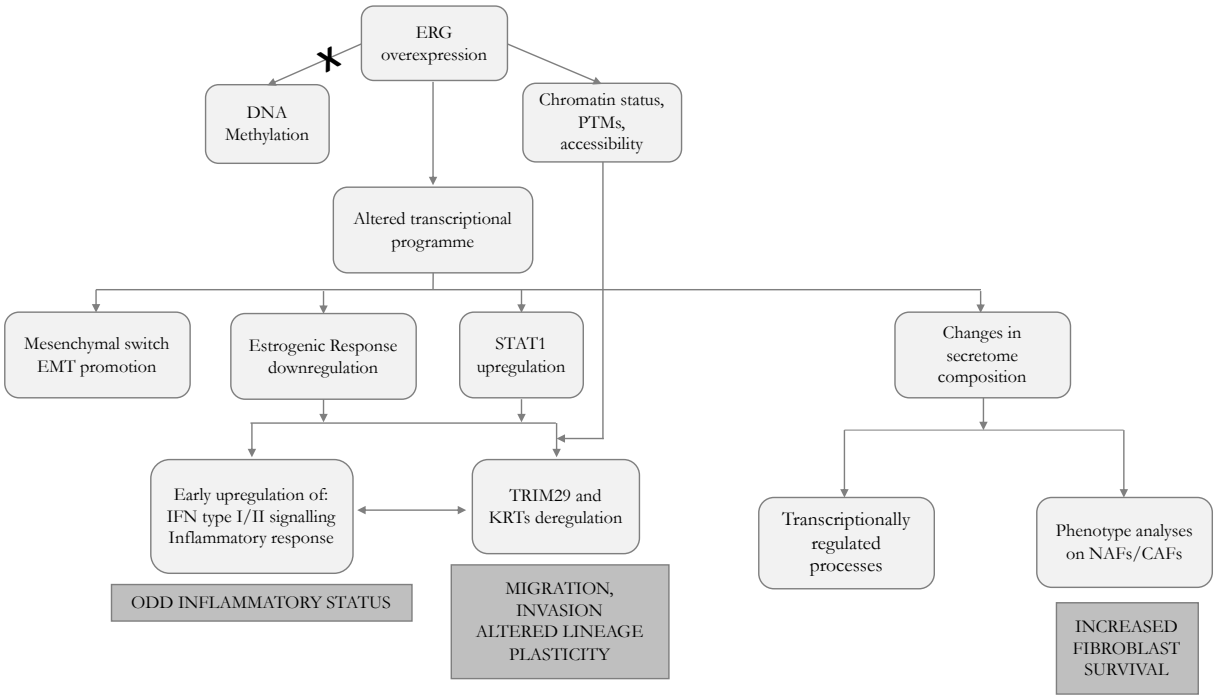


Figure 66| Schematization of results obtained throughout the thesis project.

References

[1] H. M. Wise, M. A. Hermida, and N. R. Leslie, "Prostate cancer, PI3K, PTEN and prognosis," *Clin. Sci.*, vol. 131, no. 3, pp. 197–210, Feb. 2017.

[2] S. Punnen, N. Pavan, and D. J. Parekh, "Finding the Wolf in Sheep's Clothing: The 4Kscore Is a Novel Blood Test That Can Accurately Identify the Risk of Aggressive Prostate Cancer.," *Rev. Urol.*, vol. 17, no. 1, pp. 3–13, 2015.

[3] K. Fujita and N. Nonomura, "Urinary biomarkers of prostate cancer," *Int. J. Urol.*, vol. 25, no. 9, pp. 770–779, Sep. 2018.

[4] I. Kohaar, G. Petrovics, and S. Srivastava, "A Rich Array of Prostate Cancer Molecular Biomarkers: Opportunities and Challenges.," *Int. J. Mol. Sci.*, vol. 20, no. 8, Apr. 2019.

[5] H. Moch, A. L. Cubilla, P. A. Humphrey, V. E. Reuter, and T. M. Ulbright, "The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours.," *Eur. Urol.*, vol. 70, no. 1, pp. 93–105, Jul. 2016.

[6] G. A. Giannico and O. Hameed, "Evaluation of Prostate Needle Biopsies," Springer, Cham, 2018, pp. 69–86.

[7] L. Egevad, B. Delahunt, J. R. Srigley, and H. Samaratunga, "International Society of Urological Pathology (ISUP) grading of prostate cancer - An ISUP consensus on contemporary grading," *APMIS*, vol. 124, no. 6, pp. 433–435, Jun. 2016.

[8] P. Harnden, M. D. Shelley, B. Coles, J. Staffurth, and M. D. Mason, "Should the Gleason grading system for prostate cancer be modified to account for high-grade tertiary components? A systematic review and meta-analysis," *Lancet Oncology*, vol. 8, no. 5. Elsevier, pp. 411–419, 01-May-2007.

[9] M. Andreoiu and L. Cheng, "Multifocal prostate cancer: biologic, prognostic, and therapeutic implications," *Hum. Pathol.*, vol. 41, no. 6, pp. 781–793, Jun. 2010.

[10] P. C. Boutros *et al.*, "Spatial genomic heterogeneity within localized, multifocal prostate cancer," *Nat. Genet.*, vol. 47, no. 7, pp. 736–745, Jul. 2015.

[11] L. Wei and J. Wang, "Intratumoral and Intertumoral Genomic Heterogeneity of Multifocal Localized Prostate Cancer Impacts Molecular Classifications and Genomic Prognosticators," *Eur Urol*, 2017.

[12] J. Lindberg *et al.*, "Exome Sequencing of Prostate Cancer Supports the Hypothesis of Independent Tumour Origins," *Eur. Urol.*, vol. 63, no. 2, pp. 347–353, Feb. 2013.

[13] M. Chrisofos, A. G. Papatsoris, A. Lazaris, and C. Deliveliotis, "Precursor Lesions of Prostate Cancer," *Crit. Rev. Clin. Lab. Sci.*, vol. 44, no. 3, pp. 243–270, Jan. 2007.

[14] M. J. Putzi and A. M. De Marzo, "Morphologic transitions between proliferative inflammatory atrophy and high-grade prostatic intraepithelial neoplasia," *Urology*, vol. 56, no. 5, pp. 828–832, Nov. 2000.

[15] K. S. Sfanos and A. M. De Marzo, "Prostate cancer and inflammation: the evidence.," *Histopathology*, vol. 60, no. 1, pp. 199–215, Jan. 2012.

[16] S. A. Tomlins, M. A. Rubin, and A. M. Chinnaiyan, "Integrative Biology of Prostate Cancer Progression," *Annu. Rev. Pathol. Mech. Dis.*, vol. 1, no. 1, pp. 243–271, Feb. 2006.

[17] T. Mitchell and D. E. Neal, "The genomic evolution of human prostate cancer," *Br. J. Cancer*, vol. 113, no. 2, pp. 193–198, Jul. 2015.

[18] A. M. De Marzo *et al.*, "Inflammation in prostate carcinogenesis," *Nat. Rev. Cancer*, vol. 7, no. 4, pp. 256–269, Apr. 2007.

[19] Y. Cheng *et al.*, "Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials," *Signal Transduct. Target. Ther.*, vol. 4, no. 1, pp. 1–39, Dec. 2019.

[20] G. Attard *et al.*, "Prostate cancer," *Lancet*, vol. 387, no. 10013, pp. 70–82, Jan. 2016.

[21] J. R. Packer and N. J. Maitland, "The molecular and cellular origin of human prostate cancer," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1863, no. 6, pp. 1238–1260, Jun. 2016.

[22] V. Velcheti, S. Karnik, S. F. Bardot, and O. Prakash, "Pathogenesis of prostate cancer: lessons from basic research.," *Ochsner J.*, vol. 8, no. 4, pp. 213–8, 2008.

[23] A. P. Rybak, R. G. Bristow, A. Kapoor, A. P. Rybak, R. G. Bristow, and A. Kapoor, "Prostate cancer stem cells: deciphering the origins and pathways involved in prostate tumorigenesis and aggression," *Oncotarget*, vol. 6, no. 4, pp. 1900–1919, Feb. 2015.

[24] D. Zhang, S. Zhao, X. Li, J. S. Kirk, and D. G. Tang, "Prostate Luminal Progenitor Cells in Development and Cancer," *Trends in Cancer*, vol. 4, no. 11. Cell Press, pp. 769–783, 01-Nov-2018.

[25] I. Sansal and W. R. Sellers, "The biology and clinical relevance of the PTEN tumor suppressor pathway.," *J. Clin. Oncol.*, vol. 22, no. 14, pp. 2954–63, Jul. 2004.

[26] E. M. Ciruelos Gil, "Targeting the PI3K/AKT/mTOR pathway in estrogen receptor-positive breast cancer.," *Cancer Treat. Rev.*, vol. 40, no. 7, pp. 862–71, Aug. 2014.

[27] J. A. McCubrey *et al.*, "Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance.," *Biochim. Biophys. Acta*, vol. 1773, no. 8, pp. 1263–84, Aug. 2007.

[28] R. E. Bakin, D. Gioeli, R. A. Sikes, E. A. Bissonette, and M. J. Weber, "Constitutive Activation of the Ras/Mitogen-activated Protein Kinase Signaling Pathway Promotes Androgen Hypersensitivity in LNCaP Prostate Cancer Cells 1," 1981.

[29] D. Robinson *et al.*, "Integrative Clinical Genomics of Advanced Prostate Cancer," *Cell*, vol. 161, no. 5, pp. 1215–1228, May 2015.

[30] M. Kluth *et al.*, "Clinical significance of different types of *p53* gene alteration in surgically treated prostate cancer," *Int. J. Cancer*, vol. 135, no. 6, pp. 1369–1380, Sep. 2014.

