



**A combined biological and bioinformatics approach
for the analysis of transcription factors regulated
by PDGF**

Ahmed ESSAGHIR

Thèse présentée en vue de l'obtention du grade de
Docteur en Sciences Biomédicales et Pharmaceutiques

Promoteur: Pr. Jean-Baptiste Demoulin

2010

École doctorale de Biochimie, Biologie cellulaire et Microbiologie
Université catholique de Louvain (UCL)
Institut de Duve

Promoter: Pr. Jean-Baptiste Demoulin

Jury

President: Pr. Fred Opperdoes

External members:

- Pr. Lodewyk Wessels, Netherlands Cancer Institute (NKI-AVL)
- Pr. Jacques van Helden, Université Libre de Bruxelles

Internal members:

- Pr. Stefan N. Constantinescu
- Pr. Laurent Knoops

This thesis was supported by:

- transMaze project from *la région Wallonne de Belgique*
- *Bourse du patrimoine de la faculté de médecine* (UCL)
- *Bourse d'excellence Wallonie-Bruxelles International* (WBI)

PREFACE

To Anir and Hanane,

“L’aveugle se détourne de la fosse où le clairvoyant se laisse tomber.” [Averroès]

“Quelque part, quelque chose d’incroyable attend d’être connu.” [Carl Sagan]

ABBREVIATIONS

AKT/PKB: protein kinase B (AKT)

ANOVA: analysis of variance

AP-1: activator Protein 1

DF: degree of freedom

Eval: e-value

ER: endoplasmic reticulum

ERK: elk-related tyrosine kinase

EST: expressed sequence tag

FAS: fatty acid synthase

FDR: false discovery rate

FGF: fibroblast growth factor

FOXO: forkhead box, sub-group O

FWER: family-wise error rate

GO: gene ontology

HMM: hidden markov model

HMG-COA: hydroxymethylglutaryl-Coenzyme A

bHLH: basic helix-loop-helix

IGF-1: insulin like growth factor 1

INSIG: Insulin induced gene

JAK: Janus kinase

JNK: c-jun N terminal kinase

MAPK: mitogen activated protein kinase

MITF: microphthalmia-associated transcription factor

MM: mismatch

MSE: mean square error

PCR: polymerase chain reaction

PDGF: platelet-derived growth factor

PDGFR: platelet-derived growth factor receptor

PI3K: phosphatidylinositol-3 Kinase

PFM: position frequency matrix

PM: perfect match

Pval: p-value

PWM: position weight matrix

Qval: q-value

RC: random control

RTK: receptor tyrosine kinase

SAM: significance analysis of microarrays

SCAP: SREBP cleavage-activating protein

SOX10: SRY (sex determining region Y)-box 10

SREBP: sterol-regulatory element binding protein

STAT: signal-transducer and activator of transcription

TF: transcription factor

TFBS: transcription factor binding site

TG2: transglutaminase 2

CONTENTS

Preface	i
Abbreviations	iii
List of figures	vi
List of tables	vii
1 Introduction	2
1.1 Gene expression mechanism overview	2
1.2 Platelet-Derived Growth Factor and regulated transcription factors	6
1.2.1 Platelet-Derived Growth Factor	6
1.2.2 Transcription factors regulated by PDGF	7
1.3 Measuring gene expression: microarrays	13
1.3.1 Overview on microarrays	13
1.3.2 Microarray manufacturing	13
1.3.3 Expression microarray experimental design	15
1.3.4 Microarray image acquisition and analysis	19
1.3.5 Normalization of microarray data	20
1.3.6 Statistical analysis of gene expression microarrays	25
1.3.7 Perspectives on gene expression measuring : RNA-Seq	28
1.4 Functional data mining in gene expression	30
1.4.1 Gene function, metabolic and signalling pathway mining	30
1.4.2 Transcription factor and regulatory network mining	31
2 Objectives	36
3 Results	38
3.1 Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data	38
3.1.1 Abstract	38
3.1.2 Introduction	38
3.1.3 Materials and methods	39
3.1.4 Results and discussion	43
3.1.5 Conclusions	55
3.2 The Transcription of FOXO Genes is Stimulated by FOXO3 and Repressed by Growth Factors	57
3.2.1 Abstract	57

3.2.2	Introduction	57
3.2.3	Materials and methods	58
3.2.4	Results and discussion	61
3.2.5	Conclusions	74
4	General Discussion	78
4.1	Gene expression induced by PDGF	78
4.2	TFactS tool	79
4.3	Predicted TFs regulated by PDGF receptor	80
4.3.1	STAT	80
4.3.2	SREBP	80
4.3.3	FOXO	80
4.3.4	Other transcription factors	81
4.4	TFactS potential applications	81
4.4.1	TFactS cancer screening	81
4.4.2	FOXO regulation in cancer	82
4.4.3	<i>De novo</i> TFBS predictions	84
4.5	TFactS amelioration	84
4.5.1	More regulations	84
4.5.2	Distinct species	84
4.5.3	Additional biological information mining	85
4.5.4	Alternative tests to Fisher's	85
	Bibliography	86
	Annexes	107
	Summary	118
	Résumé	119
	Acknowledgements	120

LIST OF FIGURES

1.1	Eukaryotic transcription mechanism	4
1.2	PDGF-induced signalling pathways and their effects.	7
1.3	A typical gene expression microarray experiment workflow.	14
1.4	Examples of typical cDNA microarray experimental designs.	17
1.5	cDNA microarrays flowchart for choosing normalization method.	22
3.1	Data integration in TFactS catalogues.	43
3.2	SOX10, MITF and JUN transcription factors are regulated in melanoma and associated to mutations in RAS and RAF.	50
3.3	STAT1 and STAT3 are phosphorylated in EOL1 cells and inhibited by imatinib.	51
3.4	Robustness analysis of TFactS.	52
3.5	Hierarchical clustering of STAT or SREBP target genes significantly regulated by PDGF-BB and b-FGF in human fibroblasts.	53
3.6	STAT phosphorylation is specifically induced by PDGF-BB.	55
3.7	FOXO expression is down-regulated by growth factors in human fibroblasts.	63
3.8	PDGF induces FOXO phosphorylation and translocation from the nucleus to the cytosol.	64
3.9	Regulation of FOXO target genes by growth factors in fibroblasts.	65
3.10	Activation of FOXO3 prevents fibroblast proliferation upon growth factor stimulation.	66
3.11	FOXO1-specific shRNAs enhance fibroblast growth.	67
3.12	Role of PI3 kinase in the FOXO repression by PDGF.	68
3.13	FOXO3 can induce FOXO1 and FOXO4 expression.	69
3.14	Comparison of the human, mouse and rat FOXO1 promoter sequences.	71
3.15	Identification of a FOXO-binding site in the FOXO1 promoter.	72
3.16	PDGF inhibits FOXO binding to the FOXO1 promoter.	73
3.17	Regulation of FOXOs by growth factors at the transcriptional and post-translational levels.	74
3.18	The proteasome inhibitor MG132 blocks the effect of PDGF on FOXO gene expression.	75
4.1	Transcription factor regulation by growth factors in fibroblasts as predicted by TFactS.	82
4.2	FOXO transcription factors could be regulated at different levels.	83

LIST OF TABLES

1.1	Predicted transcription factor counts by organism.	3
1.2	Gene expression microarray experimental designs.	18
1.3	Affymetrix GeneChip normalization procedure comparison	24
3.1	Example for contingency tables built under the three hypotheses tested by TFactS . .	44
3.2	TFactS validation.	47
3.3	TFactS comparison with other tools.	48
3.4	Analysis of genes regulated by growth factors in human fibroblasts using TFactS. . .	54

Introduction

CHAPTER 1

INTRODUCTION

1.1 Gene expression mechanism overview

The regulation of gene expression is fundamental to each living organism. Along this process, protein-coding genes are transcribed to mRNA, which are translated to the corresponding proteins. Transcription and its control in eukaryotes is a complex mechanism, which involves many well-identified molecules [191].

The general high-resolution chromatin structural findings describe DNA as wrapped around histone octamers forming nucleosomes [43]. Nucleosomes repress all genes except those whose transcription is brought about by specific positive regulatory mechanisms [191]. The N- and C-terminal tails of histones are flexible and contain several residues that can be subjected to post-translational modifications, such as acetylation of lysine residues [43]. Acetylation of the histone tails is a characteristic of transcribed chromatin. Co-activators, like CBP and p300, acetylate the histone tails while co-repressors deacetylate it. Acetylation can disable transcription repression by chromatin and deacetylation can enable repression [191]. Acetylation associated with nucleosome remodeling factors and chromatin accessibility complex, which can remodel the structure of chromatin in an ATP-dependent manner, contributes to the formation of nucleosome-free regions, known as “nuclease-hypersensitive sites.” These sites contain the enhancers, transcription factors and promoters of transcriptionally active genes [191]. The DNA sequences in and around promoters dictate the time, space and kinetics of gene transcription. A distinction can be made between: the core promoter, the proximal region to the core promoter, and the more distant enhancer sequences [123].

Transcription is mediated by three kinds of enzymes called RNA polymerases (Pol I, Pol II and Pol III). Pol II is responsible for all mRNA synthesis. It has many general factors, the most important being TFIIB, TFII-D, TFII-E, TFII-F and TFII-H [301]. They cover the core promoter sequences. In addition, promoter-proximal and the more distant sequences direct the binding of specific transcription factors, which execute their regulation predominantly through co-regulators, although they can interact directly with the general factors. Some of the co-regulators can interact directly with Pol II and the general factors influencing transcription. Others can reorganize nucleosomes or covalently modify chromatin, and change its structure. This can in turn influence transcription-factor associations and

the transcriptional status of Pol II [123].

Transcription factor proteins possess distinct DNA-binding and activation domains. The DNA-binding domains define several major transcription factor (TF) families. The affinity and specificity of DNA binding is regulated, by interactions with accessory proteins bound to adjacent DNA sites [191]. Based on their predicted DNA binding domains, TFs could be annotated and classified. DNA binding domains could be modeled using *Hidden Markov Models* (HMMs). HMMs are build from known domains and used to scan unknown protein sequences to infer a putative TF activity. Using this approach, the DBD database annotated TFs from all the transcripts, including alternative splicing, in many genomes. Table 1.1 summarized the number of predicted TFs from human, mouse and rat genomes [370]. DBD reported that HMMs from SUPERFAMILY database, which is a collection of those models, represent 37 superfamilies and 87 families of TFs according to the SCOP domain definitions [137, 370]. Gene ontology (GO) database reported 984 TF and 693 in the human and mouse genomes respectively [256].

Table 1.1: **Predicted transcription factor counts by organism.**

	number of Genes	number of TFs	Percentage of Genes predicted as TFs
<i>Homo sapiens</i>	22, 997	1, 508	6.56%
<i>Mus musculus</i>	23, 873	1, 426	5.97%
<i>Rattus norvegicus</i>	22, 503	1, 202	5.34%

Transcription regulation can implicate epigenetic modifications. Indeed, in addition to the histone chemical modifications, such as histone 3 methylation on lysine 9 and 27, the promoters can be silenced by cytosine methylation, especially in the CpG islands [81, 143]. Recently, epigenetic changes induced by small inhibitory RNAs (siRNAs) in gene promoters was shown in many organisms, including human [241]. These siRNA-induced repression of transcription is a longer-lasting, if not permanent mode of silencing gene transcription [143].

The three main phases of the transcription cycle summarized in Figure 1.1, are known as initiation, elongation and termination, and each stage is subject to regulation [301].

Transcription initiation: the TFIID complex recognizes the promoter and binds to the TATA box by its TATA-binding protein (TBP) subunit. The remaining subunits interact with flanking sequences and confer additional specificity. TFIIB positions the polymerase on the promoter. TFIIF includes ATP-dependent helicases that unwind the promoter around the start site to trigger the initiation of transcription. Another factor called Mediator interact with Pol II to form a holoenzyme complex. Mediator is an important interface between transcription factors and Pol II. It transduces regulatory information from enhancers to promoters in organisms ranging from yeast to human [191]. Enhancers are positioned at the more distant sites in chromatin and modify the rate of initiation complex forma-

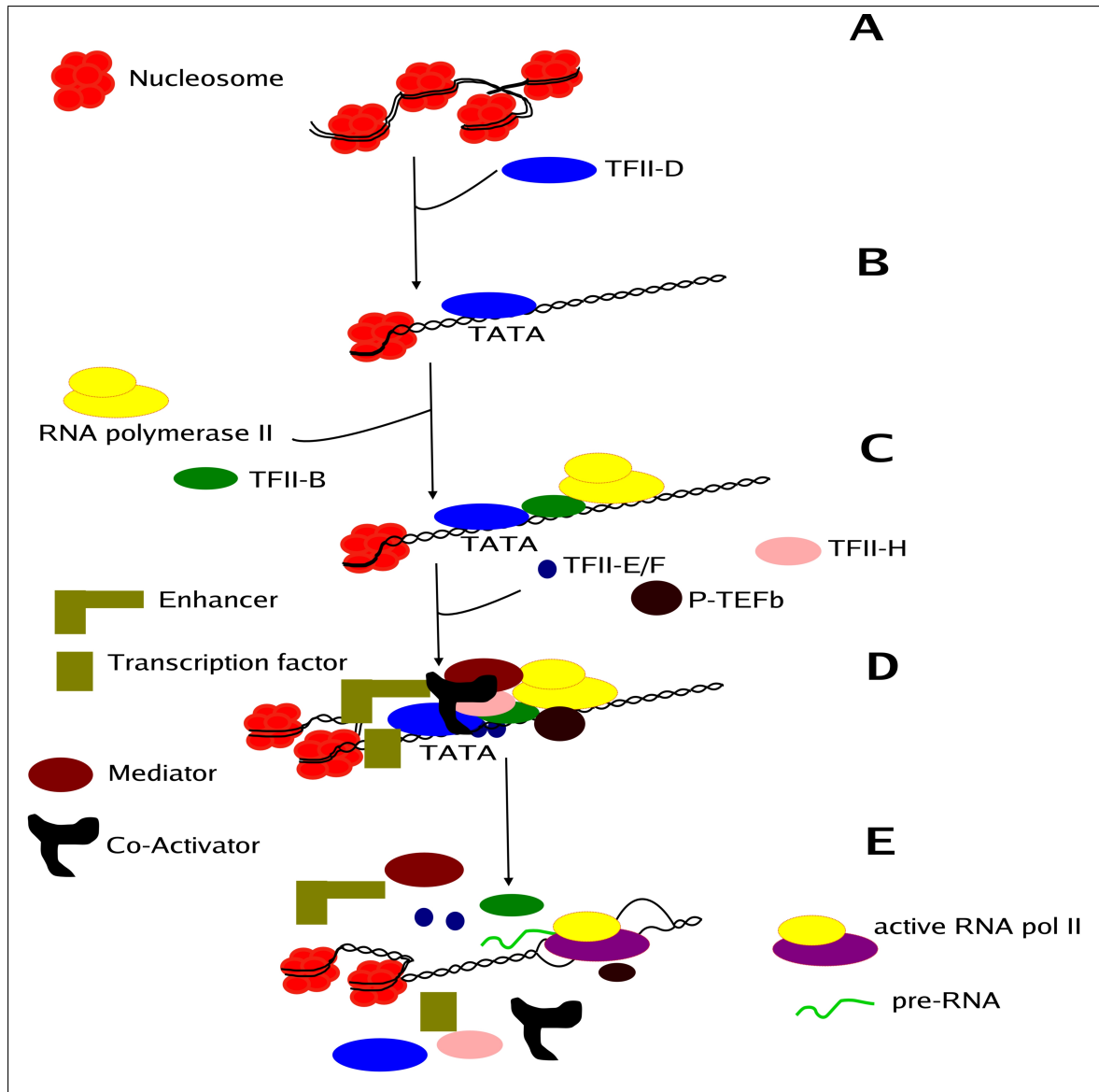


Figure 1.1: *Eukaryotic transcription mechanism.*

A- wrapped DNA within transcriptionally inactive chromatin. **B-** Histones are modified by acetyl transferases and chromatin is remodeled by nucleosome remodeling factors, this leads to TFII-D binding to TATA box. **C-** TFII-B binds to TFII-D and positions the RNA polymerase II on the promoter. **D-** Other TFII factors and the mediator form a complex with the TFII-D and the co-activator integrate the proximal transcription factor and the distant enhancer. P-TEFb (CDK9) joins the complex and together with the TFII-H phosphorylate the C-terminal domain (CTD) of the RNA polymerase II. **E-** CTD phosphorylation leads to disassociation of the complex and the elongation process of RNA polymerase II is started together with RNA transcription.

tion by mechanisms that are still poorly understood [139].

Transcription elongation: this step is controlled by P-TEFb (CDK9), which is a cyclin-dependent kinase that associates with TFII-H (CDK7) to phosphorylate two serine residues in the C-terminal domain (CTD) of the Pol II. These phosphorylations lead to the increase in diameter and rigidity of CTD. Thereafter, mediator and many other general initiation factors are removed from the promoter. Certain repressors block the recruitment of CDK9 to the transcription unit. The control of CDK9 is done via the regulation of its subunits expression, and by reversible associations to RNA binding proteins and a small nuclear RNA [267].

Transcription termination: the RNA transcript is released and the transcription complex dissociates from the DNA template. Two models have been proposed: (i) the “allosteric model,” which involves a conformational change in Pol II after the transcription of the poly-A site; (ii) the “torpedo model,” which involves a 5' → 3' RNA exonuclease that cleaves and releases the RNA transcript, and degrades the rest of the RNA associated with RNA polymerase II [301].

Gene transcription can also be regulated at the post-transcriptional level by RNA interference process. This gene regulation inhibitory mechanism involves double-stranded (dsRNA) or single-stranded (miRNA) small RNAs, which provoke mRNA degradation by homology binding to the mRNA 3' UTR. They also inhibit mRNA translation [143].

1.2 Platelet-Derived Growth Factor and regulated transcription factors

Gene transcription regulation in cells is influenced by many mechanisms, including extra-cellular stimuli, such as growth factors. These factors are secreted proteins that bind to specific membrane receptors, which include several families. Among these families, the Receptor Tyrosine Kinase (RTK) family is well studied. Receptors belonging to this family have an extra-cellular part that binds to the ligand and an intra-cellular part that contains a specific domain with a tyrosine kinase activity. These receptors are anchored in the plasma membrane by a single trans-membrane domain.

1.2.1 Platelet-Derived Growth Factor

Platelet-Derived Growth Factor (PDGF) is produced by platelets, macrophages, vascular endothelium cells, fibroblasts and keratinocytes [31]. It was shown to play a major role in cell proliferation and cell migration, in processes such as wound healing and angiogenesis [31]. It is also involved in the development of the cranial and cardiac neural crest, gonads, kidney, lung, intestine, skin, CNS, and skeleton. PDGF is implicated in many diseases, including gliomas, sarcomas, leukemias, epithelial cancers, atherosclerosis, restenosis, pulmonary hypertension, and retinal diseases, as well as in fibrotic diseases [14, 300].

PDGF is encoded by four different genes: A, B, C and D. Their products form homo-dimers, including PDGF-AA, PDGF-BB, PDGF-CC and PDGF-DD, and one hetero-dimer PDGF-AB. PDGF-C and D are synthesized as inactive precursors that are activated upon cleavage of their CUB domain [131, 197].

There are two known isomers of PDGF receptors: PDGFR- α and PDGFR- β . PDGF binding induces the receptor dimerization. *In vitro*, PDGFR- $\alpha\alpha$ homo-dimer can be formed upon PDGF-AA, PDGF-AB, PDGF-CC or PDGF-BB binding. PDGFR- $\alpha\beta$ hetero-dimer is formed upon the binding of PDGF-AB, PDGF-CC, PDGF-BB or PDGF-DD, while PDGFR- $\beta\beta$ homo-dimer is formed upon PDGF-DD or PDGF-BB binding. Thus, PDGF-BB is able to bind to the three types of receptor dimers [332].

When activated, the receptor autophosphorylates itself on tyrosines, which will serve as docking sites for proteins with specific domains, such as SH2 and PTB domains. This is the basis of signal transduction by PDGF receptors, including PI3K, Ras-MAPK and PLC γ pathways [332]. Activation of these pathways ends up, in most cases, by regulating transcription factors, which activate or repress transcription of gene sets leading to specific biological response. Chemotaxis induced by PDGF is independent of transcription. Specificity in PDGFR signalling is achieved through a combination of cell type-specific expression and differential engagement of downstream signalling pathways [14]. Figure 1.2 summarizes the main pathways induced by PDGF receptors and their biological effect.

These signalling pathways are balanced by feed-back control mechanisms. The ultimate effect of PDGFR depends on the balance between induced signals and the controls [146]. The control mechanisms include: (i) inhibitory molecules, which are usually phosphatases promoting dephosphorylation of the receptor and its substrates; (ii) receptor internalization and degradation [30, 61, 146, 147].

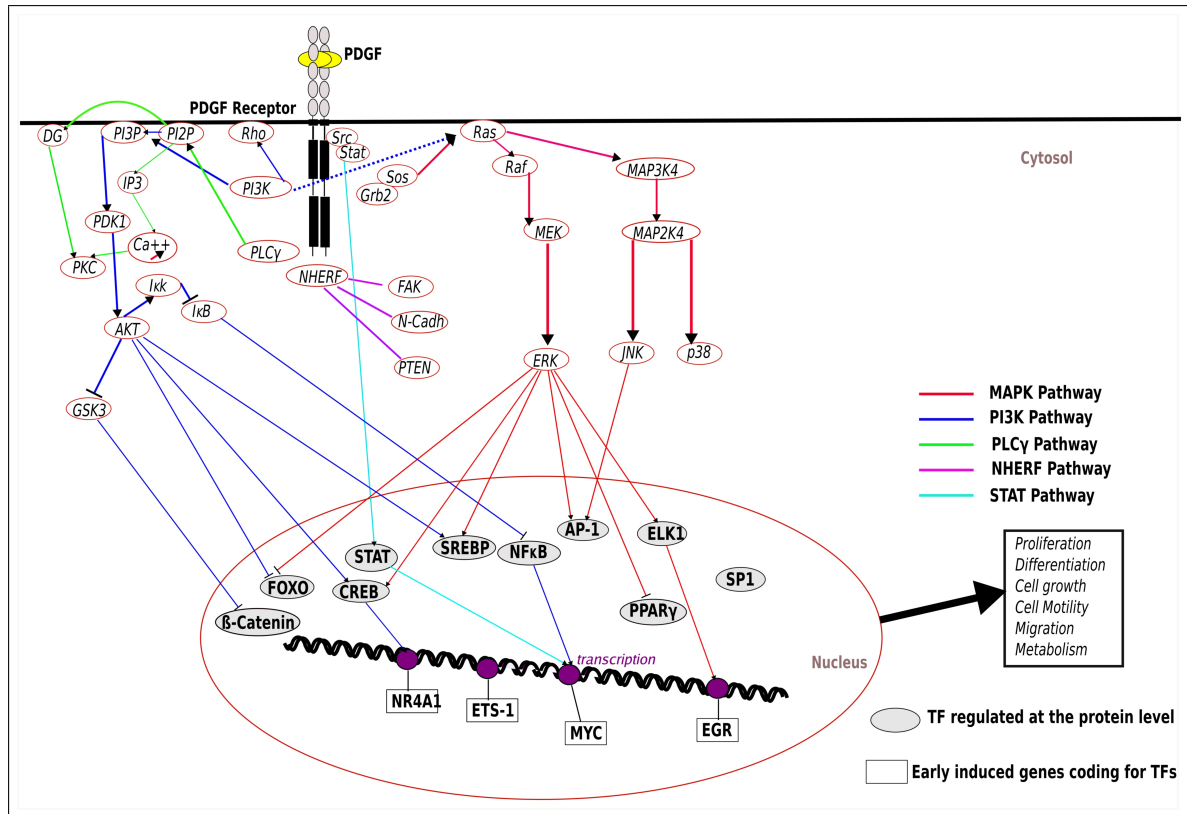


Figure 1.2: *PDGF-induced signalling pathways and their effects.*

1.2.2 Transcription factors regulated by PDGF

TFs regulated by PDGF receptors could be separated in two groups (Figure 1.2): (i) a group of TFs that are directly regulated at the protein level by post-translational modification; (ii) TFs regulated at the transcriptional level by inducing or inhibiting their gene transcription.

1.2.2.1 Transcription factors regulated at the protein level by PDGF

This category includes the following TFs: AP-1, STAT, FOXO, SREBP, NFκB, CREB, SP1, β-catenin and PPARγ TFs.

Activating complex Protein-1 (AP-1): the AP-1 TF is a direct target of MAPKs. AP-1 is composed of members of Jun and Fos proteins. The relative amount of expressed Jun and Fos members depends on cell types, which makes AP-1 activity cell-specific [231]. Fos is directly phosphorylated by ERK [236], while Jun is phosphorylated by JNK. Both Fos and Jun are also regulated by transcription [232]. Phosphorylated Jun and Fos are grouped together to form AP-1 complex and then bind to DNA to promote gene transcription. Target genes of AP-1 are essentially involved in proliferation and differentiation [15].

Signal Transducers and Activators of Transcription (STAT): the activation of STAT is well established for receptors lacking tyrosine kinase activity, such as interleukin receptors. In these receptors, Janus Kinases (JAKs) phosphorylate tyrosines of the receptor, leading to STAT binding via its SH2 domain. When STAT binds the receptor, it is then phosphorylated by JAK. This leads to STAT dimerization and migration to the nucleus where it binds DNA and stimulates gene transcription [127]. STAT-5 and -3 have been shown to promote proliferation, while STAT1 induces cell growth arrest [142, 190]. However, STAT3 can also have a tumor suppressor activity [104]. These contradictory roles of STAT could be related to a dose-dependent mechanism, as suggested by the findings that higher dose of active STAT5 promotes differentiation, while intermediate doses induce proliferation in CD34+ cells [367].

RTKs, including PDGF receptor, have also been shown to activate STAT. In this case, this activation seems to be JAK independent [198]. Since the discovery of STAT3 activation by the oncogenic v-Src [57], the debate between reports associating PDGF receptor-induced STAT activation to Src [46, 359] and those defending a direct activation of STAT by the receptor [261, 296] is still open. Nevertheless, JAK1 was shown to be phosphorylated by the PDGF receptor [64]. Simon *et al.* have reported an activation of JAK-STAT pathway by PDGFR in the human airway smooth muscle cells [314]. Moreover, in pancreatic stellate cells, this JAK-STAT pathway is reported to be dependent on Src activation by PDGFR [228]. In addition, Jahn *et al.* have recently proposed a model by which Kit receptor, another RTK of the PDGF receptor family, induces an indirect activation of JAK-STAT pathway via a direct interaction and activation of the IL-7 receptor [165]. Reactive oxygen species (ROS) were shown to induce JAK (JAK2 and TYK2) activation and mediate PDGF activation of STAT [313]. But this observation awaits confirmation.

Most reports of STAT activation by PDGF studied the PDGF receptor β . In an attempt to study the differential effects of PDGFR α and β , human mesangial cells were stimulated with PDGF-BB and PDGF-DD. These ligands were found to induce their biological activity almost exclusively via the PDGFR β , which was able to induce STAT3 and STAT1 activation [350]. In addition, JAKs were also shown to interact preferentially with the PDGFR β compared to the α receptor [347]. However, the hybrid oncogene FIP1L1-PDGFR α expressed in hypereosinophilic syndrome was shown to induce STAT5 activation [388].

Forkhead box transcription factors: Forkhead box (FOX) TF family consists at least on 43 members, and characterized by their winged-helix DNA-binding domain [179]. The group O of the FOX family, named FOXO, contains FOXO1, FOXO3, FOXO4 and FOXO6. FOXO were firstly studied in *Caenorhabditis elegans* and *Drosophila*, in which they were shown to be involved in development, lifespan longevity and oxidative stress resistance [154, 195, 207, 244, 255, 361, 393]. Chromosomal translocation involving the trans-activating domain of FOXO1 and DNA-binding domain of PAX3 were shown to be expressed in pediatric alveolar rhabdomyosarcoma [110], suggesting a role of FOXO in cancer.

FOXO are regulated at the protein level by RTKs. They were reported to be inhibited by acti-

vated insulin receptor in *Caenorhabditis* [255]. In *Drosophila*, this inhibition is achieved via dFOXO phosphorylation by dAKT [274]. This mechanism has been well studied in mouse hepatocytes, where Nakae *et al.* showed that Insulin and Insulin-like growth factor (IGF-1) receptors activate AKT, which phosphorylates a specific serine (S253) in the DNA-binding domain of FOXO1. This serine phosphorylation is required for two additional phosphorylations at T24 and S316. T24 phosphorylation is important for FOXO1 exclusion from the nucleus [247]. FOXO exclusion from the nucleus requires additional phosphorylations, interaction with protein 14-3-3 and the nuclear export machinery [349]. In the cytosol, FOXO interacts with the ubiquitin ligase SKP2 and is targeted to proteasome degradation [352, 353, 356]. PDGF, like many other growth factors was shown to inhibit FOXO in various cell types [109, 171, 187, 189, 246, 284, 366, 392]. PDGF receptors inhibit FOXO via the PI3K/AKT pathway, preventing apoptosis and cell cycle arrest [63]. It also induces phosphorylation of FOXO1 and its exclusion from the nucleus in Hepatic stellate cells [2]. Inhibition of AKT downstream PDGFR by specific inhibitors activates FOXO3 and increases apoptosis in vascular smooth muscle cells [385]. PDGFR was shown to repress p27, a FOXO target involved in cell cycle arrest [1].

Beside AKT inhibition of FOXO, JNK phosphorylates FOXO4 on two threonines T447 and T451. These phosphorylations are induced by oxidative stress. For instance, H₂O₂ activates Ral GTPase that activates JNK. This process provokes FOXO4 transactivation and translocation to the nucleus [113]. FOXO4 was also shown to be acetylated on lysines K186, K189, and K408 by CREB-binding protein (CBP). These acetylations lead to mitigated activation and sequestration in the nucleus of FOXO4 [124]. In *Caenorhabditis*, FOXO interacts with β -catenin, leading to its activation in an oxidative stress dependent condition [112].

Surprisingly, FOXO6, which is highly expressed in the brain, seems to be highly translocated to the nucleus after growth factor stimulation. In contrast to other FOXO members, this translocation is promoted by activation of the PI3K-AKT pathway [164]. Phosphorylation of FOXO by ERK and p38 was also shown downstream EGF receptor [1, 22]. ERK inhibits FOXO3 by direct interaction, phosphorylating Serines S294, S344 and S425. This leads to FOXO3 degradation via an MDM2-mediated ubiquitin-proteasome pathway [376].

Sterol Regulatory Element Binding Protein (SREBP): The SREBP family contains three proteins: SREBP-1a and SREBP-1c, which are encoded by two alternative transcripts of the SREBF1 gene; and SREBP2 is encoded by the SREBF2 gene. SREBP was first described as a TF involved in cholesterol biosynthesis. It regulates genes coding for many key enzymes in this process, such as HMG-CoA reductase, the rate limiting enzyme in cholesterol synthesis. SREBP regulates also the LDL receptor, which is involved in cholesterol uptake [51]. In addition, it is regulating fatty acid synthesis enzymes, such as Fatty Acid Synthase (FAS) and Stearoyl CoA Desaturase-1 (SCD) [51, 84].

SREBP is embedded in the endoplasmic reticulum (ER) membrane in an inactive form, where the N-terminal part that contains an acidic domain and the DNA-binding domain, is in the cytosol together with the C-terminal domain (regulatory domain), whereas a loop domain is in the ER lumen. SREBP is escorted from ER to Golgi by the SCAP protein. INSIG-1 and INSIG-2 proteins retain

the SREBP/SCAP complex in the ER in a sterol-dependent manner. To be active, SREBP must be cleaved in two sites, the first site is cleaved by Site-1 protease (S1P) in the luminal loop, and the second cleavage is made by S2P in the transmembrane domain linked to the N-terminal domain [99]. This leads to the release of the N-terminal part of SREBP, which translocates to the nucleus. Its bHLH-zip domain binds DNA E-boxes and SRE-boxes to activate target genes [51]. PDGF induced activation of SREBP was shown to be related to the MAPK pathway [193] and AKT pathway [84]. However, the detailed mechanism leading to activation of SREBP by growth factors is not known yet. AKT can phosphorylate SREBP *in vitro* (Kallin, Demoulin unpublished data) and promote its transcriptional activity [103].

Nuclear Factor of Kappa light polypeptide gene enhancer in B-cells (NF κ B): NF κ B is a transcription factor composed of five proteins: rel A (p65), rel B, c-rel, p52, and p50. These proteins can form either hetero or homodimers. The transcription factor NF κ B plays a role in inflammation, cell proliferation, and apoptosis [73]. NF κ B is over-expressed and mediates proliferation in gliomas in response to PDGF [319]. In its inactive form, NF κ B is sequestered in the nucleus by I κ B. PDGF signal induces I κ B degradation via PI3K pathway and leads to NF κ B translocation to the nucleus. PDGF-stimulated activation of NF κ B is mediated by transient association of Akt with IKK. This activates IKK, leading to I κ B phosphorylation on serines S32 and S36. Phosphorylated I κ B is targeted to ubiquitination and degradation via the proteasome pathway [288]. It is also shown that AKT phosphorylates IKK α on the threonine T23. This is required for IKK to phosphorylate p65 on S534 [26].

cAMP response element-binding protein (CREB): CREB was shown to be important for PDGF-induced cell proliferation and migration [166, 168]. PDGF-induced Ser-133 phosphorylation of CREB leads to its activation, translocation to the nucleus and subsequent binding to the CRE DNA boxes in target gene promoters [169, 251, 321, 331, 384]. This phosphorylation on S133 was not required for the activation of CREB by PDGF via RAF-MAPK pathway [310]. In addition, a chronic stimulation with PDGF has the opposite effect: it provokes CREB phosphorylation on serines S103 and S107 leading to its exclusion from the nucleus and its proteasomal degradation. This regulation is dependent on PI3K/AKT pathway [126].

Specificity Protein 1 / Trans-acting transcription factor 1 (SP1): SP1 is a zinc finger transcription factor involved, among many other roles, in the regulation of wound healing by growth factors, including PDGF-A and B genes [233, 280, 364]. Platelet-derived growth factor enhances SP1 binding to the promoters of LDL receptor gene [32], KFL gene [79] and VEGF gene [118]. Phosphorylated Rb could play a role in SP1 regulation by PDGF [240].

β -catenin: β -catenin was shown to be associated to NHERF, cadherin and the PDGF receptor β and mediates cell migration [337]. It regulates gene transcription associated with cell cycle progression. β -catenin is also a major mediator of Wnt signalling. It is inactivated in the cytosol upon phosphorylation

by GSK-3, forming a complex with adenomatous polyposis coli (APC) and axin. PDGF receptors induce the phosphorylation and degradation of GSK-3 leading to an increase in the level of β -catenin in the nucleus [253]. The phosphorylation of GSK-3 is mediated via PI3K/AKT pathway [119]. PDGF receptor could also phosphorylate the p68 RNA-helicase on tyrosine Y593. This facilitates β -catenin translocation to the nucleus by blocking GSK-3 and displacing Axin from β -catenin independently from Wnt signalling [144, 377, 378].

PPAR γ : Peroxisomal Proliferator-activated Receptor γ (PPAR γ) is a regulator of adipocyte differentiation. Many studies have reported interferences between the proliferative effect of PDGF and the activation of PPAR γ [20, 125, 129, 206]. Moreover, Camp *et al.* have shown that PDGF treatment decreased ligand-activated PPAR γ transcriptional activity, which was dependent on Serine (S82) phosphorylation via MAP kinase pathway [56].

ELK1: ELK-1 is activated via MAPK pathway. PDGF receptor activates MAPK pathway molecules, including Raf-1, MEK and ERK. Activated ERK translocates to the nucleus and activates ELK-1, inducing a proliferative response [269].

1.2.2.2 Transcription factors regulated at the RNA level by PDGF

Myc, FOS, JUN, ETS-1, N4RA and EGR are among TF genes that are immediate early genes (IEGs) induced by PDGF. IEGs are directly regulated by PDGF stimulation and do not need *de novo* protein synthesis.

MYC transcription factor: Myc mRNA transcription is induced by PDGF in a cell-cycle-dependent manner [182]. PDGF-activated Src kinases control the transcriptional activation of Myc [30, 60]. STAT activation by Src leads to Myc transcription [46]. In addition, NF κ B activated by PDGF induce the transcriptional activation of Myc via the PI3K pathway [288]. The RNA-helicase p68, which is phosphorylated and activated by PDGFR, participates in the transcription of Myc gene [379]. Myc plays an important role in cell proliferation and is over-expressed in many cancers [53]. Myc negatively regulates PDGFR β by repressing its gene expression [257].

Early growth response 1 (EGR-1): EGR-1, a zinc finger TF, is involved in duodenal ulcer healing, cell growth and proliferation [87, 186, 291, 292, 330]. It is highly expressed in atherosclerosis [101, 230]. PDGF induces the gene transcription of EGR-1 [243, 294, 295]. Early induced transcription of EGR-1 by PDGF is independent of PKC [290] and could be mediated by ELK1 via ERK1/ERK2 pathway [151, 157, 204, 220, 304]. Serum response element (SRE) consensus sequences, to which SRF binds, were found in EGR-1 gene promoter. These SRE boxes were shown to be important for the mediated activation of EGR-1 by PDGF [293].

ETS-1: ETS-1 is involved in neoangiogenesis [279]. It activates expression of matrix-degrading proteinases, such as collagenase. PDGF was shown to regulate ETS-1 gene transcription in different cell types [74, 132, 245, 279]. ETS-1 is upregulated by PDGF-BB via PI3K and mTOR pathways [214].

NOR1/NUR77/NR4A1: The NR4A1 orphan nuclear receptor mediates cell proliferation. NR4A1-induced gene transcription by PDGF occurs early and depends on cAMP response elements present in its promoter. This transcription involves Ca⁺⁺ mobilization, PKC activation, the MAP kinase pathway and CREB [227, 251].

1.3 Measuring gene expression: microarrays

1.3.1 Overview on microarrays

The availability of the full genome sequence of many species, including *Homo sapiens*, has revolutionized molecular biology. One of the most spectacular advances is the ability to screen the expression of all the genes in the genome using “gene expression microarray.” A typical gene expression microarray experiment consists of hybridizing labeled cDNA or cRNA obtained from mRNA extracted from a given biological sample to a chip containing probes for a whole or partial genome library. We can distinguish four parts in a typical gene expression microarray experiment (Figure 1.3): (i) extraction of mRNA and preparation of labeled probes; (ii) hybridization to the array containing probes for a referenced genome; (iii) scanning of the hybridized array; (iv) image analysis and transformation to intensities that will be the basis for statistical analysis to select regulated genes.

How these microarrays are made? How gene expression microarray studies could be planned? How the generated data are analyzed? And how to extract meaningful information from processed data in regard to regulated genes and their biological functions? This introduction will try to answer these questions.

There are other microarray techniques in the proteomic and genetic fields. They deal, for instance, with protein-protein and protein-DNA interactions, genotyping and DNA copy number. These kind of microarrays are out of this thesis scope.

1.3.2 Microarray manufacturing

Gene expression microarrays contained few spots or probes at the beginning and evolved to thousands nowadays [6, 303]. Each of these spots corresponds to a specific gene. Arrays can be either home made or purchased from private companies. They can be classified mainly into spotted or *in situ* synthesized, depending on the way the probes are produced and fixed to the array solid surface.

Spotted microarrays were the first to be developed. They are more commonly made in academic labs due to their lower cost and their ability to be customized to represent only desired transcripts. A computer-controlled robotic machinery spots the probes in well-identified XY coordinates onto the array surface coated with poly-L-lysine, aminosilane [145], alumina or gold [59]. The array could be slides made of glass [92], ceramic [219], plastic [391], FAST slides [323] or nylon membrane [216]. The spotted area of the whole genome is so dense that it could be easily covered by a simple cover slip [91]. The coordinates of each spot are stored in the system and could be easily mapped to the corresponding gene. The length of spotted oligonucleotides (probes) is ranging from 20 nucleotides [65] to the full-length cDNA [308]. A variant approach in which probes of 50 nucleotides are attached to beads prior to fixation into a silicon wafer surface is used in the Illumina® technology [160].

In contrast to the spotting technique, Affymetrix has developed an approach guided by *in situ* light-directed synthesis of the probes [120]. It is based on two techniques: photolithography and solid phase DNA-synthesis [209, 215]. The most commonly used Affymetrix arrays in gene expression

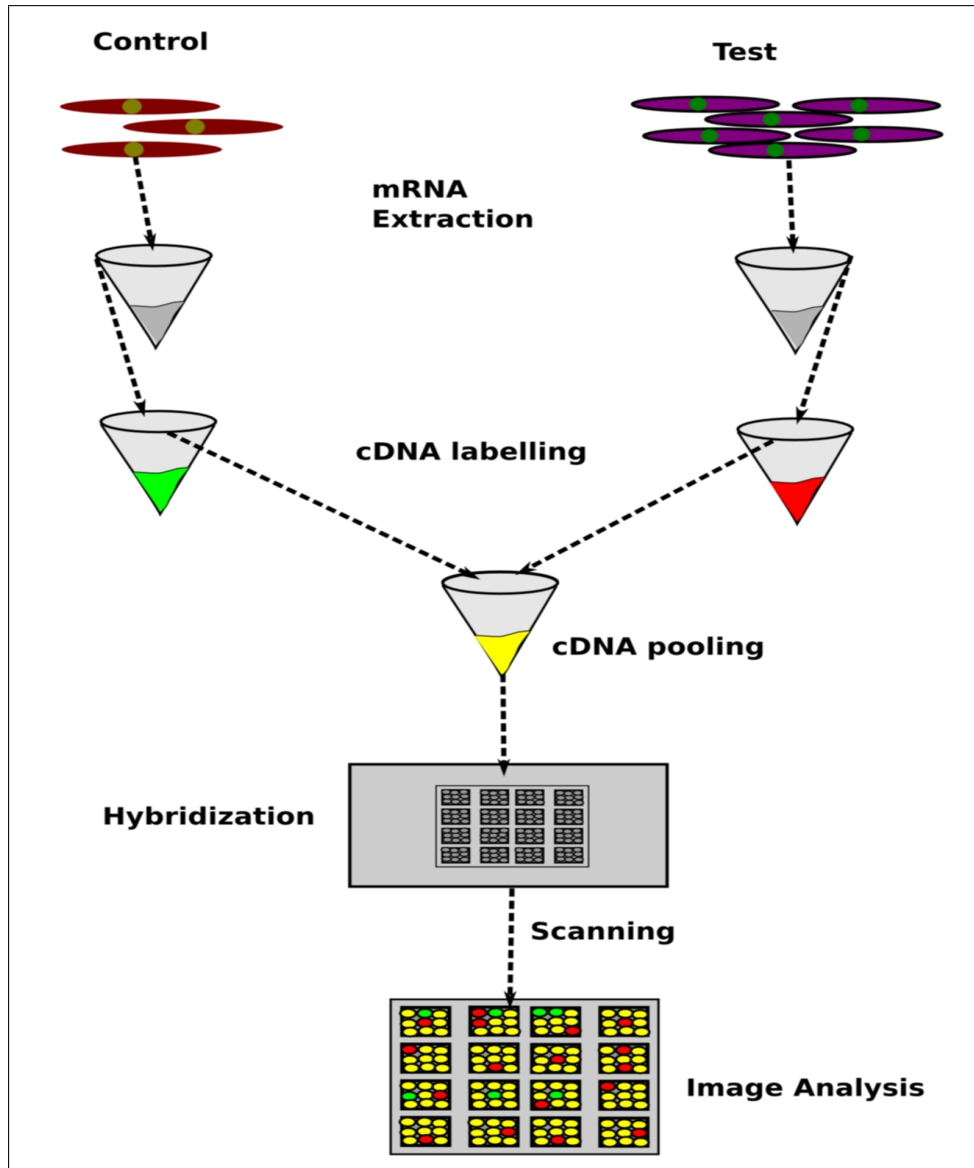


Figure 1.3: A typical gene expression microarray experiment workflow.

Gene expression from two conditions are compared. mRNA is extracted from each of the test and the control conditions. It is then reverse transcribed to the complementary DNA. This step include labeling with two different fluorescent dyes, usually cy5 and cy3. Both cDNAs are then mixed and hybridized to a microarray containing spots for the studied organism genome. After the hybridization, the slide is scanned in two different wave-lengths, exciting either cy-5 or cy-3. The two resulted images are merged. The spots colored in green and red correspond to down- and up-regulated genes, respectively. In the Affymetrix platform only one color hybridization is allowed per chip, therefore in this example there will be a need for two chips, one for the control condition and the other one for the test condition.

monitoring is the GeneChip[®] array that contains probes for ten of thousands of genes (e.g, 28,869 human genes are interrogated in HG-U1.0ST chip with 764,885 distinct probes). In GeneChip, each gene is referenced by one or more set of probes (probe sets). In each probe set, there are 16-20 (HG-U95a chips) or 11 (HG-U133 chips) pairs of 25 nucleotide probes representing perfect match (PM) and mismatch (MM) probe pairs to the target cRNA. The MM probe is identical to its corresponding PM except the nucleotide in the position 13 that is mutated to its complementary base [4, 6, 44].

1.3.3 Expression microarray experimental design

The objective of a cDNA microarray experiment is to monitor the gene expression in a particular condition at the genome level. Figure 1.3 summarizes the steps of a typical double-dyed microarray experiment. Double-dyed microarrays has the advantage to allow a direct comparison of two samples in the same chip. This was particularly important to compensate for irregularities in the array spotting, which was a problem for the first arrays. There are also single-dyed microarray experiments. The Affymetrix GeneChip[®] arrays, for instance, do not allow the use of double dyes and cRNA is biotin-labeled instead of cy3- or cy5-labeled prior to hybridization. The design of a microarray experiment to generate meaningful data is crucial [66]. To understand how to perform this design, one should keep in mind the sources of variation in a microarray gene expression study.

1.3.3.1 Sources of variation in a microarray experiment

In a typical microarray experiment workflow (Figure 1.3), we can distinguish three sources of variability: a biological variation (σ_B^2) intrinsic to the studied organisms; a technical variation (σ_A^2) introduced during the mRNA extraction, cDNA labeling and hybridization and a variation (σ_e^2) that could be associated to the measurement when scanning the fluorescence [66, 97]. The estimation of these variances is linked to a certain error. This error can be measured by the *mean square error* (MSE) criterion described in the Formula 1.1 for an experiment with n samples from k individuals (experimental units) and m technical replicates using microarray slides in which the clones are repeated r times. The constant $0 < a < 1$ is linked to the sample pooling, it is null in case of no RNA pooling was performed [66].

$$MSE = \sqrt{\frac{\frac{\sigma_B^2}{k^a} + \frac{\sigma_e^2/r}{m}}{n}} \quad (1.1)$$

The precision of the estimated MSE is measured by the experiment residual degree of freedom (df), which is obtained by subtracting the number of distinct treatments from the number of independent experimental units [66]. Minimizing the MSE and maximizing the df is a motivation for a good microarray experimental design [196].

In a large microarray gene expression profiling study, Bakay *et al.* found that the greatest sources of variability were the cell type content and the inter-patient variation while the experimental variation was minor [29]. Generally, biological replicates should be performed. Nevertheless, a common

technical source of variation in cDNA microarrays is the dye bias or unbalanced fluorescences over the two channels. Dye bias can depend on spot overall intensity and/or spatial location within the array [382].

1.3.3.2 Planning a microarray experiment

Designing a microarray experiment, in general, implies deciding on the number of biological and technical replicates, and on the distribution of samples over hybridizations [289]. Over the following sections, we use the word “class” to refer to a group of samples having biological characteristics that can be used to distinguish them among others. The experiment plan is guided by the study’s objectives that could be either a class comparison, a class prediction or a class discovery [315, 316]. It is also guided by the microarray platform to be used [383]. When using dual labeled arrays, for instance, one must decide on a design for pairing and labeling samples [316, 383]. In addition, there are constraints such as the number of available arrays, RNA quantity and other “logistical” considerations that can affect the experimental design [383]. This leads to different microarray experimental designs such as the “reference design,” the “loop design” and the “dye-swapping design” [66, 184] (Figure 1.4).

Table 1.2 enumerates the most commonly used dual channel microarray experimental designs with their characteristics and recommended applications [66, 96, 184, 289, 315, 316]. This table shows that, overall, the reference design can be used efficiently to reach most microarray study objectives such as class discovery and class prediction. This design uses a large number of arrays and is less suitable for class comparison [97, 383]. However, Altman et al. proposed an extended loop design as a model of choice in dual channel microarray experiments [11]. A web-based tool was developed to assess the efficient loop design for such experiments [270].

Since Affymetrix arrays are single-label based, they are less subject to dye bias as dual color microarrays. However, precautions must be taken when performing technical replicates, by using the same technical material and reagents by the same technician in the same lab conditions [25]. Affymetrix standard protocols were designed to reduce technical variability limiting the level of replicates to be controlled at the biological level. Affymetrix propose to control the number of replicates at the *coefficient of variation* (cv) level by introducing more replicates until the cv stabilizes. They also introduce a control at each critical step to control variability, such as RNA quality, *in vitro* transcription (IVT) yield, hybridization, and data quality controls [3].

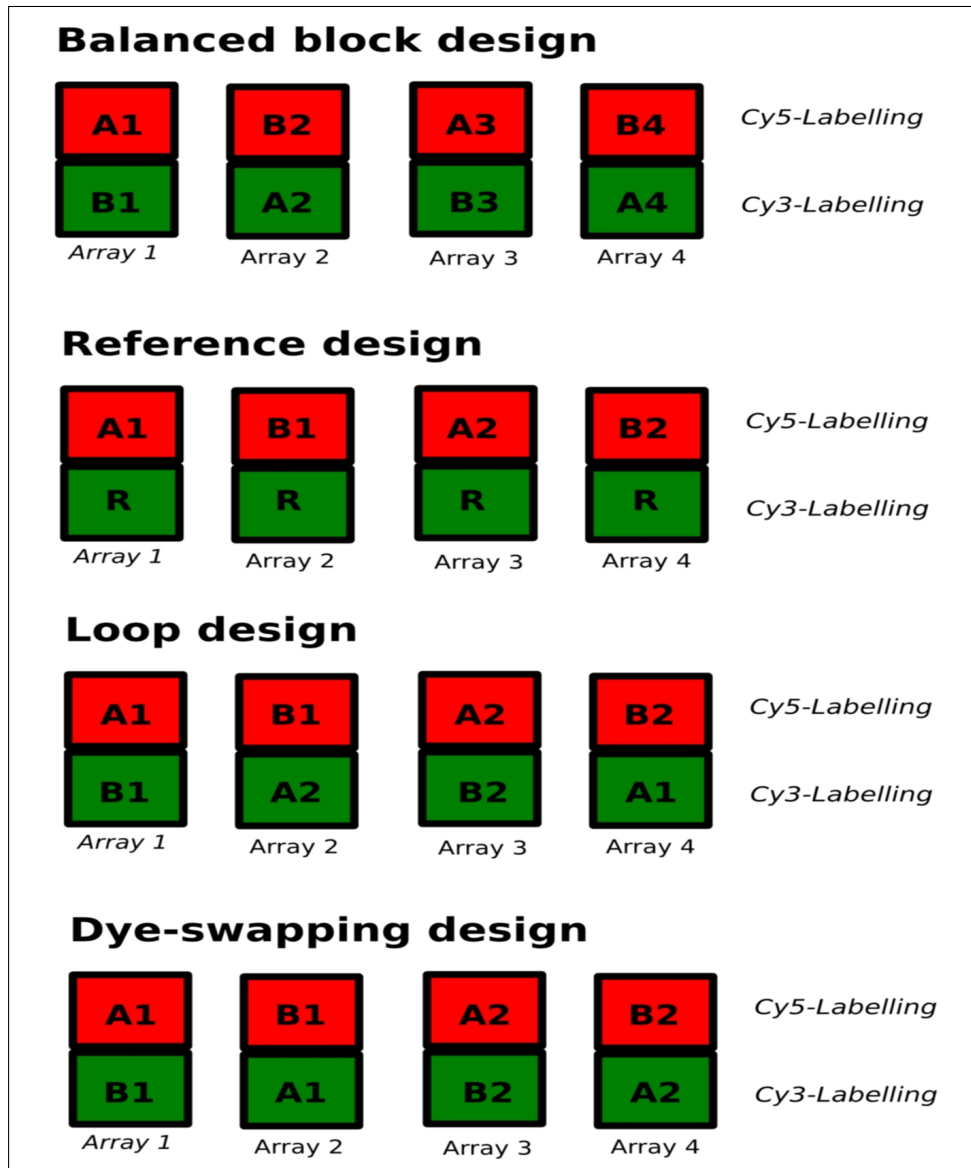


Figure 1.4: *Examples of typical cDNA microarray experimental designs.*

A and B denote different biological classes of the study's interest (e.g., tumors, treatments). From these classes, samples were derived (A1, A2, ... B3, B4). R denotes a reference RNA. Cy-5 and Cy-3 are represented in red and green colors, respectively. In each array, two samples are hybridized one from the class A or R and the second from the class B or R. cDNA from samples to be hybridized together are differentially labeled one with cy5 and the other one with cy-3.

Table 1.2: Gene expression microarray experimental designs.

Experimental Design	Characteristics	Limitations	Applications
Balanced block design	- Each sample from each class is directly compared to one in the other class. - No reference RNA is needed. - One RNA aliquot per sample. - The class samples are equally labeled through the arrays by both dyes. - Efficient in use of arrays. - Spot-to-spot effect is eliminated - Most effective in a single type of class comparison.	- Non efficient in cluster analysis. - Less effective when large inter-sample variability is present. - Cannot compare more than two classes.	- 2 class comparison.
Reference design	- Common reference cDNA hybridized to each array and is labeled with the same dye across all arrays. - Relative comparison of each gene in the tested sample to its corresponding in the reference.	- Indirect comparison. - Half of the hybridizations used for the reference. - Dye and treatment effects confounded. - Any subset of samples could be compared to any other one.	- Class discovery. - Class prediction. - Class comparison.
Loop design	- Samples are compared one to another in a loop chain pattern. - Needs two aliquots of RNA from each class sample. - Each aliquot is labeled by different dye and hybridized to the other class sample aliquots. - Dye and treatment effects uncounfounded. - Small loops give good average precision. - Additive robustness if combined with reference design or with another loop design. - Collect more data than the reference design using the same number of arrays.	- Doubles the number of the labelling reactions. - Efficient for less than 10 samples. - Less efficient than reference design in cluster analysis. - Less efficient than balanced block design in class comparison. - Less robust against the presence of bad quality arrays. - Deconvolution of relative expression values is not always intuitive.	- Class comparison. - Class discovery.
Dye-swapping design	- Each hybridization is done twice with the dye assignment reversed in one of the hybridizations. - Reduces technical and biological variation when performed on sample replicates. - Reduce dye bias. - Adequate for direct comparison hybridizations.	- Doubles the number of hybridizations. - Less important when a reference RNA is used.	- Class comparison.

1.3.3.3 The number of replicates in a microarray experiment

Replicates in microarrays are crucial, since they reduce variability and allow a formal statistical analysis of obtained data [383]. The level (biological or technical) on which replicates are made depends on the goals of the study [96, 316]. A simple rule of thumb method to assess the efficiency of the number of replicates in a microarray study is to determine the degree of freedom (df), which should be above or equal to 5 [66]. On the basis of the Formula 1.1, Churchill suggested to add more biological replicates when the variability of the measurement is lower than the variability between individuals (experimental units) and to add more technical replicates when it is greater. When the biological variability is large, it is worth pooling samples [66]. For rigorous statistical approach, one should use the power calculation to determine the number of biological or technical replicates needed in the experiment [383].

1.3.4 Microarray image acquisition and analysis

Once microarrays are hybridized and scanned, the next step is to analyze the obtained images for spot recognition, segmentation and extraction of intensities [201]. The goal of microarray image analysis is to extract spot intensity. This intensity represents an estimator of gene expression level. techniques used to analyze microarray images aim to extract as much as possible data for genes expressed at the low level while avoiding the influence of saturated intensities.

Glatigny *et al.* [134] proposed a general approach for scanning. It consists on averaging two, or at most three, acquisitions of microarrays scanned at moderate photomultiplier settings, they showed that this was significantly improving the quality of the data. They provide free software to align and average obtained images in their web site.

The spot recognition or gridding is done either manually or via fully automated softwares and aims to mark all bad quality spots [27, 48, 52, 130, 133, 387]. The segmentation helps to differentiate background from foreground intensities, it relays on several algorithms like *fixed circle*, *adaptive circle*, *adaptive shape* and *histogram* segmentation [381]. This is a subject of intense research and new algorithms are developed each year [23, 24, 180]. In a comparison study, Daskalakis *et al.* [76] found that *Histogram Concavity* algorithm was performing better than others, especially in low quality images.

Background intensity subtraction is done either by using the median or the mean background intensity of each spot [145] or by using the overall array background intensity or by applying a background adjustment filter called *morphological opening (MO)* [381]. At the end of this step, saturated spots and those with low intensities are filtered or marked to be discarded from any further analysis. A statistical filter consisting on the elimination of all spots with intensities lower than the background intensity plus 2 times its variance is recommended [201, 276].

In Affymetrix GeneChip array image acquisition, grid alignment and refinement is aided by checkerboards, which are squares (4×4 each) of alternating dim and bright probe cells, and are represented in each corner of the array hybridization surface. Each probe cell is mapped to an array of

pixels in a resolution of 64 (8×8) pixels. Pixels on the perimeter of the probe cell are discarded to avoid grid misalignment artefact. Then the array of pixels will represent only the center of the probe cell. From these pixels, the 75th percentile is reported as the equivalent intensity of the probe [19, 44]. Another variant is used in the dChip software: after alignment, the probe cell pixels with high intensities are masked in an iterative process. The set of pixels for which the removal highly reduces the coefficient of variation of the remaining pixels are removed from the pixel array. This pixel array is used to estimate the probe cell intensity. The remaining pixels should not be below a (4×4) array. The mean of the selected pixels is reported as the intensity of the corresponding probe [302].

The recorded pixel value, for each probe is stored in 16 bits, which means that for saturated pixels the intensity will be truncated to fit in the maximum value 65,536 ($= 2^{16}$). To detect defects in GeneChip images, a comparison with a reference value is done for each probe. The reference value could be defined either by the average value from replicates, or from other sources like GEO or using a single array by an adaptation of z statistics [19].

1.3.5 Normalization of microarray data

After image acquisition and analysis, data are ready to be processed and the obtained intensities are assumed to represent the expression level of each arrayed gene. This processing is dependent on the study's objectives and usually involves normalization of raw data before any downstream analysis [276].

We will refer in this section to the intensity of a spot (i) by R_i and G_i for its intensities in red and green channels, respectively; to the number of genes used in the normalization by N ; and to ϵ as the error associated to each of the normalization fitted models. ϵ is assumed to be independent and identically distributed [163]. In single-dyed microarrays, the red and green could be replaced by two different arrays.

Microarray normalization, in principle, is similar to the one used in quantitative reverse transcription PCR (RT-PCR), where measured expression levels were adjusted relatively to the expression of one or more reference genes whose levels are constant between samples [276]. The process of normalization aims to make hybridized genes and microarrays comparable by removing systematic variations affecting them, such as, dye bias effect [163, 201, 382]. Dye bias is defined as a difference in intensities attributable to the dye effect instead of gene effect. This bias is hard to eliminate without appropriate experimental design and normalization and can affect result interpretations as suggested by a large study from Dobbin *et al.* [98]. Systematic sources of variation are reported in both oligonucleotide microarray (Affymetrix GeneChips) and cDNA microarray data [374]. Dudoit [102] proposed to use the MA-plot, which presents $M(= \log_2(\frac{R_i}{G_i}))$ versus $A(= \log_2(\sqrt{R_i \cdot G_i}))$, as a method to visualize systematic effects in microarray data. If only a small number of genes is differentially expressed ($\log_2(\frac{R_i}{G_i}) \sim 0$, for the majority), it is expected to see a horizontal cloud around zero [140].

Microarray normalization methods could be divided into global normalization, linear and non-

linear normalizations [40, 201, 260, 374]. These methods could either use the whole spots in the genome or a partial set of spots [276]. All of these methods could be modeled by either global linear models, such as regression models [260] and ANOVA models [185], or linear mixed models [140]. Flowcharts for choosing adequate normalization methods were suggested (Figure 1.5) [140, 260]. The normalization could be done for each chip separately. This is called within-chip normalization. Between-chip normalization is performed by taking into account all the chips in the experiment. *Scalarization* of microarray data might be required after within-chip normalization for a consequent comparison among different experimental arrays to adjust for dyes unbalance [116, 276].

1.3.5.1 Set of genes used in normalization

Two types of genes can be used to normalize microarray data: spike-in genes or house keeping genes. Spike-in mRNA controls are generally from bacterial origin, purified and commercially available. They are mixed to mRNA samples prior to cDNA synthesis and labeling. The cDNAs of these spike-in controls are arrayed in the used chip. They are assumed to have no similarity with the studied genome [116]. House keeping genes are a set of genes that are presumed not to be differentially expressed between conditions [102]. Normalization using external spike-in controls is shown to perform better than house keeping genes, because of variability in the expression of the latter [382]. Instead of house keeping genes and external spike-in controls, Yang *et al.* proposed the use of microarray sample pool (MSP), which consist on arraying serial dilutions of pooled ESTs representing all the genes in the slide [382]. Nevertheless, some authors argued against using a set of predefined genes in the normalization. They propose to search for invariable genes (rank-invariant) from the expression data instead [55, 183, 264, 345].

1.3.5.2 Normalization methods used in cDNA microarrays

Global intensity normalization: This method is based on the assumption that there is no dependence between M and A [260, 276, 382]. The normalization factor could be the ratio between the sum of red intensities over the the sum of green intensities for all the spots in the array or in the normalization gene set. But, in general it is the mean or the median intensity of the normalization gene set that is used [135, 276].

Linear methods for microarray normalization: The methods included in this category assume a linear dependence between M and A . This dependence could be detected in MA plots.

Linear regression normalization: The dependence between M and A can be modeled by a linear regression Equation 1.2 [260]. Both β_0 and β_1 are determined by least squares estimation [260].

$$M = \beta_0 + \beta_1 A + \epsilon \quad (1.2)$$

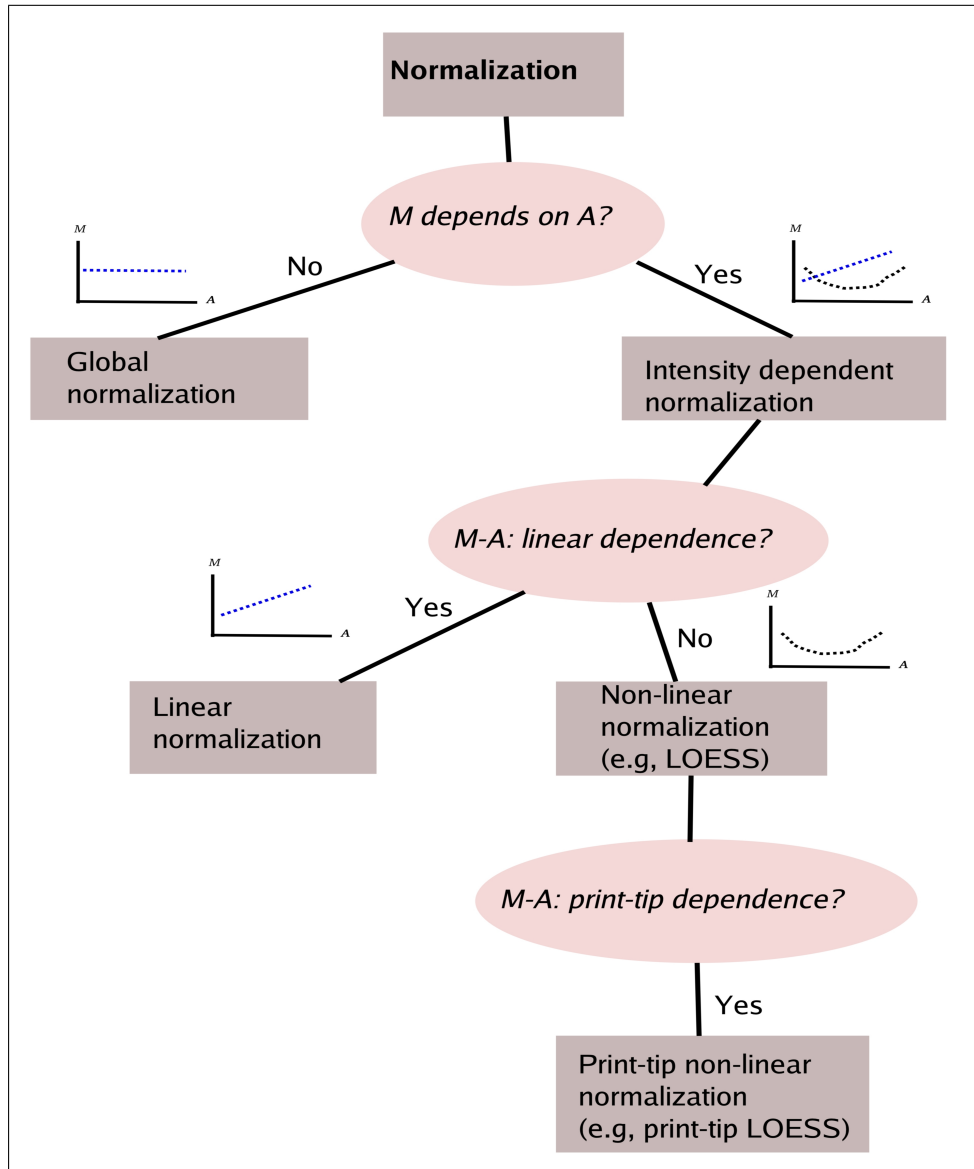


Figure 1.5: *cDNA microarrays flowchart for choosing a normalization method.*

On the basis of MA plot, we can detect a dependence between M ($\log_2(R/G)$) and A ($\log_2(\sqrt{RG})$). Consequent normalization method could be then chosen according to nature of dependence (linear or non-linear).

ANOVA models: Kerr *et al.* [185] proposed ANOVA models adapted to each experimental design. In their model (Equation 1.3) they considered each spot log intensity ($\log(y)$, not the log-ratio) as a deviation from the overall average (μ) by additive effects of the gene itself (G), array (Ar), dye (D), treatment (T) and interactions between gene and array ($G \times Ar$) or treatment ($G \times T$). They fitted their model and estimate its parameters using least squares method. Non zero differences of the $G \times T$ values across treatments are considered as reflective of differential expression. In this model, both normalization and differential gene expression are combined [185].

$$\log(y) = \mu + Ar + D + T + G + (Ar \times G) + (T \times G) + \epsilon \quad (1.3)$$

Wolfinger *et al.* [373] used a modified form of Kerr’s ANOVA model to propose a normalization method for the log-ratio intensities, in which they separated the “normalization model” from the “gene model.” The normalization model, which does not consider the gene and its interaction effects is fitted to the log-ratio intensities and the residual values are used as input to the gene model in which the *gene* $\times$ *treatment* parameter is estimated for each spot and used for differential expression analysis [373].

Non-linear methods for microarray normalization: In order to detect systematic non-linear effects, Dudoit *et al.* [102] introduced the use of MA plots. This non-linear effect could be modeled by the Equation 1.4, where $c(A)$ is the non-linear function to be estimated [260]. Different algorithms were proposed to smoothen this M - A dependence, including the most used *loess* [382] and some others like, local regression [183] and *Qspline* [374].

$$M = c(A) + \epsilon \quad (1.4)$$

LOESS normalization: Yang *et al.* proposed a normalization method that can correct the intensity-dependent dye bias by fitting a local scatter smoother function (*loess*) to the MA plot [382]. Measured M were then adjusted for each spot according to the fitted *loess*. The fitting is done either for the whole chip or for each print-tip (the area covered by the printer tip when gene were spotted) separately [382]. *Loess* can also be used in single channel microarrays [105]. *Loess* normalization assume equal number of genes to be up- and down-regulated. However, this is not always true and alternatives have been proposed [357].

1.3.5.3 Normalization methods for Affymetrix GeneChip arrays

Although the methods above could also be adapted to Affymetrix gene expression chips, special methods, which take into account the specificity of the platform were developed. These methods make use of the Perfect match (PM) and/or MisMatch (MM) probes in each probe set. A comparison summary of the most popular methods in Affymetrix GeneChip arrays are listed in Table 1.3 and developed below [5, 158, 161, 162, 202, 203, 205, 212, 213, 302, 375].

Table 1.3: Affymetrix GeneChip normalization procedure comparison.

Algorithm	Background correction	Normalization	Intensity estimation	Multiple Array
MAS5	$PM - MM$	Linear	Tukey's Biweight	No
RMA	- normal background distribution fitting	Quantile	Median polish	Yes
	- exponential signal distribution fitting			
GC-RMA	MM estimation by GC content probe affinity	Quantile	Median polish	Yes
Li-Wong	None	Invariant set	Multiplicative model fitting	Yes

Microarray Affymetrix Suite (MAS) normalization: The MAS5 suite of algorithms was developed by Affymetrix as a replacement for MAS4. The latter was based on intensity estimates that can yield negative values [158, 161, 375]. In MAS5, Hubbell's algorithm provides always positive values for intensities that are assumed to be linear estimates of target cRNA concentrations [158]. Liu's algorithm provides p-values for absolute "detection calls" (whether a gene is present or absent), as well as "comparison calls" to detect whether a gene is increasing or decreasing or without any significant change in a condition compared to another [212]. The detection call utilizes the discrimination scores and the comparison call uses intensity differences. These p-values are computed using the one-sided upper-tail Wilcoxon's signed rank test from PM and MM intensities [212, 213].

In a multiple chip comparison purpose, a scaling factor is applied to each probe set in the array. It is estimated from the target intensity specified by the user (500 by default) [4]. A weakness of Mas5 is related to its bad performance on multi-chip comparison. Affymetrix proposed PLIER algorithm, which is more adapted to between-chip normalization[5].

The development of alternative algorithms is motivated by: (i) the use of MM subtraction from PM intensities as background adjustment; (ii) the variance for probe sets with low intensities is much larger than for probe-sets with large intensities [375].

Model-Based Expression Index (MBEI) normalization: Li and Wong [202, 203] proposed a general model to better estimate PM and MM intensities. Their model (Equation 1.6) considers that PM and MM difference in array i , probe j of a given probe set is the product of the gene expression index (θ_i) and the probe-sensitivity index (ϕ_j) plus random error (ϵ). Least square estimates for the parameters are carried out by iteratively fitting the set of θ s and ϕ s, keeping the other set as known [202]. The normalization process is based on intensity values of "invariant set" of genes across multiple arrays. This "invariant set" is identified through an iterative ranking procedure [203]. This procedure is implemented in dChip software [302].

Robust Multi-array Average (RMA) normalization: RMA assumes that the real PM intensity is an addition of the true signal and a background signal. RMA starts by background-correcting intensities assuming that the background signal follows a normal distribution and that the signal follows an exponential distribution. Then, RMA normalizes the arrays using quantile normalization. Then, for each probe set RMA fits a linear model (Equation 1.5, where α represent the probe affinity effect and ϵ the error and μ the log-scaled expression level) to the background-corrected, normalized and log2 transformed probe intensities. RMA uses a robust procedure, median polish, to estimate the model parameters and consider the estimate of μ in the model (Equation 1.5) as the log scale measure of expression. The MM intensities are not used in this procedure [161], which makes RMA more precise (excellent correlation between replicates) but less accurate (the measured signal may not perfectly reflect the true RNA concentration) due to the inappropriate adjustment for the presence of nonspecific hybridization (case where $MM \geq PM$) [162, 375].

$$Y_{ijn} = \mu_{ijn} + \alpha_{jn} + \epsilon_{ijn} \quad (1.5)$$

GC content Robust Multi-array Average (GC-RMA) normalization: This algorithm is a variation of RMA that uses a different approach for background estimation. GC-RMA uses subsets of MM probes with the same G-C content as a population that can be considered in background correction for all subsets of PM probes having the same G-C content. Thus, the distribution of background noise could be estimated from observed intensities of probes with similar G-C content. Using the logarithm transformation, this distribution is considered as normal. Estimation of the background noise distribution parameters (mean and standard deviation) is performed either by *maximum likelihood estimate* or *Bayesian estimate* [375].

$$y_{ij} = PM_{ij} - MM_{ij} = \theta_i \phi_j + \epsilon_{ij} \quad (1.6)$$

1.3.6 Statistical analysis of gene expression microarrays

When expression data from a microarray experiment are normalized, they are ready to be processed depending on the initial study objectives in order to predict significant biological outcomes concerning regulated genes. As stated before, microarray study purposes can be distinguished in three categories: (i) class comparison; (ii) class prediction; (iii) class discovery. A class refers to any characteristic shared by one group of samples but not other samples: e.g, cancer versus normal tissue, metastatic versus primary tumor, responders to treatment versus nonresponders. A “known class” is any class that a researcher can distinguish using any differentiating feature [266]. Many of the methods cited in this section are implemented in the widely used open source softwares TM4 [297] and R/Bioconductor [283].

1.3.6.1 Statistical methods for class comparison

Methods used in this category assume that the classes to be compared are known and the investigation concerns the differences at the gene expression level [266]. Their goal is to detect differentially expressed (DE) or differentially co-expressed genes (DC) between compared classes.

Genes are DE if their mean expressions are significantly different between compared classes and they are DC if their correlation changes between compared conditions [62]. Traditionally, DE genes are inferred by a fixed threshold cut off method (i.e. a two-fold increase or decrease) [201]. More relevant statistical methods such as Student's t-test, ANOVA or Mann-Whitney test, can be used to rank the genes from replicated conditions according to the rejection or the acceptance of the null hypothesis, which considers the mean expression of a gene being equal between classes [10, 72, 97]. T-test and ANOVA assume the normality of the data, Mann-Whitney and other non parametric statistics are used as alternatives to this assumption. For time-series microarray, regression methods, Bayesian methods and meta-analysis strategies were proposed [16, 17, 21, 69, 211, 329]. To unravel DC genes, procedures were developed based on identifying groups of genes where pairwise correlations are necessarily high in at least one condition. Examples of these methods include differential clustering algorithm (DCA), DC scoring algorithm and coXpress algorithm [159, 192, 363]. Most of these methods begin by identifying co-expressed genes within a condition followed by a comparison to other conditions, but they could also integrate prior information, such as, groups of genes sharing common function [62].

Setting a cut-off to find genes differentially expressed is statistically challenging due to the balance between the false positives (Type I error) and the false negatives (Type II error) [201]. Furthermore, the number of genes to test generate a multiple hypothesis-testing problem as false positives increase with the number of performed tests [286]. To tackle this issue, many procedures were suggested, such as family-wise error rate (FWER) control and False Discovery Rate (FDR) control. The FWER is the probability of having at least one false positive among all testing hypotheses, it is controlled classically by the Bonferroni method that correct the p-values by the number of performed tests. However, this approach can be too stringent and limits the power to identify significantly differentially expressed genes [201]. FDR, which is the expected proportion of false positives is often more appropriate [286]. Methods to achieve this aim were described and adapted, such as, FDR controlling procedure by Benjamini and Hochberg [36, 286] and the q-value by Storey and Tibshirani [324]. By incorporating the concept of FDR, many procedures were implemented based on modified t-test to set gene selection cut-off. Such programs include Significance Analysis of Microarrays (SAM) and Ranking Analysis of Microarray data (RAM). SAM uses permutations of the repeated measurements to estimate the FDR [346], and RAM is based on a "randomly splitting" approach instead of a permutation approach [333].

1.3.6.2 Statistical methods for class discovery

In this category, no *a priori* knowledge of the investigated classes is assumed. Class discovery attempts to discover new sub-classes where the actual classification needs more definition [266]. The methods used in this category are called unsupervised, they cluster samples (or genes) depending only on their global gene (or sample) expression [266, 275]. These unsupervised methods will find the predominant relationship between genes or samples, which may not be biologically relevant [266].

Clustering techniques could be either “divisive” or “agglomerative.” A divisive method begins with all elements considered as one cluster and broke it down gradually into smaller groups. On the other hand, agglomerative techniques begin with single-member clusters and gradually associate them together [275]. The “closest” genes or samples are grouped together in a cluster. Approximation of the distance between two genes or two experiments is calculated via multiple distance methods.

There is often a need for data dimensionality reduction before any clustering. Some commonly used methods in this purpose as well as for clustering include principal component analysis (PCA) [201, 282]. PCA is a statistical technique for determining the key variables in a multidimensional data set that explain the differences in the observations, and can be used to simplify the analysis and visualization of microarray data sets [281, 282].

Hierarchical clustering, K-means and self organizing maps (SOMs) are examples for clustering methods [106, 201, 275]. Hierarchical clustering is an agglomerative method and produces intuitive visual dendrograms [106]. K-means assume that the number of clusters is known in advance, and assign each individual (gene or sample) to a cluster if its distance is close to the average distance of that cluster. SOM is a divisive approach based on neural networks [275]. SOM starts with defined number of clusters and defined geometric configuration of reference vectors, which are generated randomly in a training step. Then, individuals are assigned to clusters on the basis of the similarity of their expression vectors to the reference vectors that are adjusted iteratively according to the assigned individual expression vector [275].

It remains difficult to statistically determine the number of clusters needed by some algorithms such as SOMs and K-means [386]. Although methods were proposed to deal with this issue [342, 386]. The quality and stability of discovered clusters is assessed by resampling techniques with replacement such as bootstrapping [39] or without replacement such as jackknife [386]. These resampling techniques could be incorporated in multiple clustering runs to derive a clustering consensus [239].

1.3.6.3 Statistical methods for class prediction

Since its successful application in cancer microarray gene expression data by Golub et al., class prediction is widely used in cancer diagnosis and toxicology [135, 277]. Class prediction methods are known as supervised methods. They are based on a “predictor,” which is a gene expression-based multivariate mathematical function [266].

A predictor is build upon feature selection output methods, which identify genes differentially

regulated between compared classes. Numerous feature selection methods have been applied to build predictors, as well as in class comparison, including SAM, ANOVA, empirical Bayes t-statistic, template matching, maxT, between group analysis (BGA), area under the receiver operating characteristic (ROC) curve, the Welch t-statistic, fold change and rank products [167].