[31] B. S. Taylor *et al.*, "Integrative genomic profiling of human prostate cancer.," *Cancer Cell*, vol. 18, no. 1, pp. 11–22, Jul. 2010.

[32] A. M. Udager *et al.*, "Concurrent nuclear ERG and MYC protein overexpression defines a subset of locally advanced prostate cancer: Potential opportunities for synergistic targeted therapeutics.," *Prostate*, vol. 76, no. 9, pp. 845–53, 2016.

[33] C. S. Grasso *et al.*, "The mutational landscape of lethal castration-resistant prostate cancer.," *Nature*, vol. 487, no. 7406, pp. 239–43, Jul. 2012.

[34] T. Senda, A. Shimomura, and A. Iizuka-Kogo, "Adenomatous polyposis coli (Apc) tumor suppressor gene as a multifunctional gene," *Anat. Sci. Int.*, vol. 80, no. 3, pp. 121–131, Sep. 2005.

[35] S. S. Karhadkar *et al.*, "Hedgehog signalling in prostate regeneration, neoplasia and metastasis," *Nature*, vol. 431, no. 7009, pp. 707–712, Oct. 2004.

[36] D. O. Walterhouse, M. L. G. Lamm, E. Villavicencio, and P. M. Iannaccone, "Emerging Roles for Hedgehog-Patched-Gli Signal Transduction in Reproduction1," *Biol. Reprod.*, vol. 69, no. 1, pp. 8–14, Jul. 2003.

[37] J. T. Hua *et al.*, "Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19," *Cell*, vol. 174, no. 3, pp. 564-575.e18, Jul. 2018.

[38] J. C. Francis, A. Capper, J. Ning, E. Knight, J. de Bono, and A. Swain, "SOX9 is a driver of aggressive prostate cancer by promoting invasion, cell fate and cytoskeleton alterations and epithelial to mesenchymal transition," *Oncotarget*, vol. 9, no. 7, pp. 7604–7615, Jan. 2018.

[39] K. H. Stopsack *et al.*, "Low Expression of the Androgen-Induced Tumor Suppressor Gene PLZF and Lethal Prostate Cancer.," *Cancer Epidemiol. Biomarkers Prev.*, vol. 28, no. 4, pp. 707–714, 2019.

[40] I. V Tereshchenko *et al.*, "ERG and CHD1 heterogeneity in prostate cancer: use of confocal microscopy in assessment of microscopic foci.," *Prostate*, vol. 74, no. 15, pp. 1551–9, Nov. 2014.

[41] S. Huang, Z. G. Gulzar, K. Salari, J. Lapointe, J. D. Brooks, and J. R. Pollack, "Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness," *Oncogene*, vol. 31, no. 37, pp. 4164–4170, Sep. 2012.

[42] C. Geng *et al.*, "Prostate cancer-associated mutations in speckle-type POZ protein (SPOP) regulate steroid receptor coactivator 3 protein turnover," *Proc. Natl. Acad. Sci.*, vol. 110, no. 17, pp. 6997–7002, Apr. 2013.

[43] K. Takayama and S. Inoue, "Transcriptional network of androgen receptor in prostate cancer progression," *Int. J. Urol.*, vol. 20, no. 8, pp. 756–768, Aug. 2013.

[44] K. K. Waltering and A. Urbanucci, "Androgen receptor (AR) aberrations in castration-resistant prostate cancer," *Mol. Cell. Endocrinol.*, vol. 360, no. 1–2, pp. 38–43, Sep. 2012.

[45] S. A. Tomlins *et al.*, "Integrative molecular concept modeling of prostate cancer progression," *Nat. Genet.*, vol. 39, no. 1, pp. 41–51, Jan. 2007.

[46] J. C. Brenner *et al.*, "Mechanistic Rationale for Inhibition of Poly(ADP-Ribose) Polymerase in ETS Gene Fusion-Positive Prostate Cancer," *Cancer Cell*, vol. 19, no. 5, pp. 664–678, May 2011.

[47] M. P. Craig and S. Sumanas, "ETS transcription factors in embryonic vascular development.," *Angiogenesis*, vol. 19, no. 3, pp. 275–85, 2016.

[48] A. V. Shah, G. M. Birdsey, and A. M. Randi, "Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG," *Vascul. Pharmacol.*, vol. 86, pp. 3–13, 2016.

[49] A. D. Sharrocks, "The ETS-domain transcription factor family," *Nat. Rev. Mol. Cell Biol.*, vol. 2, no. 11, pp. 827–837, Nov. 2001.

[50] S. A. Tomlins *et al.*, "Role of the TMPRSS2-ERG Gene Fusion in Prostate Cancer 1,2,3," *Neoplasia*, vol. 10, pp. 177–188, 2008.

[51] Tomlins et al., "Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer," *Science (80-.).*, vol. 310, 2005.

[52] J. P. Clark and C. S. Cooper, "ETS gene fusions in prostate cancer," *Nat. Rev. Urol.*, vol. 6, no. 8, pp. 429–439, Aug. 2009.

[53] S. A. Tomlins *et al.*, "ETS Gene Fusions in Prostate Cancer: From Discovery to Daily Clinical Practice," *Eur. Urol.*, vol. 56, no. 2, pp. 275–286, Aug. 2009.

[54] R. Singareddy *et al.*, "Transcriptional Regulation of CXCR4 in Prostate Cancer: Significance of TMPRSS2-ERG Fusions," *Mol. Cancer Res.*, vol. 11, no. 11, pp. 1349–1361, Nov. 2013.

[55] C. Song and H. Chen, "Predictive significance of TMRPSS2-ERG fusion in prostate cancer: a meta-analysis," *Cancer Cell Int.*, vol. 18, no. 1, p. 177, Dec. 2018.

[56] S.-R. Lee, Y.-D. Choi, and N.-H. Cho, "Association between pathologic factors and ERG expression in prostate cancer: finding pivotal networking," *J. Cancer Res. Clin. Oncol.*, vol. 144, no. 9, pp. 1665–1683, Sep. 2018.

[57] G. M. Birdsey *et al.*, "Transcription factor Erg regulates angiogenesis and endothelial apoptosis through VE-cadherin," *Blood*, vol. 111, no. 7, p. 3498, 2008.

[58] J. A. I. Thoms *et al.*, "ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer," *Blood*, vol. 117, no. 26, pp. 7079–7089, Jun. 2011.

[59] X. Zhang, M. V. Fournier, J. L. Ware, M. J. Bissell, A. Jacoub, and Z. E. Zehner, "Inhibition of Vimentin or β1-integrin Reverts Morphology of Prostate Tumor Cells Grown in Laminin-rich ECM gels and Reduces Tumor Growth in vivo," *Mol. Cancer Ther.*, vol. 8, no. 3, p. 499, 2009.

[60] S. C *et al.*, "TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation.," *Oncogene*, vol. 27, no. 40, pp. 5348–5353, Jun. 2008.

[61] J. Yu *et al.*, "An Integrated Network of Androgen Receptor, Polycomb, and TMPRSS2-ERG Gene Fusions in Prostate Cancer Progression," *Cancer Cell*, vol. 17, no. 5, p. 443, 2010.

[62] L. Yang *et al.*, "An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B.," *Biochem. J.*, vol. 369, no. Pt 3, p. 651, 2003.

[63] J. Kim, L. Wu, J. C. Zhao, H.-J. Jin, and J. Yu, "TMPRSS2–ERG gene fusions induce prostate tumorigenesis by modulating microRNA miR-200c," *Oncogene*, vol. 33, no. 44, pp. 5183–5192, Oct. 2014.

[64] C.-J. Kao *et al.*, "miR-30 as a tumor suppressor connects EGF/Src signal to ERG and EMT," *Oncogene*, vol. 33, no. 19, pp. 2495–2503, May 2014.

[65] R. Kalluri and R. A. Weinberg, "The basics of epithelial-mesenchymal transition.," *J. Clin. Invest.*, vol. 119, no. 6, pp. 1420–8, Jun. 2009.

[66] G. Barriere, P. Fici, G. Gallerani, F. Fabbri, W. Zoli, and M. Rigaud, "Circulating tumor cells and epithelial, mesenchymal and stemness markers: characterization of cell subpopulations.," *Ann. Transl. Med.*, vol. 2, no. 11, p. 109, Nov. 2014.

[67] M. A. Nieto, "The Ins and Outs of the Epithelial to Mesenchymal Transition in Health and Disease," *Annu. Rev. Cell Dev. Biol.*, vol. 27, no. 1, pp. 347–376, Nov. 2011.

[68] J. Condeelis, R. H. Singer, and J. E. Segall, "THE GREAT ESCAPE: When Cancer Cells Hijack the Genes for Chemotaxis and Motility," *Annu. Rev. Cell Dev. Biol.*, vol. 21, no. 1, pp. 695–718, Nov. 2005.

[69] H. Wallerand *et al.*, "The epithelial-mesenchymal transition-inducing factor TWIST is an attractive target in advanced and/or metastatic bladder and prostate cancers," *Urol. Oncol. Semin. Orig. Investig.*, vol. 28, no. 5, pp. 473–479, Sep. 2010.

[70] H. Wang, N. C. McKnight, T. Zhang, M. L. Lu, S. P. Balk, and X. Yuan, "SOX9 Is Expressed in Normal Prostate Basal Cells and Regulates Androgen Receptor Expression in Prostate Cancer Cells," *Cancer Res.*, vol. 67, no. 2, pp. 528–536, Jan. 2007.