Once feature selection is performed, the predictor is “trained” on the identified group of genes to distinguish between compared classes, then it is “tested” on new unknown samples. A good predictor has to be able to assign new unknown samples into the correct class [266, 277]. To achieve this objective, a myriad of algorithms have been described in literature, including k-nearest neighbors (kNN) [135], support vector machines (SVM) [50], nearest shrunken centroids in PAM algorithm [340, 341] and neural networks (NNs) [13, 37, 108, 287].

A common method used for estimating classification error rates is cross-validation. In cross-validation, some observations are hold out from the data set before building the predictor, which will be tested to the held-out observations [49]. When only one observation is hold out for testing, the cross-validation is called “leave-one-out”. The choice of feature selection method, the number of genes in the gene list, the number of samples and the noise in the dataset, substantially influence classification error rate [167].

Supervised methods suffer from two major limitations: “over-fitting” and “bias selection.” Over-fitting designates the fact that a predictor performs better in the samples from which it was “trained” in comparison to independent ones [266]. Bias selection is relative to the use of the whole data set in the feature selection before splitting it into training and test sets [12]. Due to these limitations, class prediction methods need careful validation that should be performed by dividing the data into distinct “training set” and “test set” clearly independent from each other [266]. To correct for selection bias, it is essential that cross-validation be used external to the gene-selection process [12]. Another limitation concerning cancer classification, is the integration of mis-diagnosed samples in the “training set,” which is inevitable since current diagnostic tests are neither perfectly sensitive nor specific [266].

1.3.7 Perspectives on gene expression measuring : RNA-Seq

The classical gene expression microarrays can miss unknown transcripts and splice-junction information. Tiling and exon arrays were developed to circumvent these drawbacks. However, they still rely on the hybridization process and use an indirect measurement of the RNA abundance [224, 327]. As an alternative to microarrays, Expressed Sequence Tags (EST) and Serial Analysis of Gene Expression (SAGE) directly measure the number of RNA molecules by counting a short sequenced portion of the mRNA (tag) as an indication of gene expression. However, these techniques require tremendous experimental effort and are subject to cloning bias. In addition, SAGE is not informative about splice-junction events [222, 327].

Deep sequencing, also known as next generation sequencing (NGS), has been recently developed and present a number of advantages compared to gene expression microarray. This method allows the sequencing of minute amounts of DNA in a high throughput manner. The primary application of NGS

was to count genomic DNA abundance. Recently, it has been successfully applied to transcriptomics by adapting the genomic DNA sequencing protocols to RNA [224, 368].

The workflow of a typical RNA-Seq experiment can be summarized as follows [222, 344, 360]: (i) messenger RNA isolation from the sample; (ii) reverse transcription of mRNA to cDNA; (iii) cDNA fragmentation and production of sequencing libraries with adaptors that will be able to identify each fragment separately; (iv) run parallel sequencing of all these fragments; (v) base signal pre-processing and quality control; (vi) mapping of reads to the reference genome and calculation of the RNA expression from the mapping.

RNA-Seq read lengths depend on platforms and range from 30-100 bp from Illumina and SOLiD to 200-500 bp for FLX [222]. The count of the mapped reads to each gene is indicative about its expression level [222, 223, 360, 368].

The major technical issues raised by this technique are: (i) the cDNA library may be biased; (ii) it requires huge computer resources and (iii) sequencing errors are particularly challenging, as they are difficult to distinguish from polymorphisms. Costs are going down but equipment remains expensive [89, 223, 360, 369].

In order to find differentially expressed genes between two conditions, the RNA-Seq expression levels are normalized. RPKM procedure computes the number of reads for each gene divided by its length and by the number of total reads [242]. Other methods fit Poisson distribution model to the read counts per gene to derive the expression levels [170, 224]. Normalized results can then be analyzed using the methods developed for microarrays, such as SAM [358].

1.4 Functional data mining in gene expression

The list of genes identified as relevant by one of the methods developed above needs to be considered in its biological context in order to clearly understand the way their products are involved in the studied process. Bioinformatics offers a considerable choice of methods to achieve this purpose by comparing the data with knowledge-based databases, including gene function, metabolic and signalling pathways, regulatory elements, protein domains, protein-protein interactions, chromosomal location, genetic mapping disease, structure activity relationship, dose response curves and phenotypic or clinical information [277].

1.4.1 Gene function, metabolic and signalling pathway mining

The Gene ontology (GO) database is widely used in gene functional mining. It consists on a collection of genes with their respective published functional annotations, which are organized in three levels, each of them has a hierarchical structure. The three top levels are: (i) Molecular Function (MF), which is the activity of the corresponding genes; (ii) Biological Process (BP), which groups series of events accomplished by one or more MF; Cellular Component (CC), which describes the anatomical structure of the cell where the gene product can be localized [256].

Beyond GO, There is a large collection of metabolic and signalling pathway data maintained from the literature and freely available to the research community. Such collection is present in KEGG database [175], GeneMapp pathways [128], PantherDB pathways [338] and Biocarta pathways [41]. These databases usually map each gene to the corresponding pathway. They contain only a small fraction of the wealth of data available in the literature.

An increasing number of tools were developed to mine these biological information and find functional enrichment in a significant list of genes generated by microarrays. They are based mainly on Fisher's test, χ^2 and hypergeometric distribution tests. They compare submitted gene lists (query list) with the reference list catalogues (GO, pathways) and generate enrichment p-values with multiple testing corrections (FDR, FWER, ...). Examples of such tools include the GO-specialized softs: GoMiner [117], Gostat [34] and FatiGO [8]. More general tools that combine GO and pathways among other biological information mining, include the popular DAVID web tool [88].

The statistical tests used by these tools have two major limitations: (i) they use only gene counts in the overlap between the query and the reference lists and cannot directly integrate gene expression intensities; (ii) they depend largely on the number of genes identified as significantly regulated between studied conditions [45, 262]. Therefore, alternative approaches were developed to circumvent these limits. Gene Set Enrichment Analysis (GSEA), for instance, has the particularity to mine against other published microarray gene sets in addition to other biological information [325]. The pre-ranked GSEA algorithm encompasses the first limit by taking as input a top ranked list of gene with their expression intensities. The ranking of such list could be done using different metrics, such as signal to noise ratio between sample classes. The algorithm seeks for enrichment of gene sets in the top and the bottom of the ranked list. GSEA computes a Kolmogorov-Smirnov score, which is assessed by

the comparison to scores obtained with random permuted gene sets [238]. However, this approach has the limitation of using pre-determined gene lists. The use of the whole expression data for data mining instead of pre-selected lists is shown to be more robust. In this approach, all the genes of the functional category are used to build the enrichment score rather than missing some genes because of the used cut-off [262]. Tools that are built on this approach were developed, such as T-profiler, which uses t-test to compare the mean expression of functional gene set with the mean expression of the rest of the genes in the whole microarray data [45]. GSEA algorithm was also adapted to use the rank of the whole expression data set [325].

1.4.2 Transcription factor and regulatory network mining

As transcription factors (TFs) are the master regulators of gene expression, a functional analysis often starts with the identification of the TF involved in a given model system.

Some TFs bind to either one or multiple Transcription Factor Binding Sites (TFBS) in a gene promoter, others do not bind directly to DNA and need adaptors (co-factors). TFBS could be *palindromic*, which means that the motif is the same as its reverse complement, or *spaced dyad*, which describes a TFBS organized in two closed motifs that are separated by any other DNA sequence [75].

Databases of TFs and their target genes with the experimentally proved corresponding binding sites are also available, such as JASPAR [299], public TRANSFAC [372], TRRD [181], ORegAnno [237], PAZAR [272], TRED [390] and NFIREgulomeDB [249]. In contrast to biological information, transcription factor mining in gene expression data do not use a direct comparison to these databases. Instead, TF enrichment is done with *in silico* derived TF-gene regulations as used in GSEA and Babelomics for instance [9, 325].

Two types of *In silico* identification of regulatory elements in gene promoters can be distinguished, depending on the objectives to be achieved. *Pattern matching*, aims to find TFBS in DNA sequences for a given TF, knowing that this TF can bind to a DNA consensus sequence. On the other hand, *pattern discovery* is the process by which frequently occurring DNA motifs are discovered from a set of co-expressed gene promoters. This is built in the assumption that these more frequent motifs could be linked to biologically relevant TFBS [75].

Pattern matching: Pattern matching is done in two steps: (i) build TFBS motif consensus from published regulated promoters; (ii) use this consensus to scan new promoters for putative TFBS. Consensus motifs are defined such as they represent all of the known TFBS for a given TF. There are many different methods to define the consensus motif with the great majority of them not performing well [78]. Beyond a simple consensus motif, Position Weight Matrices (PWMs) were proposed to scan gene promoters [226]. For each TF, a Position Frequency Matrix (PFM) is built by aligning known promoters and counting the frequency of each base in a given position. PWM is derived from the logarithmic transformation of PFM after dividing the nucleotide frequency in each position by its random frequency, which is built from random sequence references [67]. Since promoter alignment is used to

calculate the nucleotide frequencies and a limited number of known promoter sequences are available for TFs, there is a risk to miss one of the four nucleotides in a given position. This is an obvious problem when doing the logarithm transformation to build the PWM for such unobserved nucleotide. Therefore, the notion of *pseudo-counts* is introduced to fix this constraint. Pseudo-counts correct the calculation of the nucleotide frequency to avoid obtaining null values for unobserved bases. Depending on which approach is used, pseudo-counts could be considered as equal for all the nucleotides, or dependent on the nucleotide background frequency, or dependent on nucleotide substitution matrices [67]. By comparing the real PFM and pseudo-counts corrected PFM, Nishidia et al. recently proposed a guidance to choose pseudo-counts [250]. When scanning a new promoter sequence using PWM, an *aggregate score* is computed, which is the sum of the weights of the aligned nucleotides given their respective positions [67]. To be biologically significant, a TFBS found by PWM gene promoter scanning needs to have a higher score with a low probability to find such a score by chance. This probability depends on the target sequence length and its nucleotide composition. Using the probability density of random scores, which are observed from random sequences, can help to fix a statistically significant score threshold above which matched TFBS are significant [67]. As TF can bind to less specific sequences (degenerate), the built PWMs could be less specific, which results in a high false-positive TFBS discovery rate. Thus, the use of multiple PWMs instead of one is proposed [141]. In addition, some TFs are shown to act in a modular way to activate transcription, which means that they are clustered together near the promoter. These modules need special approach to be discovered, including PWMs combined to phylogeny conservation and machine learning techniques, such as *logistic regression* [362]. It is also known that TF binding to DNA is context-specific. It depends on limiting factors, such as DNA bending, the accessibility of the binding site, or the positive or negative effects of neighboring TFBSs. To control this, Huang et al. proposed a *Markov* background model [155].

Pattern discovery: The basis of pattern discovery is to start from a set of promoters presumed to contain common unknown regulatory elements. It is usually the situation where promoters for known co-expressed genes are aligned to derive *ab initio* consensus motifs or PWMs and use them back to perform pattern matching on these promoters for elucidating their localizations [339]. Pattern discovery methods could be classified in two categories: (i) string (motif) counting based methods; (ii) probabilistic model based methods [75].

Although tools cited in the gene function mining can be applied to search for enrichment in any kind of annotated data as stated above, there are many tools specialized only in TFBS mining. The major common point between them is their relying on *in silico* predictions of regulations by gene promoter scanning, either pre-compiled or computed *ad hoc* using TRANSFAC PWMs and JASPAR profiles. Their differences are mainly in the implemented background filtering methods [80, 148, 200, 210, 318, 326].

1.4.2.1 Regulatory networks

The classical analysis of regulated genes, such as GO, do not provide informations about how these genes globally interact [277, 365]. Building associations between regulated genes can provide a global view for the mechanistic interactions involving these genes. These associations are inferred from expression data. The result of such analysis is a co-expressed gene network that has the advantage to be built for the specific studied biological condition and not derived from pre-compiled networks from the literature [365]. Attempts to build such networks are described recently in the literature and it is a field of intense research.

The problem of network construction from microarray expression data could be summarized as follows: considering a set of nodes (genes) in the represented graph (network) and a set of observations of these nodes (expression data), find a set of edges (arrows in the network: link between genes) that could link the nodes together. Edges are based on known associations between genes. Therefore, the network is just a graphical representation of the solution, although graphs can be used also for the elaboration of this solution using *graph theory* approach [225]. The start is to find genes with high similarity according to their expression. This similarity is measured by the correlation, which is based on the linear assumption of the similarity between genes, but non linear assumption could also be addressed using machine learning [221] or probabilistic methods based on prior information [95]. The built networks need to be assessed to be significantly different from random networks that will be found just by chance. The inferred network cannot at this stage inform about the real dependencies from observed gene-to-gene relationship. To be informative about dependencies, either perturbation experiments should be made and see their influence on the network or try to infer the dependencies from the network based on *conditional independence* methods [225]. Many tools exist allowing the construction of such networks, including the recent user interactive tool Miner [172].

An interesting application of the co-expressed gene networks is to associate TF to genes and derive a *regulatory network*. In this line, some studies associate TFBS enrichment with clusters in the co-expressed gene network and thus add the TF as a node to the graph where relationships between this node to the gene nodes represent inferred putative regulations [351]. Attempt to construct regulatory networks from gene expression clustering coupled to promoter clustering, which is inferred from TFBS enrichment, was also reported [354]. In another approach, Segal *et al.* tried to model the TF occupancy on gene promoters and predict expression rate, which is compared to the observed expression and then to infer the most probable TF binding configuration in the set of input gene promoters [305]. Literature and functional mining could also be used as a guidance for gene clustering and TF association to grouped genes [307].

Recent efforts were made to improve *in silico* TF-gene regulation inference, but many pitfalls can still be enumerated, including: (i) the high background they generate, especially in mammalian genomes where intergenic regions are long, leading to high probabilities to find regulatory elements, which are usually short-sized [298]; (ii) they rely on alignment of known promoters for a given species or their orthologues, when performing phylogenetic conservation, so they prone to error in misalign-

ment [136]; (iii) They rely on clusters of co-expressed genes, thus mis-clustering could be a problem; (iv) TFs that do not bind directly to DNA could not be discovered; (v) Epigenetics and DNA accessibility.

The most important factors affecting *in silico* TFBS identification is the input promoter sequences. Identification of the promoter region and its sequence length is not trivial. Computational methods were developed for this task guided by inter-species promoter conservation and promoter signal (e.g. TATA-box) identification [28]. However, the complexity of this task increases with the size of genomes. In the high organisms, such as mammals, the promoter is spread in nearby sequences. In addition, the modularity of some TFs, which act in clusters complicate the task of promoter delimitation [177]. Moreover, in mammalian genomes, transcription was shown to be context-specific dependent on the epigenetic modifications and driven by alternative transcription start sites [77, 263]. The high-throughput identification of nuclease hypersensitive sites, transcription start sites and nucleosome binding sites are used as experimental approaches to better define promoter localization [47, 70, 306].

Nowadays, the general tendency is to integrate data from diverse sources with the gene expression data, such ChIP-ChIP, ChIP-seq, protein-protein interactions and chromatin structure. This kind of approach produces a more complete understanding of the biological response of a cell, organ, or individual to a particular stimulus [277, 328]. In a recent successful example, FANTOM consortium proceeds by sequencing transcription start sites (TSS) in time course experiments on human monocytic cell line THP-1 and mapped these TSS to the genome, then the regions around these TSS in the gene promoters were analysed for TFBS enrichment combining PWMs with phylogeny conservation. The over-represented TFs were linked to genes according to their expression. In that way, they could infer TF activity based on compared conditions in a microarray data set. Their algorithm is called MARA and it is freely available [70]. However, much more development is needed to build robust and accurate TF regulatory networks.

Objectives

CHAPTER 2

OBJECTIVES

The main objective is to develop a tool to predict TF regulation from microarray experiments, and to apply it on a biological system that is much studied in the laboratory: the stimulation of cell proliferation by growth factors such as PDGF.

The aims of the present work are the following:

- Understand how PDGF regulates transcription factors (TF) and genes. We will perform microarray experiments to study gene expression regulation by PDGF in primary fibroblasts as well as gene expression induced by constitutive active forms of PDGFR hybrid proteins in cancer cell lines. We will also compare the effects of PDGF with other growth factors such as FGF.
- Develop a computational approach that will be able to assign TF regulation and activation status starting from gene expression data. In contrast to the existing methods, our approach will be based on comparing lists of microarray regulated genes to the experimentally well-established TF-gene regulations. This would enable us to detect the TF regulated by PDGF.
- Confirm experimentally some of the findings and link the TFs regulated by PDGF to their upstream signalling pathways.

Results

CHAPTER 3

RESULTS

3.1 Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data¹

3.1.1 Abstract

Deciphering transcription factor networks from microarray data remains difficult. This study presents a simple method to infer the regulation of transcription factors from microarray data based on well-characterized target genes. We generated a catalogue containing 343 transcription factors associated with 2,720 target genes and 6,401 experimentally validated regulations. When it was available, a distinction between transcriptional activation and inhibition was included for each regulation. Next, we built a tool (www.tfacts.org) that compares submitted gene lists with target genes in the catalogue to detect regulated transcription factors. TFactS was validated with published lists of regulated genes in various models and compared to tools based on *in silico* promoter analysis. We next analyzed the NCI60 cancer microarray dataset and showed the regulation of SOX10, MITF and JUN in melanomas. We then performed microarray experiments comparing gene expression response of human fibroblasts stimulated by different growth factors. TFactS predicted the specific activation of STAT factors by PDGF-BB, which was confirmed experimentally. Our results show that the expression levels of transcription factor target genes constitute a robust signature for transcription factor regulation, and can be efficiently used for microarray data mining.

3.1.2 Introduction

Data mining in gene lists generated by microarray experiments remains a major challenge. A particularly intricate issue is to identify the regulatory network responsible for gene regulation in a given biological system. Several tools like GSEA offer the possibility to search for enriched transcription factor (TF) targets in lists of co-expressed genes [80, 148, 153, 210, 325]. They mostly use pre-computed putative cis-regulatory elements predicted in gene promoters and compare their results to the submitted gene lists. A drawback of binding site prediction is the high level of noise

¹Essaghir A, Toffalini F, Knoops L, Kallin A, van Helden J, Demoulin JB; Nucleic Acids Research; 2010; In Press

when it is applied to mammalian genomes [298]. Several improvements have been proposed, such as context-sensitive promoter analysis and promoter clustering based on microarray co-expressed genes [70, 107, 208, 317, 354, 365].

We and others have observed that the activation of TFs like SREBP1 [84, 174] and p53 [188] can easily be detected in microarray gene lists based on the fact that the expression of their target genes is changed. It is however not clear as to whether this approach is valid for other TFs.

The aim of this work was to test whether a catalogue of well-characterized TF targets can be used to predict the activity of transcription factors in various biological systems, based on microarray data.

As a model system to test this method, we studied gene regulation by growth factor receptors in normal and leukemia cells. Growth factors, such as Platelet-Derived Growth Factor (PDGF) and Fibroblast Growth Factor (FGF), are key regulators of cell proliferation, differentiation and survival. They exert these effects through specific tyrosine kinase receptors, which activate signalling pathways that ultimately regulate transcription factors and gene expression. Previous reports using expression microarrays suggested that growth factors regulate gene sets involved in wound healing [138], fatty acid and cholesterol synthesis [84, 174], cell division and differentiation [83]. Various tyrosine kinase receptors are thought to induce overlapping gene expression, suggesting that growth factors activate a common transcriptional program associated with cell growth [115].

3.1.3 Materials and methods

3.1.3.1 Data

To build a database linking TFs to their target genes, we started with lists of genes obtained from curated TRED [390], TRRD [181], PAZAR [273] and NFIREgulomeDB databases, and completed some of these factor-gene lists with the regulation type (“up” or “down”) based on the original publications. SREBP and p53 target gene signatures were based on our previous reports [84, 188] and were completed by data from TRED. The FOXO1, FOXO3, TCF/ β -Catenin, GLI, ID, AR, GR, LXR, NOTCH/RBP-J, PPAR α , ATF6, HNF, SF1 and STAT transcription factors gene target lists were collected from published papers. On the basis of these data, we built two catalogues: The “sign-sensitive” catalogue that takes into account the sign of the regulation (“up” or “down”) and the “sign-less” catalogue that neglects this information (see Figure 3.1 and supplementary file 1 for catalogues and references).

3.1.3.2 Statistics

The number of target genes for a particular TF in the query gene lists and in the catalogue are compared using right-sided Fisher’s test. Three different hypotheses are tested for each TF. These hypotheses are respectively : activation and inhibition (using sign-sensitive catalogue), and regulation (using sign-less catalogue). For each hypothesis we compute the p-value ($Pval$), e-value ($Eval$), q-value ($Qval$), False Discovery Rate control (FDR_c), as well as a non-parametric Random Control (RC).

The p-value is computed according to the hypergeometric Formula 3.1, where m is the number of target genes annotated for the TF under consideration, n the number of query genes, N the number of regulations in the catalogue (sign-less or sign-sensitive) and k varies with the tests as follows.

- sign-less regulation: k is the number of query genes that are annotated as regulated by TF (i.e. the intersection between the query and the TF signature);
- sign-sensitive activation: k is the number of query genes for which the sign of the transcriptional response (+ for up-regulated, – for down-regulated) is the same as the sign of their regulation by TF (+ for activation, – for repression);
- sign-sensitive inhibition: k is the number of query genes for which the sign of the transcriptional response is the opposite of the sign of their regulation by TF.

$$Pval = \sum_{i=k}^{i=n} \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} \quad (3.1)$$

The nominal p-value ($Pval$) represents the risk of false positive for a single test. Since the list of query genes is systematically compared to each target gene signature, a multi-testing condition is required. The $Eval$ represents the expected number of false positives for a given nominal p-value. It is computed using the Formula 3.2, , where T is the number of tests.

$$Eval = Pval \cdot T \quad (3.2)$$

The q-value ($Qval$), which is an estimate of the False Discovery Rate (FDR), i.e. the fraction of predictions expected to be false (false predictions / all predictions) for a given level of nominal p-value, is computed using the R package from Storey et al. [324].

The Benjamini-Hochberg procedure to control the FDR is implemented as described in Benjamini et al. [36].

To evaluate empirically the probability to select a TF by chance, we run Fisher’s tests with random gene selections. The program computes the “random control” (RC) as stated in the Formula 3.3, where λ is a user-specified threshold on $Eval$ ($\lambda = 0.05$ by default), N is the number of repetitions to perform ($N = 100$ by default), and $N(Eval_{TF} \leq \lambda)$ is the number of random simulations returning an $Eval \leq \lambda$.

$$RC_{TF} = \frac{N(Eval_{TF} \leq \lambda)}{N} \cdot 100 \quad (3.3)$$

3.1.3.3 Implementation and availability

The tool is available via a Web page www.tfacts.org. It is implemented in HTML, JavaScript, PHP, R and Perl, the catalogue of TF signatures is stored and maintained under MySQL.

3.1.3.4 Validation/comparison

NCI60 dataset [311] was retrieved from cellMiner web site. In the e-value matrix constructed from NCI60 dataset analysis, prior to the $-\log_{10}$ transformation, we arbitrarily set all the null e-values to $1.00e^{-4}$, and put the non detected TFs e-values to the matrix maximum value.

Tools that can detect TF binding sites over-representation in a set of co-expressed gene promoters are compared here with TFactS (sign-less catalogue). Results are summarized in Table 2.

The GseaPreRanked tool of GSEA (v2 GUI version) was run using the motif gene sets: “c3.all.v2.5.symbols.gmt” or TFactS sign-less catalogue. The parameters used with GSEA are as follows:

`xtools.gsea.GseaPreranked`, with `plot top x = 10`, `norm = meandiv`, `scoring scheme = weighted`, `make sets = true`, `mode = Max probe`, `gmx = c3.all.v2.5.symbols.gmt`, `set min = 2`, `nperm = 1000`, `rnd seed = timestamp` and `set max = 500`. Output ranking in GSEA is based on FWER p-value. Outputs from GSEA are considered here as significant when having the nominal p-value ≤ 0.05 . The minimum rank from either positive or negative output lists is considered in ranking comparison with other tools.

TFM-Explorer web server from <http://bioinfo.lifl.fr/TFME/> was run using both TRANSFAC and JASPAR vertebrate matrices on upstream promoter sequences for each gene located from -2000 to 0 with a ratio (density of clusters) of 2.5. Input gene sets were based on refSeq names. Significant TF binding sites were those with p-value ≤ 0.05 . The minimum rank based on p-value is considered for a given TF in the comparison with other tools.

CORE_TF web server was fed with 2,000bp sequences upstream each gene first exon. The extraction of these sequences was done using Biomart web service (<http://www.biomart.org/biomart/martview/>). The random set generation was specified to be from Ensembl using 200 human gene promoters with the same base pair length upstream as the input promoters. The match cut-off selection was done with “Min False Sum” and significant TF binding sites were those with p-value ≤ 0.05 from the output list without processing for evolution conservation. The minimum rank based on p-value ranking is considered for a given TF when comparing with other tools.

oPOSSUM web server version 2 was run by selecting the following inputs: Human as “Species”, HUGO gene names for each set of co-expressed genes, vertebrates as “JASPAR core profiles”, top 10% of conserved regions (min. conservation 70%) as the “Level of conservation”, 80% as “Matrix match threshold” and 2000/0 as the “Amount of upstream / downstream sequence”. Profiles with Fisher-Score ≤ 0.05 were considered as significant. The minimum rank based on Fisher-score is considered for a given TF for the comparison with other tools.

CRSD Genome-wide Iterative Enrichment Analysis web server from <http://140.120.213.10:8080/crsd/> was used with default values. And refSeq gene names were given as input. Significant TF binding sites were chosen based on 0.05 p-value threshold. The minimum rank based on p-value is considered for a given TF for the comparison with other tools.

For all these tools, the output lists that were chosen in this study were restricted to maximum 100 TFs if the generated significant lists contains more than 100 TFs.

3.1.3.5 Cell cultures and microarray analysis

Human fibroblasts (AG01518) were cultured ($1.5 \cdot 10^6$ cells/10 cm dish) in MEM (with fetal calf serum 10% and L-Glutamine) for 24h. Cells were starved for 24h in MEM (0.05% BSA and L-Glutamine) and then stimulated for another 24h or 1h by PDGF-BB (25ng/ml) or b-FGF (10ng/ml) in the presence of heparin (50 μ g/ml). The control condition was left without stimulation in the starvation medium for 48h.

The human fibroblast microarray experiment was done in two biological replicates. RNA extraction was performed using RNeasy (Qiagen). RNA quality was tested in agarose gel electrophoresis and by real-time PCR for two known PDGF-BB target genes: NR4A1 (after 1h) and SCD (after 24h) compared to a control gene RPLPO [84]. The human Affymetrix chip HGU133A2 was used as described [111, 199]. Results from Affymetrix chips were analyzed using the GCOS suite software, and in each replicate the stimulated conditions were normalized using the control condition as baseline for the comparison. The MAS5 algorithm in GCOS was used as follows: the scaling factor using all probe sets was set to 100 and the normalize factor was set to 1. Using MeV [297], the Significant Analysis of Microarray (SAM) algorithm was applied to detect the most significant probe sets differentially expressed between the four conditions normalized to the control condition (b-FGF stimulated after 1h, b-FGF stimulated after 24h, PDGF-BB stimulated after 1h and PDGF-BB stimulated after 24h), with 6.73 as the median number of falsely significant genes (a False discovery rate of 5.85%) and a delta value equal to 1.01. The hierarchical analysis of these probe sets and the samples was done using the Euclidian distance and the average linkage clustering. Fold change analysis was done between each condition and the control after the absent calls filtering and taking only probe sets that are marked by GCOS software as Increased or Decreased having the mean log₂-ratio absolute value of the two replicates greater or equal to 1.

EOL1 (human acute myeloid eosinophilic leukemia cell line [71]) were purchased from DSMZ (ACC386, Germany). Cells were maintained in RPMI medium with 10% fetal bovine serum (FBS) and supplemented with penicillin-streptomycin solution (Sigma). Cells were treated with 100 nM imatinib mesylate for 4 hours and then washed once in ice-cold PBS before being used for further experiments. RNA extraction was done as above from $5 \cdot 10^6$ cells and the human Affymetrix Chip U133 Plus 2.0 was used according to manufacturer's instruction (Affymetrix). GCOS was used with the same parameters as above with the imatinib-treated condition as baseline. Lists of regulated genes used here were obtained by applying a 3-fold expression filter.

These data sets are available in Gene Expression Omnibus database (ref: GSE14256 and GSE15237).

Western blots were performed as described [58, 85, 111, 343] with antibodies from Cell Signalling Technology.

3.1.4 Results and discussion

3.1.4.1 Construction of the catalogues

In an effort to build a catalogue linking transcription factors to their known target genes, we integrated data from the literature and freely available databases (Figure 3.1). This catalogue contains 6,401 regulations linking 343 distinct TFs to 2,720 distinct genes. We considered all TF target genes independently from conditions such as tissue, treatment or cell type. Since gene regulation is highly conserved in mammals, data from human, mouse and rat was pooled. This is justified by the fact that many published experiments were performed in heterogeneous biological systems, for instance human promoters studied in mouse cells transfected with cDNA from different species. We reasoned that including the sign of the TF-gene associations (“up-” or “down-” regulation) would be useful. However, although it is usually available from experimental data, this information is surprisingly not systematically included in mammalian TF databases. Therefore, in our first catalogue, we did not take into account the sign of each regulation (i.e. activation or repression). We will refer to it as the “sign-less” catalogue.

To build a “sign-sensitive” catalogue, we annotated the regulations from original publications and from databases when available. Altogether, this second catalogue contains 63 TFs (Figure 3.1). Both catalogues are provided in supplementary file 1.

For some TFs, the “specific target gene signature” was defined as the list of target genes known to be regulated by a given TF. We further define “generic target gene signatures” by regrouping closely related TFs sharing many target genes. For instance, the E2F generic signature groups target genes of the four activating isoforms of E2F (E2F1-4).

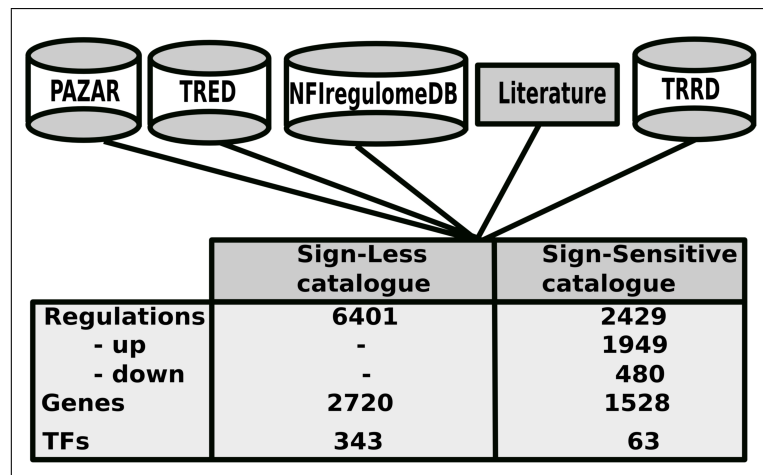


Figure 3.1: *Data integration in TFactS catalogues.*

3.1.4.2 Enrichment in transcription factor target genes: statistical analysis

The comparison between TF signatures and the query gene lists is based on Fisher’s test. For each TF, a contingency table is created which compares the gene query lists and the TF signature in the catalogue. Using the sign-sensitive catalogue we calculate p-values ($Pval$) for two hypotheses respectively: activation (activated target genes are “up” and repressed are “down”) and TF inhibition (activated target genes are “down” and repressed target genes are “up”), see methods section for more details. Using the sign-less catalogue we calculated target gene enrichment $Pval$ for each TF (this is called regulation hypothesis). Examples of contingency tables for the three hypotheses (“regulation”, “activation” and “inhibition”, respectively) are presented in Table 3.1.4.2.

Table 3.1: **Example for contingency tables built under the three hypotheses tested by TFactS.**

In this example, 43 regulated genes from Terragni et al. [335] were submitted to TFactS. Eleven of them belong to the foxo3 signature (defined by 68 regulations) in the sign-less catalogue (6, 401 regulations in total). Among these 11 genes, 8 matched the regulation type (“up” and “down”), whereas 3 were of the opposite type in the sign-sensitive catalogue (containing 66 foxo3 signatures among 2, 429 regulations).

transcription factor (foxo3)	user: present	user: absent	total
sign-less regulation hypothesis:			
catalogue: present	11	57	68
catalogue: absent	32	6301	6333
total	43	6358	6401
$Pval$ Fisher’s test			$0.00e^{+0}$
sign-sensitive activation hypothesis:			
catalogue: present	8	58	66
catalogue: absent	35	2328	2363
total	43	2386	2429
$Pval$ Fisher’s test			$1.40e^{-4}$
sign-sensitive inhibition hypothesis:			
catalogue: present	3	63	66
catalogue: absent	40	2323	2363
total	43	2386	2429
$Pval$ Fisher’s test			$2.15e^{-1}$

The corresponding e-value ($Eval$), q-value ($Qval$) and False Discovery Rate (FDR_c) for each p-value ($Pval$) are calculated to control the rate of false positives in multiple testing conditions.

As a negative control, we included a non-parametric false positive percentage estimation, which we call “random control” (RC). It is based on the selection of random sets of genes from the catalogue according to the number of genes in the query lists.

In this study, we considered a TF as regulated, activated or inhibited, respectively, if, under the corresponding hypothesis, the $Pval$ was significant (≤ 0.05) combined with either $Eval \leq 0.05$ or $Qval \leq 0.05$ or significant FDR_c or $RC \leq 5\%$.

3.1.4.3 User interface

TFactS is available at www.tfacts.org. The main page includes three frames. The query frame allows the user to submit the lists of up-regulated and/or down-regulated genes, and to select the analysis parameters (statistical thresholds, number of random selections). By default, in this version, only HGNC standard official gene names and Entrez gene IDs are accepted in the list of query genes. The user is allowed to enter a custom catalogue, as an alternative to the TFactS catalogue proposed by default.

Results are displayed in the main frame. The results of each run are reported in two pages. Depending on which catalogue the user chooses, the first page contains a table for each type of hypothesis: regulation (sign-less catalogue), activation and inhibition (sign-sensitive catalogue), each one contains a list of regulated TFs, the corresponding $Pval$, $Eval$, $Qval$, FDR_c , the number of genes in the intersection, the number of target genes for that TF and the RC (see methods). The second page shows for each submitted gene the corresponding TFs in the catalogue and its type of regulation. An export link to the raw results is also provided.

The tool is documented at various levels. The statistical methods and options are explained in a detailed help page. A “Demo” button loads the tool with a study case data set, to illustrate the process for a typical submission. An additional frame allows browsing the list of target genes for a particular TF.

3.1.4.4 Validation with published microarray data

TFactS validation was first performed by re-analyzing microarray-based studies in which TFs present in our catalogues were shown to be activated or inhibited. Sixteen such studies, independent from those used to build the database, were found in PubMed [33, 68, 82, 90, 100, 121, 176, 218, 265, 271, 334–336, 348, 355, 389], covering 18 transcription factors in total (14 activations and 4 inhibitions). These experiments consisted in detecting genes that are differentially expressed between tumor and normal cells, in different cell types, or that respond to cytokines (interferons) or drugs affecting specific signalling pathways (Table 3.2). Those studies cover Human and Mouse species and a broad range of biological processes and conditions. In each report, the regulation of one or more TF was inferred from microarray and experimental data.

To evaluate the ability of TFactS to detect the relevant TFs, we submitted the genes reported by the authors as showing a significant response in their respective microarray analysis. When regulated genes were not listed in the paper, we re-analyzed the raw data obtained from GEO database and we selected genes significantly regulated more than two-fold.

Even though these studies were based on very different biological systems, the results summarized in Table 3.2 (details in supplementary file 1) show that TFactS identified all (18/18) of the relevant TFs. For example, Terragni *et al.* [335] showed that inhibition of the AKT pathway provokes the activation of FOXO3 and the inhibition of NF- κ B. Consistently, TFactS identified FOXO3 as regulated ($Pval \simeq 0.00e^{+0}$) and activated ($Pval = 1.40e^{-4}$) and NF- κ B as regulated ($Pval \simeq 0.00e^{+0}$) and inhibited

($P_{val} = 1.16e^{-3}$).

For some studies, the association is detected both in the sign-less test and in the sign-sensitive tests (activation or inhibition). The associations are sometimes returned both at the generic (e.g. FOXO) and specific (e.g. FOXO3) levels.

The sign-sensitive catalogue allowed inferring the type of TF regulation (activated or inhibited). As shown in Table 3.2 (details in supplementary file 1), TFactS failed to recognize the sign of the regulation in two cases. Another limitation of the sign-sensitive catalogue is the more restricted number of TFs that is included. As expected, tests using that catalogue returned a lower RC compared to sign-less predictions, indicating that it is more conservative.

In summary, using the sign-less catalogue associated with Fisher statistics, TFactS results matched the conclusions of published experimental studies. Taking into account the sign of regulation provided a useful piece of information, but did not improve the predictions, at least in these initial validation tests. However, much more sign annotations would be required to provide a fair comparison between the two catalogues.

3.1.4.5 Comparison with other methods

We attempted to compare the results returned by TFactS with those of other freely available tools, which are generally based on *in silico* cis regulatory motif predictions in gene promoters.

Among those tools, we used GSEA [325] based on (“c3”) signatures generated from predictions of cis-regulatory motifs that are conserved across the human, mouse, rat and dog genomes and matched to known TFs by their consensus DNA binding motifs. We also chose four other tools: TFM-Explorer [80], CORE_TF [148], CRSD [210] and oPOSSUM [153], which are based on over-representation of TF binding sites using TRANSFAC [371] and JASPAR [299] position weight matrices (PWM) for gene promoter scanning. TRANSFAC contains 365 and presents a large overlap with TFactS database, including all TFs used in the validation step. Each tool has a specific method for background filtering. CRSD pre-filters results by q-value ranking; TFM-Explorer uses a threshold based on the parameters determined by the local genomic context; CORE_TF compares the submitted gene promoter set with a random set; and oPOSSUM is based on phylogenetic footprinting.

To compare the different tools, we have used similar settings for all of them in terms of statistical cut-off and promoter length. Lists of enriched transcription factors generated using these tools were ranked according to p-values or FWER p-value (GSEA). We considered only significant results (nominal p-value ≤ 0.05) and limited the number of TFs in the output lists to maximum 100, even though such long lists are not suitable for experimental validation. Using these parameters, CRSD and CORE_TF found 12 out of 18 expected TFs, TFM-Explorer 8, oPOSSUM 7 and GSEA 2. Three TFs were absent from JASPAR profiles used by oPOSSUM. When used with TFactS sign-less catalogue instead of “c3” signatures, GSEA performed better (5/18). We did not use other GSEA gene set signatures as they are not centred on TFs. Detailed results are shown in Table 3.1.4.5. Compared to TFactS, these tools produced much longer lists of regulated TFs, but it is not clear whether these rep-

Table 3.2: **TFactS validation.**

Eighteen TFs (14 activated and 4 inhibited) reported in 16 studies were used to test TFactS. Under each hypothesis, TFs are considered significant on the basis of multiple score thresholds (*Pval* and *Eval* or *Qval* or *FDRc* or *RC*) as described in the text. The expected TFs that were described in the studies are represented here with names in **bold** character. Symbols: (y): found regulated, (+): found activated, (-): found inhibited and (n): not found. act.: activation, inh.: inhibition and reg.: regulation.

Ref.	Model	Expected TFs	Sign-Less Catalogue	Sign-Sensitive Catalogue
			Reg.	Act./Inh.
[82]:	colon cancer	FOXO3 act.	y	+
		MYC inh.	y	n
[68]:	fibroblasts	MYC act.	y	+
[90]:	fibrosarcoma	STAT1 act.	y	+
[176]:	lung cancer	p53 act.	y	+
[218]:	glioblastoma	SREBP act.	y	+
[336]:	granulocytes	NF-κB act.	y	+
[33]:	B Lymphocytes	NF-κB act.	y	+
[265]:	B lymphoma	SREBP inh.	y	-
[335]:	glioblastoma	FOXO3 act.	y	+
		NF-κB inh.	y	-
[100]:	adrenocortical cells	SF-1 act.	y	n
[121]:	endothelial cells	EGR-1 act.	y	+
[271]:	synovial fibroblasts	SMAD act.	y	+
[348]:	Mouse Commd1 Null	HIF-1 act.	y	+
[334]:	Mouse TG2 KO hepatocytes	SP-1 act.	y	+
[355]:	osteosarcoma	E2F act.	y	+
[389]:	keratinocytes	AP-1 inh.	y	-
Expected hits (Found/Total)			18/18	16/18

resent background or previously unrecognized regulations. In summary, TFactS was able to identify expected transcription factor regulations, which, at least in some cases, were not found by tools based on PWM or consensus motifs, using standard settings.

Table 3.3: **TFactS comparison with other tools.**

Numbers represent the rank of the expected TF for each study, based on p-value or FWER p-value (GSEA) among the top 100 list of predicted TFs. For GSEA the rank represented here is the best rank in both positive and negative output lists. GSEA was run using either “c3” signatures or TFactS sign-less catalogue. (-) symbol means: not found and (N.A): not applicable (CRSD does not support mouse data; oPOSSUM does not integrate profiles for SREBP, SMAD and SF-1). Only significant TFs having nominal p-value ≤ 0.05 are listed in this table. See Table 2 (manuscript) for references of the studies, which are listed in the same order.

Ref.	Expected TFs	TFactS	Tfm-Explorer	Core_ TF	Crstd	oPOS-SUM	Gsea C3	Gsea TFactS
[82]	FOXO3	1 st	-	32 nd	-	-	-	-
	MYC	17 th	77 th	22 nd	3 rd	7 th	-	-
[68]	MYC	1 st	61 st	6 th	9 th	-	-	-
[90]	STAT1	1 st	-	-	-	2 nd	-	-
[176]	P53	1 st	94 th	-	3 rd	-	-	-
[218]	SREBP	3 rd	-	-	42 nd	N.A	-	1 st
[336]	NF- κ B	1 st	86 th	22 nd	90 th	1 st	16 th	1 st
[33]	NF- κ B	2 nd	48 th	1 st	1 st	1 st	-	1 st
[265]	SREBP	3 rd	-	55 th	20 th	N.A	-	-
[335]	FOXO3	2 nd	-	21 st	-	-	-	-
	NF- κ B	1 st	-	-	59 th	11 th	-	-
[100]	SF-1	3 rd	-	43 rd	-	N.A	3 rd	-
[121]	EGR-1	3 rd	62 nd	12 th	26 th	-	-	-
[271]	SMAD	1 st	-	-	12 th	N.A	-	-
[348]	HIF-1	1 st	-	3 rd	N.A	-	-	-
[334]	SP-1	15 th	1 st	56 th	N.A	12 th	-	21 st
[355]	E2F	3 rd	-	7 th	1 st	6 th	-	-
[389]	AP-1	13 th	30 th	-	67 th	-	-	6 th
Total hits		18/18	8/18	12/18	12/16	7/14	2/18	5/18

3.1.4.6 Analysis of NCI60 cancer cell line dataset

To test whether TFactS could be used to screen cancer microarray datasets, we analyzed the NCI60 [311] cancer collection, which contains 60 human cancer cell lines clustered in 9 tumor types (melanoma, ovarian, lung, leukemia, prostate, renal, colon, glioma and breast). On the basis of normalized log-ratio, we selected genes that were significantly regulated at least two-fold compared to the reference provided by Shankavaram et al. These lists of genes were analyzed by TFactS, which predicted a number of regulated TF in each cell line. In order to compare the different cell lines, a matrix of e-values (*Eval*) was constructed from these results with TFs in rows and cell lines in columns. We observed that some TFs, such as β -catenin, FOXO and NF- κ B, were found significantly regulated in

a large number of cell lines, as expected from their widespread role in cancer. Next, we applied the SAM algorithm to identify TFs differentially regulated between the 9 tumor groups. Among these TFs, SOX10, MITF and JUN were significantly regulated in melanomas in comparison with other cancer types (Figure 3.2). The specific role of MITF in melanocytes and melanoma is well known. By contrast, SOX10 is widely expressed and its unique association with melanoma was more surprising. Future work will have to unravel whether SOX10 may be a specific target for this cancer type.

Since the NCI60 dataset contains information about the presence of cancer mutations in each cell line, we looked for significant associations between these mutations and TF regulations. As the number of cell lines harbouring one particular mutation was low, we grouped mutations according to signal transduction pathways. Figure 3.2 B shows that mutations in RAS or RAF genes, key mediators of the MAP kinase pathway which are frequently mutated in melanoma and other tumors, were associated with the regulation of SOX10, MITF and JUN. The regulation of JUN and MITF by MAP kinases has already been described [178, 234]. In contrast, association of SOX10 with this pathway was not expected.

3.1.4.7 Analysis of novel microarray data sets

As an experimental validation for our approach we performed two microarray experiments in two distinct well-defined biological models: the human eosinophilic leukemia cell line EOL1 and primary human fibroblast cells stimulated with growth factors.

EOL1 cells express a constitutively activated FIP1L1-PDGFR α fusion protein [71]. This oncogene can be efficiently inhibited by the drug imatinib mesylate (Gleevec®). We compared gene expression profiles of EOL1 cells treated with imatinib for 4 hours to those of untreated cells. We selected genes that were significantly regulated at least 3-fold in control compared to imatinib-treated cells (supplementary Table 2). TFactS indicated a significant activation of STAT1, STAT3 and STAT5 (see Figure 3.3).

The FIP1L1-PDGFR α has previously been reported to activate STAT5 [71] and STAT3 [259], but to our knowledge the activation of STAT1 has not been established yet. We tested this prediction by Western blotting using antibodies that are able to recognize the active, phosphorylated form of STATs. Figure 3.3 shows that STAT3 and STAT1 are indeed phosphorylated in EOL1 cells and that this phosphorylation is inhibited by imatinib suggesting that it may also play a role in the transformation process. These results confirmed the validity of the TFactS target gene analysis.

We next tested the impact of the length of the input gene list on TFactS results. Starting from the list of significantly regulated genes ranked according to absolute fold-change, we submitted increasing numbers of top-ranked genes to TFactS (from 10 to 2500). Figure 3.3 shows that TFactS found STAT1 regulated if the number of submitted genes was superior to 50. Similar data were obtained for STAT3 and STAT5. We then performed a robustness analysis by adding randomly selected genes to the same data set. The addition of a background of up to 500 random genes affected the final result significantly

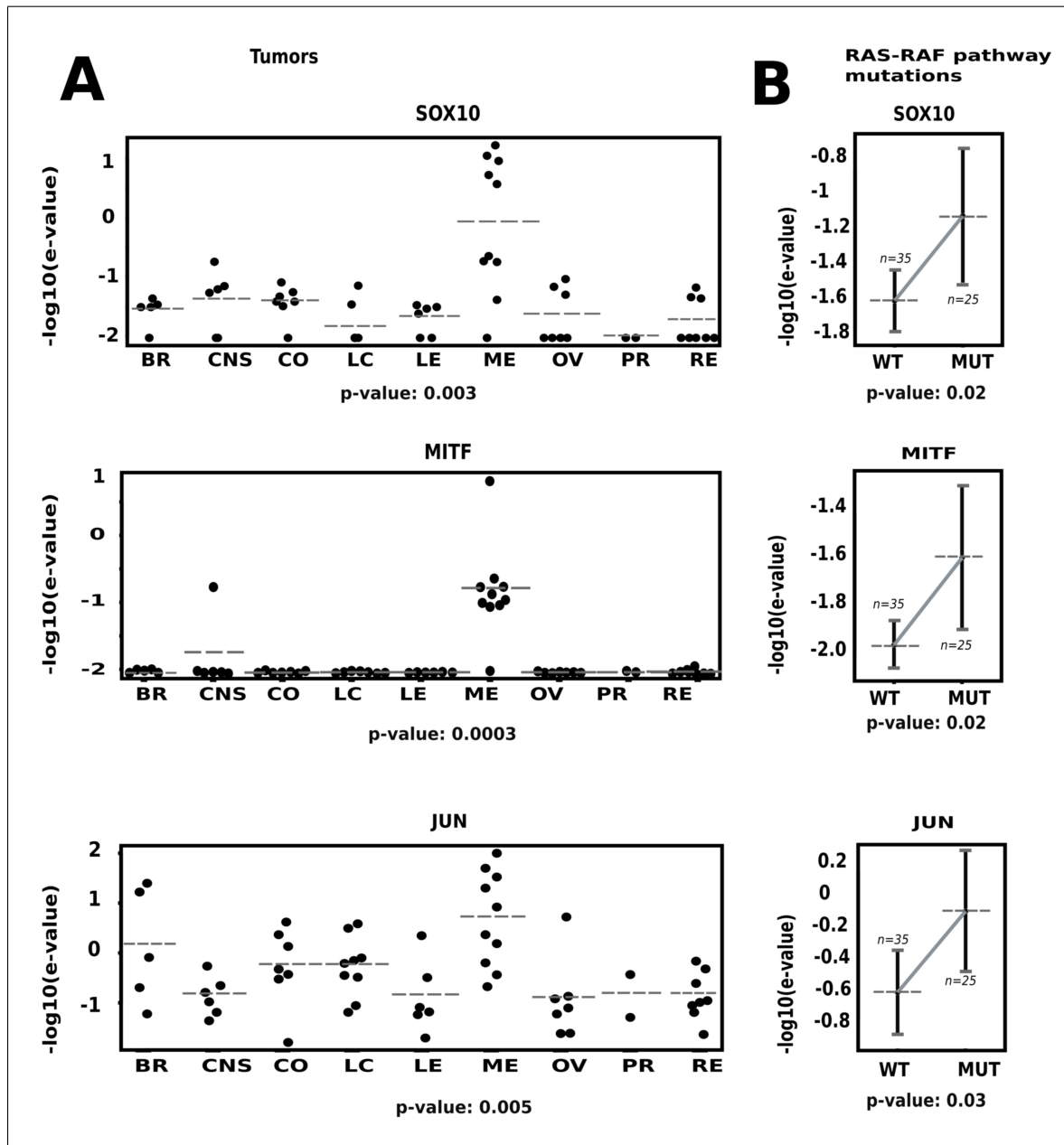


Figure 3.2: *SOX10*, *MITF* and *JUN* transcription factors are regulated in melanoma and associated to mutations in *RAS* and *RAF*.

E-values of TFactS-predicted TF regulations for each NCI60 cell line are transformed into scores ($-\log_{10}(\text{e-value})$). These scores were plotted by cancer type (A) or by pathway mutations (B) for *SOX10*, *MITF* and *JUN*. Mutations clustered into *RAF-RAS* pathway targeted *BRAF*, *KRAS*, *HRAS* and *NRAS*. *P*-values were obtained by Kruskal-Wallis test (A) and Student *t*-test (B). WT: “wild type”; MUT: “mutant”; BR: breast; CNS: central nervous system; CO: colon; LC: lung cancer; LE: leukemia; ME: melanoma; OV: ovarian; PR: prostate; RE: renal.

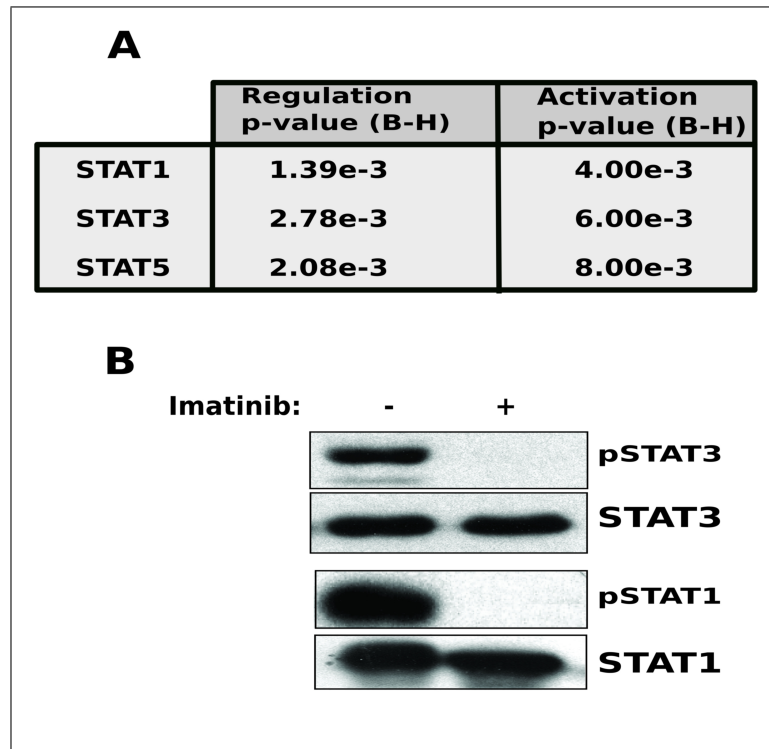


Figure 3.3: *STAT1 and STAT3 are phosphorylated in EOL1 cells and inhibited by imatinib.*

A- STAT activation Benjamini-Hochberg corrected p-values predicted by TFactS. B- Cell lysates from imatinib-treated or control EOL1 cells were used in western blot probed with antibodies against phospho-STAT3 and phospho-STAT1. As a control we used antibodies against total STAT3 and STAT1.

only when short lists of regulated genes were submitted (Figure 3.4), suggesting that TFactS can detect a TF signature even when it is diluted in a large number of unrelated genes.

We also analyzed gene expression data from human fibroblasts stimulated by either PDGF-BB or b-FGF for 1h and 24h. SAM analysis identified 115 relevant probe sets. Hierarchical clustering of these probe sets showed divergent early gene regulation by PDGF and FGF but overlapping late response (supplementary Figure 1).

We first analyzed genes commonly regulated by PDGF-BB and b-FGF more than 2-fold after 24h of stimulation. TFactS indicated that these two growth factors activate SREBP and inhibit FOXO (Table 3.4 and supplementary Table 3), as expected from our previous publications [84, 111, 174]. Our previous *in silico* promoter analysis did not point to any known TF regulation (unpublished). However, in a similar model, SREBP activation had been detected using a combination of promoter analysis and literature mining [307].

We then focused on the early gene expression response induced by both growth factors. A combined analysis with Affymetrix GCOS and a 2-fold threshold on expression ratio, led to the selection of 114 probe sets responding to PDGF-BB and 42 probe sets responding to b-FGF, 37 of which were shared between the two gene lists. We analyzed genes specifically regulated by PDGF-BB (and not

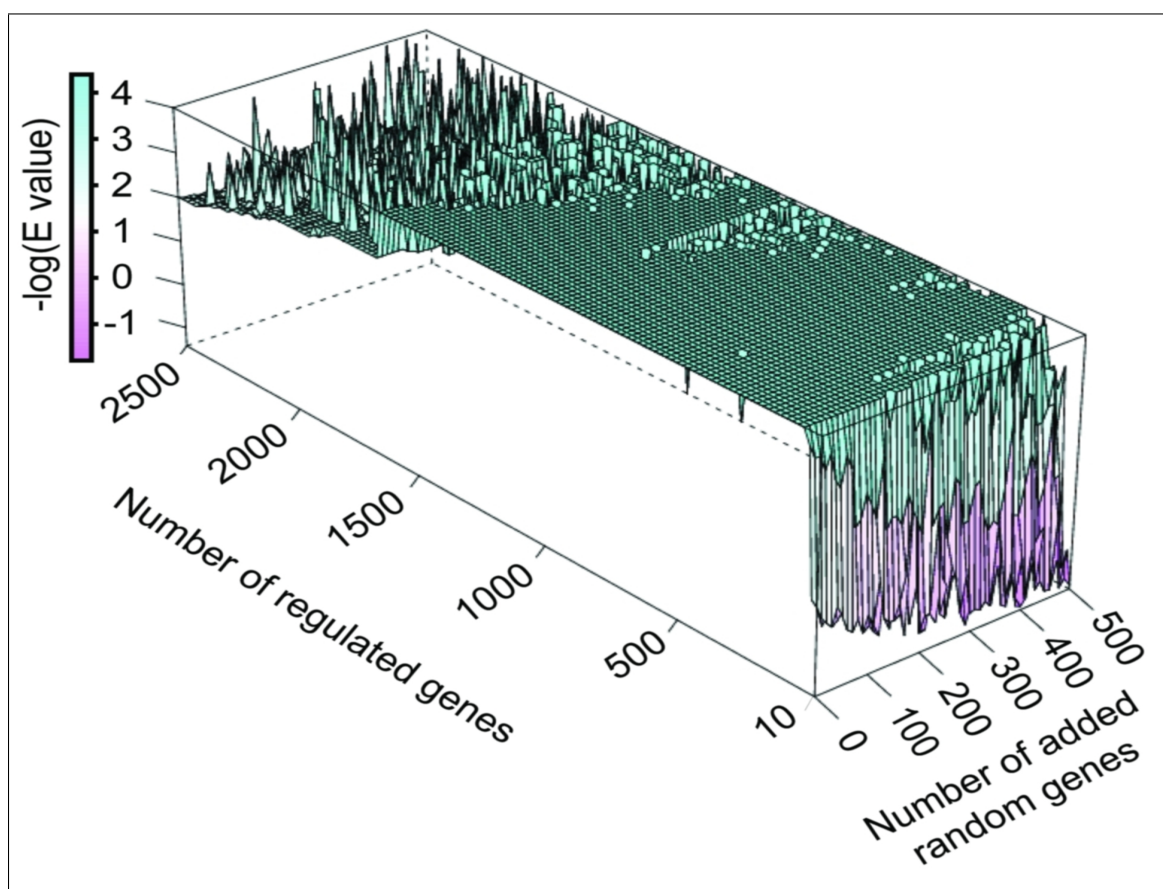


Figure 3.4: **Robustness analysis of TFactS.**

Significantly regulated genes in the EOL1 experiment were ranked according to the absolute value of the log2-ratio (untreated/treated). Then, the top 10 to 2,500 regulated genes were submitted to TFactS (sign-less). Each run was repeated after addition of 10 to 500 genes randomly selected. We plotted the number of regulated genes and the number of randomly added genes against a score ($-\log(e\text{-value})$) obtained for the STAT1 transcription factor.

b-FGF). TFactS detected several significant STAT signatures: STAT1, STAT3 and STAT5 (Table 3.4).

To further confirm these TFactS findings, we sought to cluster the conditions in this microarray experiment using the expression of predicted TF target genes that were also present in submitted lists. Figure 3.5 shows a time dependent clustering of the PDGF-BB and b-FGF conditions based on SREBP target genes, which is consistent with the late (24h) activation of SREBP by both growth factors. It shows also a successful clustering of the two PDGF-BB 1h stimulation replicates based on STAT target genes, confirming the predicted early (1h) specific activation of STATs by PDGF-BB.

It is well known that growth factors activate STAT [46, 149, 278, 380], but the specific regulation of these transcription factors by PDGF and not by FGF was unexpected. To test the specific action of PDGF-BB, we performed western blots against phosphorylated STAT1 and STAT3 in human fibroblasts after short term (15 min) stimulation with PDGF-BB or b-FGF. Figure 3.6 shows that PDGF-BB

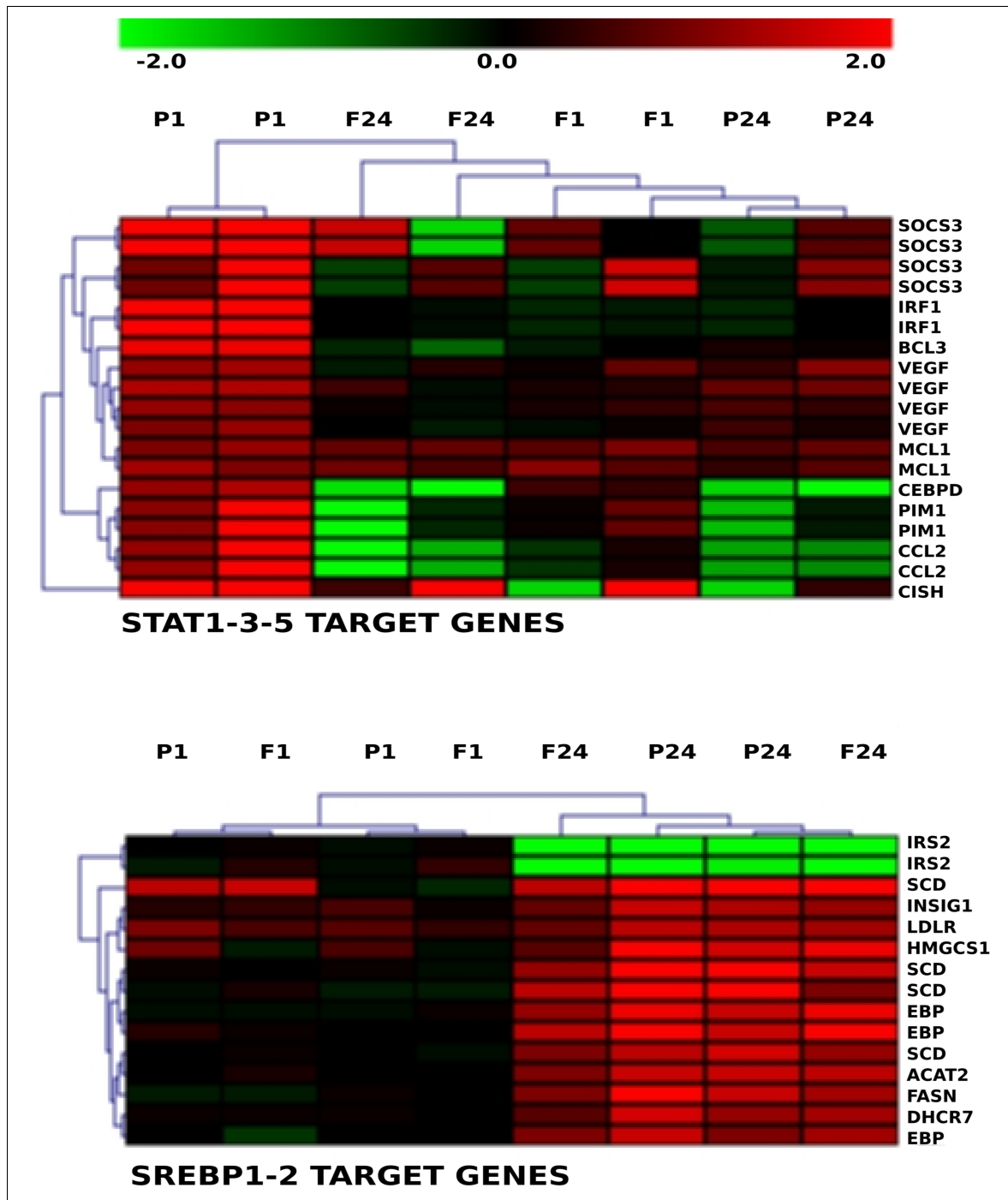


Figure 3.5: *Hierarchical clustering of STAT or SREBP target genes significantly regulated by PDGF-BB and b-FGF in human fibroblasts.*

STAT(1, 3 and 5) target gene signatures were pooled as well as SREBP(1 and 2) target genes. Several reports have shown that IRS2 gene expression is down-regulated by SREBP while other targets are up-regulated [174]. The intensities are in log2-ratios (color scale). Two replicate experiments are shown. F1: b-FGF(1h); F24: b-FGF(24h); P1: PDGF-BB(1h); P24: PDGFBB(24h).

Table 3.4: Analysis of genes regulated by growth factors in human fibroblasts using TFactS.

Analysis of significant gene lists regulated more than 2-fold by PDGF-BB and b-FGF after 24h (upper Table) or only by PDGF-BB after 1h (lower Table) of stimulation. For each test: regulation (sign-less), activation and inhibition (sign-sensitive), only transcription factors verified experimentally in this study or elsewhere are listed with corresponding statistics as reported by TFactS (detailed results are shown in supplementary Tables 3, 4 and 5). $\cap$: number of genes in common between TF target genes and the query list, # TG: number of Target genes for the corresponding TF.

Genes regulated by both PDGF and FGF after 24h						
	TF	<i>Pval</i>	<i>Eval</i>	$\cap$	# TG	<i>RC</i>
Regulation	SREBP	$0.00e^{+0}$	$0.00e^{+0}$	9	46	0%
	FOXO1	$0.00e^{+0}$	$0.00e^{+0}$	15	145	4%
	FOXO3	$2.00e^{-5}$	$1.90e^{-3}$	8	68	1%
Activation	SREBP	$1.00e^{-5}$	$3.00e^{-4}$	9	46	0%
Inhibition	FOXO1	$2.00e^{-4}$	$6.00e^{-3}$	14	144	0%
	FOXO3	$1.29e^{-3}$	$3.87e^{-2}$	8	66	0%
Genes regulated specifically by PDGF after 1h						
Regulation	STAT3	$0.00e^{+0}$	$0.00e^{+0}$	8	67	2%
	STAT1	$2.50e^{-4}$	$1.70e^{-2}$	4	61	1%
	STAT5	$6.00e^{-5}$	$4.08e^{-3}$	4	42	1%
Activation	STAT3	$0.00e^{+0}$	$0.00e^{+0}$	8	48	2%
	STAT1	$5.00e^{-5}$	$1.05e^{-3}$	4	23	1%
	STAT5	$5.40e^{-4}$	$1.13e^{-2}$	4	42	2%

induced a strong phosphorylation of STAT1 and STAT3, while b-FGF did not. We also included other fibroblast mitogens, such as IGF1, PDGF-AA, PDGF-AB, FGF-4 and insulin. None of these growth factors activated STAT1 or STAT3 to a significant extent. As a control for b-FGF activity, we checked other signalling pathways and showed JNK phosphorylation by both PDGF and FGF.

The current hypothesis is that growth factors activate largely overlapping signal transduction pathways [115]. This was confirmed for late gene regulation in our model. However, in this model, PDGF-BB stimulation specifically produced a STAT transcriptional signature, by contrast to other growth factors. A STAT3-dependent pathway has been suggested to contribute to PDGF-driven proliferation [46], but whether this cascade is shared by multiple growth factors is not clear. The mechanism by which PDGF receptors specifically activates STAT requires further investigations.

In conclusion, TFactS allowed us to pinpoint an unexpected specificity in the activation of STAT transcription factors by PDGF receptors in fibroblasts and leukemia cells.

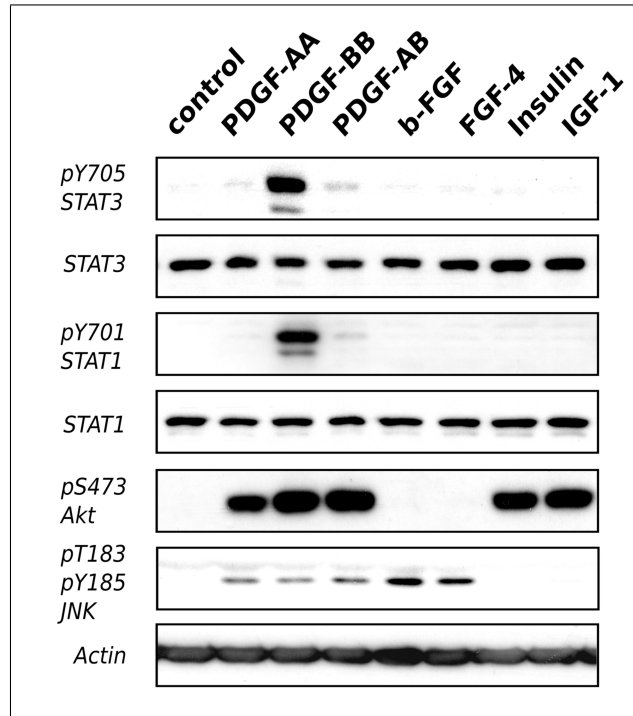


Figure 3.6: **STAT phosphorylation is specifically induced by PDGF-BB.**

Human fibroblasts were starved for 48h in serum free medium and stimulated with the indicated growth factors for 15 min. Total cell lysate was used for western blot with antibodies against specific phospho-STAT, phospho-JNK, phospho-AKT, total STAT and total JNK as indicated. Actin was detected as a control.

3.1.5 Conclusions

In this study, we showed that the regulation of TFs can be accurately predicted from transcriptome data, by comparing the lists of up- and down-regulated genes to a list of curated target gene signatures. We validated the approach using published microarray studies and new experimental data. It proved able to propose valid candidate TFs involved in the transcriptional response to a specific treatment (i.e. imatinib or growth factors). These predictions were relatively robust, although it is likely that the quality of the experimental microarray data and the normalization steps may significantly affect the final results. This applies to all the methods dealing with sets of co-expressed genes as input.

We also showed that TFactS can accurately predict the activity of TFs if based on a catalogue that is annotated with the sign of the regulation (activation or repression). However, we could not fully assess the performance of this feature, by comparing with other methods, because of the limited number of signatures annotated in the sign-sensitive TFactS catalogue.

TF signature validations were performed on well characterized TF for which a significant number of target genes are known (at least 21 in Table 1). In the NCI60 cancer data set analysis, MITF and SOX10 were represented by only 3 and 10 targets in the database, respectively, indicating that shorter signatures can also be useful. We are currently working on extending the number of regulations and

increasing the number of TFs in the catalogue. In addition, we offer the possibility to the scientific community to add their own annotated target genes either temporarily or permanently. This option is already implemented in the online version of TFactS. A limitation of our approach is the availability of high quality experimental target gene signatures. A few databases, such as TRANSFAC [371], TRED [390], TRRD [181], ORegAnno [237] and PAZAR [273] tackle this issue by gathering published promoter analysis. As mentioned above, it would be valuable to systematically annotate the sign of regulation in these databases, which has not been done so far. Obviously, much more work will be needed to identify the gene signature of all transcription factors present in mammalian genomes.

At this stage, we pooled data from experimental systems differing in terms of species (rat, mouse and human), cell types and environmental conditions, in order to get consensus signatures. The results from TFactS validation, based on multiple tissues, species and cell types, suggest the existence of robust core target signatures for the transcription factors that were analyzed. Nevertheless, the importance of these parameters will be carefully tested in future versions of the tool, in order to account for inter-species variations. Some of these parameters are already included in other databases [181, 237, 273, 371, 390].

Given the diversity in the parameters that can be optimized for each tool, it is difficult to produce an unbiased, fair comparison with the other tools that are able to predict TF activities, such as those based on *in silico* promoter analysis. In our hands, TFactS was able to predict TF regulations that, in a number of cases, could not be found using a few popular tools based on consensus motifs or matrices. We did not test more complex analysis, which were reported to improve results significantly [107, 208, 307, 317, 365]. Obviously, TFactS is not suitable for the identification of novel target genes, by contrast to PWM-based tools. Our data suggest that TFactS is a viable, simple alternative to promoter analysis methods to study differential activity of known TFs when sufficient known targets of the TF are represented in the database.

Altogether, we show here that comparing groups of genes showing a response in microarray data with experimental target gene signatures is an efficient way to predict the regulation and the level of activity (activation or inhibition) of well-characterized trans-acting factors. As a proof of concept, we validated a simple tool (TFactS) that combines Fisher's test and a curated target catalogues, which do or do not take into account the sign of the regulations. We also suggest that TFactS may contribute to the functional analysis of cancer microarray data.

3.2 *The Transcription of FOXO Genes is Stimulated by FOXO3 and Repressed by Growth Factors*¹

3.2.1 Abstract

Forkhead box O (FOXO) transcription factors induce cell growth arrest and apoptosis, which can be prevented by FOXO phosphorylation by AKT in response to growth factors such as platelet-derived growth factors (PDGF). Beside this well-characterized post-translational modification, we showed that FOXO1, FOXO3 and FOXO4 were also regulated at the transcriptional level. PDGF, FGF and IGF-I repressed the expression of FOXO genes in human fibroblasts. This process was sensitive to phosphatidylinositol-3 kinase inhibition by LY294002. FOXO1-specific shRNA decreased FOXO1 mRNA expression and enhanced fibroblast proliferation, mimicking the effects of growth factors. Conversely, ectopic FOXO3 activation blocked the proliferation of fibroblasts, and induced the expression of FOXO1, FOXO4 and p27-KIP1. Using luciferase reporter assays and chromatin immunoprecipitations, we identified a conserved FOXO-binding site in the promoter of the FOXO1 gene, which was required for regulation by PDGF, and mediated the up-regulation of FOXO1 by itself and by FOXO3. Altogether, our results suggest that the expression of FOXO1 and FOXO4 genes is stimulated by FOXO3 and possibly by other FOXO factors in a positive feedback loop, which is disrupted by growth factors.

3.2.2 Introduction

Forkhead transcription factors, which were initially described in *Drosophila melanogaster*, constitute a family of transcription factors that share a conserved DNA binding domain, the so-called forkhead box [156, 349]. The forkhead box O (FOXO) group comprises four homologous mammalian proteins, namely FOXO1 (also called FKHR), FOXO3 (FKHR-L1), FOXO4 (AFX) and FOXO6. Activation of these factors induces cell cycle arrest, DNA damage repair, differentiation and apoptosis. They also increase the resistance to oxidative stress, which was shown to be particularly important in hematopoietic stem cells, and regulate glucose metabolism in various organs. In *Caenorhabditis elegans*, the FOXO orthologue DAF-16 prolongs life span.

The activity of FOXO proteins is tightly controlled by multiple post-translational modifications [254, 356]. Growth factors, insulin and other cell stimuli induce FOXO phosphorylation and inactivation by AKT (also called protein kinase B or PKB), a serine/threonine kinase which is activated via the phosphatidylinositol-3 (PI3) kinase pathway [156, 254]. All FOXO proteins are substrates of AKT, which phosphorylates three conserved sites, resulting in the exclusion of FOXOs from the nucleus and in their subsequent ubiquitination and degradation. Phosphorylation by AKT may also regulate FOXO ability to bind to DNA [254]. In addition, the mitogen-activated protein kinases ERK and p38, as well as serum- and glucocorticoid-inducible kinase (SGK), DIRK1 and IKK β also inactivate FOXO1 and/or FOXO3 by direct phosphorylation [22, 156, 376]. By contrast, phosphorylation

¹Essaghir A, Dif N, Marbehant CY, Coffier PJ, Demoulin JB; J Biol Chem. 2009 Apr 17;284(16):10334-42.

by JNK kinases upon cell stress activates FOXO4 [113].

In the absence of growth factors, FOXOs reside in the nucleus and up-regulate genes which inhibit the cell cycle (p27 KIP1, p21 WAF1), promote apoptosis (Fas ligand, Bim, TRAIL) and decrease oxidative stress (superoxide dismutase, catalase) [156]. A number of genes are also repressed by activated FOXOs, including cyclin D1 and D2. The different members of the FOXO family share a common DNA binding site and regulate overlapping sets of target genes. FOXO transcriptional activity is further regulated by acetylation in the nucleus [254].

In line with their ability to block cell growth and to induce apoptosis, FOXO genes act as tumor suppressors. Deletion of all FOXO1, FOXO3 and FOXO4 alleles in adult mice induces a cancer prone condition, characterized by hemangiomas and thymic lymphomas [258]. In human cancers, several chromosomal translocations disrupt FOXO genes, producing hybrid proteins in which the forkhead DNA binding domain and the AKT phosphorylation sites are lost [122]. Pax3-FOXO1 and Pax6-FOXO1 were described in alveolar rhabdomyosarcomas, and the fusion of MLL with FOXO3 or FOXO4 was found in acute myeloid leukemias. In many tumor types, the constitutive activation of PI3 kinase, AKT and ERK is expected to inhibit FOXO proteins [122].

Not much is known about the regulation of FOXO mRNA expression. The transcription factor E2F1 was shown to induce FOXO1 and FOXO3 expression [252], and FoxC1 up-regulates FOXO1 in the eye [38]. In the present report, we observed that the RNA expression of FOXO1, FOXO3 and FOXO4 was repressed by growth factors such as platelet-derived growth factors (PDGF) and fibroblast growth factors (FGF), which induce cell proliferation. These growth factors play a key role in the development of the embryo and in various human diseases, including cancer [14, 114, 268]. PDGFs are dimeric ligands, which bind to specific receptor tyrosine kinases, PDGFR α and β , forming homo- or hetero-dimers [147]. They activate multiple signaling effectors, including PI3 kinase, AKT, ERK and phospholipase C γ [85, 147, 173, 332]. These signal transduction pathways regulate the activity of numerous transcription factors, controlling the expression of more than one hundred target genes [83, 84, 115, 174]. PDGF was reported to induce the phosphorylation and inactivation of FOXO1 protein in hepatic stellate cells, in fibroblasts and in vascular smooth muscle cells [1, 2, 18, 352, 353]. Here, we suggest that the inactivation of FOXO3 protein by PDGF disrupts a positive feedback mechanism, whereby FOXO3 stimulates the expression of FOXO genes.

3.2.3 Materials and methods

3.2.3.1 Cells and reagents

AG01518 human foreskin fibroblasts were grown in Quantum 333 fibroblast growth medium (PAA, Austria). PDGF-AA, -BB, FGF-2 (basic-FGF), FGF-4 and IGF-I were obtained from Peprotech (London, UK). Antibodies against FOXO and AKT were from Cell Signaling Technology and Santa Cruz. Anti- β -actin antibodies and cycloheximide were from Sigma. LY294002 was purchased from Calbiochem. BJ-hTert human fibroblasts, HEK293T and NIH3T3 were cultured as described [84].

Thymidine incorporation assays were performed as described with minor modifications [86]. Fibroblasts were seeded in triplicates in a 96-well plate in Quantum 333 (4000 cells/well). After one day, cells were washed and starved in serum-free Minimum Eagle medium (MEM, Sigma) for 24h. Growth factors (25 ng/ml) were then added with [³H]thymidine (0.5 μ Ci/well, GE Healthcare) for 24 h. Microtiter plates were harvested after a brief incubation with trypsin using a cell harvester (Perkin Elmer). The incorporation of [³H]thymidine was quantified using a TopCount instrument (Perkin Elmer).

3.2.3.2 Gene expression analysis

Subconfluent AG01518 fibroblasts were incubated for 48h in MEM with bovine serum albumin (0.05%). PDGF-BB (25 ng/ml) or b-FGF (10 ng/ml, heparin 50 μ g/ml) were added 1h or 24h before harvest. RNA extraction was performed using the RNeasy kit (Qiagen). The RNA quality was first tested in agarose gel. As a further control, the expression of known target genes NR4A1 and SCD was checked by quantitative RT-PCR [84]. Human Affymetrix chips HGU133A2 were hybridized following manufacturer instructions.

Results were analyzed using GCOS® suite software, and normalized using the control condition as baseline to calculate ratios [199]. In the MAS5 algorithm, the scale factor using all probe sets was set to 100 and the normalize factor was set to 1. The data was submitted to GEO (accession #GSE14256). We performed two biological replicates. Probe sets that were flagged ‘absent’ across all conditions in the two experiments were discarded. Quantitative PCR analysis was performed as described [83, 84, 199] using ABgene SYBR green-based kits (Thermo Scientific) and the oligonucleotides shown in supplementary table 1.