[71] V. Ivanović *et al.*, "Elevated plasma levels of transforming growth factor-β1 (TGF-β1) in patients with

advanced breast cancer: association with disease progression," *Eur. J. Cancer*, vol. 39, no. 4, pp. 454–461, Mar. 2003.

[72] M. E. Tan, J. Li, H. E. Xu, K. Melcher, and E. Yong, "Androgen receptor: structure, role in prostate cancer and drug discovery," *Acta Pharmacol. Sin.*, vol. 36, no. 1, pp. 3–23, Jan. 2015.

[73] B. Ciruna and J. Rossant, "FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak.," *Dev. Cell*, vol. 1, no. 1, pp. 37–49, Jul. 2001.

[74] S. Brabletz and T. Brabletz, "The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer?," *EMBO Rep.*, vol. 11, no. 9, p. 670, Sep. 2010.

[75] S. Slavin *et al.*, "Estrogen receptor α in cancer-associated fibroblasts suppresses prostate cancer invasion via modulation of thrombospondin 2 and matrix metalloproteinase 3," *Carcinogenesis*, vol. 35, no. 6, pp. 1301–1309, Jun. 2014.

[76] P. Dey *et al.*, "Estrogen Receptor β2 Induces Hypoxia Signature of Gene Expression by Stabilizing HIF-1α in Prostate Cancer.," *PLoS One*, vol. 10, no. 5, p. e0128239, 2015.

[77] Y. E. Zhang, "Non-Smad pathways in TGF-β signaling," *Cell Res.*, vol. 19, no. 1, pp. 128–139, Jan. 2009.

[78] D. M. Gonzalez and D. Medici, "Signaling mechanisms of the epithelial-mesenchymal transition," *Sci. Signal.*, vol. 7, no. 344, p. re8, 2014.

[79] S. L. Shiao, G. C.-Y. Chu, and L. W. K. Chung, "Regulation of prostate cancer progression by the tumor microenvironment," *Cancer Lett.*, vol. 380, no. 1, pp. 340–348, Sep. 2016.

[80] X. Lu and Y. Kang, "Epidermal growth factor signalling and bone metastasis," *Br. J. Cancer*, vol. 102, no. 3, p. 457, 2010.

[81] G. Di Lorenzo *et al.*, "Expression of Epidermal Growth Factor Receptor Correlates with Disease Relapse and Progression to Androgen-independence in Human Prostate Cancer," *Clin. Cancer Res.*, vol. 7, no. 10, pp. 2958–2970, Nov. 2002.

[82] Z. Liu *et al.*, "Snail regulated by PKC/GSK-3β pathway is crucial for EGF-induced epithelial-mesenchymal transition (EMT) of cancer cells," *Cell Tissue Res.*, vol. 358, no. 2, pp. 491–502, Nov. 2014.

[83] K. H. Cho *et al.*, "A ROS/STAT3/HIF-1α signaling cascade mediates EGF-induced TWIST1 expression and prostate cancer cell invasion," *Prostate*, vol. 74, no. 5, pp. 528–536, May 2014.

[84] M. Montanari *et al.*, "Epithelial-mesenchymal transition in prostate cancer: An overview," *Oncotarget*, vol. 8, no. 21. Impact Journals LLC, pp. 35376–35389, 2017.

[85] P.-C. Lin *et al.*, "Epigenomic Alterations in Localized and Advanced Prostate Cancer," *Neoplasia*, vol. 15, no. 4, pp. 373-IN5, Apr. 2013.

[86] H. Suzuki, T. Ueda, T. Ichikawa, and H. Ito, "Androgen receptor involvement in the progression of prostate cancer.," *Endocr. Relat. Cancer*, pp. 209–216, Jun. 2003.

[87] B. G. Suzuki H, Freije D, Nusskern DR, Okami K, Cairns P, Sidransky D, Isaacs WB, "Interfocal heterogeneity of PTEN/MMAC1 gene alterations in multiple metastatic prostate cancer tissues. - PubMed - NCBI," *Cancer Res.*, 1998.

[88] P. Pakneshan, R. H. Xing, and S. A. Rabbani, "Methylation status of uPA promoter as a molecular mechanism regulating prostate cancer invasion and growth in vitro and in vivo," *FASEB J.*, vol. 17, no. 9, pp. 1081–1088, Jun. 2003.

[89] R. Goyal, R. Reinhardt, and A. Jeltsch, "Accuracy of DNA methylation pattern preservation by the Dnmt1

methyltransferase," *Nucleic Acids Res.*, vol. 34, no. 4, p. 1182, 2006.

[90] C. Pistore *et al.*, "DNA methylation variations are required for epithelial-to-mesenchymal transition induced by cancer-associated fibroblasts in prostate cancer cells," *Nat. Publ. Gr.*, vol. 36, pp. 5551–5566, 2017.

[91] P. A. Jones, "Functions of DNA methylation: islands, start sites, gene bodies and beyond," *Nat. Rev. Genet.*, vol. 13, no. 7, pp. 484–492, Jul. 2012.

[92] D. B. Seligson *et al.*, "Global histone modification patterns predict risk of prostate cancer recurrence," *Nature*, vol. 435, no. 7046, pp. 1262–1266, Jun. 2005.

[93] J. Ellinger *et al.*, "Global Histone H3K27 Methylation Levels are Different in Localized and Metastatic Prostate Cancer," *Cancer Invest.*, vol. 30, no. 2, pp. 92–97, Feb. 2012.

[94] K. Ruggero, S. Farran-Matas, A. Martinez-Tebar, and A. Aytes, "Epigenetic Regulation in Prostate Cancer Progression," *Curr. Mol. Biol. Reports*, vol. 4, no. 2, pp. 101–115, Jun. 2018.

[95] H. Chen, S. Tu, and J.-T. Hsieh, "Down-regulation of Human DAB2IP Gene Expression Mediated by Polycomb Ezh2 Complex and Histone Deacetylase in Prostate Cancer," *J. Biol. Chem.*, vol. 280, no. 23, pp. 22437–22444, Jun. 2005.

[96] I. A. Asangani *et al.*, "Characterization of the EZH2-MMSET Histone Methyltransferase Regulatory Axis in Cancer," *Mol. Cell*, vol. 49, no. 1, p. 80, 2013.

[97] J. Kim *et al.*, "Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator," *Cell Rep.*, vol. 25, no. 10, pp. 2808-2820.e4, Dec. 2018.

[98] C. Soekmadji, P. Russell, and C. Nelson, "Exosomes in Prostate Cancer: Putting Together the Pieces of a Puzzle," *Cancers (Basel).*, vol. 5, no. 4, pp. 1522–1544, Nov. 2013.

[99] A. Mantovani, S. Sozzani, M. Locati, P. Allavena, and A. Sica, "Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes," *Trends Immunol.*, vol. 23, no. 11, pp. 549–555, Nov. 2002.

[100] A. Mantovani, P. Allavena, A. Sica, and F. Balkwill, "Cancer-related inflammation," *Nature*, vol. 454, no. 7203, pp. 436–444, Jul. 2008.

[101] D. J. Campbell and M. A. Koch, "Treg cells: patrolling a dangerous neighborhood," *Nat. Med.*, vol. 17, no. 8, pp. 929–930, Aug. 2011.

[102] C. Mauri and A. Bosma, "Immune Regulatory Function of B Cells," *Annu. Rev. Immunol.*, vol. 30, no. 1, pp. 221–241, Apr. 2012.

[103] C. Murdoch, A. Giannoudis, and C. E. Lewis, "Mechanisms regulating the recruitment of macrophages into hypoxic areas of tumors and other ischemic tissues," *Blood*, vol. 104, no. 8, pp. 2224–2234, Oct. 2004.

[104] M. L. Taddei, E. Giannoni, G. Comito, and P. Chiarugi, "Microenvironment and tumor cell plasticity: An easy way out," *Cancer Lett.*, vol. 341, no. 1, pp. 80–96, Nov. 2013.

[105] M. M. Meredith *et al.*, "Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage," *J. Exp. Med.*, vol. 209, no. 6, pp. 1153–1165, Jun. 2012.

[106] Z. Granot, E. Henke, E. A. Comen, T. A. King, L. Norton, and R. Benezra, "Tumor Entrained Neutrophils Inhibit Seeding in the Premetastatic Lung," *Cancer Cell*, vol. 20, no. 3, pp. 300–314, Sep. 2011.

[107] J. E. De Larco, "The Potential Role of Neutrophils in Promoting the Metastatic Phenotype of Tumors Releasing Interleukin-8," *Clin. Cancer Res.*, vol. 10, no. 15, pp. 4895–4900, Aug. 2004.

[108] P. Carmeliet and R. K. Jain, "Molecular mechanisms and clinical applications of angiogenesis," *Nature*, vol.

473, no. 7347, pp. 298–307, May 2011.

[109] V. G. Cooke *et al.*, "Pericyte Depletion Results in Hypoxia-Associated Epithelial-to-Mesenchymal Transition and Metastasis Mediated by Met Signaling Pathway," *Cancer Cell*, vol. 21, no. 1, pp. 66–81, Jan. 2012.

[110] B. Weigelt and M. J. Bissell, "Unraveling the microenvironmental influences on the normal mammary gland and breast cancer," *Semin. Cancer Biol.*, vol. 18, no. 5, pp. 311–321, Oct. 2008.

[111] H. Y. Chang *et al.*, "Diversity, topographic differentiation, and positional memory in human fibroblasts," *Proc. Natl. Acad. Sci.*, vol. 99, no. 20, pp. 12877–12882, Oct. 2002.

[112] R. Kalluri and M. Zeisberg, "Fibroblasts in cancer," *Nat. Rev. Cancer*, vol. 6, no. 5, pp. 392–401, May 2006.