3.2.3.3 Lentivirus-mediated shRNA silencing

A set of 5 shRNA-specific clones for human FOXO1 was obtained from Sigma. All of them were packaged for viral production, infection and tested for target knockdown. Two constructs: TRCN0000039578 and TRCN0000039581 were used for further studies (Supplementary Table 1). A negative “scrambled” control was obtained from Addgene (#1864).

The shRNA-expressing and control lentiviral plasmids were transfected with the packaging plasmid pCMV-dr8.2 dvpr and the envelop plasmid pCMV-VSV-G [Addgene, #8455 and #8454, [322]] into HEK293T cells using Fugene 6 (Roche). Virus-containing supernatants were collected at 36 h and 60 h after transfection and used to infect fibroblasts in the presence of 8 μ g/ml polybrene (Sigma). Then, cells were incubated with standard growth medium containing 2 μ g/ml puromycin for 3 days prior to cell analysis.

3.2.3.4 Immunofluorescence experiments

NIH3T3 fibroblasts (200.000 per well) were grown on glass coverslips for 24h in 6 well plate. Then cells were co-transfected using Lipofectamine 2000 (Invitrogen) with pEGFP-N1 and pCMV6- -FOXO1 (Origene) or pCMV5-FOXO3 [Addgene, #14937, [309]] or empty vector. After 24h, cells were washed with PBS and starved in DMEM in the presence of BSA (0.05%) for 48h. Cells were stimulated with PDGF-BB (25ng/ml) for one hour, washed with cold PBS, and fixed in paraformaldehyde (4% in PBS) for 15 min. Cells were washed and incubated for 1 h at room temperature in PBS containing 5% fetal calf serum and 0.5% saponin. Cells were incubated overnight at 4°C with anti-FOXO1 or anti-FOXO3 antibodies in the presence of DAPI. Cells were then washed and incubated with secondary antibodies coupled to Alexa-Fluor (Molecular Probes). After 2 h, cells were washed, mounted on slides and visualized using a fluorescent microscope. Transfected cells were identified by GFP fluorescence.

3.2.3.5 Luciferase assays

The FOXO1 promoter was cloned by PCR from the BAC clone RP11-181D10 (BACPAC, <http://bacpac.chori.org/>) as template (see supplementary table 1 for primer sequences). The promoter was sequenced and subcloned into pGL3-Luciferase, using SmaI and KpnI restriction sites. The 588 bp mutant resulted from the digestion with EcoRI (within the FOXO1 promoter) and KpnI (within the plasmid). The 200bp mutant was generated similarly by digestion with ApaI and KpnI. The 132bp and 79bp mutants were amplified by PCR and subcloned into pGL3 (KpnI and HindIII sites). The F mutant was obtained by QuickChange mutagenesis (Stratagene). All constructs were sequenced.

Plasmids were linearized by BamHI and KpnI digestion before transfection. Subconfluent mouse NIH3T3 fibroblasts were transfected using Lipofectamine 2000 with 1 μ g pGL3 construct and 1.5 μ g pEF- β -galactosidase per well (6 wells plates) in the presence of 10% calf serum. After 24h of incubation, the cells were washed with PBS and starved in medium with 0.05% BSA for 24h and then stimulated with PDGF-BB (25 ng/ml) for 24 h.

Alternatively, subconfluent HEK293T cells were transfected using calcium phosphate with 0.125 μ g pGL3, 0.25 μ g pEF- β -galactosidase and FOXO1, FOXO3 or FOXO3-A3-ER [94]. Cells were incubated for 4h, washed with PBS and then stimulated with 4-hydroxy-tamoxifen 100 nM or left untreated. The luciferase activity was monitored using Bright Glo (Promega) and normalized using β -galactosidase activity as described [84, 174].

3.2.3.6 Chromatin immunoprecipitation (ChIP) assay

Chromatin cross-linking was performed with 1% formaldehyde and was stopped by the addition of glycine to a final concentration of 125 mM. Cells were harvested with cold PBS buffer containing protease inhibitors. Chromatin was fragmented by enzymatic digestion (Enzymatic shearing kit; Ac-

tive Motif, Belgium). An aliquot was used to check fragmentation on an agarose gel and to quantify the amount of DNA. Digested chromatin (50 μ g) was first precleared for 1 h with Ultralink protein A/G beads (Pierce). After centrifugation, supernatants were incubated overnight at 4 °C with anti-FOXO1 (H-128; Santa Cruz Biotechnology) or anti-FOXO3 (Cell Signaling Technologies) or anti-STAT5 (Santa Cruz Biotechnology), used as a control (3 μ g). The immunoprecipitated complexes were bound to protein A/G beads for 3 h at 4°C and washed successively in a radioimmunoprecipitation buffer (10 mM Tris, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 140 mM NaCl, 0.1% sodium deoxycholate and 1 mM PMSF, pH 7.4), LiCl buffer (1 mM Tris, 250 mM LiCl, 0.5% Nonidet P40, 0.5% sodium deoxycholate and 1 mM EDTA, pH 7.4) and Tris–EDTA buffer (1 mM EDTA and 10 mM Tris, pH 7.4). Proteins were then eliminated using proteinase K in the presence of 10% (w/v) SDS overnight at 37°C. After phenol extraction, DNA was precipitated, dissolved in water, and used as template for PCR with the primers shown in supplementary table 1. PCR products were visualized on 3% agarose gels. Alternatively, quantitative PCR was performed using the same primers in the presence of SYBER-green (ABgene kit) on an iCycler instrument (Bio-Rad). Enrichment of specific promoter regions after immunoprecipitation was expressed as the percentage of the total input chromatin.

3.2.3.7 Statistical analysis

Unless otherwise stated, one representative experiment out of at least three replicates is shown with standard deviations. Statistical significance was calculated according to Student *t*-test (*, p -value $\leq$ 0.05; **, p -value $\leq$ 0.01).

3.2.4 Results and discussion

3.2.4.1 Growth factors repress the expression of FOXO1, FOXO3 and FOXO4 genes.

In a microarray experiment using Affymetrix Genechips, we analyzed the regulation of gene expression by PDGF and FGF in AG01518 normal human fibroblasts. The detailed microarray results will be described elsewhere. We noticed that the expression of the genes FOXO1, FOXO3 and FOXO4 was decreased 24h after growth factor stimulation. To confirm this observation, we measured the RNA levels of FOXO genes by quantitative PCR. Figure 3.7 A shows that the expression of FOXO1 is repressed threefold after stimulation of AG01518 fibroblasts with PDGF-BB, FGF-2 or FGF-4. FOXO4 expression was also strongly repressed, while FOXO3 was regulated to a lesser extent by PDGF. FOXO6, which was described as a brain-specific isoform, was not expressed in these cells (data not shown). Similar results were obtained in BJ dermal fibroblasts immortalized with telomerase (BJ-hTert, Figure 3.7 A, right panels). IGF1 also repressed FOXO1 expression in a reproducible manner. In time course experiments, the expression of FOXO1 and FOXO4 was decreased 6 hours after the addition of PDGF (Figure 3.7 B). Shorter treatment periods did not significantly lower FOXO expression (data not shown). Next, the level of FOXO1 protein was visualized by Western blot in AG01518

cells stimulated for 24 h with PDGF-BB (Figure 3.7 C). The result confirmed the down-regulation of FOXO1 in PDGF-treated cells.

3.2.4.2 PDGF inactivates FOXO.

The activity of FOXO proteins is inhibited by phosphorylation upon growth factor stimulation. We confirmed the phosphorylation of FOXO1, FOXO3 and FOXO4 in AG01518 fibroblasts stimulated with PDGF by Western blot using antibodies directed against phospho-serine 256 of FOXO1 (which cross-react with FOXO4) or antibodies against serine 253 of FOXO3 (Figure 3.8 A). This conserved site is phosphorylated by AKT [254, 349], which was also activated in these cells upon PDGF treatment, as judged from blots with the anti-phospho-serine 473 antibodies.

It is well established that FOXO phosphorylation induces its translocation from the nucleus to the cytosol. Interestingly, this was reported in vascular smooth muscle cells treated by PDGF [1]. To confirm that PDGF affects FOXO localization in fibroblasts, we transfected NIH3T3 cells with FOXO1 or FOXO3, and studied the localization of the proteins by immunofluorescence. Figure 3.8 B shows that FOXO1 and FOXO3 were excluded from the nucleus upon treatment with PDGF for one hour.

To evaluate FOXO transcriptional activity in cells treated by growth factors, we analyzed the regulation of known FOXO target genes by PDGF-BB and FGF-2 in AG01518 fibroblasts using our microarray data (Figure 3.9). Based on published work, we identified 58 genes which were reported to be up-regulated by at least one FOXO isoform and were present in our initial microarray data, as well as 36 down-regulated genes. Most of the genes known to be induced by FOXO were repressed by stimulation of fibroblasts with PDGF and FGF-2 for 24 hours. Well-characterized FOXO target genes, such as p27/CDKN1B, SOD2, CITED2, GADD45A, CASP8 and BAX, were all repressed by growth factors. Conversely, the majority of the genes known to be repressed by FOXO were induced by growth factors, including cyclin D (CCND1, CCND2) and BIRC5. To exclude the possibility that this result was obtained by chance, we performed Fisher's statistical test, which gave a p-value of 1.6×10^{-4} . No significant effect of growth factors was observed after one hour of treatment (Figure 3.9). Altogether, these results suggested that the expression of FOXO targets was affected by PDGF and FGF-2 in a manner consistent with FOXO inhibition.

3.2.4.3 FOXOs play a key role in the proliferation of fibroblasts.

It is well established that FOXO factors induce cell growth arrest and apoptosis [152]. To test whether this also applies to human fibroblasts, we used a FOXO3-oestrogen receptor fusion protein (FOXO3-A3-ER), which can be activated by 4-hydroxy-tamoxifen [94, 235]. In this construct, the three AKT phosphorylation sites of FOXO3 are mutated to alanine, preventing the inactivation of the protein by growth factors. FOXO3-A3-ER was expressed in AG01518 using a retroviral vector. The number of viable cells was estimated by measuring ATP levels or by counting cells in the presence of Trypan

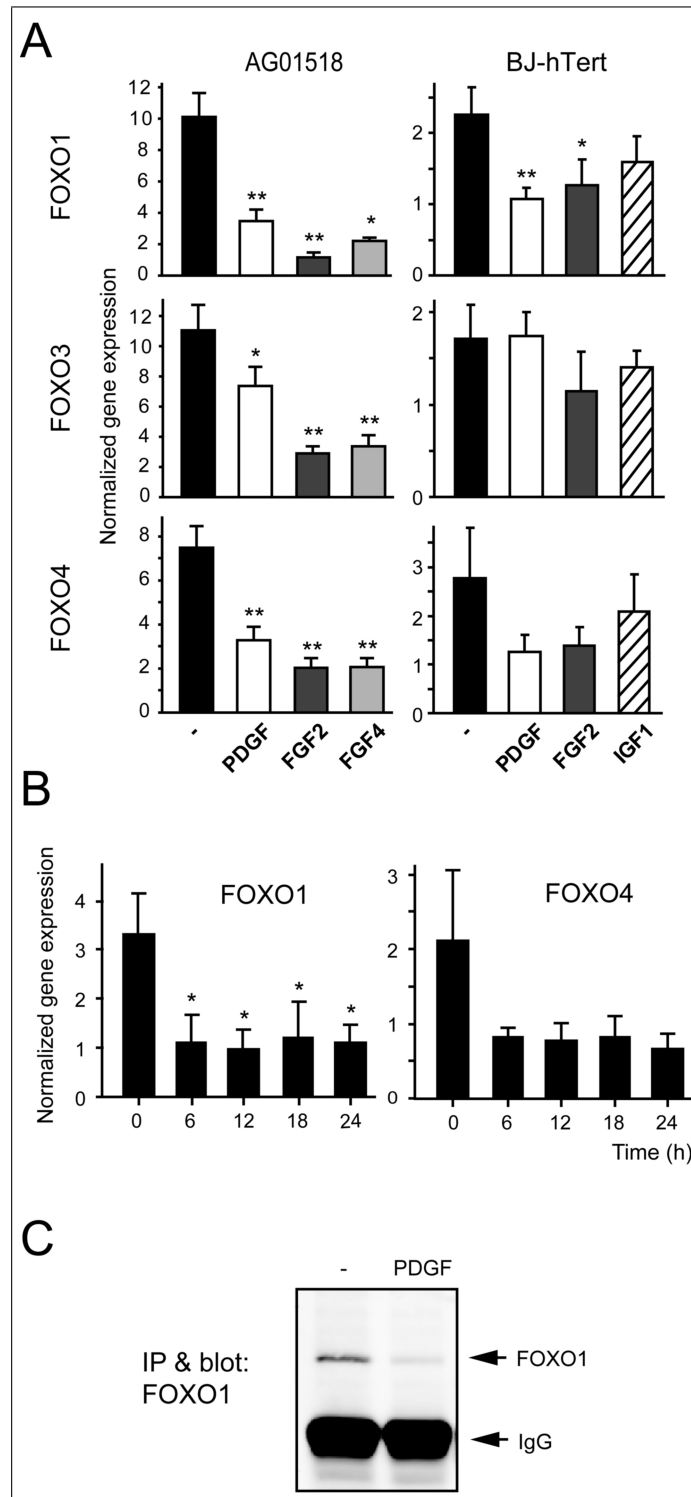


Figure 3.7: FOXO expression is down-regulated by growth factors in human fibroblasts.

A- AG01518 human fibroblasts (left panels) were starved for 24 hours in serum-free medium, and then stimulated for 24 hours with the indicated growth factor before RNA extraction. The expression of FOXO genes was monitored by quantitative PCR. The same experiment was performed with BJ-hTert cells stimulated for 16 hours (right panels). **B-** Starved AG01518 cells were stimulated with PDGF-BB for the indicated periods of time and analyzed as above. **C-** Starved AG01518 cells were treated with PDGF-BB or control medium for 24 hours. Cells were lysed in the presence of 1% Triton-X100. FOXO1 was immunoprecipitated and analyzed by western blot with specific anti-FOXO1 antibodies as described [84].

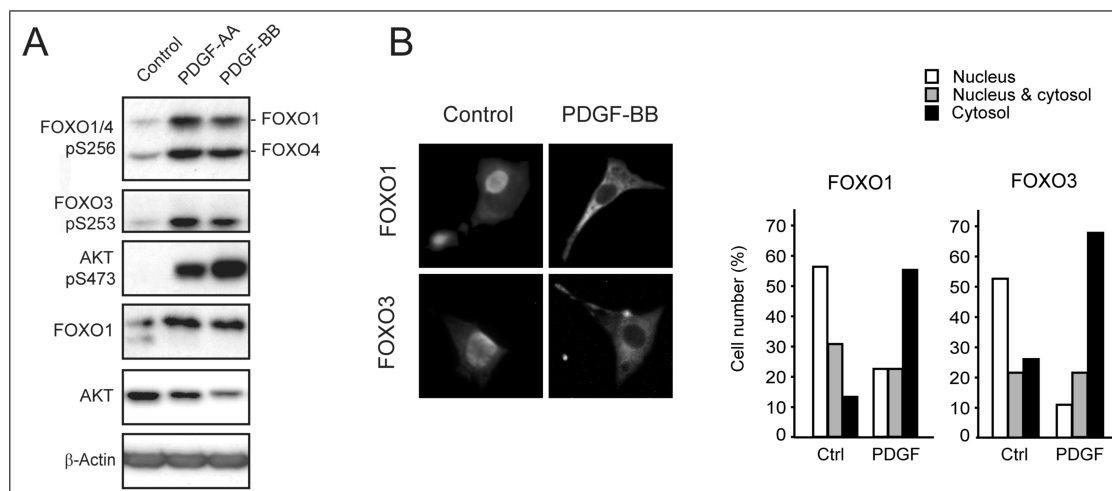


Figure 3.8: PDGF induces FOXO phosphorylation and translocation from the nucleus to the cytosol.

A- AG01518 were starved for 48h and stimulated with PDGF-AA or -BB for 15 min. Total cell lysates were analyzed by Western blotting with the indicated specific antibodies as described [173, 320]. **B-** NIH3T3 cells were grown on glass coverslips and transfected with FOXO1 or FOXO3. Cells were starved for 48 hours and then stimulated with PDGF-BB or control medium for 1 hour. Staining was performed with anti-FOXO1 or anti-FOXO3 antibodies. For each condition, the localization of FOXO was reported in more than 200 individual cells. The percentage of cells in which FOXO was exclusively located in the nucleus (white bar), exclusively in the cytosol (black bars) or diffuse in the whole cell (gray bars) is shown for a representative experiment.

blue (Figure 3.10). Activation of FOXO3-A3-ER by 4-hydroxy-tamoxifen blunted the effect of PDGF and FGF-2 on fibroblast proliferation.

As FOXO proteins are tightly regulated by several post-translational mechanisms, it was not clear whether a change in FOXO RNA level affects cell growth. To tackle this question, we used the pLKO lentiviral vector, which is designed to encode short inhibitory RNA hairpins (shRNA). Two specific FOXO1 shRNA reproducibly reduced FOXO1 mRNA expression measured by quantitative PCR, while a control sequence had no effect (Figure 3.11 A). When fibroblasts were infected with both constructs together, FOXO1 expression was repressed by 63%. Remarkably, this is comparable to the effect of PDGF on FOXO1 expression. Transduction of these two FOXO1 shRNA alone or in combination increased the basal proliferation of AG01518 fibroblasts, as demonstrated by the thymidine incorporation experiment shown in Figure 3.11 B. PDGF stimulated fibroblast proliferation more than twofold in the presence of control shRNA, but did not significantly increase cell growth in the presence of FOXO1 shRNA. In conclusion, these data suggested that decreasing FOXO1 expression is enough to enhance fibroblast cell growth.

3.2.4.4 The regulation of FOXO1 by PDGF is sensitive to PI3 kinase inhibition.

To decipher the mechanism of regulation of FOXO expression by PDGF, we first tested whether *de novo* protein synthesis was required. As shown in Figure 3.12 A, PDGF was able to decrease FOXO1 expression in the presence of cycloheximide, a potent protein synthesis inhibitor, suggesting

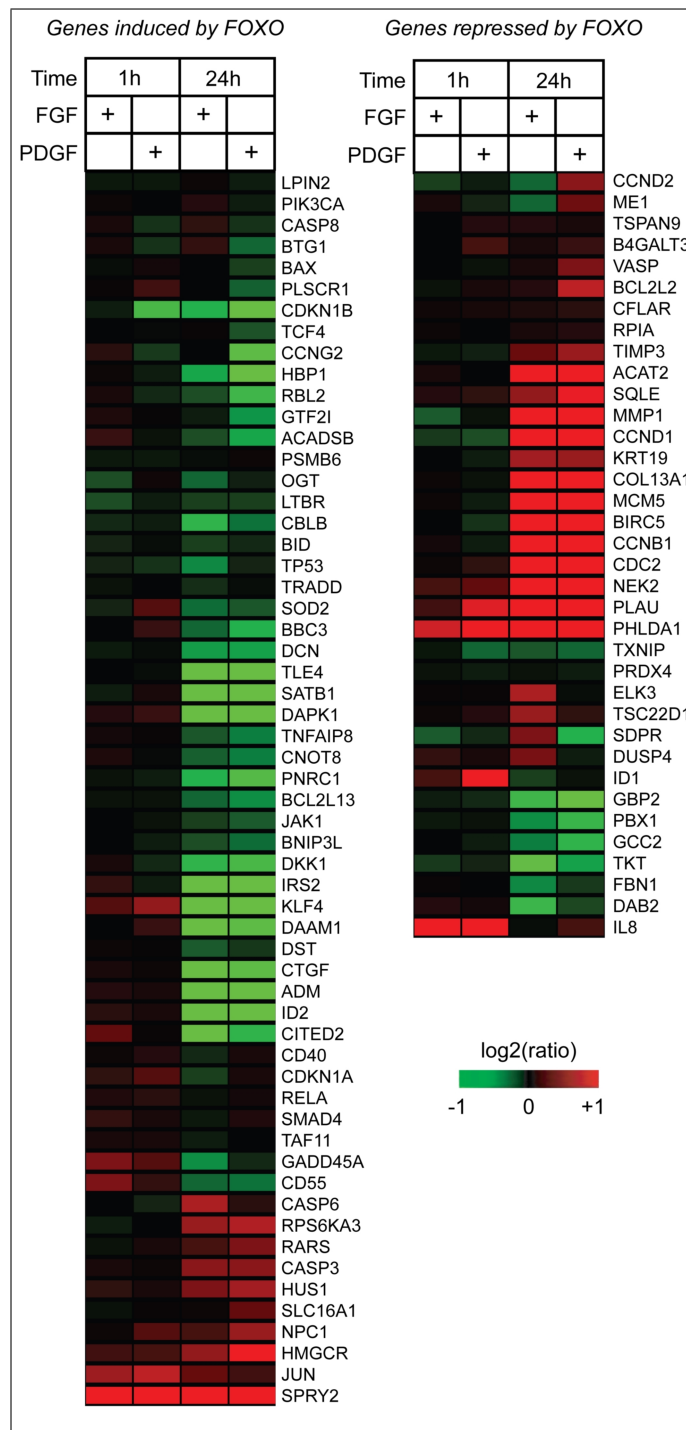


Figure 3.9: Regulation of FOXO target genes by growth factors in fibroblasts.

Starved AG01518 fibroblasts were stimulated with PDGF-BB or FGF-2 for one or 24 hours. RNA was extracted to perform Affymetrix Genechip hybridizations. Results from two independent experiments were expressed as average log₂ ratios, using untreated cells as reference. Only published target genes of FOXOs are shown.

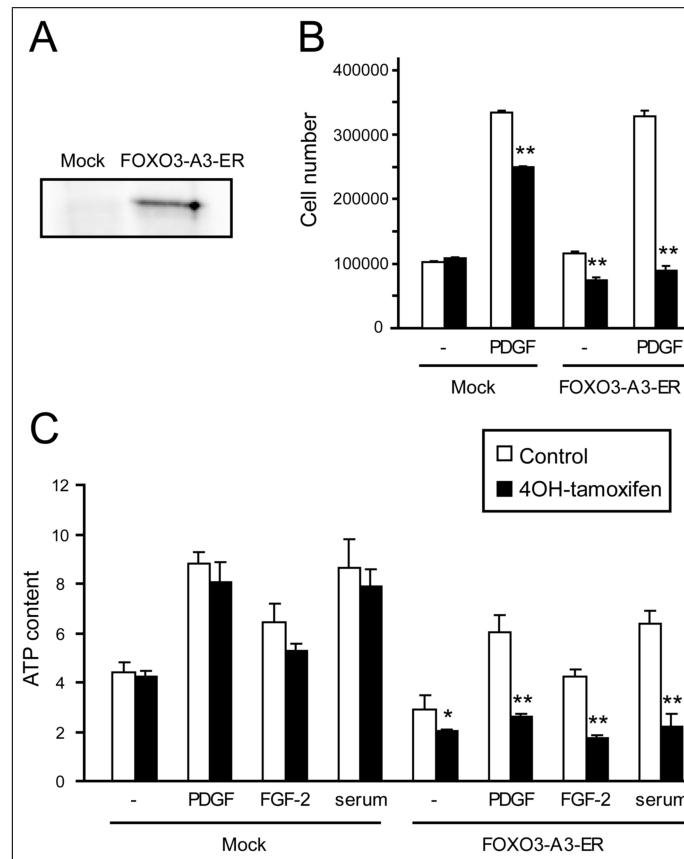


Figure 3.10: Activation of FOXO3 prevents fibroblast proliferation upon growth factor stimulation. AG01518 cells were infected with retroviruses expressing FOXO3-A3-ER or control retroviruses (empty pBABE vector) and selected in the presence of puromycin. **A-** FOXO3-A3-ER expression was detected by Western blot using specific anti-FOXO3 antibodies. **B-** Cells were incubated in the presence of PDGF-BB for 5 days in serum-free medium with 4-hydroxy-tamoxifen (black bars) or vehicle (white bars). Living cells were counted in the presence of Trypan blue. **C-** Cells were incubated in the presence of control medium, FGF-2, PDGF-BB or serum (10%) with 4-hydroxy-tamoxifen (black bars) or control medium (white bars) for three days (triplicate cultures). The ATP content of each well was tested by Cell-Titer Glow (Promega).

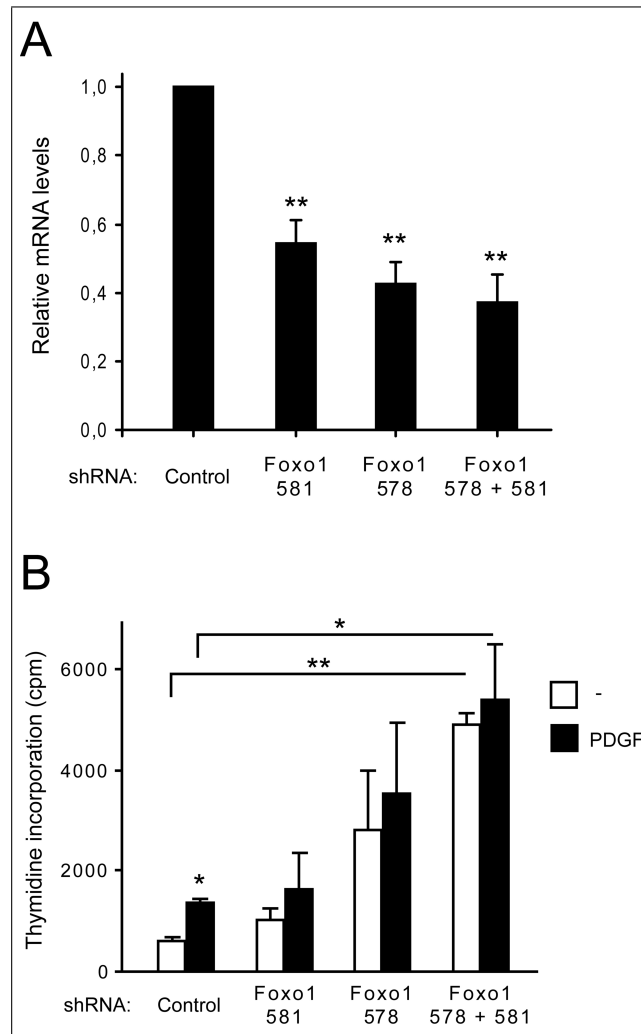


Figure 3.11: *FOXO1*-specific shRNAs enhance fibroblast growth.

A- AG01518 cells were infected with lentiviral particles expressing control shRNA, or one of the two *FOXO1* shRNA or both at the same time. RNA was isolated 72 h after infection, and quantitative PCR was performed to detect *FOXO1* and *RPLP0* as a control. **B-** Proliferation of infected fibroblasts was assessed by a [3 H]thymidine incorporation assay. Starved cells were stimulated with PDGF (black bars) or control medium (white bars) in the presence of [3 H]thymidine. In each panel, the average of three independent experiments is shown with SEM.

the existence of a direct regulatory mechanism.

We next sought to identify which signal transduction pathway was required in this process, using small inhibitory molecules. The PI3 kinase inhibitor LY294002 abolished the regulation of *FOXO1* by PDGF in AG01518 cells, while the mitogen-activated protein kinase pathway inhibitor U0126 had no effect (Figure 3.12 B). Similar results were obtained in BJ-hTert cells (Figure 3.12 C). LY294002 also blocked the down-regulation of *FOXO4*. We had observed previously that LY294002 prevents the activation of AKT by PDGF in AG01518, as expected [[84] and unpublished data]. As PI3 kinase is also required for *FOXO* phosphorylation and inactivation by PDGF [2, 353], our data raised the

possibility that FOXO proteins stimulate the expression of their own genes, a process that would be disrupted upon AKT activation by growth factors.

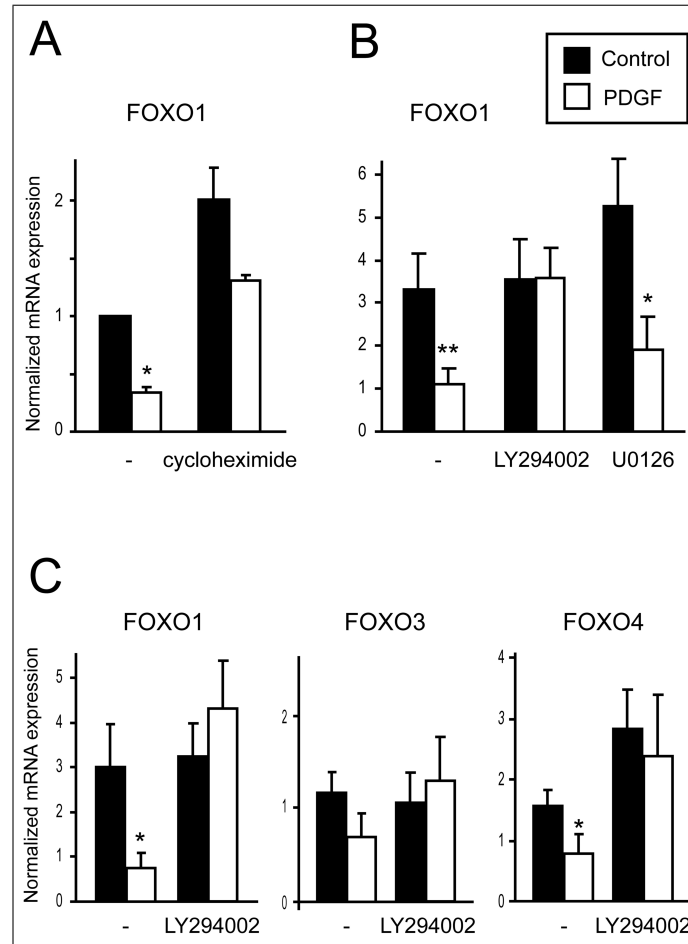


Figure 3.12: Role of PI3 kinase in the FOXO repression by PDGF.

A- BJ-hTert fibroblasts were starved for 24 h, and incubated with cycloheximide (15 ng/ml) for 1h. Cells were treated with PDGF-BB (white bars) or control medium (black bars) for 8 h. RNA was extracted and the expression of FOXO1 and RPLP0 was quantified by PCR. The average of two independent experiments is shown with SEM. **B-** AG01518 fibroblasts were starved for 24 h and stimulated with PDGF-BB for 24 h in the presence of LY294002 (20 μ M), U0126 (10 μ M) or control medium for 24 h. **C-** BJ-hTert cells were treated for 6 hours as described in panel B.

3.2.4.5 FOXO1 transcription is regulated by FOXO3.

To test this hypothesis, we first analyzed the expression of FOXO genes in AG01518 fibroblasts transduced with FOXO3-A3-ER. The activation of FOXO3-A3-ER by 4-hydroxy-tamoxifen induced the expression of FOXO1 and FOXO4, as shown by quantitative PCR (Figure 3.13). The p27-KIP1 gene, a well characterized target of FOXO3, was used as a positive control [94]. As shown in Figure 3.13, it was regulated to a similar extent as FOXO1 and FOXO4. By contrast, we did not observe any regu-

lation of endogenous FOXO3, which was amplified with probes hybridizing with the 3' untranslated region of FOXO3, absent in the FOXO3-A3-ER construct. The expression of FOXO3-A3-ER was controlled using FOXO3 oligonucleotides that were able to amplify the retroviral construct. Collectively, these data indicated that FOXO3 activation can induce the expression of FOXO1 and FOXO4.

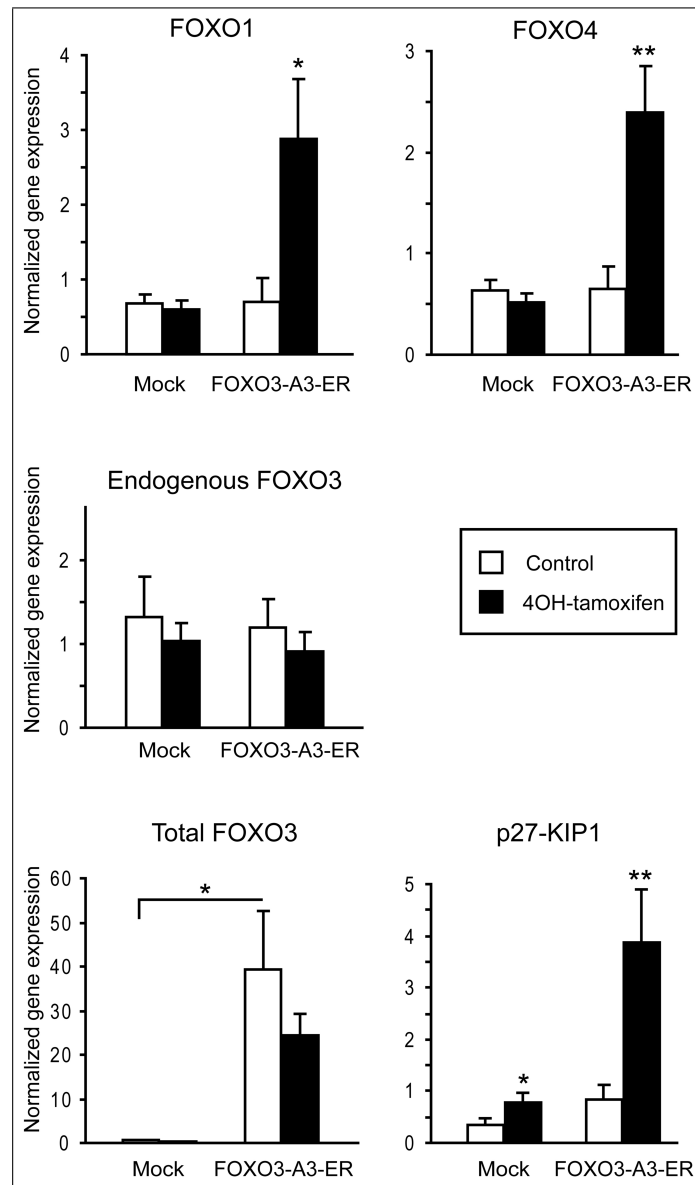


Figure 3.13: **FOXO3 can induce FOXO1 and FOXO4 expression.**

AG01518 cells were infected with FOXO3-A3-ER retroviruses or controls. After 48 hours, cells were treated with 4-hydroxy-tamoxifen (black bars) or control medium (white bars) for 24 hours. RNA was extracted and the expression of FOXO genes, p27-KIP1 and RPLP0 was measured by quantitative PCR.

We next analyzed the FOXO1 promoter. We isolated a 1238-base pair fragment of the human FOXO1 gene promoter (Figure 3.15 A), which presents a relatively high level of homology with the

mouse sequence. This fragment was cloned in front of a luciferase reporter gene and co-transfected with FOXO3-A3-ER in HEK293T cells. In the presence of 4-hydroxy-tamoxifen, FOXO3-A3-ER enhanced FOXO1 promoter activity (Figure 3.15 B). We next performed serial 5' deletions of the promoter (Figure 3.15 A) and analyzed the effect of FOXO3 activation. Deletion of 650 nucleotides did not affect the promoter activity or its response to FOXO3 activation (Figure 3.15 B). By contrast, FOXO3 activation had a much reduced impact on a 200-base pair promoter, and did not stimulate the activity of promoters retaining only 132 or 79 base pairs. The *in silico* analysis of the FOXO1 promoter revealed a putative FOXO binding site located 370 nucleotides before the first exon. This sequence is conserved in the mouse and rat genome (see Figure 3.14). Based on the consensus FOXO binding site, we mutated two key adenines into cytosines (Figure 3.15 A). A full length 1238-bp promoter harboring this mutation responded only weakly – but significantly ($p < 0.01$) – to FOXO3-A3-ER activation, compared to wild-type (Figure 3.15 B).

To test if FOXO1 could regulate the expression of its own gene, we transfected HEK293T cells with wild-type FOXO1 and compared it with wild-type FOXO3, in a luciferase assay (Figure 3.15 C). Transfection of FOXO3 strongly stimulated FOXO1 promoter activity, confirming the experiments performed with FOXO3-A3-ER. FOXO1 stimulated the promoter of its own gene significantly, but to a lesser extent compared to FOXO3. Mutation of the FOXO binding site of the FOXO1 promoter abolished the effect of FOXO1 and decreased the effect of FOXO3.

To assess whether FOXO1 and FOXO3 were binding directly to the FOXO1 promoter, we performed chromatin immunoprecipitation experiments with anti-FOXO antibodies. A 108-bp DNA fragment that included the FOXO consensus site of the promoter was specifically associated with FOXO1 in HEK293T cells (Figure 3.15 D). This effect was increased by FOXO1 transfection. FOXO3 was also bound to the same promoter fragment. Like in luciferase experiments described above, the highest signal was observed in cells transfected with FOXO3-A3-ER and stimulated with 4-hydroxy-tamoxifen. In conclusion, our data suggest that FOXO1 expression is up-regulated upon binding of FOXO1 and FOXO3 proteins to the FOXO1 gene promoter. Nevertheless, a second regulatory mechanism may contribute to FOXO1 regulation independently of the consensus FOXO binding site that we identified, as suggested by the effect of FOXO3 on the mutant promoter.

3.2.4.6 PDGF blunts the activation of the FOXO1 promoter by FOXO factors.

To investigate whether this mechanism could account for the regulation of FOXO1 by PDGF, luciferase reporter assays were performed in transfected NIH3T3 fibroblasts. We observed that PDGF significantly repressed the transcriptional activity of the FOXO1 promoter (Figure 3.16 A). The activity of the mutant FOXO1 promoter, which lacks the consensus FOXO binding site, was not stimulated by PDGF. Next, we tested whether PDGF affected FOXO recruitment to the FOXO1 promoter by chromatin immunoprecipitation. We observed that FOXO1 and FOXO3 were associated to the FOXO1 promoter in AG01518 fibroblasts (Figure 3.16 B), in agreement with the results that we obtained in HEK293T cells. Fibroblast stimulation with PDGF decreased the binding of FOXO1 and

FOXO1A-human	GCCCTGCCAGCCCGCCGGGCTCG-GCCCGCAGCGGAGGTCAGGGAGCAGC	-423
Foxo1-mouse	CCACCGCCACGTCGCGTCCGGGAGCGCGCAGCGGCC-CCGCGCCATCCC	-403
Foxo1a-rat	CCACCGCCACGTCGCGTCCGGGAGCGCGCAGCGGCT-CCGCGCCATCCC	-355
	* * * * *	
FOXO1A-human	GAGCGCTCCTCCACCCGGGCTTCCCGAGTACTCGGCTCTGCTGCTCCG	-373
Foxo1-mouse	GGGAGTCAGGGTCTGGGCTC-TGCCCTGGAGTACTCGATTGCACT--CTG	-357
Foxo1a-rat	GGGAGTCAGGGTCTGGGCAC-TGCCCTGGAGTACTCGATTGCACT--CCG	-309
	* * * * *	
FOXO1A-human	<u>TAGTAAACAAA</u> -GTGTCGCCCGCGCCTCCACGCTGGTTTGCTTCCTAGCA	-324
Foxo1-mouse	TAGTAAACAAAAGTATGGTAGCCGCTCCACGCGGTTTGCTTCCTAGCA	-307
Foxo1a-rat	TAGTAAACAAAAGTATGGTCGCCCGCTCCACGCGGTTTGCTTCCTAGCA	-259

FOXO1A-human	<u>ATCAAAACACTGAGAAGGCAGAGAAATCACAGAAACACTC</u> GAGAATTACC	-274
Foxo1-mouse	ATCCGATCGCTGA--GACGCCAGAGACTCAGAAACACTTGAGAATTACA	-260
Foxo1a-rat	ATCCGAACACTGA--GGCGCCAGAGACTCAGAAACACTTGAGAATTACA	-212
	*** * * * *	
FOXO1A-human	AGAAAATAATAGAGATCCAAAAAAGGAGGTGAGTGTGTGAAAG	-224
Foxo1-mouse	GAAAATAATAACGAGATTGGAAAAAAGGAGGTGAGTGTGTGAGAA	-214
Foxo1a-rat	GAAAATAATAACGAGATTAGAAAAAAGGAGGTGAGTGTGTGAGAA	-171
	*** * * *	
FOXO1A-human	AGAAAAACACCCCACTACCCCAACCCAGCCACCGCCGCTCCCGTGGA	-174
Foxo1-mouse	AGAAAAATACCCCAACCCCAACCAACCTACTGCCGCTCCC-GAGGG	-165
Foxo1a-rat	AGAAAAACACCCCGCTGCCCAACCAACCTACTGCCGCTCCC-GAGGG	-122

FOXO1A-human	AAACCGGGCCCCACCCAGCCCGGCGCCCACTGGCTGCCCGGGCGGCGGTG	-124
Foxo1-mouse	AAGGCAGCGCCCGCCAGCCCGCGTTCATTGGCTGCTTCGGCGGAGACT	-115
Foxo1a-rat	AAAGGAGCTCCCTCCAGGCGCGTTCATTGGTTGCTTAGGCAGAGACT	-72
	** * * * *	
FOXO1A-human	CCGCATGCCCATTTGGCCGCGCGGCTGTCTGGTCAAGGGCGGGCGGCGCG	-74
Foxo1-mouse	TCGCGCGTCCATTGGCCGCGCAGCCTGTCAGTCAGGGCGGGCGGCGCG	-65
Foxo1a-rat	TGGCGCGTCCATTGGTTCGCGCAGCCTGTCAGTCAAGGGCGGGCGGCGCG	-22
	** * * * *	
FOXO1A-human	CGCGCCCGCGGGCGGGGGCG-GCGGCAGATCCCGTAAGTCGGGCGGC	-25
Foxo1-mouse	CGCGCCCGCGGGCGGGGAGCCGCTGCAGATCCCGTAAGACGGGAGTC	-15
Foxo1a-rat	CGCGCCCGCGGGCGGGGAG-----	-1

Figure 3.14: *Comparison of the human, mouse and rat FOXO1 promoter sequences.*
The alignment was performed using Clustal 2.0. The consensus forkhead box is shown in red. The DNA fragment amplified by PCR in ChIP experiments is underlined.

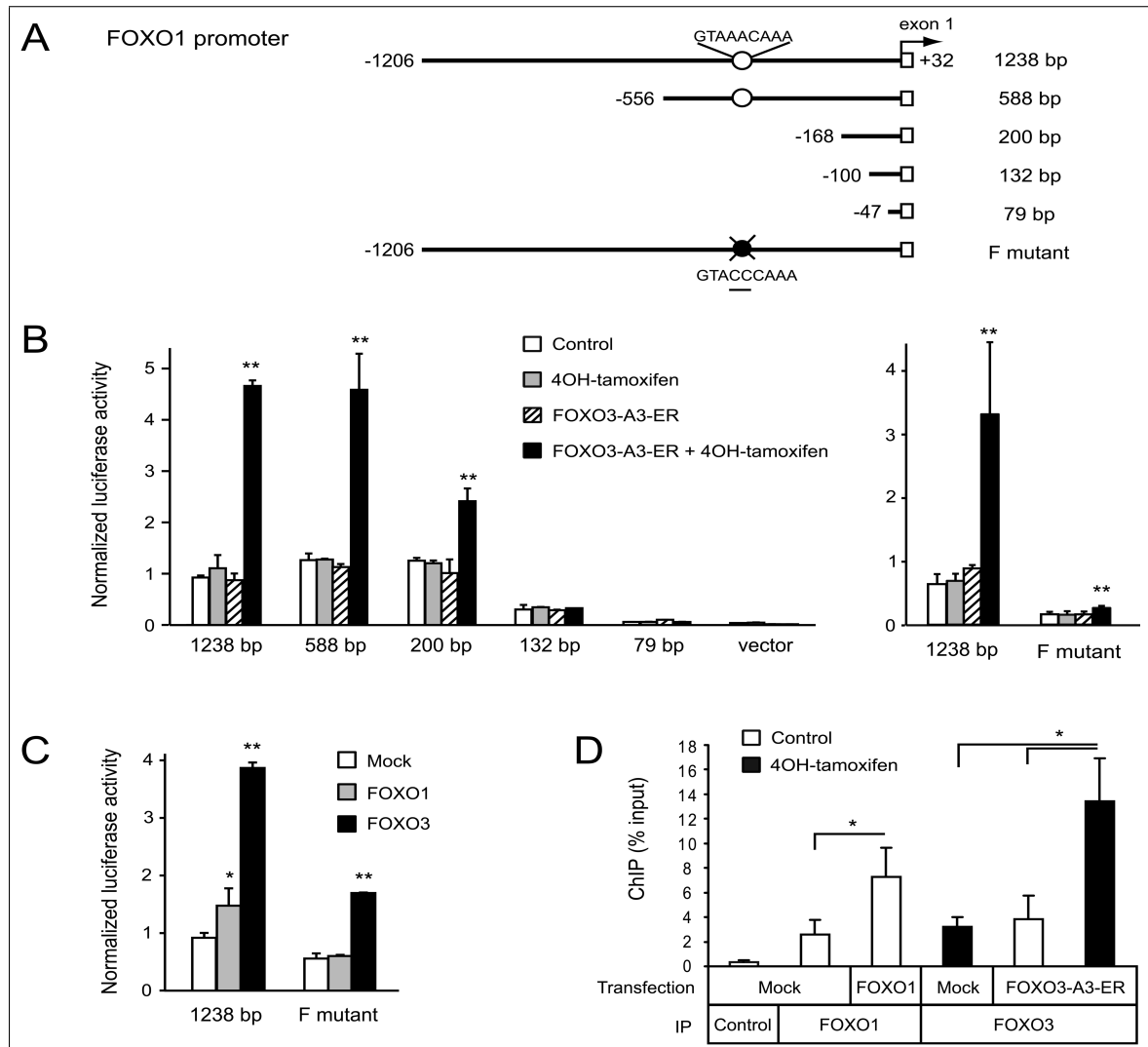


Figure 3.15: Identification of a FOXO-binding site in the FOXO1 promoter.

A- Fragments of FOXO1 promoter were cloned upstream of the luciferase reporter gene. **B-** Linearized constructs were transfected in HEK293T cells with Foxo3-A3-ER. Cells were treated with 4-hydroxy-tamoxifen or left untreated for 24h before measuring luciferase and β -galactosidase activities. **C-** HEK293T cells were co-transfected with FOXO1 promoter constructs and either wild-type FOXO1 or FOXO3. **D-** Chromatin immunoprecipitation assays for FOXO1 and FOXO3 binding to the FOXO1 gene promoter. After cross-linking chromatin to DNA-interacting proteins, immunoprecipitations of FOXO1 or FOXO3 proteins using specific antibodies were performed. Binding of FOXO proteins was measured by quantitative PCR amplification of the FOXO1 gene promoter; as shown in Figure 3.14. The average values of three independent experiments, expressed as percentages of the input DNA, are shown with SEM.

FOXO3 significantly (Figure 3.16 B). Altogether, our data suggested that PDGF dampen the activation of FOXO1 promoter by FOXO proteins.

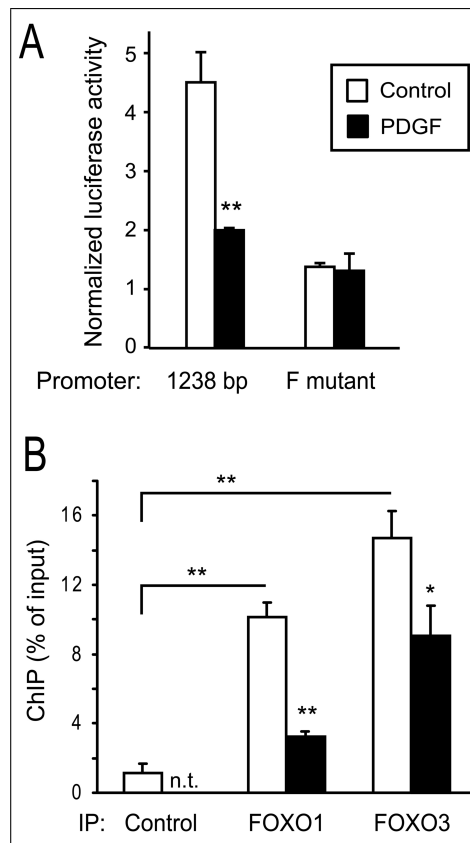


Figure 3.16: PDGF inhibits FOXO binding to the FOXO1 promoter.

A- NIH3T3 mouse fibroblasts were transfected with the indicated FOXO1 promoter construct and stimulated with PDGF-BB for 24h. Luciferase activities were measured as described in material and methods. One representative experiment is shown with SD. **B-** AG01518 fibroblasts were starved for 48h and stimulated for 24h in the presence of PDGF (black bars) or control medium (white bars). Chromatin immunoprecipitations were performed as described in Figure 3.15. The average values of three independent experiments, expressed as percentages of the input DNA, are shown with SEM. N.t., not tested.

As mentioned above, the inactivation of FOXO proteins by growth factors has been ascribed to the phosphorylation of FOXOs and their subsequent translocation from the nucleus to the cytosol [54, 156]. This well-established pathway likely accounts for the inhibition of FOXO proteins by PDGF in the present study (Figure 3.17). In addition, Aoki *et al.* showed that PDGF induces the degradation of FOXO1 protein by proteasomes [18]. We observed that treating fibroblasts with the proteasome inhibitor MG132 blocked the repression of FOXO1 mRNA expression by PDGF (Figure 3.18), suggesting that the degradation of FOXO by proteasomes may play a role in our model. Altogether, these observations are in accordance with the model proposed by Matsuzaki and colleagues, who showed that both cytosolic retention and degradation of FOXO1 are required for its efficient inactivation by insulin [229]. We can not rule out the involvement of additional layers of

regulation, at the level of FOXO DNA binding or transcriptional activity, for instance [54].

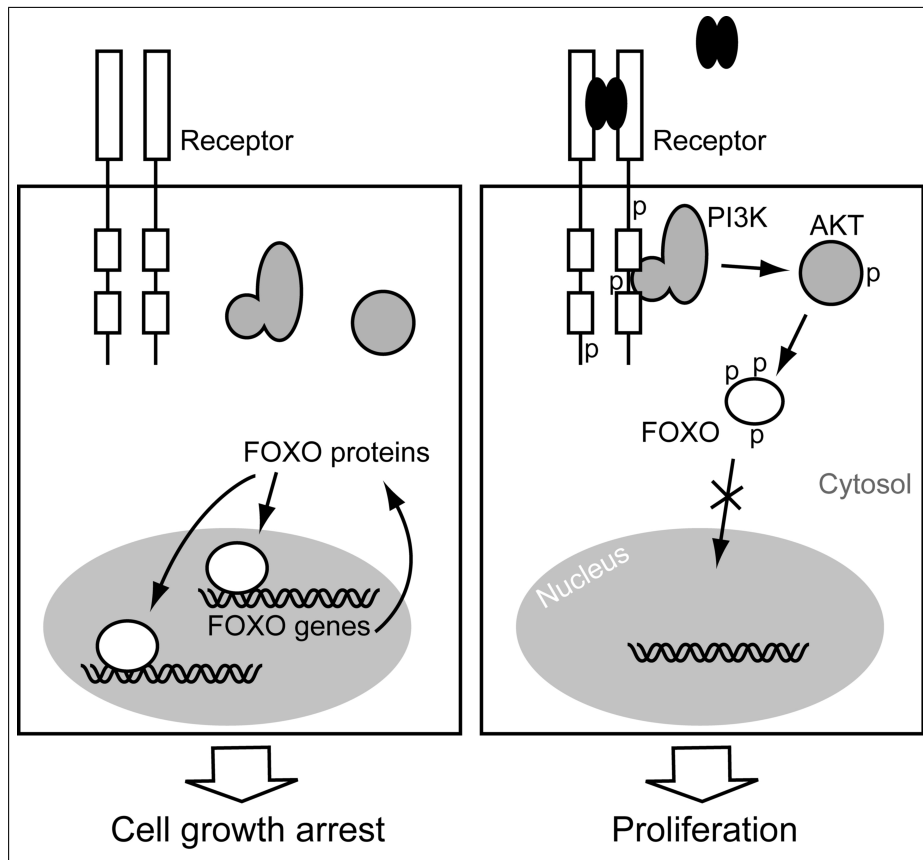


Figure 3.17: **Regulation of FOXOs by growth factors at the transcriptional and post-translational levels.** In the absence of growth stimulus (left panel), FOXO are activated and regulate the expression of a number of target genes including FOXO1 and FOXO4. In the presence of growth factors (right), AKT inactivates FOXO, switching off target gene expression and disrupting the positive feedback on FOXO transcription. P, phosphorylated site.

3.2.5 Conclusions

Our data demonstrate that FOXO1 expression is stimulated by activated FOXO3, based on the following observations: (i) FOXO3-A3-ER activation leads to increased FOXO1 expression and increased FOXO1 promoter activity, (ii) wild-type FOXO3 over-expression also increases FOXO1 promoter activity, (iii) FOXO3 binds to at least one consensus site in FOXO1 promoter and (iv) inactivation of FOXO3 by phosphorylation in the presence of PDGF correlates with FOXO1 down-regulation. Likewise, FOXO3 may also regulate the expression of FOXO4, as its expression is induced by activated FOXO3-A3-ER and decreased by PDGF. However, no conserved consensus FOXO binding site could be identified in the promoter of FOXO4. In addition, luciferase experiments and chromatin immunoprecipitations suggested that FOXO1 induces the expression of its own gene. Since all FOXO

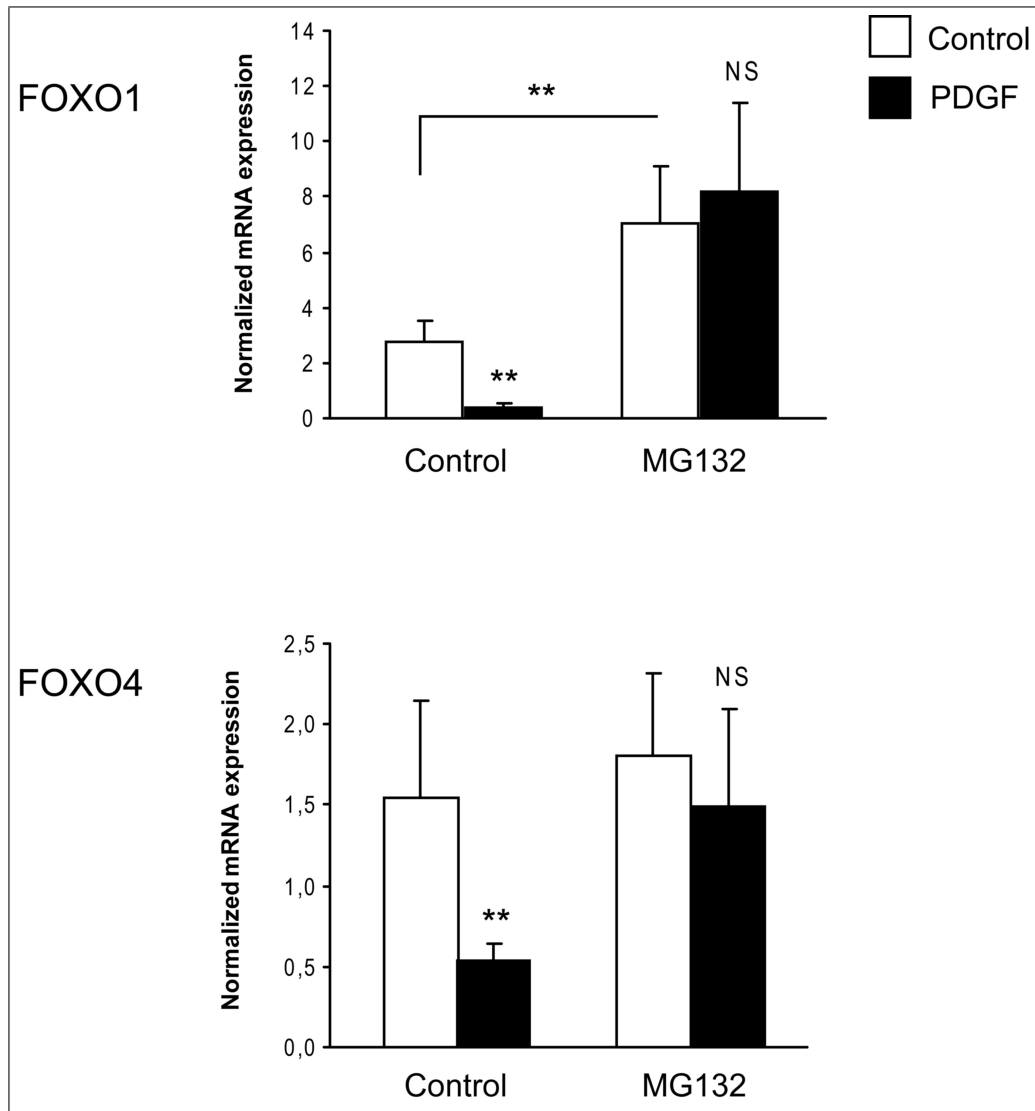


Figure 3.18: **The proteasome inhibitor MG132 blocks the effect of PDGF on FOXO gene expression.** AG01518 fibroblasts were starved for 48 h in serum-free medium. Cells were treated with PDGF-BB (25 ng/ml), MG132 (20 μ M) or control medium for 15h. Cells were harvested and RNA was extracted. Quantitative PCR was performed to monitor FOXO1, FOXO4 and RPLP0 expression. Visual inspection of the cells revealed significant toxicity induced by MG132, which might be caused by FOXO stabilization. Results are expressed as a ratio between the expression of FOXO and RPLP0 with SD. **, p -value < 0.01.

transcription factors bind to the same promoter elements and share a number of target genes, it is possible that they all induce the expression of FOXO1 and FOXO4, at least to some extent, creating a positive feedback network controlling FOXO gene expression. Such positive feedback regulation has been described for other transcription factors such as SREBP [93]. Understanding the regulation of the FOXO3 gene by growth factors such as FGF requires further studies.

In accordance to the decrease in FOXO1 mRNA, we also observed a decrease in FOXO1 protein

expression in cells treated with PDGF for 24 hours. However, FOXO1 degradation by proteasomes may also contribute the decrease in FOXO1 protein. Since the transcriptional repression of FOXO1 gene also depends on protein degradation (Figure 3.18), it was not possible to discriminate between these two possibilities.

In line with previous reports on other cell types [18, 42, 152], we show that activation of FOXO3 blocked the proliferation of human fibroblasts in response to growth factors. Conversely, reducing the expression of the FOXO1 transcript with specific shRNA enhanced fibroblast proliferation. This latter experiment indicated that the level of FOXO1 mRNA expression is an important determinant of cell proliferation, in addition to the well characterized post-translational modifications of FOXO proteins by AKT and other signaling kinases.

The regulation of FOXO mRNA expression may sustain or amplify the anti-proliferative and anti-apoptotic effects that are initiated by the direct and rapid control of FOXO proteins by phosphorylation in the presence of growth factors. Conversely, up-regulation of FOXO genes by activated FOXO3 may also help to restore FOXO expression level upon growth factor withdrawal or stress.

In a microarray study, Zhu et al suggested that FOXO1 expression was down-regulated in mouse B lymphocytes by several mitogens, such as interleukin-4, lipopolysaccharide, anti-IgM and anti-CD40 [394]. Decreased mRNA expression of FOXO1, FOXO3 and FOXO4 has also been reported in B cells upon activation of the B-cell receptor, which also induced FOXO1 protein phosphorylation [150]. The decrease in FOXO1, FOXO3 and FOXO4 expression in these cells was sensitive to PI3 kinase inhibition, like in our study. However, FOXO1 was not down-regulated in cells deficient in Bruton's kinase, a key mediator of BCR signaling, although it could still be phosphorylated on serine 256, suggesting that the two processes were uncoupled. Thus, a different mechanism may account for FOXO down-regulation in B lymphocytes.

As the constitutive activation of growth factor receptors and the PI3 kinase pathway is a frequent event in cancer cells, our data predict that FOXO gene expression may be decreased in tumors, which may contribute to uncontrolled cell growth. Interestingly, a decreased expression of FOXO3 correlates with prostate cancer cell progression toward androgen independence [217]. Future studies will have to determine whether the mRNA expression of FOXO genes is reduced in tumor cells.

In conclusion, we identified a new mechanism of regulation of FOXOs at the transcriptional level, which modulates fibroblast proliferation.

General Discussion

CHAPTER 4

GENERAL DISCUSSION

Gene transcription is a major mechanism by which cells regulate and maintain their functions. This process is subjected to a tight control involving many general and specific transcription factors. Depending on cell type, the regulation of these factors can be influenced by cell signalling pathways induced by circulating agents, such as hormones, growth factors and immune complexes.

4.1 Gene expression induced by PDGF

In this study, we used microarrays to explore the gene expression response induced by growth factors on primary human fibroblasts and a cancer cell line. Growth factors influence cells by inducing many signalling cascades. The general action of these pathways is to monitor TF activation and repression to coordinate biological effects, such as growth, cell division, metabolism and immune response. It is thought that different growth factors regulate largely overlapping rather than independent set of genes [115].

Platelet-Derived Growth Factor (PDGF) is the most studied growth factors in our laboratory, especially its action on primary fibroblasts. A particular issue that is covered by this thesis, is the gene transcription response of fibroblasts when stimulated with PDGF. Demoulin *et al.*, in our laboratory, have already shown that fibroblasts turn on cholesterol and fatty acid metabolism, when stimulated by this growth factor for a long time (24h), based on gene expression data. From the differentially regulated list of genes, targets of Sterol Regulatory Element Binding Protein (SREBP) were recognized. SREBP is a transcription factor involved in sterol biosynthesis. They further confirmed that this transcription factor is induced by PDGF on primary fibroblasts via the PI3K signalling pathway.

Starting from these observations, our objective was to study in detail the fibroblast transcriptional response induced by PDGF stimulation, and compare it to other growth factors, including Fibroblast Growth Factor (FGF). To achieve this goal, we performed microarray experiments using Affymetrix GeneChip arrays. In these experiments, we compared two time-point stimulations: short-term (1h) and long-term (24h). After RNA extraction, hybridization, data normalization and condition comparison, we end up with lists of genes significantly regulated in each time-point for both growth factors compared to the non-stimulated control condition.

A primary analysis of the significant list of genes differentially regulated between all conditions as defined by SAM algorithm has revealed that a difference could be observed between PDGF and FGF in term of gene regulation at the short time-point stimulation, however this difference is adjusted after 24h of stimulation, in which both growth factors induce a commonly regulated set of genes. These findings suggest that a short action of growth factors may induce different signalling pathways, but a long action ends toward overlapping pathways.

4.2 *TFactS tool*

The next step, was to analyze genes regulated by each growth factor to extract meaningful information with respect to transcription factors and their activity. When analyzed by *in silico* promoter scanning algorithms, SREBP binding elements were not found in the set of gene promoters regulated by PDGF after 24h-stimulation. As an alternative approach, we sought to build a tool that can infer both the regulation and the activity of TFs, in any biological system, based on lists of regulated genes as those generated by microarrays. The idea is to automate, by a computer program, the process used by many biologists, which consists on recognizing known target genes of a particular TF in the regulated gene list, and perform a statistical analysis. We called this tool “TFactS.”

Strikingly, although Fisher’s test is used in gene function enrichment, TFactS approach had not been previously used in TF mining. Databases of experimentally reported TF-gene regulations were only used to build Position Weight Matrices (PWMs). These PWMs are the basis of the majority of Transcription Factor Binding Site (TFBS) *in silico* discovery. Motivated by the availability of increasing genome sequences, and the challenge to predict TFBS based only on the sequence-information content, this field is rapidly expanding with an increasing number of published algorithms. The most successful methods were reported for lower organisms, such as yeast, however the analysis of mammalian genomes is still challenging. The promoter sequence identification, the intergenic sequence length and nucleotide content, in addition to the complexity of the gene transcription, involving indirect TF DNA binding and the epigenetic modifications are the major sources of problems for these algorithms.

TFactS compares submitted lists of genes (up- and down- regulated) with a catalogue of TF-gene regulations. Its primary objective is to detect not only significant TFs responsible for regulating these lists, but also the activity of those TFs. Thereafter, two distinct catalogues were built from published TF-gene regulations gathered from both the literature and TF databases for human, mouse and rat experimentally known regulations. The first catalogue contains only associations between TFs and their target genes without any information on the regulation-type (up- or down-regulation). This catalogue is called “sign-less,” and used as a basis to detect TF regulation in the studied system. The second catalogue, which takes into account the type of the regulation linking TFs to their targets, is called “sign-sensitive.” It is the basis to infer TF activity from the submitted gene lists and needs a clear separation between down- and up-regulated gene list submissions. The comparison is performed according to Fisher’s test as inspired from the majority of biological information enrichment-based

bioinformatics' tools. When compared to some of the *in silico* prediction based tools on selected microarray studies, for which authors defined experimentally the TFs involved, TFactS performed better, although the compared tools use different background filtering methods, including phylogeny conservation, context-sensitive and false positive rigorous filtering. This comparison can be criticized, with respect to the choice of validation studies (although some were suggested by the reviewers), and the parameters used for each tool (which may not be optimal). Importantly, TFactS is unable to find novel target genes for a given TF.

4.3 Predicted TFs regulated by PDGF receptor

4.3.1 STAT

TFactS predicted an early specific activation of STAT by PDGF (Figure 4.1), when compared to other growth factors in human primary fibroblasts. This was unexpected and requires a more detailed study to unravel the mechanism behind. STAT activation by PDGF receptors plays a role in growth and motility in vessel wall remodeling [194]. This activation is also involved in the proliferation induced by PDGF receptor [312]. There is a possible role of Myc in STAT induced proliferation [46]. STAT activation by PDGF receptors is not fully understood as many mechanisms were proposed according to studied cell lines.

The specific activation of STAT by PDGFR β and not by PDGFR α is in line with literature, in which the major role of PDGFR β in STAT activation is shown in comparison to PDGFR α [350]. Nevertheless, the PDGFR α expressed as a hybrid protein FIP1L1-PDGFR α in the EOL-1 cancer cell line was suggested by TFactS to induce STAT activation, which was confirmed experimentally. Thus, a careful study of the mechanisms by which STATs are activated preferentially by normal PDGFR β should be undertaken.

4.3.2 SREBP

When fed with the regulated gene lists in our microarray experiment, TFactS predicted significant regulation and activation of SREBP by both PDGF and FGF after 24h-stimulation (Figure 4.1), confirming the findings of our laboratory that were previously described [84]. Since SREBP is responsible for sterol biosynthesis, its induction by growth factors is essentially needed, in the context of *in vitro* cultured primary fibroblasts, for the maintenance and the regeneration of cell membranes, especially when cells are proliferating, which is the major action induced by growth factors. Thus, SREBP activation may be part of the cell cycle as also suggested by the others [35].

4.3.3 FOXO

Interestingly, TFactS also predicted a significant regulation and inhibition of FOXO by both PDGF and FGF after long-term stimulation (Figure 4.1). It is well known that FOXO are phosphorylated by

AKT. This post-translational modification is responsible for their inhibition and exclusion from the nucleus.

In this study, we have shown that in primary fibroblasts: (i) FOXO maintain their own gene expression in a feed-back positive loop, which could be mastered mainly by FOXO3; (ii) FOXO interfere with proliferation induced by PDGF; (iii) PDGF interrupts this FOXO feed-back positive loop, by inducing AKT phosphorylation, which leads to inhibition of FOXO and their exclusion from the nucleus. Our findings were confirmed by a recent paper [7].

FGF-induced inhibition of FOXO needs further investigations since this growth factor is not known to activate AKT on human fibroblasts (see Figure 3.6). One possibility is that FGF inhibits FOXO via the ERK pathway, since it is shown that ERK can also inhibit FOXO by phosphorylation [376].

Based on these findings, we can assert that inhibition of FOXO at both protein and transcriptional levels contribute to cell proliferation induced by growth factors. This is in line with SREBP and E2F activation, suggesting cell entrance in division and proliferation.

4.3.4 Other transcription factors

TFactS predicted the regulation and activation of E2F, which is widely involved in cell division (Figure 4.1). Other TFs were predicted by TFactS as regulated, suggesting their additive role in the proliferation induced by PDGF. The majority of them are in line with the literature. Some of them are widely involved in cell division and proliferation, such as MYC, AP-1, *beta*-catenin and NF κ B. We also observed the regulation of cell cycle genes by growth factors in neural stem cells (see Paper 5 in Annexes). PPAR γ could be inhibited, since many studies have shown an interference between proliferative-induction by PDGF and this TF [20, 125, 129, 206]. Some others represent new findings like ATF, PPAR α , GLI and FLI. These TFs need further investigations to confirm exactly their activation state and the way they could be involved in this process.

4.4 TFactS potential applications

4.4.1 TFactS cancer screening

A potential use of TFactS could be in the oncology field. Indeed, many studies reported significant differences in term of gene regulation between cancer classes, which opened an area of research in the mechanisms underlying cancer development, progression and metastasis. These mechanisms have been studied for diagnosis and prognosis purposes as well as for treatment development. The differences in gene regulation that govern these processes, could be a reflect of TF differential regulation between cancer classes or between normal and malignant cells. To test the ability of TFactS to discover new TF regulations on cancer microarray data set, NCI60 collection was analyzed for gene regulations and then submitted to sign-less TFactS. TFactS predicted a group of TFs as regulated in most cancer types. These TFs are almost known as involved in cancer, including β -catenin, GLI,

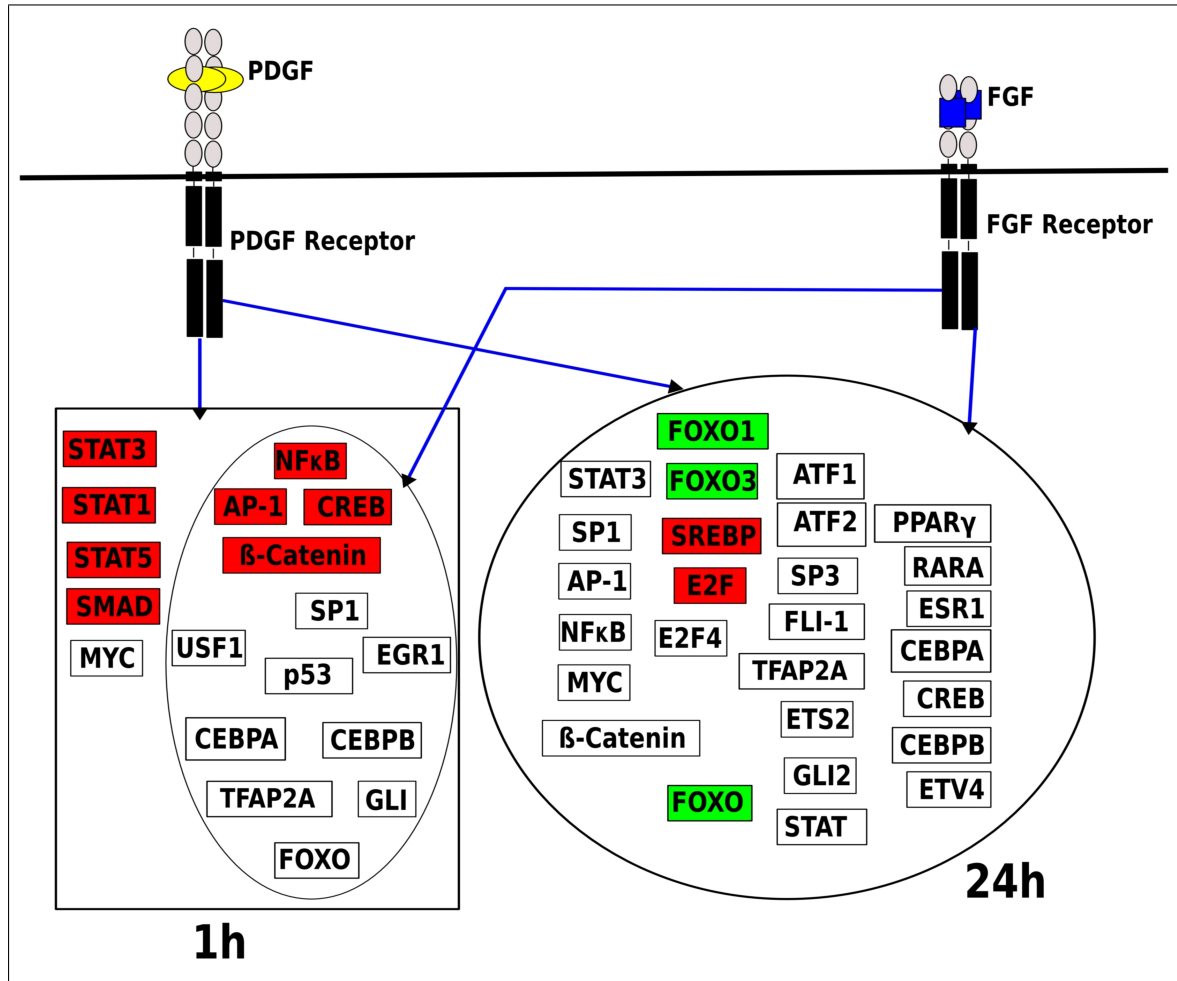


Figure 4.1: *Transcription factor regulation by growth factors in fibroblasts as predicted by TFactS.* Significantly predicted TF regulations by TFactS are shown in white boxes, and predicted activations and inhibitions are shown in red and green, respectively. The raw list was filtered by discarding TFs with less than 3 genes in the intersection between submitted genes and target genes.

NFκB, FOXO and SMAD.

We plan to screen larger cancer expression data sets from public databases (like GEO, ArrayExpress, Oncomine, . . .), first to confirm TFactS findings on NCI60 data set, then to find new regulations in other cancer types. We will particularly focus on FOXO regulation in cancer.

4.4.2 FOXO regulation in cancer

In the literature, FOXO have been associated with cancer and can act as tumor suppressors. FOXO loss of function leads to increased cellular survival and a predisposition to neoplasia [285]. In addition, deletion of FOXO genes in mice engendered thymic lymphomas and hemangiomas [258]. An indirect implication is also suggested from the position of FOXO downstream of AKT. Indeed, in

many cancers, AKT pathway is activated, which leads to FOXO inhibition. However, there was no systematic check for FOXO activity status in tumors. Therefore, a careful study should be done to screen different tumor types for FOXO status, especially at the mRNA level.

The results of TFactS, on NCI60 cancer data set, suggest an important role of FOXO in tumors. Our findings on the way FOXO could be silenced by growth factors, are in line with these prediction results. Indeed, the feed-back positive loop by which FOXO maintain their expression and activation, could be interrupted in tumors. This interruption would be induced either by an intervention at the protein level and/or at the RNA level (Figure 4.2).

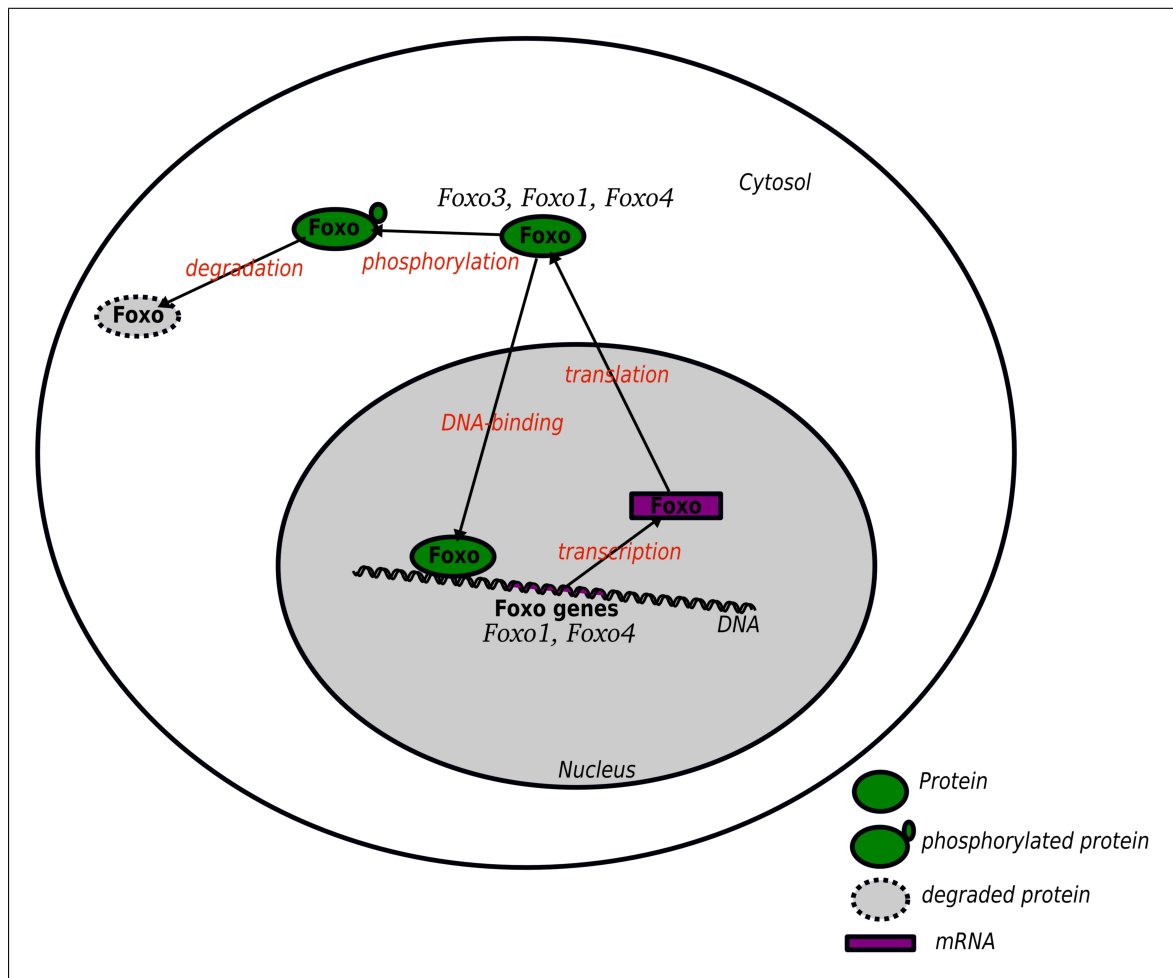


Figure 4.2: FOXO transcription factors could be regulated at different levels.

FOXO TFs may be regulated at different levels. At the protein level, they could be regulated through phosphorylation and degradation. At the RNA level, they could be regulated by transcription repressors and small inhibitory RNA affecting RNA stability and translation.