[113] P. Cirri and P. Chiarugi, "Cancer-associated-fibroblasts and tumour cells: a diabolic liaison driving cancer progression," *Cancer Metastasis Rev.*, vol. 31, no. 1–2, pp. 195–208, Jun. 2012.

[114] E. Giannoni *et al.*, "Reciprocal Activation of Prostate Cancer Cells and Cancer-Associated Fibroblasts Stimulates Epithelial-Mesenchymal Transition and Cancer Stemness," *Cancer Res.*, vol. 70, no. 17, pp. 6945–6956, Sep. 2010.

[115] R. Kalluri, "The biology and function of fibroblasts in cancer," *Nature Reviews Cancer*, vol. 16, no. 9. Nature Publishing Group, pp. 582–598, 01-Sep-2016.

[116] G. Comito *et al.*, "Cancer-associated fibroblasts and M2-polarized macrophages synergize during prostate carcinoma progression," *Oncogene*, vol. 33, no. 19, pp. 2423–2431, May 2014.

[117] P. G *et al.*, "Analysis of Secretome Changes Uncovers an Autocrine/Paracrine Component in the Modulation of Cell Proliferation and Motility by c-Myc," *J. Proteome Res.*, vol. 10, no. 12, 2011.

[118] L. MC, M. EA, G. J, O. L, and S. R, "Bi-directional Protein Transport Between the ER and Golgi," *Annu. Rev. Cell Dev. Biol.*, vol. 20, 2004.

[119] R. C, "Pathways of Unconventional Protein Secretion," *Trends Cell Biol.*, vol. 27, no. 3, 2017.

[120] J. Kim, H. Y. Gee, and M. G. Lee, "Unconventional protein secretion – new insights into the pathogenesis and therapeutic targets of human diseases," *J. Cell Sci.*, vol. 131, no. 12, p. jcs213686, Jun. 2018.

[121] C. Rabouille, "Pathways of Unconventional Protein Secretion," *Trends Cell Biol.*, vol. 27, no. 3, pp. 230–240, Mar. 2017.

[122] P. JL, B. L, and K.-G. Y, "The Secretome in Cancer Progression," *Biochim. Biophys. Acta*, vol. 1834, no. 11, 2013.

[123] M. RA *et al.*, "Extracellular Remodelling During Oncogenic Ras-induced Epithelial-Mesenchymal Transition Facilitates MDCK Cell Migration," *J. Proteome Res.*, vol. 9, no. 2, 2010.

[124] S. AG *et al.*, "Extracellular Vesicles as Drivers of Epithelial-Mesenchymal Transition and Carcinogenic Characteristics in Normal Prostate Cells," *Mol. Carcinog.*, vol. 57, no. 4, 2018.

[125] P. Rawla, "Epidemiology of Prostate Cancer.," *World J. Oncol.*, vol. 10, no. 2, pp. 63–89, Apr. 2019.

[126] S. M. Falzarano and C. Magi-Galluzzi, "ERG protein expression as a biomarker of prostate cancer," *Biomark. Med.*, vol. 7, no. 6, pp. 851–865, Dec. 2013.

[127] G. Wang, D. Zhao, D. J. Spring, and R. A. Depinho, "Genetics and biology of prostate cancer," 2018.

[128] F. Müller *et al.*, "RnBeads 2.0: Comprehensive analysis of DNA methylation data," *Genome Biol.*, vol. 20, no. 1, p. 55, Mar. 2019.

[129] J. R. Wiśniewski, A. Zougman, N. Nagaraj, and M. Mann, "Universal sample preparation method for proteome analysis," *Nat. Methods*, vol. 6, no. 5, pp. 359–362, 2009.

[130] J. Rappsilber, M. Mann, and Y. Ishihama, "Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips," *Nat. Protoc.*, vol. 2, no. 8, pp. 1896–1906, 2007.

[131] A. E. Laria *et al.*, "Secretome analysis of hypoxia-induced 3T3-L1 adipocytes uncovers novel proteins potentially involved in obesity," *Proteomics*, vol. 18, no. 7, Apr. 2018.

[132] S. Tyanova, T. Temu, and J. Cox, "The MaxQuant computational platform for mass spectrometry-based shotgun proteomics," *Nat. Protoc.*, vol. 11, no. 12, pp. 2301–2319, Dec. 2016.

[133] O. Leshem *et al.*, "TMPRSS2/ERG promotes epithelial to mesenchymal transition through the ZEB1/ZEB2 axis in a prostate cancer model," *PLoS One*, vol. 6, no. 7, 2011.

[134] A. Ślusarz *et al.*, "Aggressive Prostate Cancer Is Prevented in ERαKO Mice and Stimulated in ERβKO TRAMP Mice," *Endocrinology*, vol. 153, no. 9, pp. 4160–4170, Sep. 2012.

[135] "Expression of steroid and peptide hormone receptors, metabolic enzymes and EMT-related genes in prostate tumors in relation to the presence of the TMPRSS2/ERG fusion | Experimental oncology." [Online]. Available: https://exp-oncology.com.ua/article/11260. [Accessed: 04-Mar-2021].

[136] G. D. Bader and C. W. V. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC Bioinformatics*, vol. 4, no. 1, p. 2, Jan. 2003.

[137] A. Naba, K. R. Clauser, S. Hoersch, H. Liu, S. A. Carr, and R. O. Hynes, "The matrisome: In silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices," *Mol. Cell. Proteomics*, vol. 11, no. 4, pp. 1–18, 2012.

[138] A. M. Socovich and A. Naba, "The cancer matrisome: From comprehensive characterization to biomarker discovery," *Seminars in Cell and Developmental Biology*, vol. 89. Elsevier Ltd, pp. 157–166, 01-May-2019.

[139] M. W. Pickup, J. K. Mouw, and V. M. Weaver, "The extracellular matrix modulates the hallmarks of cancer," *EMBO Rep.*, vol. 15, no. 12, pp. 1243–1253, 2014.

[140] D. M. Gilkes, G. L. Semenza, and D. Wirtz, "Hypoxia and the extracellular matrix: Drivers of tumour metastasis," *Nat. Rev. Cancer*, vol. 14, no. 6, pp. 430–439, 2014.

[141] S. H. Oram *et al.*, "A previously unrecognized promoter of LMO2 forms part of a transcriptional regulatory circuit mediating LMO2 expression in a subset of T-acute lymphoblastic leukaemia patients," *Oncogene*, vol. 29, no. 43, pp. 5796–5808, 2010.

[142] E. A. Simonik *et al.*, "LIM-only protein 4 (LMO4) and LIM domain binding protein 1 (LDB1) Promote growth and metastasis of human head and neck cancer (LMO4 and LDB1 in head and neck cancer)," *PLoS One*, vol. 11, no. 10, pp. 1–18, 2016.

[143] V. Karantza, "Keratins in health and cancer: More than mere epithelial cell markers," *Oncogene*, vol. 30, no. 2. NIH Public Access, pp. 127–138, 13-Jan-2011.

[144] Y. Xue, F. Smedts, F. M. J. Debruyne, J. J. M. C. H. de la Rosette, and J. A. Schalken, "Identification of intermediate cell types by keratin expression in the developing human prostate," *Prostate*, vol. 34, no. 4, pp. 292–301, Mar. 1998.

[145] F. Li *et al.*, "ERG orchestrates chromatin interactions to drive prostate cell fate reprogramming," *J. Clin. Invest.*, vol. 130, no. 11, pp. 5924–5941, Oct. 2020.

[146] S. N. Mapelli *et al.*, "A novel prostate cell type-specific gene signature to interrogate prostate tumor differentiation status and monitor therapeutic response (Running title: Phenotypic classification of prostate tumors)," *Cancers (Basel).*, vol. 12, no. 1, Jan. 2020.

[147] A. Nadu *et al.*, "Changes in collagen metabolism in prostate cancer: A host response that may alter progression," *J. Urol.*, vol. 166, no. 5, pp. 1698–1701, 2001.

[148] C. L. Hall, C. W. Dubyk, T. A. Riesenberger, D. Shein, E. T. Keller, and K. L. Van Golen, "Type I collagen receptor (α2β1) signaling promotes prostate cancer invasion through RhoC GTPase," *Neoplasia*, vol. 10, no. 8, pp. 797–803, 2008.

[149] C. Delliaux *et al.*, "TMPRSS2:ERG gene fusion expression regulates bone markers and enhances the osteoblastic phenotype of prostate cancer bone metastases," *Cancer Lett.*, vol. 438, pp. 32–43, Dec. 2018.

[150] C. L. E. Massie *et al.*, "New androgen receptor genomic targets show an interaction with the ETS1 transcription factor," *EMBO Rep.*, vol. 8, no. 9, pp. 871–878, Sep. 2007.

[151] A. S. Yang, M. R. H. Estécio, K. Doshi, Y. Kondo, E. H. Tajara, and J. P. J. Issa, "A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements.," *Nucleic Acids Res.*, vol. 32, no. 3, pp. e38–e38, Feb. 2004.

[152] T. Chen, Y. Ueda, S. Xie, and E. Li, "A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation," *J. Biol. Chem.*, vol. 277, no. 41, pp. 38746–38754, Oct. 2002.

[153] L. Yang, R. Rau, and M. A. Goodell, "DNMT3A in haematological malignancies," *Nat. Rev. Cancer*, vol. 15, no. 3, pp. 152–165, Feb. 2015.

[154] H. Cui *et al.*, "DNA methyltransferase 3A isoform b contributes to repressing E-cadherin through cooperation of DNA methylation and H3K27/H3K9 methylation in EMT-related metastasis of gastric cancer," *Oncogene*, vol. 37, no. 32, pp. 4358–4371, Aug. 2018.