4.4.3 *De novo* TFBS predictions

TFactS uses already known TF-gene regulations, which makes the search for new regulations out of context with this tool. It is designed to infer TF regulations but not gene regulations. However, TFactS could be used as a guide to gene promoter analysis by *in silico* based methods. This could be done, for example, by running clustering algorithms in gene expression data and detecting significantly regulated TFs using TFactS on each cluster. Then, using the PWMs of these predicted TFs, one can run pattern matching programs on the promoters of clustered genes. The newly discovered TFBS in the promoters of genes that belong to a cluster and that are not known targets of the predicted TF on that cluster could be the basis of experimental studies, such as luciferase activity tests and ChIP experiments.

4.5 *TFactS* amelioration

4.5.1 More regulations

Adding more data in both sign-sensitive and sign-less catalogues is our priority. For sign-less catalogue, we will add new regulations from TF databases as soon as they are available. However, sign-sensitive catalogue will need manual curation of the literature. This is the only possibility while waiting for a systematic annotation of the regulation type in the current TF databases. This option is starting to be included in the recent TF databases, such as PAZAR.

It would be interesting to include regulations generated by high-throughput experiments, such as microarrays, ChIP-on-chip and ChIP-Seq. However, since these methods are more noisy, a real strategy to incorporate their results in the catalogue should be defined and tested. In addition, only microarrays generated results could be incorporated into the sign-sensitive catalogue, because they can give information on the regulation type. One should be careful about incorporating raw microarray data, since there is no evidence that regulated genes are direct targets of the studied TF.

The next versions of TFactS will integrate regulation data from high-throughput techniques. The user will be given a choice to select either curated catalogues with high quality data or the high-throughput generated catalogues with much more data of lesser quality.

4.5.2 Distinct species

TFactS catalogues, as implemented in the current version, are built in one-shot from three organisms (human, rat and mouse). This does not prevent a good performance on the validation studies, which contained both human and mouse models. However, separating catalogues in a species-relative fashion is one of our goal. This should be relatively easy to implement since TF databases already include this feature.

4.5.3 Additional biological information mining

TFactS can also be extended to signalling pathway mining. This could be achieved by building catalogues of well-established gene-pathway and/or TF-pathway associations. In that way, each predicted TF could be linked to one or more pathways. An addition of experimentally curated microRNA (miR) target catalogue to TFactS will be a plus. Thus, regulated miRs in the system could also be predicted. This will permit to build a global network view of the processes underlying the observed system transcriptome. This network would be used to deduce the molecular events (signalling, TFs and miRs) that could link a treatment, for instance, and the related mRNA signature.

This task will be surmounted by the availability of databases, such as NetPath, which contain information on experimentally proven gene-pathway and TF-pathway associations [248]. In addition, curated miR target databases, such as mir2disease, are also available [170].

4.5.4 Alternative tests to Fisher's

TFactS uses Fisher's test, which relies on the submitted gene list size. The robustness analysis showed that when TFactS is fed with long lists, it still predicts accurately expected TF regulation. The addition of noisy data to these lists did not affect the performance. However, the submitted gene list has to contain sufficient number of target genes for a given TF to be predicted. In some situations, depending on the studied system and the statistics used to analyze microarray data, it is impossible to have enough regulated gene list. In this case, rather than Fisher's test, other methods could be used. These methods, such as Kolmogorov-Smirnov's test, make use of intensities instead of gene counts. In addition, they can use the whole expression data set. Thus, the performance of these alternatives using TFactS catalogues should be tested.

Bibliography

BIBLIOGRAPHY

- [1] ABID, M. R., YANO, K., GUO, S., PATEL, V. I., SHRIKHANDE, G., SPOKES, K. C., FERRAN, C., AND AIRD, W. C. Forkhead transcription factors inhibit vascular smooth muscle cell proliferation and neointimal hyperplasia. J Biol Chem 280, 33 (Aug 2005), 29864–29873.
- [2] ADACHI, M., OSAWA, Y., UCHINAMI, H., KITAMURA, T., ACCILI, D., AND BRENNER, D. A. The forkhead transcription factor foxo1 regulates proliferation and transdifferentiation of hepatic stellate cells. Gastroenterology 132, 4 (Apr 2007), 1434–1446.
- [3] AFFYMETRIX. Affymetrix technical support. http://www.affymetrix.com/support/technical/sample_data/ht_cartridge_sampledata_readme.pdf.
- [4] AFFYMETRIX. whitepapers. <http://www.affymetrix.com/support/technical/whitepapers/sadd.whitepaper.pdf>.
- [5] AFFYMETRIX. Guide to probe logarithmic intensity error (plier) estimation. http://www.affymetrix.com/support/technical/technotes/plier_technote.pdf, December 2009.
- [6] AFFYMETRIX. Affymetrix web page. <http://www.affymetrix.com>, April 2010.
- [7] AL-MUBARAK, B., SORIANO, F. X., AND HARDINGHAM, G. E. Synaptic nmdar activity suppresses foxo1 expression via a cis-acting foxo binding site: Foxo1 is a foxo target gene. Channels (Austin) 3, 4 (2009), 233–238.
- [8] AL-SHAHROUR, F., DIAZ-URIARTE, R., AND DOPAZO, J. Fatigo: a web tool for finding significant associations of gene ontology terms with groups of genes. Bioinformatics 20, 4 (Mar 2004), 578–580.
- [9] AL-SHAHROUR, F., MINGUEZ, P., VAQUERIZAS, J. M., CONDE, L., AND DOPAZO, J. Babelomics: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. Nucleic Acids Res 33, Web Server issue (Jul 2005), W460–W464.
- [10] ALLISON, D. B., CUI, X., PAGE, G. P., AND SABRIPOUR, M. Microarray data analysis: from disarray to consolidation and consensus. Nat Rev Genet 7, 1 (Jan 2006), 55–65.
- [11] ALTMAN, N. S., AND HUA, J. Extending the loop design for two-channel microarray experiments. Genet Res 88, 3 (Dec 2006), 153–163.
- [12] AMBROISE, C., AND MCLACHLAN, G. J. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 99, 10 (May 2002), 6562–6566.
- [13] ANDO, T., SUGURO, M., HANAI, T., KOBAYASHI, T., HONDA, H., AND SETO, M. Fuzzy neural network applied to gene expression profiling for predicting the prognosis of diffuse large b-cell lymphoma. Jpn J Cancer Res 93, 11 (Nov 2002), 1207–1212.
- [14] ANDRAE, J., GALLINI, R., AND BETSHOLTZ, C. Role of platelet-derived growth factors in physiology and medicine. Genes Dev 22, 10 (May 2008), 1276–1312.
- [15] ANGEL, P., AND KARIN, M. The role of jun, fos and the ap-1 complex in cell-proliferation and transformation. Biochim Biophys Acta 1072, 2-3 (Dec 1991), 129–157.
- [16] ANGELINI, C., CUTILLO, L., DE CANDITHS, D., MUTARELLI, M., AND PENSKY, M. Bats: a bayesian user-friendly software for analyzing time series microarray experiments. BMC Bioinformatics 9 (2008), 415.
- [17] ANGELINI, C., DE CANDITHS, D., MUTARELLI, M., AND PENSKY, M. A bayesian approach to estimation and testing in time-course microarray experiments. Stat Appl Genet Mol Biol 6 (2007), Article24.
- [18] AOKI, M., JIANG, H., AND VOGT, P. K. Proteasomal degradation of the foxo1 transcriptional regulator in cells transformed by the p3k and akt oncoproteins. Proc Natl Acad Sci U S A 101, 37 (Sep 2004), 13613–13617.
- [19] ARTEAGA-SALAS, J. M., ZUZAN, H., LANGDON, W. B., UPTON, G. J. G., AND HARRISON, A. P. An

- overview of image-processing methods for affymetrix genechips. *Brief Bioinform* 9, 1 (Jan 2008), 25–33.
- [20] ARTEMENKO, Y., GAGNON, A., AUBIN, D., AND SORISKY, A. Anti-adipogenic effect of pdgf is reversed by pkc inhibition. *J Cell Physiol* 204, 2 (Aug 2005), 646–653.
- [21] ARYEE, M., J.A., G.-P., KRAMNIK, I., MAITI, T., AND QUACKENBUSH, J. An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: Betr (bayesian estimation of temporal regulation). *BMC Bioinformatics* 10 (2009), 409.
- [22] ASADA, S., DAITOKU, H., MATSUZAKI, H., SAITO, T., SUDO, T., MUKAI, H., IWASHITA, S., KAKO, K., KISHI, T., KASUYA, Y., AND FUKAMIZU, A. Mitogen-activated protein kinases, erk and p38, phosphorylate and regulate foxo1. *Cell Signal* 19, 3 (Mar 2007), 519–527.
- [23] ATHANASIADIS, E. I., CAVOURAS, D. A., GLOTSOS, D. T., GEORGIADIS, P. V., KALATZIS, I. K., AND NIKIFORIDIS, G. C. Segmentation of complementary dna microarray images by wavelet-based markov random field model. *IEEE Trans Inf Technol Biomed* 13, 6 (Nov 2009), 1068–1074.
- [24] ATHANASIADIS, E. I., CAVOURAS, D. A., SPYRIDONOS, P. P., GLOTSOS, D. T., KALATZIS, I. K., AND NIKIFORIDIS, G. C. Complementary dna microarray image processing based on the fuzzy gaussian mixture model. *IEEE Trans Inf Technol Biomed* 13, 4 (Jul 2009), 419–425.
- [25] AUER, H., NEWSOM, D. L., AND KORNACKER, K. Expression profiling using affymetrix genechip microarrays. *Methods Mol Biol* 509 (2009), 35–46.
- [26] BAI, D., UENO, L., AND VOGT, P. K. Akt-mediated regulation of nfkapab and the essentialness of nfkapab for the oncogenicity of pi3k and akt. *Int J Cancer* 125, 12 (Dec 2009), 2863–2870.
- [27] BAJCSY, P. Gridline: automatic grid alignment in dna microarray scans. *IEEE Trans Image Process* 13, 1 (Jan 2004), 15–25.
- [28] BAJIC, V. B., BRENT, M. R., BROWN, R. H., FRANKISH, A., HARROW, J., OHLER, U., SOLOVYEV, V. V., AND TAN, S. L. Performance assessment of promoter predictions on encode regions in the egasp experiment. *Genome Biol* 7 Suppl 1 (2006), S3.1–S3.13.
- [29] BAKAY, M., CHEN, Y. W., BORUP, R., ZHAO, P., NAGARAJU, K., AND HOFFMAN, E. P. Sources of variability and effect of experimental approach on expression profiling data interpretation. *BMC Bioinformatics* 3 (2002), 4.
- [30] BARONE, M. V., AND COURTNEIDGE, S. A. Myc but not fos rescue of pdgf signalling block caused by kinase-inactive src. *Nature* 378, 6556 (Nov 1995), 509–512.
- [31] BARRIENTOS, S., STOJADINOVIC, O., GOLINKO, M. S., BREM, H., AND TOMIC-CANIC, M. Growth factors and cytokines in wound healing. *Wound Repair Regen* 16, 5 (2008), 585–601.
- [32] BASHEERUDDIN, K., LI, X., RECHTORIS, C., AND MAZZONE, T. Platelet-derived growth factor enhances spl binding to the ldl receptor gene. *Arterioscler Thromb Vasc Biol* 15, 8 (Aug 1995), 1248–1254.
- [33] BASSO, K., KLEIN, U., NIU, H., STOLOVITZKY, G. A., TU, Y., CALIFANO, A., CATTORETTI, G., AND DALLA-FAVERA, R. Tracking cd40 signaling during germinal center development. *Blood* 104, 13 (Dec 2004), 4088–4096.
- [34] BEIBARTH, T., AND SPEED, T. P. Gostat: find statistically overrepresented gene ontologies within a group of genes. *Bioinformatics* 20, 9 (2004), 1464–1465.
- [35] BENGOCHEA-ALONSO, M. T., AND ERICSSON, J. Cdk1/cyclin b-mediated phosphorylation stabilizes srebp1 during mitosis. *Cell Cycle* 5, 15 (Aug 2006), 1708–1718.
- [36] BENJAMINI, Y., AND HOCHBERG, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57, 01 (1995), 289–300.
- [37] BERRAR, D. P., DOWNES, C. S., AND DUBITZKY, W. Multiclass cancer classification using gene expression profiling and probabilistic neural networks. *Pac Symp Biocomput* (2003), 5–16.
- [38] BERRY, F. B., SKARIE, J. M., MIRZAYANS, F., FORTIN, Y., HUDSON, T. J., RAYMOND, V., LINK, B. A., AND WALTER, M. A. Foxc1 is required for cell viability and resistance to oxidative stress in the eye through the transcriptional regulation of foxo1a. *Hum Mol Genet* 17, 4 (Feb 2008), 490–505.
- [39] BHATTACHARJEE, A., RICHARDS, W. G., STAUNTON, J., LI, C., MONTI, S., VASA, P., LADD, C., BEHESHTI, J., BUENO, R., GILLETTE, M., LODA, M., WEBER, G., MARK, E. J., LANDER, E. S., WONG, W., JOHNSON, B. E., GOLUB, T. R., SUGARBAKER, D. J., AND MEYERSON, M. Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 98, 24 (Nov 2001), 13790–13795.

- [40] BILBAN, M., BUEHLER, L. K., HEAD, S., DESOYE, G., AND QUARANTA, V. Normalizing dna microarray data. *Curr Issues Mol Biol* 4, 2 (Apr 2002), 57–64.
- [41] BIOCARTA. <http://www.biocarta.com/>, April 2010.
- [42] BIRKENKAMP, K. U., AND COFFER, P. J. Regulation of cell survival and proliferation by the foxo (forkhead box, class o) subfamily of forkhead transcription factors. *Biochem Soc Trans* 31, Pt 1 (Feb 2003), 292–297.
- [43] BJÖRKLUND, S., ALMOUZNI, G., DAVIDSON, I., NIGHTINGALE, K. P., AND WEISS, K. Global transcription regulators of eukaryotes. *Cell* 96, 6 (Mar 1999), 759–767.
- [44] BOES, T., AND NEUHÄUSER, M. Normalization for affymetrix genechips. *Methods Inf Med* 44, 3 (2005), 414–417.
- [45] BOORSMA, A., FOAT, B. C., VIS, D., KLIS, F., AND BUSSEMAKER, H. J. T-profiler: scoring the activity of predefined groups of genes using gene expression data. *Nucleic Acids Res* 33, Web Server issue (Jul 2005), W592–W595.
- [46] BOWMAN, T., BROOME, M. A., SINIBALDI, D., WHARTON, W., PLEDGER, W. J., SEDIVY, J. M., IRBY, R., YEATMAN, T., COURTNEIDGE, S. A., AND JOVE, R. Stat3-mediated myc expression is required for src transformation and pdgf-induced mitogenesis. *Proc Natl Acad Sci U S A* 98, 13 (Jun 2001), 7319–7324.
- [47] BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S., AND CRAWFORD, G. E. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 2 (Jan 2008), 311–322.
- [48] BOZINOV, D. Autonomous system for web-based microarray image analysis. *IEEE Trans Nanobioscience* 2, 4 (Dec 2003), 215–220.
- [49] BRAGA-NETO, U., HASHIMOTO, R., DOUGHERTY, E. R., NGUYEN, D. V., AND CARROLL, R. J. Is cross-validation better than resubstitution for ranking genes? *Bioinformatics* 20, 2 (Jan 2004), 253–258.
- [50] BROWN, M. P., GRUNDY, W. N., LIN, D., CRISTIANINI, N., SUGNET, C. W., FUREY, T. S., ARES, M., AND HAUSSLER, D. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A* 97, 1 (Jan 2000), 262–267.
- [51] BROWN, M. S., AND GOLDSTEIN, J. L. The srebp pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* 89, 3 (May 1997), 331–340.
- [52] BURGESS, J. K. Gene expression studies using microarrays. *Clin Exp Pharmacol Physiol* 28, 4 (Apr 2001), 321–328.
- [53] BUTT, A. J., CALDON, C. E., MCNEIL, C. M., SWARBRICK, A., MUSGROVE, E. A., AND SUTHERLAND, R. L. Cell cycle machinery: links with genesis and treatment of breast cancer. *Adv Exp Med Biol* 630 (2008), 189–205.
- [54] CALNAN, D. R., AND BRUNET, A. The foxo code. *Oncogene* 27, 16 (Apr 2008), 2276–2288.
- [55] CALZA, S., VALENTINI, D., AND PAWITAN, Y. Normalization of oligonucleotide arrays based on the least-variant set of genes. *BMC Bioinformatics* 9 (2008), 140.
- [56] CAMP, H. S., AND TAFURI, S. R. Regulation of peroxisome proliferator-activated receptor gamma activity by mitogen-activated protein kinase. *J Biol Chem* 272, 16 (Apr 1997), 10811–10816.
- [57] CAO, X., TAY, A., GUY, G. R., AND TAN, Y. H. Activation and association of stat3 with src in v-src-transformed cell lines. *Mol Cell Biol* 16, 4 (Apr 1996), 1595–1603.
- [58] CHIARA, F., BISHAYEE, S., HELDIN, C. H., AND DEMOULIN, J. B. Autoinhibition of the platelet-derived growth factor beta-receptor tyrosine kinase by its c-terminal tail. *J Biol Chem* 279, 19 (May 2004), 19732–19738.
- [59] CHIARI, M., CRETICH, M., DAMIN, F., CARLO, G. D., AND OLDANI, C. Advanced polymers for molecular recognition and sensing at the interface. *J Chromatogr B Analyt Technol Biomed Life Sci* 866, 1–2 (Apr 2008), 89–103.
- [60] CHIARIELLO, M., MARINISSEN, M. J., AND GUTKIND, J. S. Regulation of c-myc expression by pdgf through rho gtpases. *Nat Cell Biol* 3, 6 (Jun 2001), 580–586.
- [61] CHIARUGI, P., CIRRI, P., MARRA, F., RAUGEI, G., FIASCHI, T., CAMICI, G., MANAO, G., ROMANELLI, R. G., AND RAMPONI, G. The src and signal transducers and activators of transcription pathways as specific targets for low molecular weight phosphotyrosine-protein phosphatase in platelet-derived growth factor signaling. *J Biol Chem* 273, 12 (Mar 1998), 6776–6785.

- [62] CHOI, Y. J., AND KENDZIORSKI, C. Statistical methods for gene set co-expression analysis. *Bioinformatics* 25, 21 (Nov 2009), 2780–2786.
- [63] CHOUDHURY, G. G., LENIN, M., CALHAUN, C., ZHANG, J. H., AND ABOUD, H. E. Pdgf inactivates forkhead family transcription factor by activation of akt in glomerular mesangial cells. *Cell Signal* 15, 2 (Feb 2003), 161–170.
- [64] CHOUDHURY, G. G., MARRA, F., KIYOMOTO, H., AND ABOUD, H. E. Pdgf stimulates tyrosine phosphorylation of jak 1 protein tyrosine kinase in human mesangial cells. *Kidney Int* 49, 1 (Jan 1996), 19–25.
- [65] CHUN, S. M., KIM, Y. L., CHOI, H. B., OH, Y. T., KIM, Y. J., LEE, S., KIM, T. G., YANG, E. G., PARK, Y. K., KIM, D. W., AND HAN, B. D. Identification of leukemia-specific fusion gene transcripts with a novel oligonucleotide array. *Mol Diagn Ther* 11, 1 (2007), 21–28.
- [66] CHURCHILL, G. A. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 32 Suppl (Dec 2002), 490–495.
- [67] CLAVERIE, J. M., AND AUDIC, S. The statistical significance of nucleotide position-weight matrix matches. *Comput Appl Biosci* 12, 5 (Oct 1996), 431–439.
- [68] COLLIER, H. A., GRANDORI, C., TAMAYO, P., COLBERT, T., LANDER, E. S., EISENMAN, R. N., AND GOLUB, T. Expression analysis with oligonucleotide microarrays reveals that myc regulates genes involved in growth, cell cycle, signaling, and adhesion. *PNAS* 97, 7 (Mar 2000), 3260–3265.
- [69] CONESA, A., NUEDA, M., FERRER, A., AND TAL'ON, M. masigpro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* 22, 9 (May 2006), 1096–1102.
- [70] CONSORTIUM, F. A. N. T. O. M., SUZUKI, H., FORREST, A. R. R., VAN NIMWEGEN, E., DAUB, C. O., BALWIERZ, P. J., IRVINE, K. M., LASSMANN, T., RAVASI, T., HASEGAWA, Y., DE HOON, M. J. L., KATAYAMA, S., SCHRODER, K., CARNINCI, P., TOMARU, Y., KANAMORI-KATAYAMA, M., KUBOSAKI, A., AKALIN, A., ANDO, Y., ARNER, E., ASADA, M., ASAHARA, H., BAILEY, T., BAJIC, V. B., BAUER, D., BECKHOUSE, A. G., BERTIN, N., BJERKEGREN, J., BROMBACHER, F., BULGER, E., CHALK, A. M., CHIBA, J., CLOONAN, N., DAWE, A., DOSTIE, J., ENGSTRM, P. G., ESSACK, M., FAULKNER, G. J., FINK, J. L., FREDMAN, D., FUJIMORI, K., FURUNO, M., GOJOBORI, T., GOUGH, J., GRIMMOND, S. M., GUSTAFSSON, M., HASHIMOTO, M., HASHIMOTO, T., HATAKEYAMA, M., HEINZEL, S., HIDE, W., HOFMANN, O., HRNQUIST, M., HUMINIECKI, L., IKEO, K., IMAMOTO, N., INOUE, S., INOUE, Y., ISHIHARA, R., IWAYANAGI, T., JACOBSEN, A., KAUR, M., KAWAJI, H., KERR, M. C., KIMURA, R., KIMURA, S., KIMURA, Y., KITANO, H., KOGA, H., KOJIMA, T., KONDO, S., KONNO, T., KROGH, A., KRUGER, A., KUMAR, A., LENHARD, B., LENNARTSSON, A., LINDOW, M., LIZIO, M., MACPHERSON, C., MAEDA, N., MAHER, C. A., MAQUINGO, M., MAR, J., MATTIGIAN, N. A., MATSUDA, H., MATTICK, J. S., MEIER, S., MIYAMOTO, S., MIYAMOTO-SATO, E., NAKABAYASHI, K., NAKACHI, Y., NAKANO, M., NYGAARD, S., OKAYAMA, T., OKAZAKI, Y., OKUDA-YABUKAMI, H., ORLANDO, V., OTOMO, J., PACHKOV, M., PETROVSKY, N., PLESSY, C., QUACKENBUSH, J., RADOVANOVIC, A., REHLI, M., SAITO, R., SANDELIN, A., SCHMEIER, S., SCHNBACH, C., SCHWARTZ, A. S., SEMPLER, C. A., SERA, M., SEVERIN, J., SHIRAHIGE, K., SIMONS, C., LAURENT, G. S., SUZUKI, M., SUZUKI, T., SWEET, M. J., TAFT, R. J., TAKEDA, S., TAKENAKA, Y., TAN, K., TAYLOR, M. S., TEASDALE, R. D., TEGNER, J., TEICHMANN, S., VALEN, E., WAHLESTEDT, C., WAKI, K., WATERHOUSE, A., WELLS, C. A., WINTHER, O., WU, L., YAMAGUCHI, K., YANAGAWA, H., YASUDA, J., ZAVOLAN, M., HUME, D. A., CENTER, R. O. S., ARAKAWA, T., FUKUDA, S., IMAMURA, K., KAI, C., KAIHO, A., KAWASHIMA, T., KAWAZU, C., KITAZUME, Y., KOJIMA, M., MIURA, H., MURAKAMI, K., MURATA, M., NINOMIYA, N., NISHIYORI, H., NOMA, S., OGAWA, C., SANO, T., SIMON, C., TAGAMI, M., TAKAHASHI, Y., KAWAI, J., AND HAYASHIZAKI, Y. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* 41, 5 (May 2009), 553–562.
- [71] COOLS, J., QUENTMEIER, H., HUNTLY, B. J., MARYNEN, P., GRIFFIN, J. D., DREXLER, H. G., AND GILLILAND, D. G. The eol-1 cell line as an in vitro model for the study of fip111-pdgfra-positive chronic eosinophilic leukemia. *Blood* 103, 7 (Apr 2004), 2802–2805.
- [72] CUI, X., AND CHURCHILL, G. A. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* 4, 4 (2003), 210.
- [73] CZYZ, M. [specificity and selectivity of the nfκappab response]. *Postepy Biochem* 51, 1 (2005), 60–68.
- [74] DANDRÉ, F., AND OWENS, G. K. Platelet-derived growth factor-bb and ets-1 transcription factor negatively regulate transcription of multiple smooth mus-

- cle cell differentiation marker genes. *Am J Physiol Heart Circ Physiol* 286, 6 (Jun 2004), H2042–H2051.
- [75] DAS, M. K., AND DAI, H. K. A survey of dna motif finding algorithms. *BMC Bioinformatics* 8 Suppl 7 (2007), S21.
- [76] DASKALAKIS, A., GLOTSOS, D., KOSTOPOULOS, S., CAVOURAS, D., AND NIKIFORIDIS, G. A comparative study of individual and ensemble majority vote cdna microarray image segmentation schemes, originating from a spot-adjustable based restoration framework. *Comput Methods Programs Biomed* 95, 1 (Jul 2009), 72–88.
- [77] DAVULURI, R. V., SUZUKI, Y., SUGANO, S., PLASS, C., AND HUANG, T. H. M. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet* 24, 4 (Apr 2008), 167–177.
- [78] DAY, W. H., AND MCMORRIS, F. R. Critical comparison of consensus methods for molecular sequences. *Nucleic Acids Res* 20, 5 (Mar 1992), 1093–1099.
- [79] DEATON, R. A., GAN, Q., AND OWENS, G. K. Sp1-dependent activation of klf4 is required for pdgf-bb-induced phenotypic modulation of smooth muscle. *Am J Physiol Heart Circ Physiol* 296, 4 (Apr 2009), H1027–H1037.
- [80] DEFRANCE, M., AND TOUZET, H. Predicting transcription factor binding sites using local overrepresentation and comparative genomics. *BMC Bioinformatics* 7 (2006), 396.
- [81] DEHAN, P., KUSTERMANS, G., GUENIN, S., HORION, J., BONIVER, J., AND DELVENNE, P. Dna methylation and cancer diagnosis: new methods and applications. *Expert Rev Mol Diagn* 9, 7 (Oct 2009), 651–657.
- [82] DELPUECH, O., GRIFFITHS, B., EAST, P., ESSAFI, A., LAM, E. W., BURGERING, B., DOWNWARD, J., AND SCHULZE, A. Induction of mx1-sr by foxo3a contributes to repression of myc-dependent gene expression. *Molecular and cellular Biology* 27, 13 (Jul 2007), 4917–4930.
- [83] DEMOULIN, J. B., ENARSSON, M., LARSSON, J., ESSAGHIR, A., HELDIN, C. H., AND FORSBERG-NILSSON, K. The gene expression profile of pdgf-treated neural stem cells corresponds to partially differentiated neurons and glia. *Growth Factors* 24, 3 (Sep 2006), 184–196.
- [84] DEMOULIN, J. B., ERICSSON, J., KALLIN, A., RORSMAN, C., RÖNNSTRAND, L., AND HELDIN, C. H. Platelet-derived growth factor stimulates membrane lipid synthesis through activation of phosphatidylinositol 3-kinase and sterol regulatory element-binding proteins. *J Biol Chem* 279, 34 (Aug 2004), 35392–35402.
- [85] DEMOULIN, J. B., SEO, J. K., EKMAN, S., GRAPENGIESSER, E., HELLMAN, U., RÖNNSTRAND, L., AND HELDIN, C. H. Ligand-induced recruitment of na⁺/h⁺-exchanger regulatory factor to the pdgf (platelet-derived growth factor) receptor regulates actin cytoskeleton reorganization by pdgf. *Biochem J* 376, Pt 2 (Dec 2003), 505–510.
- [86] DEMOULIN, J. B., UYTENHOVE, C., VAN ROOST, E., DELESTRÉ, B., DONCKERS, D., VAN SNICK, J., AND RENAULD, J. C. A single tyrosine of the interleukin-9 (il-9) receptor is required for stat activation, antiapoptotic activity, and growth regulation by il-9. *Mol Cell Biol* 16, 9 (Sep 1996), 4710–4716.
- [87] DENG, X., SZABO, S., KHOMENKO, T., JADUS, M. R., YOSHIDA, M., AND CHEN, L. Detection of duodenal ulcer-associated genes in rats. *Dig Dis Sci* 53, 2 (Feb 2008), 375–384.
- [88] DENNIS, G. J., SHERMAN, B. T., HOSACK, D. A., YANG, J., GAO, W., LANE, H. C., AND LEMPICKI, R. A. David: Database for annotation, visualization, and integrated discovery. *Genome Biology* 4, 5 (2003), 3.
- [89] DENOEUDE, F., AURY, J. M., DA SILVA, C., NOEL, B., ROGIER, O., DELLEDONNE, M., MORGANTE, M., VALLE, G., WINCKER, P., SCARPELLI, C., JAILLON, O., AND ARTIGUENAVE, F. Annotating genomes with massive-scale rna sequencing. *Genome Biol* 9, 12 (2008), R175.
- [90] DER, S. D., ZHOU, A., WILLIAMS, B. R. G., AND SILVERMAN, R. H. Identification of genes differentially regulated by interferon α , β , or γ using oligonucleotide arrays. *Proc. Natl. Acad. Sci* 95, 26 (Dec 1998), 15623–15628.
- [91] DERISI, J. Overview of nucleic acid arrays. *Curr Protoc Neurosci Chapter 4* (Nov 2001), Unit 4.25.
- [92] DERISI, J., PENLAND, L., BROWN, P. O., BITTNER, M. L., MELTZER, P. S., RAY, M., CHEN, Y., SU, Y. A., AND TRENT, J. M. Use of a cdna microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14, 4 (Dec 1996), 457–460.
- [93] DIF, N., EUTHINE, V., GONNET, E., LAVILLE, M., VIDAL, H., AND LEFAI, E. Insulin activates human sterol-regulatory-element-binding protein-1c (srebp-1c) promoter through sre motifs. *Biochem J* 400, 1 (Nov 2006), 179–188.

- [94] DIJKERS, P. F., MEDEMA, R. H., PALS, C., BANERJI, L., THOMAS, N. S., LAM, E. W., BURGERING, B. M., RAAIJMAKERS, J. A., LAMMERS, J. W., KOENDERMAN, L., AND COFFER, P. J. Forkhead transcription factor fkhr-1l modulates cytokine-dependent transcriptional regulation of p27(kip1). *Mol Cell Biol* 20, 24 (Dec 2000), 9138–9148.
- [95] DJEBBARI, A., AND QUACKENBUSH, J. Seeded bayesian networks: constructing genetic networks from microarray data. *BMC Syst Biol* 2 (2008), 57.
- [96] DOBBIN, K., SHIH, J. H., AND SIMON, R. Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. *J Natl Cancer Inst* 95, 18 (Sep 2003), 1362–1369.
- [97] DOBBIN, K., AND SIMON, R. Comparison of microarray designs for class comparison and class discovery. *Bioinformatics* 18, 11 (Nov 2002), 1438–1445.
- [98] DOBBIN, K. K., KAWASAKI, E. S., PETERSEN, D. W., AND SIMON, R. M. Characterizing dye bias in microarray experiments. *Bioinformatics* 21, 10 (May 2005), 2430–2437.
- [99] DOBROSOTSKAYA, I. Y., GOLDSTEIN, J. L., BROWN, M. S., AND RAWSON, R. B. Reconstitution of sterol-regulated endoplasmic reticulum-to-golgi transport of srebp-2 in insect cells by co-expression of mammalian scap and insigs. *J Biol Chem* 278, 37 (Sep 2003), 35837–35843.
- [100] DOGHMAN, M., KARPOVA, T., RODRIGUES, G. A., ARHATTE, M., DE MOURA, J., CAVALLI, L. R., VIROLLE, V., BARBRY, P., ZAMBETTI, G. P., FIGUEIREDO, B. C., HECKERT, L. L., AND LALLI, E. Increased steroidogenic factor-1 dosage triggers adrenocortical cell proliferation and cancer. *Mol Endocrinol* 21, 12 (Dec 2007), 2968–2987.
- [101] DU, B., FU, C., KENT, K. C., BUSH, H., SCHULICK, A. H., KREIGER, K., COLLINS, T., AND MCCAFFREY, T. A. Elevated egr-1 in human atherosclerotic cells transcriptionally represses the transforming growth factor-beta type ii receptor. *J Biol Chem* 275, 50 (Dec 2000), 39039–39047.
- [102] DUDOIT, S., YANG, Y. H., CALLOW, M. J., AND SPEED, T. P. statistical methods for identifying differentially expressed genes in replicated cdna microarray experiments. *Statistica sinica* 12, 1 (2002), 111–139.
- [103] EBERLÉ, D., HEGARTY, B., BOSSARD, P., FERRÉ, P., AND FOUFELLE, F. Srebp transcription factors: master regulators of lipid homeostasis. *Biochimie* 86, 11 (Nov 2004), 839–848.
- [104] ECKER, A., SIMMA, O., HOELBL, A., KENNER, L., BEUG, H., MORIGGL, R., AND SEXL, V. The dark and the bright side of stat3: proto-oncogene and tumor-suppressor. *Front Biosci* 14 (2009), 2944–2958.
- [105] EDWARDS, D. Non-linear normalization and background correction in one-channel cdna microarray studies. *Bioinformatics* 19, 7 (May 2003), 825–833.
- [106] EISEN, M. B., SPELLMAN, P. T., BROWN, P. O., AND BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95, 25 (Dec 1998), 14863–14868.
- [107] ELEMENTO, O., SLONIM, N., AND TAVAZOIE, S. A universal framework for regulatory element discovery across all genomes and data types. *Mol Cell* 28, 2 (Oct 2007), 337–350.
- [108] ELLIS, M., DAVIS, N., COOP, A., LIU, M., SCHUMAKER, L., LEE, R. Y., SRIKANCHANA, R., RUSSELL, C. G., SINGH, B., MILLER, W. R., STEARNS, V., PENNANEN, M., TSANGARIS, T., GALLAGHER, A., LIU, A., ZWART, A., HAYES, D. F., LIPPMAN, M. E., WANG, Y., AND CLARKE, R. Development and validation of a method for using breast core needle biopsies for gene expression microarray analyses. *Clin Cancer Res* 8, 5 (May 2002), 1155–1166.
- [109] ENGSTRÖM, M., KARLSSON, R., AND JÖNSSON, J. I. Inactivation of the forkhead transcription factor foxo3 is essential for pkb-mediated survival of hematopoietic progenitor cells by kit ligand. *Exp Hematol* 31, 4 (Apr 2003), 316–323.
- [110] EPSTEIN, J. A., SONG, B., LAKKIS, M., AND WANG, C. Tumor-specific pax3-fkhr transcription factor, but not pax3, activates the platelet-derived growth factor alpha receptor. *Mol Cell Biol* 18, 7 (Jul 1998), 4118–4130.
- [111] ESSAGHIR, A., DIF, N., MARBEHANT, C. Y., COFFER, P. J., AND DEMOULIN, J. B. The transcription of foxo genes is stimulated by foxo3 and repressed by growth factors. *J Biol Chem* 284, 16 (Feb 2009), 10334–10342.
- [112] ESSERS, M. A. G., DE VRIES-SMITS, L. M. M., BARKER, N., POLDERMAN, P. E., BURGERING, B. M. T., AND KORSWAGEN, H. C. Functional interaction between beta-catenin and foxo in oxidative stress signaling. *Science* 308, 5725 (May 2005), 1181–1184.
- [113] ESSERS, M. A. G., WEIJZEN, S., DE VRIES-SMITS, A. M. M., SAARLOOS, I., DE RUITER, N. D., BOS, J. L., AND BURGERING, B. M. T. Foxo

- transcription factor activation by oxidative stress mediated by the small gtpase ral and jnk. EMBO J 23, 24 (Dec 2004), 4802–4812.
- [114] ESWARAKUMAR, V. P., LAX, I., AND SCHLESSINGER, J. Cellular signaling by fibroblast growth factor receptors. Cytokine Growth Factor Rev 16, 2 (Apr 2005), 139–149.
- [115] FAMBROUGH, D., MCCLURE, K., KAZLAUSKAS, A., AND LANDER, E. S. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. Cell 97, 6 (Jun 1999), 727–741.
- [116] FARDIN, P., MORETTI, S., BIASOTTI, B., RICCIARDI, A., BONASSI, S., AND VARESEO, L. Normalization of low-density microarray using external spike-in controls: analysis of macrophage cell lines expression profile. BMC Genomics 8 (2007), 17.
- [117] FENG, W., WANG, G., ZEEBERG, B. R., GUO, K., FOJO, A. T., KANE, D. W., REINHOLD, W. C., LABABIDI, S., WEINSTEIN, J. N., AND WANG, M. D. Development of gene ontology tool for biological interpretation of genomic and proteomic data. AMIA Annu Symp Proc (2003), 839.
- [118] FINKENZELLER, G., SPARACIO, A., TECHNAN, A., MARMÉ, D., AND SIEMEISTER, G. Sp1 recognition sites in the proximal promoter of the human vascular endothelial growth factor gene are essential for platelet-derived growth factor-induced gene expression. Oncogene 15, 6 (Aug 1997), 669–676.
- [119] FISCHER, A. N. M., FUCHS, E., MIKULA, M., HUBER, H., BEUG, H., AND MIKULITS, W. Pdgf essentially links tgfbeta signaling to nuclear betacatenin accumulation in hepatocellular carcinoma progression. Oncogene 26, 23 (May 2007), 3395–3405.
- [120] FODOR, S. P., READ, J. L., PIRRUNG, M. C., STRYER, L., LU, A. T., AND SOLAS, D. Light-directed, spatially addressable parallel chemical synthesis. Science 251, 4995 (Feb 1991), 767–773.
- [121] FU, M., ZHU, X., ZHANG, J., LIANG, J., LIN, Y., ZHAO, L., EHRENGRUBER, M. U., AND CHEN, Y. E. Egr-1 target genes in human endothelial cells identified by microarray analysis. Gene 315 (Oct 2003), 33–41.
- [122] FU, Z., AND TINDALL, D. J. Foxos, cancer and regulation of apoptosis. Oncogene 27, 16 (Apr 2008), 2312–2319.
- [123] FUDA, N. J., ARDEHALI, M. B., AND LIS, J. T. Defining mechanisms that regulate rna polymerase ii transcription in vivo. Nature 461, 7261 (Sep 2009), 186–192.
- [124] FUKUOKA, M., DAITOKU, H., HATTA, M., MATSUZAKI, H., UMEMURA, S., AND FUKAMIZU, A. Negative regulation of forkhead transcription factor afx (foxo4) by cbp-induced acetylation. Int J Mol Med 12, 4 (Oct 2003), 503–508.
- [125] GALLI, A., CRABB, D., PRICE, D., CENI, E., SALZANO, R., SURRENTI, C., AND CASINI, A. Peroxisome proliferator-activated receptor gamma transcriptional regulation is involved in platelet-derived growth factor-induced proliferation of human hepatic stellate cells. Hepatology 31, 1 (Jan 2000), 101–108.
- [126] GARAT, C. V., FANKELL, D., ERICKSON, P. F., REUSCH, J. E. B., BAUER, N. N., MCMURTRY, I. F., AND KLEMM, D. J. Platelet-derived growth factor bb induces nuclear export and proteasomal degradation of creb via phosphatidylinositol 3-kinase/akt signaling in pulmonary artery smooth muscle cells. Mol Cell Biol 26, 13 (Jul 2006), 4934–4948.
- [127] GARCIA, R., AND JOVE, R. Activation of stat transcription factors in oncogenic tyrosine kinase signaling. J Biomed Sci 5, 2 (1998), 79–85.
- [128] GENEMAPP. <http://www.genmapp.org/>, April 2010.
- [129] GHOSH, S. S., GEHR, T. W. B., GHOSH, S., FAKHRY, I., SICA, D. A., LYALL, V., AND SCHOOLWERTH, A. C. Ppargamma ligand attenuates pdgf-induced mesangial cell proliferation: role of map kinase. Kidney Int 64, 1 (Jul 2003), 52–62.
- [130] GIANNAKEAS, N., AND FOTIADIS, D. I. An automated method for gridding and clustering-based segmentation of cDNA microarray images. Comput Med Imaging Graph 33, 1 (Jan 2009), 40–49.
- [131] GILBERTSON, D. G., DUFF, M. E., WEST, J. W., KELLY, J. D., SHEPPARD, P. O., HOFSTRAND, P. D., GAO, Z., SHOEMAKER, K., BUKOWSKI, T. R., MOORE, M., FELDHAUS, A. L., HUMES, J. M., PALMER, T. E., AND HART, C. E. Platelet-derived growth factor c (pdgf-c), a novel growth factor that binds to pdgf alpha and beta receptor. J Biol Chem 276, 29 (Jul 2001), 27406–27414.
- [132] GILLES, F., RAES, M. B., STÉHELIN, D., VANDENBUNDER, B., AND FAFEUR, V. The c-ets-1 proto-oncogene is a new early-response gene differentially regulated by cytokines and growth factors in human fibroblasts. Exp Cell Res 222, 2 (Feb 1996), 370–378.
- [133] GLASBEY, C. A., AND GHAZAL, P. Combinatorial image analysis of dna microarray features. Bioinformatics 19, 2 (Jan 2003), 194–203.