[155] A. Portela and M. Esteller, "Epigenetic modifications and human disease," *Nat. Biotechnol.*, vol. 28, no. 10, pp. 1057–1068, 2010.

[156] M. Bibikova, "DNA Methylation Microarrays," in *Epigenomics in Health and Disease*, Elsevier Inc., 2016, pp. 19–46.

[157] A. Barski *et al.*, "High-Resolution Profiling of Histone Methylations in the Human Genome," *Cell*, vol. 129, no. 4, pp. 823–837, May 2007.

[158] B. E. Bernstein *et al.*, "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells," *Cell*, vol. 125, no. 2, pp. 315–326, Apr. 2006.

[159] Z. Wang *et al.*, "Combinatorial patterns of histone acetylations and methylations in the human genome," *Nat. Genet.*, vol. 40, no. 7, pp. 897–903, Jul. 2008.

[160] I. Gyory, J. Wu, G. Fejér, E. Seto, and K. L. Wright, "PRDI-BF1 recruits the histone H3 methyltransferase G9a in transcriptional silencing," *Nat. Immunol.*, vol. 5, no. 3, pp. 299–308, Mar. 2004.

[161] J. Yu, C. Angelin-Duclos, J. Greenwood, J. Liao, and K. Calame, "Transcriptional Repression by Blimp-1 (PRDI-BF1) Involves Recruitment of Histone Deacetylase," *Mol. Cell. Biol.*, vol. 20, no. 7, pp. 2592–2603, Apr. 2000.

[162] S.-T. Su, H.-Y. Ying, Y.-K. Chiu, F.-R. Lin, M.-Y. Chen, and K.-I. Lin, "Involvement of Histone Demethylase LSD1 in Blimp-1-Mediated Gene Repression during Plasma Cell Differentiation," *Mol. Cell. Biol.*, vol. 29, no. 6, pp. 1421–1431, Mar. 2009.

[163] L. Tao, G. Huang, H. Song, Y. Chen, and L. Chen, "Cancer associated fibroblasts: An essential role in the tumor microenvironment (review)," *Oncology Letters*, vol. 14, no. 3. Spandidos Publications, pp. 2611–2620,

01-Sep-2017.

[164] B. S. Carver *et al.*, "ETS genetic rearrangements do not initiate cancer in the prostate," *Nature*, vol. 457, no. 7231, pp. 1–3, 2009.

[165] D. D. Becker-Santos *et al.*, "Integrin-linked kinase as a target for ERG-mediated invasive properties in prostate cancer models," *Carcinogenesis*, vol. 33, no. 12, pp. 2558–2567, Dec. 2012.

[166] J. F. Bromberg, C. M. Horvath, Z. Wen, R. D. Schreiber, and J. E. Darnell, "Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon α and interferon γ," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 15, pp. 7673–7678, Jul. 1996.

[167] S. G. Ayub and D. Kaul, "miR-2909 regulates ISGylation system via STAT1 signalling through negative regulation of SOCS3 in prostate cancer," *Andrology*, vol. 5, no. 4, pp. 790–797, 2017.

[168] J. Zhang, F. Wang, F. Liu, and G. Xu, "Predicting STAT1 as a prognostic marker in patients with solid cancer," *Ther. Adv. Med. Oncol.*, vol. 12, p. 175883592091755, Jan. 2020.

[169] Y. Verhoeven *et al.*, "The potential and controversy of targeting STAT family members in cancer," *Seminars in Cancer Biology*, vol. 60. Academic Press, pp. 41–56, 01-Feb-2020.

[170] S. Ma *et al.*, "The significance of LMO2 expression in the progression of prostate cancer," *J. Pathol.*, vol. 211, no. 3, pp. 278–285, Feb. 2007.

[171] J. M. Matthews, K. Lester, S. Joseph, and D. J. Curtis, "LIM-domain-only proteins in cancer," *Nat. Rev. Cancer*, vol. 13, no. 2, pp. 111–122, 2013.

[172] Z. Xu, X. Meng, Y. Cai, H. Liang, L. Nagarajan, and S. J. Brandt, "Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins.," *Genes Dev.*, vol. 21, no. 8, pp. 942–55, Apr. 2007.

[173] R. R. Singh, C. J. Barnes, A. H. Talukder, S. A. W. Fuqua, and R. Kumar, "Negative Regulation of Estrogen Receptor A Transactivation Functions by LIM Domain Only 4 Protein," 2005.

[174] J. Wei, H. Lian, B. Zhong, and H.-B. Shu, "Phosphorylation of STAT1 Facilitates JAK1/2-Mediated Tyrosine Triggered Signaling Pathways That − γ Parafibromin Is a Component of IFN," *J Immunol Ref. /#ref-list*, no. 1, 2015.

[175] A. A. Mostafa *et al.*, "Activation of ERα signaling differentially modulates IFN-γ induced HLA-class II expression in breast cancer cells," *PLoS One*, vol. 9, no. 1, 2014.

[176] Y. Hou *et al.*, "STAT1 facilitates oestrogen receptor α transcription and stimulates breast cancer cell proliferation," *J. Cell. Mol. Med.*, vol. 22, no. 12, pp. 6077–6086, 2018.

[177] L. C. Platanias, "Mechanisms of type-I- and type-II-interferon-mediated signalling," *Nat. Rev. Immunol.*, vol. 5, no. 5, pp. 375–386, 2005.

[178] M. Budhwani, R. Mazzieri, and R. Dolcetti, "Plasticity of Type I Interferon-Mediated Responses in Cancer Therapy: From Anti-tumor Immunity to Resistance," *Front. Oncol.*, vol. 8, p. 322, Aug. 2018.

[179] L. M. Snell, T. L. McGaha, and D. G. Brooks, "Type I Interferon in Chronic Virus Infection and Cancer," *Trends Immunol.*, vol. 38, no. 8, pp. 542–557, Aug. 2017.

[180] J. S. Nair *et al.*, "Requirement of Ca2+ and CaMKII for Stat1 Ser-727 phosphorylation in response to IFN-γ," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 9, pp. 5971–5976, 2002.

[181] J. S. Brzozowski and K. A. Skelding, "The multi-functional calcium/calmodulin stimulated protein kinase (CaMK) family: Emerging targets for anti-cancer therapeutic intervention," *Pharmaceuticals*, vol. 12, no. 1, pp.

1–29, 2019.

[182] Y. Kanno, M. Watanabe, T. Kimura, K. Nonomura, S. Tanaka, and S. Hatakeyama, "TRIM29 as a novel prostate basal cell marker for diagnosis of prostate cancer," *Acta Histochem.*, vol. 116, no. 5, pp. 708–712, 2014.

[183] T. Yanagi *et al.*, "Loss of TRIM29 alters keratin distribution to promote cell invasion in squamous cell carcinoma," *Cancer Res.*, vol. 78, no. 24, pp. 6795–6806, 2018.

[184] T. McCray, D. Moline, B. Baumann, D. J. Vander Griend, and L. Nonn, "Single-cell RNA-Seq analysis identifies a putative epithelial stem cell population in human primary prostate cells in monolayer and organoid culture conditions.," *Am. J. Clin. Exp. Urol.*, vol. 7, no. 3, pp. 123–138, 2019.

[185] S. K. Choi *et al.*, "Epigenetic landscape change analysis during human EMT sheds light on a key EMT mediator TRIM29," *Oncotarget*, vol. 8, no. 58, pp. 98322–98335, 2017.

[186] A. Brossa, L. Buono, S. Fallo, A. Fiorio Pla, L. Munaron, and B. Bussolati, "Alternative strategies to inhibit tumor vascularization," *Int. J. Mol. Sci.*, vol. 20, no. 24, 2019.

[187] L. Yuan *et al.*, "RhoJ is an endothelial cell-restricted Rho GTPase that mediates vascular morphogenesis and is regulated by the transcription factor ERG," *Blood*, vol. 118, no. 4, pp. 1145–1153, 2011.

[188] A. V. Shah *et al.*, "The endothelial transcription factor ERG mediates Angiopoietin-1-dependent control of Notch signalling and vascular stability," *Nat. Commun.*, vol. 8, no. May, 2017.

[189] M. Bartoli *et al.*, "Vascular endothelial growth factor activates STAT proteins in aortic endothelial cells," *J. Biol. Chem.*, vol. 275, no. 43, pp. 33189–33192, 2000.

[190] N. Melling *et al.*, "Overexpression of enhancer of zeste homolog 2 (EZH2) characterizes an aggressive subset of prostate cancers and predicts patient prognosis independently from pre- and postoperatively assessed clinicopathological parameters," *Carcinogenesis*, vol. 36, no. 11, pp. 1333–1340, Nov. 2015.

[191] D. S. Rickman *et al.*, "Oncogene-mediated alterations in chromatin conformation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 23, pp. 9083–9088, Jun. 2012.

[192] M. Taris, J. Irani, P. Blanchet, L. Multigner, X. Cathelineau, and G. Fromont, "ERG expression in prostate cancer: The prognostic paradox," *Prostate*, vol. 74, no. 15, pp. 1481–1487, Nov. 2014.

[193] J. A. Joyce and J. W. Pollard, "Microenvironmental regulation of metastasis," *Nat. Rev. Cancer*, vol. 9, no. 4, pp. 239–252, 2009.

Appendix

Appendix I – Ethics Committee approval protocol n. 355384628

Experimental Design

[…] Samples of normal and cancer tissue will be obtained from radical prostatectomies of high-risk prostate cancers (Gleason ≥ 4+4) performed in the surgical unit of urology at Molinette Hospital (Torino). The samples will be transferred by the personnel of the urology unit, immediately after the resection, to the Anatomopathological unit of Molinette Hospital, where they will be clinically evaluated by the authorized pathologist. The normal and tumour specimens' areas of interest will be isolated in loco in the presence of personnel of the University of Insubria or the University of Torino and processed in the following 12 hours in the laboratory of Pathology (Busto Arsizio). Normal and tumour tissues will be maintained separately in culture dishes and differently processed to allow migration and isolation of the distinctive microenvironmental populations (fibroblasts, macrophages, etc.). A portion of both fragments (normal and tumour) will be stocked frozen for future analysis of the genetic, epigenetic and transcriptional status.

[…] The fibroblasts obtained after cultivation of the surgical fragments will be maintained in controlled culture conditions (37°C and 5% CO2) and expanded, and a frozen stock will be obtained. The same will be done for other microenvironmental populations. Fibroblasts will be characterized for their activation profile: normal fibroblasts (NAFs) and cancer-associated fibroblasts (CAFs) will be used to produce conditioned media (CM). CM deriving from NAFs and CAFs, containing specific cytokines and factors determined by the different activation, will be used for cultivation of cell lines derived from prostate cancer (e.g., PC3) and of normal and tumour 3D organoid cultures generated as described in aim 2.

Tumour progression and induction of EMT and metastatic processes will be evaluated through Western Blot, qPCR (for preliminary analysis) and RNA-seq to identify genes and non-coding RNAs (for a comprehensive analysis), while the underlying epigenetic alterations (in DNA methylation and histone modifications) will be evaluated through Pyrosequencing and ChIP (for preliminary analysis), and through DNAseq techniques and ChIP-seq (for a comprehensive analysis).

For in-depth study of the epigenetic signature, in the first step we will identify the enzymes responsible for DNA methylation alterations and histone modification patterns observed during tumour progression and EMT. Knockdown of factors and enzymes responsible for DNA methylation (e.g., DNMTs, UHRF1/2 and others) and histone modifications associated with DNA methylation (e.g., histone methyl transferases (HMTs), histone demethylases (HDM), histone acetyl transferases (HAT), histone demethylases (HDACs) and others) will be performed, testing whether they are needed during the transitions. […]

Methodologies

Prostate cancer cases will be chosen based on Gleason levels: only high-risk (> 4+4) prostatectomies will be considered. Immediately after resection in the surgical unit of urology, tissue will be transferred by the personnel of the urology unit to the Anatomopathological unit. Here it will be histologically analysed (through fixation and hematoxilyn-eosin staining) by the authorized pathologist for individuation of normal and cancer sections. A fragment of each portion will be resected in loco and immediately placed in culture medium at 4°C in the presence of personnel of the University of Insubria, Torino and/or Trento. The obtained fragments will be transferred in the laboratory of Pathology, University of Insubria and processed in the following 12 hours. For the isolation of fibroblasts, both tissue samples will be fragmented, and the pieces thus obtained transferred in cell culture plates and placed at a distance of approximately 0.5-1 cm in a central stripe. Sterilized glass slides will be put on the fragments to avoid spreading, and the plates will be stored in culture medium (DMEM + 20% FBS + different antibiotics) for 2-to-3 weeks to allow fibroblasts migration on the culture dish. Culture medium will be changed after the first day and every week. The obtained fibroblasts will be maintained in culture medium (DMEM + 10% FBS) and evaluated through Western Blot. Cells will be lysed, and protein levels will be analysed to exclude epithelial contamination (markers of epithelial cells e.g., E-cadherin and cytokeratin) and to observe the status of activation (through markers of normal/cancer activation e.g., α-SMA, FAP and vimentin). […]

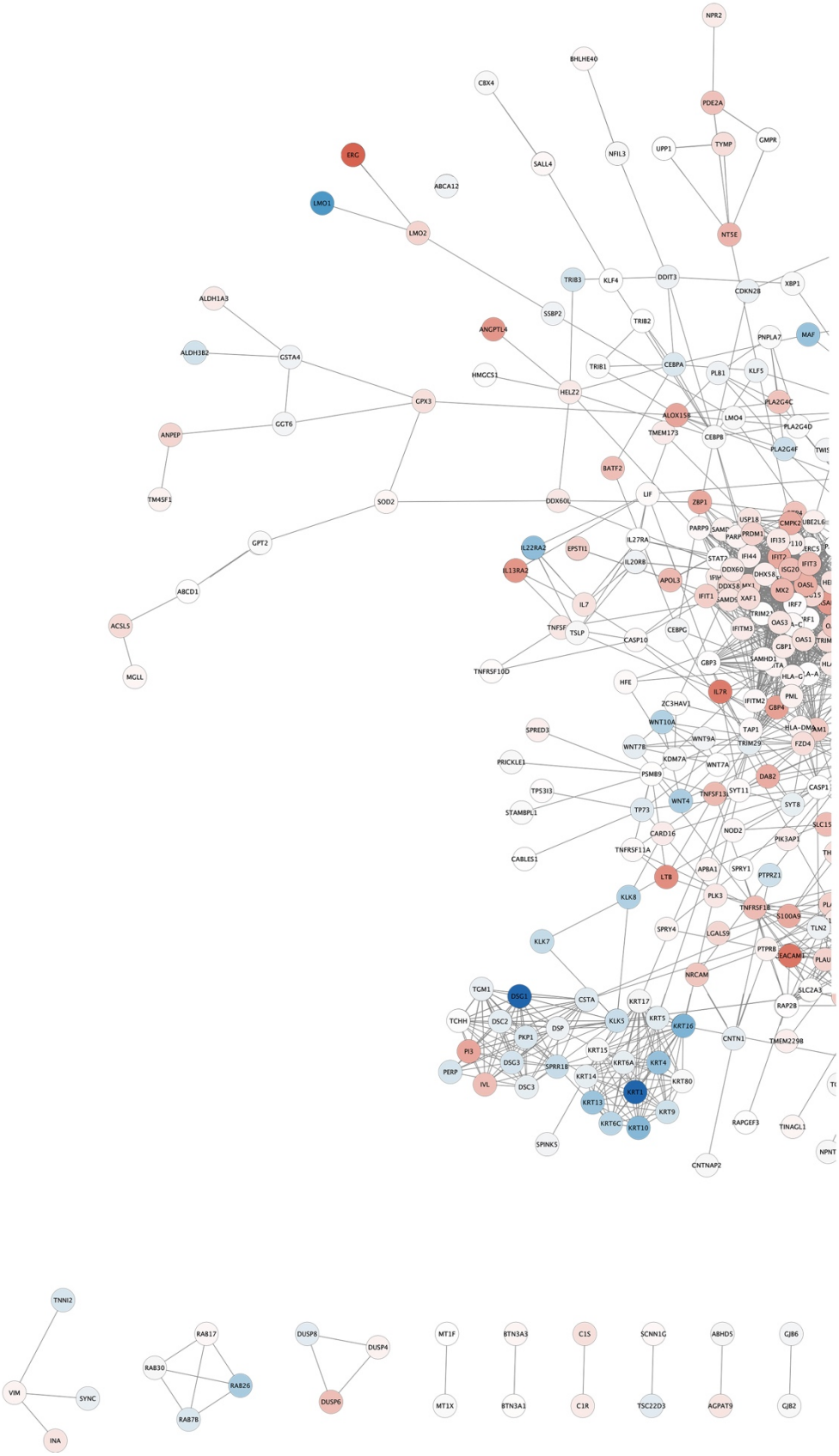
Fibroblasts presenting markers of normal activation (NAFs, positive for α-SMA and vimentin, negative for FAP) and fibroblasts presenting markers of tumour activation (CAFs, positive for α-SMA, FAP and vimentin) will be used to produce conditioned media (CM), by cultivating them in serum free medium (DMEM) for 72 hours. CMs will be clarified by centrifugation and used fresh or stocked at -80°C and used in the following week.

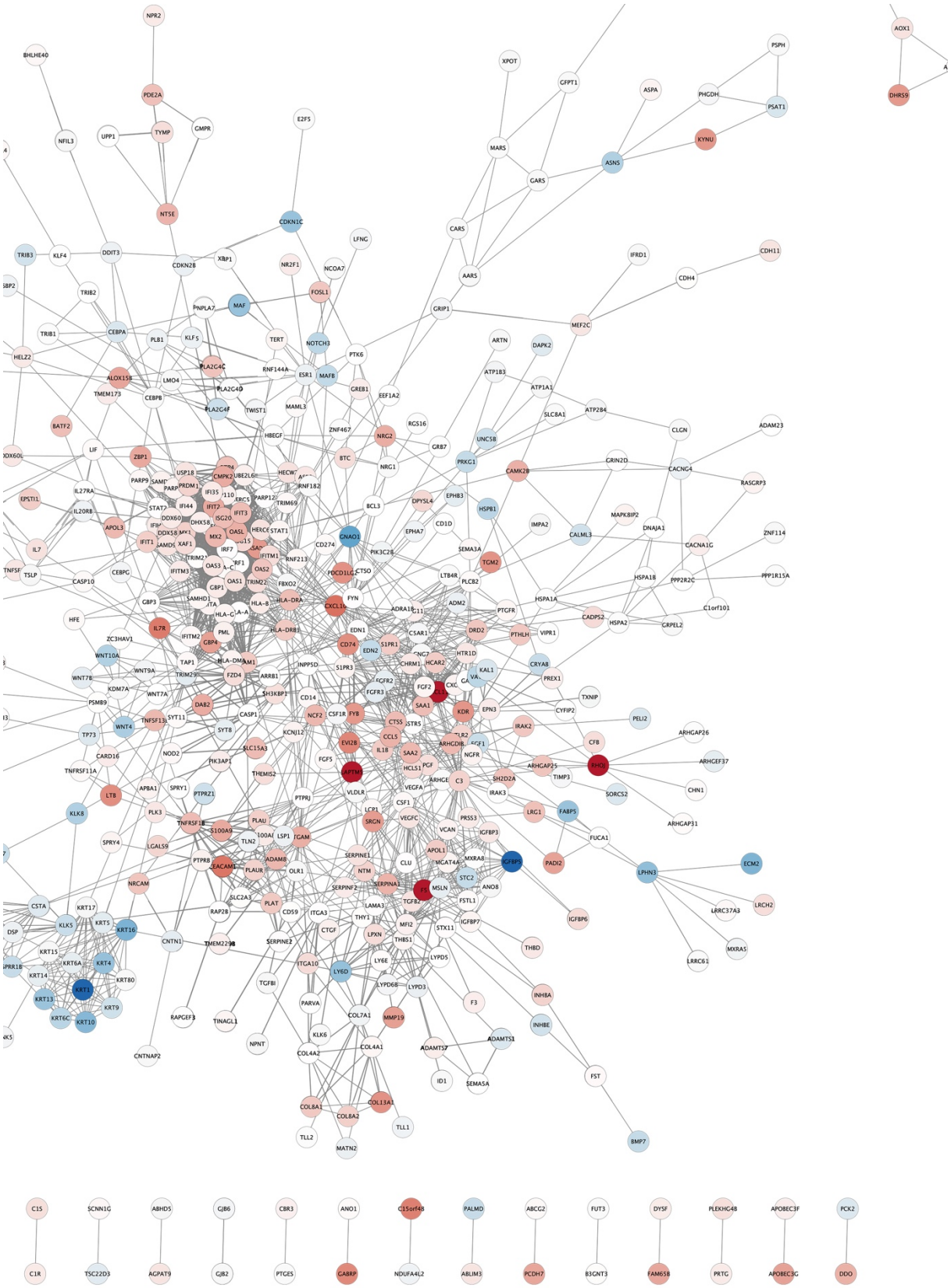
Appendix II – Primers

Gene ID	Forward	Reverse	Tm (°C)
ASPN	TTGAAGGGGTGACGGTGTTC	AGTGAAGCTCCAATAAAGTTGGT	57
CTGF	GCCTGCCATTACAACTGTCC	TGTCTCTCACTCTCTGGCTT	57
DNMT1	GAGCTACCACGCAGACATC	CGAGGAAGTAGAAGCGGTTG	57
DNMT3A Pan	CTCGAGTCCAACCCTGTGAT	CTTTGCCCTGCTTTATGGAG	58
DNMT3AA v3	GTGACACGCCAAAGGACC	CTTCTGTTCTTTGCCCGCTT	58
DNMT3AB v2	CCAAGACGGGCAGCTACTT	ATCACACTCGTCTTTCAGGC	58
DNMT3AD v6	AGAGTGAAGTATGCCTCCCC	GCTGCTCTTCCTGTCCCC	58
DNMT3B	AAGAGTTGGGCATAAAGGTAGG	GCTGGATTCACATTTGAGAGAT	57
E-CADHERIN	ATGAGTGTCCCCCGGTATCTTC	ACGAGCAGAGAATCATAAGGCG	57
ERG	AGACTTCCAAGATGAGCCCAC	CCCATCGATGTTCTGGAATAAC	57
EZH2	TTGTTGCGGAAGCGTGTAAAATC	TCCCTAGTCCCGCAATGAGC	57
G9A	CAAAATCGGGAACTTGGAGA	CTCGTTGTCAGTGAGGGTGA	57
GAPDH	GAGTCAACGGATTTGGTCGT	TTGATTTTGGAGGGATCTCG	57
GRHL2	AATACTCGGACTGGGGCATAGG	TAGGGCAGGACTGGCAAACAA	56
N-CADHERIN	GCAGAGACTTGCGAAACTCC	CACCATTAAGCCGAGTGATGG	57
PTEN	CACGACGGGAAGACAAGTTC	GGTTTCCTCTGGTCCTGGTA	57
SNAIL	CCCCAATCGGAAGCCTAACT	GACAGAGTCCCAGATGAGCA	57
UHRF1	CTGGTACGACGCGGAGAT	CGACAGTCGTTCAGAGAATCA	57
VIMENTIN	GAGTCCACTGAGTACCGGAG	ACGAGCCATTTCCTCCTTCA	57
YAP1	CGAACCCCAGATGACTTCCT	TCCATCTCCTTCCAGTGTTCC	57
ZEB1	CTGATTCCCCAGGTGGCATA	GGGCGGTGTAGAATCAGAGT	57
TARGETS			
CEACAM1	TACAGAACCCAGTGAGTGCG	GCAGGGTTAGAGGCTGCATA	57
IFI27	CTGGGAGCAACTGGACTCTC	AGAACCTCGCAATGACAGCC	57
IFI35	TGCAACAAAAGGAGCACACG	TGGCAGGAAATCCAGTGACC	57
ISG15	CTGCTGGTGGTGGACAAATG	TCAGCCAGAACAGGTCGTC	57
KRT14	CAGCTCAGCATGAAAGCATC	ACATCTCTGGATGACTGCGA	57
KRT5	ATGTCAAGAAACAGTGCGCC	GCTGCTGGAGTAGTAGCTTCC	57
OAS3	TTCCGCCTGACATCCGTA	AGACTTGTGGCTTGGGTTT	57
STAT1	TTCCGTGGACGAGGTTTTGT	GTTCCATTGGCTCTGGTGCT	57
FIBROBLASTS TARGETS			
aSMA	CCTGACCGCATGCAGAAAGA	ACCGAGTATTTGCGCTCCGA	57
FAP	CACCTGATCGGCAATTTGTA	TCCATTGCTAAGGTCATAGATG	57
FN1	GAGAATTCAAGTGTGACCCTCA	TGCCACTGTTCTCCTACGTG	57
FSP1	TTGGGGAAAAGGACAGATGAAG	TTCCTGGGCTGCTTATCTGG	57
PDGFRa	ATGAATCAGCCCAGATGGAC	TTCAACCACCTTCCCAAACG	57
PDGFRb	AATGTCTCCAGCACCTTCGT	CCAGTGAGGTTGGTCAGTGT	57
PDPN	TTGACAACTCTGGTGGCAACA	GCTGTGGCGCTTGGACTT	57
POSTN	TGCCCAGCAGTTTTGCCCAT	CGTTGCTCTCCAAACCTCTA	57
ChIP			
KRT5	GAAAACCCTGCAGCAACCT	GGGACTCTTACAGGACCAGG	57
KRT14	TTTCCAGCCGGGTATCCATC	TGCAGGCTACACTTTCCCAT	57

Protein	Company	Number	Dilution
DNMT1	Active Motif	39204	1:1000
UHRF1	IGBMC	54970	1:1000
DNMT3A1	Active Motif	39206	1:1000
DNMT3A2	Merck	54970	1:1000
DNMT3B	Active Motif	39899	1:1000
EZH2	CST	AC22	1:1000
GRHL2	Sigma Aldrich	HPA004820	1:1000
E-CADHERIN	BD Biosciences	610181	1:1000
VIMENTIN	GeneTex	GTX100619	1:1000
ERG	CST	A7L1G	1:1000
H3K27me3	Active Motif	39155	1:1000
H3K9me2	Active Motif	39753	1:1000
H3K9me3	Active Motif	39765	1:1000
H3K4me3	Active Motif	39159	1:1000
H3K12ac	Active Motif	39165	1:1000
N-CADHERIN	BD Biosciences	610920	1:1000
α-SMA	Sigma Aldrich	A2547	1:1000
GAPDH	CST	14C10	1:2000
H3K27ac	Active Motif	93133	1:1000
ERG_ChIP	Abcam	92513	1:200

Appendix IV –Network at 120 hours




NPR2
BHLHE40
PDE2A
NFIL3
UPP1
TYMP
GMPR
E2F5
NT5E
CDKN1C
TRIB3
KLF4
DDIT3
CDKN2B
XBP1
LFNG
NR2F1
NCOA7
FOSL1
TRIB2
PNPLA7
MAF
TRIB1
CEBPA
PLB1
KLF5
TERT
NOTCH3
HELZ2
PLA2G4C
RNF144A
ESR1
MAFB
PTK6
ALOX15B
LMO4
PLA2G4D
TMEM173
CEBPB
GREB1
EEF1A2
PLA2G4F
TWIST1
MAML3
BATF2
ZNF467
HBEGF
NRG2
RGS16
LIF
ZBP1
BTC
NRG1
GRB7
PRKG1
UNC5B
SLC8A1
ATP2B4
ATP1A1
ATP2B4
ARTN
DAPK2
AARS
IFRD1
CDH11
CDH4
GRIP1
MEF2C
CARS
MARS
GARS
ASNS
ASPA
PHGDH
PSAT1
PSPH
KYNU
AOX1
DHRS9
XPOT
GFPT1
CLGN
ADAM23
CACNG4
GRIN2D
CAMK2B
RASGRP3
MAPK8IP2
DNAJA1
CALML3
IMPA2
HSPB1
EPHB3
DPYSL4
BCL3
CACNA1G
ZNF114
PPP1R15A
PPP2R2C
HSPA1B
HSPA1A
HSPA2
GRPEL2
C1orf101
CADPS2
PTGFR
PTHLH
VIPR1
DRD2
ADM2
ADRA1B
LTB4R
PLCB2
CSAR1
CHRM1
HCAR2
HTR1D
KAL1
CRYAB
PREX1
TXNIP
CYFIP2
PELI2
ARHGAP26
CFB
ARHGEF37
CHN1
SORCS2
TIMP3
ARHGAP25
RHOJ
IRAK2
IRAK3
SH2D2A
ARHGEF
C3
VEGFA
PGF
HCLS1
NGFR
ARHGDIB
TLR2
CCL5
CTSS
CTSTR5
SAA2
IL1B
EVI2B
FGF5
PIK3AP1
THEMIS2
PTPN6
SLC15A3
NOD2
CARD16
TNFRSF11A
APBA1
SPRY1
PTPRZ1
LTB
KLK8
PLK3
TNFRSF1B
S100A9
S100A8
LSP1
TGAM
TLN2
PLAU
PLAUR
ADAM8
PTPRB
LGALS9
SPRY4
NRCAM
CSTA
KRT17
KRT5
KRT16
KLK5
DSP
KRT15
KRT6A
KRT4
KRT14
KRT80
KRT1
KRT13
KRT9
KRT6C
KRT10
CNTN1
TMEM229B
RAP2B
SLC2A3
PLAT
CD59
ITGA3
CTGF
SERPINE2
ITGA10
LY6D
TGFBI
PARVA
RAPGEF3
TINAGL1
CNTNAP2
NPNT
KLK6
COL4A2
COL4A1
COL7A1
MMP19
ADAMTS7
ADAMTS1
ID1
SEMA5A
F3
INHBA
INHBE
FST
BMP7
TLL2
MATN2
COL8A1
COL8A2
COL13A1
TLL1
LY6E
LYPD5
LYPD68
LYPD3
THBS1
LPXN
STX11
THBD
IGFBP7
IGFBP6
LRRC37A3
LRCH2
LRRC61
MXRA5
LPHN3
ECM2
PADI2
IGFBP5
MXRA8
MCAM
APOL1
VCAN
IGFBP3
PRSS3
VEGFC
CSF1
LCP1
SRGN
VLDLR
PTPRJ
FABP5
LRG1
FUCA1
ARHGAP31
SERPINE1
NTM
CLU
SERPINA1
SERPINF2
F2
MSLN
STC2
ANO8
FSTL1
TGFB2
LAMA3
MFI2
THY1
ITGA3
CD59
PLAT
SLC2A3
RAP2B
CNTN1
TMEM229B
CTGF
SERPINE2
ITGA10
LY6D
TGFBI
PARVA
RAPGEF3
TINAGL1
CNTNAP2
NPNT
KLK6
COL4A2
COL4A1
COL7A1
MMP19
ADAMTS7
ADAMTS1
ID1
SEMA5A
F3
INHBA
INHBE
FST
BMP7
TLL2
MATN2
COL8A1
COL8A2
COL13A1
TLL1
C1S
SCNN1G
ABHD5
GJB6
CBR3
ANO1
C15orf48
PALMD
ABCG2
FUT3
DYSF
PLEKHG48
APOBEC3F
PCK2
C1R
TSC22D3
AGPAT9
GJB2
PTGES
GABRP
NDUFA4L2
ABLIM3
PCDH7
B3GNT3
FAM65B
PRTG
APOBEC3G
DDO

Appendix V – Signature of Basal and Luminal genes

Basal signature	Luminal signature
ABI1	ABAT
ABLIM1	ACSS1
ATXN1	AP1M2
AVPI1	ARRDC1
CA12	ATP2C2
CDCA7L	ATP6V0E2
CENPL	AUH
CLCA2	BDH1
CRNDE	BSPRY
CSNK1E	C11orf80
CSTA	C19orf48
DENND2C	C8orf82
DEPDC7	CAB39L
DSC3	CD320
DUOX1	CGN
EDN1	CNDP2
ENAH	COBL
ENTPD7	CREB3L4
EPHA2	DHRS7B
ERRFI1	DMXL1
F3	DNAH5
F5	DPP4
FAM110C	EMB
FEM1B	ENPP5
FGD6	ENTPD6
FGFR2	ERBB3
FGFR3	ESRP2
FHL2	FAAH
FLRT3	FAAH2
FOXE1	FAM135A
FOXQ1	FAM136A
FRMD6	GLB1L2
FRMD6-AS1	GPRC5C
GABRP	GPT2
GAD1	GREB1
GADD45A	IDH2
GPR87	INTS12
GPX2	LAMP2
HBEGF	LCP1

HIC2
HOTAIRM1
HOXA1
HSPA4L
IFFO2
INPP1
ITGA6
ITGB6
JPH1
KANK1
KCNQ5
KRT14
KRT15
KRT16
KRT4
KRT5
KRT9
LIMA1
LNX2
LUZP1
MAST4
MEST
METAP1
MIR205HG
MPZL2
NDE1
NEDD9
NRG1
PCDH7
PEG10
PER2
PERP
PITPNM3
PKP2
PLCXD2
PLEKHA7
POLR1D
PPP1R13L
PPP1R14C
PROM2
PTPRZ1
RAP2B
MARVELD2
MCCC2
MIEN1
MRPL41
MRPS23
NCAPD3
NIPAL3
NUDT16L1
OAZ3
PCCA
PDE9A
PGM3
PLEKHH1
PPM1E
PPP2R2C
PRIM2
PRR15L
RAB17
RAB3B
RAMP1
RAP1GAP
REEP6
RGS17
RLN2
RSPH1
RWDD2A
SEC14L2
SERPINB6
SLC12A8
SLC15A2
SLC16A5
SLC31A1
SLC36A1
SLC52A3
SOCS2-AS1
STAP2
STYK1
SULT2B1
SYNE4
SYT7
TM7SF2
TMC4

RNASE7
RNF128
RNF39
SDC1
SDC4
SERPINB5
SGMS2
SLC2A1
SPATA18
ST3GAL5
ST6GALNAC2
STK17A
STON2
TAF1D
TFCP2L1
THAP9-AS1
THSD4
TIAM1
TP63
TRIM29
TUFT1
USP31
VCL
ZNF195
ZNF57
ZNF750
ZNF77
ZNRF3
TMEM125
TMEM79
TMEM87B
TPMT
TRIM36
TST
TTC39A
TTLL7
UNC13B
VPS25
YIPF1
ZG16B
ZNF350
ZNF613

Appendix VI - 77 Common genes (predicted) secreted

Gene Symbol	Gene Description
ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif 1
ADM2	Adrenomedullin 2
ANGPTL4	Angiopoietin like 4
APOL1	Apolipoprotein L1
ARTN	Artemin
BMP7	Bone morphogenetic protein 7
C1R	Complement C1r
C1S	Complement C1s
CCL5	C-C motif chemokine ligand 5
CD14	CD14 molecule
CLU	Clusterin
COL4A1	Collagen type IV alpha 1 chain
COL4A2	Collagen type IV alpha 2 chain
COL7A1	Collagen type VII alpha 1 chain
COL8A1	Collagen type VIII alpha 1 chain
COL8A2	Collagen type VIII alpha 2 chain
CSF1	Colony stimulating factor 1
CTGF	Connective tissue growth factor
CXCL10	C-X-C motif chemokine ligand 10
CXCL11	C-X-C motif chemokine ligand 11
ECM2	Extracellular matrix protein 2
EDN2	Endothelin 2
F3	Coagulation factor III, tissue factor
F5	Coagulation factor V
FGF1	Fibroblast growth factor 1
FGF2	Fibroblast growth factor 2
FGFR2	Fibroblast growth factor receptor 2
FST	Follistatin
FSTL1	Follistatin like 1
GAL	Galanin and GMAP prepropeptide
GPX3	Glutathione peroxidase 3
HBEGF	Heparin binding EGF like growth factor
IGFBP3	Insulin like growth factor binding protein 3
IGFBP5	Insulin like growth factor binding protein 5
IGFBP6	Insulin like growth factor binding protein 6
IGFBP7	Insulin like growth factor binding protein 7
INHBA	Inhibin beta A subunit
INHBE	Inhibin beta E subunit

ISG15	ISG15 ubiquitin-like modifier
KLK6	Kallikrein related peptidase 6
KLK7	Kallikrein related peptidase 7
KLK8	Kallikrein related peptidase 8
LGALS9	Galectin 9
MATN2	Matrilin 2
MGAT4A	Mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme A
MMP19	Matrix metallopeptidase 19
MXRA5	Matrix remodeling associated 5
NPNT	Nephronectin
NRG1	Neuregulin 1
NRG2	Neuregulin 2
PDCD1LG2	Programmed cell death 1 ligand 2
PGF	Placental growth factor
PLAU	Plasminogen activator, urokinase
PTHLH	Parathyroid hormone like hormone
PTPRZ1	Protein tyrosine phosphatase, receptor type Z1
S100A8	S100 calcium binding protein A8
S100A9	S100 calcium binding protein A9
SAA2	Serum amyloid A2
SEMA3A	Semaphorin 3A
SERPINA1	Serpin family A member 1
SERPINE2	Serpin family E member 2
SERPINF2	Serpin family F member 2
SPINK5	Serine peptidase inhibitor, Kazal type 5
TGFB2	Transforming growth factor beta 2
TGFBI	Transforming growth factor beta induced
TIMP3	TIMP metallopeptidase inhibitor 3
TINAGL1	Tubulointerstitial nephritis antigen like 1
TLL1	Tolloid like 1
TLL2	Tolloid like 2
TNFSF10	TNF superfamily member 10
TSLP	Thymic stromal lymphopoietin
VEGFA	Vascular endothelial growth factor A
VEGFC	Vascular endothelial growth factor C
WNT10A	Wnt family member 10A
WNT4	Wnt family member 4
WNT7A	Wnt family member 7A
WNT7B	Wnt family member 7B