- [134] GLATIGNY, A., DELACROIX, H., TANG, T., FRANÇOIS, N., AGGERBECK, L., AND MUCCHIELLI-GIORGI, M. H. Characterisation and correction of signal fluctuations in successive acquisitions of microarray images. *BMC Bioinformatics* 10 (2009), 98.
- [135] GOLUB, T. R., SLONIM, D. K., TAMAYO, P., HUARD, C., GAASENBEEK, M., MESIROV, J. P., COLLIER, H., LOH, M. L., DOWNING, J. R., CALIGIURI, M. A., BLOOMFIELD, C. D., AND LANDER, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 5439 (Oct 1999), 531–537.
- [136] GORDÂN, R., NARLIKAR, L., AND HARTEMINK, A. J. Finding regulatory dna motifs using alignment-free evolutionary conservation information. *Nucleic Acids Res* (Jan 2010).
- [137] GOUGH, J., KARPLUS, K., HUGHEY, R., AND CHOTHIA, C. Assignment of homology to genome sequences using a library of hidden markov models that represent all proteins of known structure. *J Mol Biol* 313, 4 (Nov 2001), 903–919.
- [138] GRAZUL-BILSKA, A. T., JOHNSON, M. L., BILSKI, J. J., REDMER, D. A., REYNOLDS, L. P., ABDULLAH, A., AND ABDULLAH, K. M. Wound healing: the role of growth factors. *Drugs Today (Barc)* 39, 10 (Oct 2003), 787–800.
- [139] HAGER, G. L., MCNALLY, J. G., AND MISTELI, T. Transcription dynamics. *Mol Cell* 35, 6 (Sep 2009), 741–753.
- [140] HALDERMANS, P., SHKEDY, Z., VAN SANDEN, S., BURZYKOWSKI, T., AND AERTS, M. Using linear mixed models for normalization of cdna microarrays. *Stat Appl Genet Mol Biol* 6 (2007), Article 19.
- [141] HANNENHALLI, S., AND WANG, L. S. Enhanced position weight matrices using mixture models. *Bioinformatics* 21 Suppl 1 (Jun 2005), i204–i212.
- [142] HASSEL, J. C., WINNEMÖLLER, D., SCHARTL, M., AND WELLBROCK, C. Stat5 contributes to antiapoptosis in melanoma. *Melanoma Res* 18, 6 (Dec 2008), 378–385.
- [143] HAWKINS, P. G., AND MORRIS, K. V. Rna and transcriptional modulation of gene expression. *Cell Cycle* 7, 5 (Mar 2008), 602–607.
- [144] HE, X. Unwinding a path to nuclear beta-catenin. *Cell* 127, 1 (Oct 2006), 40–42.
- [145] HEGDE, P., QI, R., ABERNATHY, K., GAY, C., DHARAP, S., GASPARD, R., HUGHES, J. E., SNESRUD, E., LEE, N., AND QUACKENBUSH, J. A concise guide to cdna microarray analysis. *Biotechniques* 29, 3 (Sep 2000), 548–50, 552–4, 556 passim.
- [146] HELDIN, C. H. Simultaneous induction of stimulatory and inhibitory signals by pdgf. *FEBS Lett* 410, 1 (Jun 1997), 17–21.
- [147] HELDIN, C. H., ÖSTMAN, A., AND RÖNNSTRAND, L. Signal transduction via platelet-derived growth factor receptors. *Biochim Biophys Acta* 1378, 1 (Aug 1998), F79–113.
- [148] HESTAND, M. S., VAN GALEN, M., VILLERUS, M. P., VAN OMMEN, G. J. B., DEN DUNNEN, J. T., AND 'T HOEN P A, C. Core.tf: a user-friendly interface to identify evolutionary conserved transcription factor binding sites in sets of co-regulated genes. *BMC Bioinformatics* 9 (2008), 495.
- [149] HIMPE, E., AND KOOIJMAN, R. Insulin-like growth factor-i receptor signal transduction and the janus kinase/signal transducer and activator of transcription (jak-stat) pathway. *Biofactors* 35, 01 (2009), 76–81.
- [150] HINMAN, R. M., BUSHANAM, J. N., NICHOLS, W. A., AND SATTERTHWAITE, A. B. B cell receptor signaling down-regulates forkhead box transcription factor class o 1 mrna expression via phosphatidylinositol 3-kinase and bruton's tyrosine kinase. *J Immunol* 178, 2 (Jan 2007), 740–747.
- [151] HJOBERG, J., LE, L., IMRICH, A., SUBRAMANIAM, V., MATHEW, S. I., VALLONE, J., HALEY, K. J., Y, G. F. H., SHORE, S. A., AND SILVERMAN, E. S. Induction of early growth-response factor 1 by platelet-derived growth factor in human airway smooth muscle. *Am J Physiol Lung Cell Mol Physiol* 286, 4 (Apr 2004), L817–L825.
- [152] HO, K. K., MYATT, S. S., AND LAM, E. W. F. Many forks in the path: cycling with foxo. *Oncogene* 27, 16 (Apr 2008), 2300–2311.
- [153] HO SUI, S. J., MORTIMER, J. R., ARENILLAS, D. J., BRUMM, J., WALSH, C. J., KENNEDY, B. P., AND WASSERMAN, W. W. opossum: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res* 33, 10 (Jun 2005), 3154–3164.
- [154] HONDA, Y., AND HONDA, S. The daf-2 gene network for longevity regulates oxidative stress resistance and mn-superoxide dismutase gene expression in caenorhabditis elegans. *FASEB J* 13, 11 (Aug 1999), 1385–1393.

- [155] HUANG, H., KAO, M. C., ZHOU, X., LIU, J. S., AND WONG, W. H. Determination of local statistical significance of patterns in markov sequences with application to promoter element identification. J Comput Biol 11, 1 (2004), 1–14.
- [156] HUANG, H., AND TINDALL, D. J. Dynamic foxo transcription factors. J Cell Sci 120, Pt 15 (Aug 2007), 2479–2487.
- [157] HUANG, R. P., NGO, L., OKAMURA, D., TUCKER, M., AND ADAMSON, E. D. V-sis induces egr-1 expression by a pathway mediated by c-ha-ras. J Cell Biochem 56, 4 (Dec 1994), 469–479.
- [158] HUBBELL, E., LIU, W., AND MEI, R. Robust estimators for expression analysis. Bioinformatics 18, 12 (Dec 2002), 1585–1592.
- [159] IHMELS, J., BERGMANN, S., BERMAN, J., AND BARKAI, N. Comparative gene expression analysis by differential clustering approach: application to the candida albicans transcription program. PLoS Genet 1, 3 (Sep 2005), e39.
- [160] ILLUMINA. illumina array qc white paper. <http://www.illumina.com>.
- [161] IRIZARRY, R. A. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 2 (2003), 249–264.
- [162] IRIZARRY, R. A., BOLSTAD, B. M., COLLIN, F., COPE, L. M., HOBBS, B., AND SPEED, T. P. Summaries of affymetrix genechip probe level data. Nucleic Acids Res 31, 4 (Feb 2003), e15.
- [163] ITTRICH, C. Normalization for two-channel microarray data. Methods Inf Med 44, 3 (2005), 418–422.
- [164] JACOBS, F. M. J., VAN DER HEIDE, L. P., WIJCHERS, P. J. E. C., BURBACH, J. P. H., HOEKMAN, M. F. M., AND SMIDT, M. P. Foxo6, a novel member of the foxo class of transcription factors with distinct shuttling dynamics. J Biol Chem 278, 38 (Sep 2003), 35959–35967.
- [165] JAHN, T., SINDHU, S., GOOCH, S., SEIPEL, P., LAVORI, P., LEIFHEIT, E., AND WEINBERG, K. Direct interaction between kit and the interleukin-7 receptor. Blood 110, 6 (Sep 2007), 1840–1847.
- [166] JALVY, S., RENAULT, M. A., LEEN, L. L. S., BELLOC, I., BONNET, J., GADEAU, A. P., AND DESGRANGES, C. Autocrine expression of osteopontin contributes to pdgf-mediated arterial smooth muscle cell migration. Cardiovasc Res 75, 4 (Sep 2007), 738–747.
- [167] JEFFERY, I. B., HIGGINS, D. G., AND CULHANE, AEDÍN, C. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics 7 (2006), 359.
- [168] JIANG, L. P., LU, Y., NIE, B. M., AND CHEN, H. Z. Antiproliferative effect of panaxynol on rasmcs via inhibition of erk1/2 and creb. Chem Biol Interact 171, 3 (Feb 2008), 348–354.
- [169] JIANG, L. P., LU, Y., NIE, B. M., AND CHEN, H. Z. Antiproliferative effect of panaxynol on rasmcs via inhibition of erk1/2 and creb. Chem Biol Interact 171, 3 (Feb 2008), 348–354.
- [170] JIANG, Q., WANG, Y., HAO, Y., JUAN, L., TENG, M., ZHANG, X., LI, M., WANG, G., AND LIU, Y. mir2disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res 37, Database issue (Jan 2009), D98–104.
- [171] KABURAGI, Y., SATOH, S., YAMAMOTO-HONDA, R., ITO, Y., AKANUMA, Y., SEKIHARA, H., YASUDA, K., SASAZUKI, T., TADOWAKI, T., AND YAZAKI, Y. Protection of insulin receptor substrate-3 from staurosporine-induced apoptosis. Biochem Biophys Res Commun 300, 2 (Jan 2003), 371–377.
- [172] KADUPITIGE, S. R., LEUNG, K. C., SELLMEIER, J., SIVIENG, J., CATCHPOOLE, D. R., BAIN, M. E., AND GAËTA, B. A. Miner: exploratory analysis of gene interaction networks by machine learning from expression data. BMC Genomics 10 Suppl 3 (2009), S17.
- [173] KALLIN, A., DEMOULIN, J. B., NISHIDA, K., HIRANO, T., RÖNNSTRAND, L., AND HELDIN, C. H. Gab1 contributes to cytoskeletal reorganization and chemotaxis in response to platelet-derived growth factor. J Biol Chem 279, 17 (Apr 2004), 17897–17904.
- [174] KALLIN, A., JOHANNESSEN, L. E., CANI, P. D., MARBEHANT, C. Y., ESSAGHIR, A., FOUFELLE, F., FERRÉ, P., HELDIN, C. H., DELZENNE, N. M., AND DEMOULIN, J. B. Srebp-1 regulates the expression of heme oxygenase 1 and the phosphatidylinositol-3 kinase regulatory subunit p55 gamma. J Lipid Res 48, 7 (Jul 2007), 1628–1636.
- [175] KANEHISA, M., GOTO, S., FURUMICHI, M., TANABE, M., AND HIRAKAWA, M. Kegg for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38, Database issue (Jan 2010), D355–D360.
- [176] KANNAN, K., AMARIGLIO, N., RECHAVI, G., JAKOB-HIRSCH, J., KELA, I., KAMINSKI, N.,

- GETZ, G., DOMANY, E., AND GIVOL, D. Dna microarrays identification of primary and secondary target genes regulated by p53. *Oncogene* 20, 18 (Apr 2001), 2225–2234.
- [177] KAPRANOV, P., WILLINGHAM, A. T., AND GINGERAS, T. R. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* 8, 6 (Jun 2007), 413–423.
- [178] KARIN, M. The regulation of ap-1 activity by mitogen-activated protein kinases. *J Biol Chem* 270, 28 (1995), 16483–16486.
- [179] KATOH, M., AND KATOH, M. Human fox gene family (review). *Int J Oncol* 25, 5 (Nov 2004), 1495–1500.
- [180] KATZER, M., KUMMERT, F., AND SAGERER, G. Methods for automatic microarray image segmentation. *IEEE Trans Nanobioscience* 2, 4 (Dec 2003), 202–214.
- [181] KEL', A. E., KOLCHANOV, N. A., KEL', O. V., ROMASHCHENKO, A. G., ANAN'KO, E. A., IGNAT'EVA, E. V., MERKULOVA, T. I., PODKOLODNAIA, O. A., STEPANENKO, I. L., KOCHETOV, A. V., KOLPAKOV, F. A., PODKOLODNY, N. L., AND NAUMOCHKIN, A. A. Trrd: a database of transcription regulatory regions in eukaryotic genes. *Mol Biol.(Mosk)* 31, 4 (Jul 1997), 626–636. Moscow.
- [182] KELLY, K., COCHRAN, B. H., STILES, C. D., AND LEDER, P. Cell-specific regulation of the c-myc gene by lymphocyte mitogens and platelet-derived growth factor. *Cell* 35, 3 Pt 2 (Dec 1983), 603–610.
- [183] KEPLER, T. B., CROSBY, L., AND MORGAN, K. T. Normalization and analysis of dna microarray data by self-consistency and local regression. *Genome Biol* 3, 7 (Jun 2002), RESEARCH0037.
- [184] KERR, M. K., AND CHURCHILL, G. A. Experimental design for gene expression microarrays. *Biostatistics* 2, 2 (Jun 2001), 183–201.
- [185] KERR, M. K., MARTIN, M., AND CHURCHILL, G. A. Analysis of variance for gene expression microarray data. *J Comput Biol* 7, 6 (2000), 819–837.
- [186] KHOMENKO, T., SZABO, S., DENG, X., JADUS, M. R., ISHIKAWA, H., OSAPAY, K., SANDOR, Z., AND CHEN, L. Suppression of early growth response factor-1 with egr-1 antisense oligodeoxynucleotide aggravates experimental duodenal ulcers. *Am J Physiol Gastrointest Liver Physiol* 290, 6 (Jun 2006), G1211–G1218.
- [187] KITAMURA, T., NAKAE, J., KITAMURA, Y., KIDO, K., BIGGS, W. H., WRIGHT, C. V. E., WHITE, M. F., ARDEN, K. C., AND ACCILI, D. The forkhead transcription factor foxo1 links insulin signaling to pdx1 regulation of pancreatic beta cell growth. *J Clin Invest* 110, 12 (Dec 2002), 1839–1847.
- [188] KNOOPS, L., HAAS, R., DE KEMP, S., MAJOOR, D., BROEKS, A., ELDERING, E., DE BOER, J. P., VERHEIJ, M., VAN OSTROM, C., DE VRIES, A., VAN'T VEER, L., AND DE JONG, D. In vivo p53 response and immune reaction underlie highly effective low-dose radiotherapy in follicular lymphoma. *Blood* 110, 4 (Aug 2007), 1116–1122.
- [189] KOMATSU, N., WATANABE, T., UCHIDA, M., MORI, M., KIRITO, K., KIKUCHI, S., LIU, Q., TAUCHI, T., MIYAZAWA, K., ENDO, H., NAGAI, T., AND OZAWA, K. A member of forkhead transcription factor fhrl1 is a downstream effector of sti571-induced cell cycle arrest in bcr-abl-expressing cells. *J Biol Chem* 278, 8 (Feb 2003), 6411–6419.
- [190] KONJEVIĆ, G. [stat proteins in cancerogenesis and therapy of malignancies]. *Srp Arh Celok Lek* 137, 1-2 (2009), 98–105.
- [191] KORNBERG, R. D. Eukaryotic transcriptional control. *Trends Cell Biol* 9, 12 (Dec 1999), M46–M49.
- [192] KOSTKA, D., AND SPANG, R. Finding disease specific alterations in the co-expression of genes. *Bioinformatics* 20 Suppl 1 (Aug 2004), i194–i199.
- [193] KOTZKA, J., MÜLLER-WIELAND, D., ROTH, G., KREMER, L., MUNCK, M., SCHÜRMANN, S., KNEBEL, B., AND KRONE, W. Sterol regulatory element binding proteins (srebp)-1a and srebp-2 are linked to the map-kinase cascade. *J Lipid Res* 41, 1 (Jan 2000), 99–108.
- [194] KUNDUMANI-SRIDHARAN, V., WANG, D., KARPURAPU, M., LIU, Z., ZHANG, C., DRONADULA, N., AND RAO, G. N. Suppression of activation of signal transducer and activator of transcription-5b signaling in the vessel wall reduces balloon injury-induced neointima formation. *Am J Pathol* 171, 4 (Oct 2007), 1381–1394.
- [195] KUO, Y. M., JONES, N., ZHOU, B., PANZER, S., LARSON, V., AND BECKENDORF, S. K. Salivary duct determination in drosophila: roles of the egf receptor signalling pathway and the transcription factors fork head and trachealess. *Development* 122, 6 (Jun 1996), 1909–1917.
- [196] LANDGREBE, J., BRETZ, F., AND BRUNNER, E. Efficient two-sample designs for microarray experiments with biological replications. *In Silico Biol* 4, 4 (2004), 461–470.

- [197] LAROCHELLE, W. J., JEFFERS, M., McDONALD, W. F., CHILLAKURU, R. A., GIESE, N. A., LOKKER, N. A., SULLIVAN, C., BOLDOG, F. L., YANG, M., VERNET, C., BURGESS, C. E., FERNANDES, E., DEEGLER, L. L., RITTMAN, B., SHIMKETS, J., SHIMKETS, R. A., ROTHBERG, J. M., AND LICHENSTEIN, H. S. Pdgf-d, a new protease-activated growth factor. *Nat Cell Biol* 3, 5 (May 2001), 517–521.
- [198] LEAMAN, D. W., PISHARODY, S., FLICKINGER, T. W., COMMANE, M. A., SCHLESSINGER, J., KERR, I. M., LEVY, D. E., AND STARK, G. R. Roles of jaks in activation of stats and stimulation of c-fos gene expression by epidermal growth factor. *Mol Cell Biol* 16, 1 (Jan 1996), 369–375.
- [199] LEJA, J., ESSAGHIR, A., ESSAND, M., WESTER, K., OBERG, K., TOTTERMAN, T. H., LLOYD, R., VASMATZIS, G., DEMOULIN, J. B., AND GIANDOMENICO, V. Novel markers for enterochromaffin cells and gastrointestinal neuroendocrine carcinomas. *Mod Pathol* 22, 2 (Feb 2009), 261–272.
- [200] LENHARD, B., AND WASSERMAN, W. W. Tfbs: Computational framework for transcription factor binding site analysis. *Bioinformatics* 18, 8 (Aug 2002), 1135–1136.
- [201] LEUNG, Y. F., AND CAVALIERI, D. Fundamentals of cdna microarray data analysis. *Trends Genet* 19, 11 (Nov 2003), 649–659.
- [202] LI, C., AND WONG, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 98, 1 (Jan 2001), 31–36.
- [203] LI, C., AND WONG, W. H. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* 2, 8 (2001), RESEARCH0032.
- [204] LIANG, K. W., TING, C. T., YIN, S. C., CHEN, Y. T., LIN, S. J., LIAO, J. K., AND HSU, S. L. Berberine suppresses mek/erk-dependent egr-1 signaling pathway and inhibits vascular smooth muscle cell regrowth after in vitro mechanical injury. *Biochem Pharmacol* 71, 6 (Mar 2006), 806–817.
- [205] LIM, W. K., WANG, K., LEFEBVRE, C., AND CALIFANO, A. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics* 23, 13 (Jul 2007), i282–i288.
- [206] LIN, J., AND CHEN, A. Activation of peroxisome proliferator-activated receptor-gamma by curcumin blocks the signaling pathways for pdgf and egf in hepatic stellate cells. *Lab Invest* 88, 5 (May 2008), 529–540.
- [207] LIN, K., DORMAN, J. B., RODAN, A., AND KENYON, C. daf-16: An hnf-3/forkhead family member that can function to double the life-span of caenorhabditis elegans. *Science* 278, 5341 (Nov 1997), 1319–1322.
- [208] LINHART, C., HALPERIN, Y., AND SHAMIR, R. Transcription factor and microrna motif discovery: the amadeus platform and a compendium of meta-zoan target sets. *Genome Res* 18, 7 (Jul 2008), 1180–1189.
- [209] LIPSHUTZ, R. J., FODOR, S. P., GINGERAS, T. R., AND LOCKHART, D. J. High density synthetic oligonucleotide arrays. *Nat Genet* 21, 1 Suppl (Jan 1999), 20–24.
- [210] LIU, C. C., LIN, C. C., CHEN, W. S. E., CHEN, H. Y., CHANG, P. C., CHEN, J. J. W., AND YANG, P. C. Crsd: a comprehensive web server for composite regulatory signature discovery. *Nucleic Acids Res* 34, Web Server issue (Jul 2006), W571–W577.
- [211] LIU, H., TARIMA, S., BORDERS, A. S., GETCHELL, T. V., GETCHELL, M. L., AND STROMBERG, A. J. Quadratic regression analysis for gene discovery and pattern recognition for non-cyclic short time-course microarray experiments. *BMC Bioinformatics* 6 (2005), 106.
- [212] LIU, W., MEI, R., BARTELL, D. M., DI, X., WEBSTER, T. A., AND RYDER, M. Rank-based algorithms for analysis of microarrays. *Proc. SPIE* 4266 (2001), 56–67.
- [213] LIU, W. M., MEI, R., DI, X., RYDER, T. B., HUBBELL, E., DEE, S., WEBSTER, T. A., HARRINGTON, C. A., HO, M. H., BAID, J., AND SREEKENS, S. P. Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics* 18, 12 (Dec 2002), 1593–1599.
- [214] LO, I. C., LIN, T. M., CHOU, L. H., LIU, S. L., WU, L. W., SHI, G. Y., WU, H. L., AND JIANG, M. J. Ets-1 mediates platelet-derived growth factor-bb-induced thrombomodulin expression in human vascular smooth muscle cells. *Cardiovasc Res* 81, 4 (Mar 2009), 771–779.
- [215] LOCKHART, D. J., DONG, H., BYRNE, M. C., FOLLETTIE, M. T., GALLO, M. V., CHEE, M. S., MITTMANN, M., WANG, C., KOBAYASHI, M., HORTON, H., AND BROWN, E. L. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14, 13 (Dec 1996), 1675–1680.

- [216] LOMAX, M. I., HUANG, L., CHO, Y., GONG, T. L., AND ALTSCHULER, R. A. Differential display and gene arrays to examine auditory plasticity. *Hear Res* 147, 1-2 (Sep 2000), 293–302.
- [217] LYNCH, R. L., KONICEK, B. W., MCNULTY, A. M., HANNA, K. R., LEWIS, J. E., NEUBAUER, B. L., AND GRAFF, J. R. The progression of Incap human prostate cancer cells to androgen independence involves decreased foxo3a expression and reduced p27kip1 promoter transactivation. *Mol Cancer Res* 3, 3 (Mar 2005), 163–169.
- [218] MA, D., NUTT, C. L., SHANEHSAZ, P., PENG, X., LOUIS, D. N., AND KAETZEL, D. M. Autocrine pdgfa dependent gene expression in glioblastoma is mediated largely by activation of srebp and is associated with altered genotype and patient survival in human brain tumors. *Cancer Research* 65, 13 (Jul 2005), 5523–5534.
- [219] MA, W. L., ZHENG, W. L., CUI, D., SONG, Y. B., AND WU, Q. H. Dna microarray chips made on surface of ceramic slides. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* 32, 3 (2000), 285–289.
- [220] MAEHARA, K., OH-HASHI, K., AND ISOBE, K. I. Early growth-responsive-1-dependent manganese superoxide dismutase gene transcription mediated by platelet-derived growth factor. *FASEB J* 15, 11 (Sep 2001), 2025–2026.
- [221] MARAZIOTIS, I. A., DRAGOMIR, A., AND BEZERRIANOS, A. Gene networks reconstruction and time-series prediction from microarray data using recurrent neural fuzzy networks. *IET Syst Biol* 1, 1 (Jan 2007), 41–50.
- [222] MARDIS, E. R. Next-generation dna sequencing methods. *Annu Rev Genomics Hum Genet* 9 (2008), 387–402.
- [223] MARGUERAT, S., AND BÄHLER, J. Rna-seq: from technology to biology. *Cell Mol Life Sci* 67, 4 (Feb 2010), 569–579.
- [224] MARIONI, J. C., MASON, J. E., MANE, S. M., STEPHENS, M., AND GILAD, Y. Rna-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18, 9 (Sep 2008), 1509–1517.
- [225] MARKOWETZ, F., AND SPANG, R. Inferring cellular networks—a review. *BMC Bioinformatics* 8 Suppl 6 (2007), S5.
- [226] MARSTRAND, T. T., FRELLSEN, J., MOLTKE, I., THIIM, M., VALEN, E., RETELSKA, D., AND KROGH, A. Asap: a framework for over-representation statistics for transcription factor binding sites. *PLoS ONE* 3, 2 (2008), e1623.
- [227] MARTÍNEZ-GONZÁLEZ, J., RIUS, J., CASTELLÓ, A., CASES-LANGHOFF, C., AND BADIMON, L. Neuron-derived orphan receptor-1 (nor-1) modulates vascular smooth muscle cell proliferation. *Circ Res* 92, 1 (Jan 2003), 96–103.
- [228] MASAMUNE, A., SATOH, M., KIKUTA, K., SUZUKI, N., AND SHIMOSEGAWA, T. Activation of jak-stat pathway is required for platelet-derived growth factor-induced proliferation of pancreatic stellate cells. *World J Gastroenterol* 11, 22 (Jun 2005), 3385–3391.
- [229] MATSUZAKI, H., DAITOKU, H., HATTA, M., TANAKA, K., AND FUKAMIZU, A. Insulin-induced phosphorylation of fkhr (foxo1) targets to proteasomal degradation. *Proc Natl Acad Sci U S A* 100, 20 (Sep 2003), 11285–11290.
- [230] MCCAFFREY, T. A., FU, C., DU, B., EKSINAR, S., KENT, K. C., BUSH, H., KREIGER, K., ROSENGART, T., CYBULSKY, M. I., SILVERMAN, E. S., AND COLLINS, T. High-level expression of egr-1 and egr-1-inducible genes in mouse and human atherosclerosis. *J Clin Invest* 105, 5 (Mar 2000), 653–662.
- [231] MECHTA-GRIGORIOU, F., GERALD, D., AND YANIV, M. The mammalian jun proteins: redundancy and specificity. *Oncogene* 20, 19 (Apr 2001), 2378–2389.
- [232] MINDEN, A., LIN, A., SMEAL, T., DÉRIJARD, B., COBB, M., DAVIS, R., AND KARIN, M. c-jun n-terminal phosphorylation correlates with activation of the jnk subgroup but not the erk subgroup of mitogen-activated protein kinases. *Mol Cell Biol* 14, 10 (Oct 1994), 6683–6688.
- [233] MOLANDER, C., HACKZELL, A., OHTA, M., IZUMI, H., AND FUNA, K. Sp1 is a key regulator of the pdgf beta-receptor transcription. *Mol Biol Rep* 28, 4 (2001), 223–233.
- [234] MOLINA, D., GREWAL, S., AND BARDWELL, L. Characterization of an erk-binding domain in microphthalmia-associated transcription factor and differential inhibition of erk2-mediated substrate phosphorylation. *J Biol Chem* 280, 51 (2005), 42051–42060.
- [235] MÖLLER, C., ALFREDSSON, J., ENGSTRÖM, M., WOOTZ, H., XIANG, Z., LENNARTSSON, J., JÖNSSON, J. I., AND NILSSON, G. Stem cell factor promotes mast cell survival via inactivation of

- foxo3a-mediated transcriptional induction and mek-regulated phosphorylation of the proapoptotic protein bim. *Blood* 106, 4 (Aug 2005), 1330–1336.
- [236] MONJE, P., MARINISSEN, M. J., AND GUTKIND, J. S. Phosphorylation of the carboxyl-terminal transactivation domain of c-fos by extracellular signal-regulated kinase mediates the transcriptional activation of ap-1 and cellular transformation induced by platelet-derived growth factor. *Mol Cell Biol* 23, 19 (Oct 2003), 7030–7043.
- [237] MONTGOMERY, S. B., GRIFFITH, O. L., SLEUMER, M. C., BERGMAN, C. M., BILENKY, M., PLEASANCE, E. D., PRYCHYNA, Y., ZHANG, X., AND JONES, S. J. Oreganno: an open access database and curation system for literature-derived promoters, transcription factor binding sites and regulatory variation. *Bioinformatics* 22, 5 (Mars 2006), 637–640.
- [238] MONTI, S., SAVAGE, K. J., KUTOK, J. L., FEUERHAKE, F., KURTIN, P., MIHM, M., WU, B., PASQUALUCCI, L., NEUBERG, D., AGUIAR, R. C. T., CIN, P. D., LADD, C., PINKUS, G. S., SALLES, G., HARRIS, N. L., DALLA-FAVERA, R., HABERMANN, T. M., ASTER, J. C., GOLUB, T. R., AND SHIPP, M. A. Molecular profiling of diffuse large b-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 105, 5 (Mar 2005), 1851–1861.
- [239] MONTI, S., TAMAYO, P., MESIROV, J., AND GOLUB, T. R. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning* 52, 1 (July 2003), 91–118.
- [240] MOON, S. K., JUNG, S. Y., CHOI, Y. H., LEE, Y. C., AND KIM, C. H. Platelet-derived growth factor induces p21/waf1 promoter in vascular smooth muscle cells via activation of an sp1 site. *FEBS Lett* 552, 2-3 (Sep 2003), 130–134.
- [241] MORRIS, K. V. sirna-mediated transcriptional gene silencing: the potential mechanism and a possible role in the histone code. *Cell Mol Life Sci* 62, 24 (Dec 2005), 3057–3066.
- [242] MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L., AND WOLD, B. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nat Methods* 5, 7 (Jul 2008), 621–628.
- [243] MUNDSCHAU, L. J., FORMAN, L. W., WENG, H., AND FALLER, D. V. Platelet-derived growth factor (pdgf) induction of egr-1 is independent of pdgf receptor autophosphorylation on tyrosine. *J Biol Chem* 269, 23 (Jun 1994), 16137–16142.
- [244] MURAKAMI, S., AND JOHNSON, T. E. The old-1 positive regulator of longevity and stress resistance is under daf-16 regulation in caenorhabditis elegans. *Curr Biol* 11, 19 (Oct 2001), 1517–1523.
- [245] NAITO, S., SHIMIZU, S., MAEDA, S., WANG, J., PAUL, R., AND FAGIN, J. A. Ets-1 is an early response gene activated by et-1 and pdgf-bb in vascular smooth muscle cells. *Am J Physiol* 274, 2 Pt 1 (Feb 1998), C472–C480.
- [246] NAKA, K., HOSHII, T., MURAGUCHI, T., TADOKORO, Y., OOSHIO, T., KONDO, Y., NAKAO, S., MOTOYAMA, N., AND HIRAO, A. Tgf-beta-foxo signalling maintains leukaemia-initiating cells in chronic myeloid leukaemia. *Nature* 463, 7281 (Feb 2010), 676–680.
- [247] NAKAE, J., BARR, V., AND ACCILI, D. Differential regulation of gene expression by insulin and igf-1 receptors correlates with phosphorylation of a single amino acid residue in the forkhead transcription factor fkh. *EMBO J* 19, 5 (Mar 2000), 989–996.
- [248] NETPATH. <http://www.netpath.org/index.html>, April 2010.
- [249] NFIREGULOMEDB. <http://nfiregulome.ccr.buffalo.edu/>, January 2010.
- [250] NISHIDA, K., FRITH, M. C., AND NAKAI, K. Pseudocounts for transcription factor binding sites. *Nucleic Acids Res* 37, 3 (Feb 2009), 939–944.
- [251] NOMIYAMA, T., NAKAMACHI, T., GIZARD, F., HEYWOOD, E. B., JONES, K. L., OHKURA, N., KAWAMORI, K., CONNEELY, O. M., AND BRUEMMER, D. The nr4a orphan nuclear receptor nor1 is induced by platelet-derived growth factor and mediates vascular smooth muscle cell proliferation. *J Biol Chem* 281, 44 (Nov 2006), 33467–33476.
- [252] NOWAK, K., KILLMER, K., GESSNER, C., AND LUTZ, W. E2f-1 regulates expression of foxo1 and foxo3a. *Biochim Biophys Acta* 1769, 4 (Apr 2007), 244–252.
- [253] NUNES, R. O., SCHMIDT, M., DUECK, G., BAARSMA, H., HALAYKO, A. J., KERSTJENS, H. A. M., MEURS, H., AND GOSENS, R. Gsk-3/beta-catenin signaling axis in airway smooth muscle: role in mitogenic signaling. *Am J Physiol Lung Cell Mol Physiol* 294, 6 (Jun 2008), L1110–L1118.
- [254] OBSIL, T., AND OBSILOVA, V. Structure/function relationships underlying regulation of foxo transcription factors. *Oncogene* 27, 16 (Apr 2008), 2263–2275.

- [255] OGG, S., PARADIS, S., GOTTLIEB, S., PATTERSON, G. I., LEE, L., TISSENBAUM, H. A., AND RUVKUN, G. The fork head transcription factor daf-16 transduces insulin-like metabolic and longevity signals in *c. elegans*. *Nature* **389**, 6654 (Oct 1997), 994–999.
- [256] ONTOLOGY, G. <http://www.geneontology.org>, April 2010.
- [257] OSTER, S. K., MARHIN, W. W., ASKER, C., FACCHINI, L. M., DION, P. A., FUNA, K., POST, M., SEDIVY, J. M., AND PENN, L. Z. Myc is an essential negative regulator of platelet-derived growth factor beta receptor expression. *Mol Cell Biol* **20**, 18 (Sep 2000), 6768–6778.
- [258] PAIK, J. H., KOLLIPARA, R., CHU, G., JI, H., XIAO, Y., DING, Z., MIAO, L., TOTHOVA, Z., HORNER, J. W., CARRASCO, D. R., JIANG, S., GILLILAND, D. G., CHIN, L., WONG, W. H., CASTRILLON, D. H., AND DEPINHO, R. A. Foxos are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis. *Cell* **128**, 2 (Jan 2007), 309–323.
- [259] PAN, J., QUINTAS-CARDAMA, A., MANSHOURI, T., GILES, F. J., LAMB, P., TEFFERI, A., CORTES, J., KANTARJIAN, H., AND VERSTOVSEK, S. The novel tyrosine kinase inhibitor exel-0862 induces apoptosis in human fip111-pdgfr-alpha-expressing cells through caspase-3-mediated cleavage of mcl-1. *Leukemia* **21**, 7 (Jul 2007), 1395–13404.
- [260] PARK, T., YI, S. G., KANG, S. H., LEE, S. Y., LEE, Y. S., AND SIMON, R. Evaluation of normalization methods for microarray data. *BMC Bioinformatics* **4** (Sep 2003), 33.
- [261] PAUKKU, K., VALGEIRSDÓTTIR, S., SAHARINEN, P., BERGMAN, M., HELDIN, C. H., AND SILVENOINEN, O. Platelet-derived growth factor (pdgf)-induced activation of signal transducer and activator of transcription (stat) 5 is mediated by pdgf beta-receptor and is not dependent on c-src, fyn, jak1 or jak2 kinases. *Biochem J* **345** Pt 3 (Feb 2000), 759–766.
- [262] PAVLIDIS, P., QIN, J., ARANGO, V., MANN, J. J., AND SIBILLE, E. Using the gene ontology for microarray data mining: a comparison of methods and application to age effects in human prefrontal cortex. *Neurochem Res* **29**, 6 (Jun 2004), 1213–1222.
- [263] PEDERSEN, A. G., BALDI, P., CHAUVIN, Y., AND BRUNAK, S. The biology of eukaryotic promoter prediction—a review. *Comput Chem* **23**, 3-4 (Jun 1999), 191–207.
- [264] PELZ, C. R., KULESZ-MARTIN, M., BAGBY, G., AND SEARS, R. C. Global rank-invariant set normalization (grsn) to reduce systematic distortions in microarray data. *BMC Bioinformatics* **9** (2008), 520.
- [265] PENG, T., GOLUB, T. R., AND SABATINI, D. M. The immunosuppressant rapamycin mimics a starvation-like signal distinct from amino acid and glucose deprivation. *Molecular and Cellular Biology* **22**, 15 (Aug 2002), 5575–5584.
- [266] PEREZ-DIEZ, A., MORGUN, A., AND SHULZHENKO, N. Microarrays for cancer diagnosis and classification. *Adv Exp Med Biol* **593** (2007), 74–85.
- [267] PETERLIN, B. M., AND PRICE, D. H. Controlling the elongation phase of transcription with p-tefb. *Mol Cell* **23**, 3 (Aug 2006), 297–305.
- [268] PIETRAS, K., SJÖBLOM, T., RUBIN, K., HELDIN, C. H., AND OSTMAN, A. Pdgf receptors as cancer drug targets. *Cancer Cell* **3**, 5 (May 2003), 439–443.
- [269] PINZANI, M. Pdgf and signal transduction in hepatic stellate cells. *Front Biosci* **7** (Aug 2002), d1720–d1726.
- [270] PIROOZNIA, M., GONG, P., YANG, J. Y., YANG, M. Q., PERKINS, E. J., AND DENG, Y. Iloop—a web application for two-channel microarray interwoven loop design. *BMC Genomics* **9** Suppl 2 (2008), S11.
- [271] POHLERS, D., BEYER, A., KOCZAN, D., WILHELM, T., THIESEN, H. J., AND KINNE, R. W. Constitutive upregulation of the transforming growth factor-beta pathway in rheumatoid arthritis synovial fibroblasts. *Arthritis Res Ther* **9**, 3 (2007), R59.
- [272] PORTALES-CASAMAR, E., ARENILLAS, D., LIM, J., SWANSON, M. I., JIANG, S., MCCALLUM, A., KIROV, S., AND WASSERMAN, W. W. The pazar database of gene regulatory information coupled to the orca toolkit for the study of regulatory sequences. *Nucleic Acids Res* **37**, Database issue (Jan 2009), D54–D60.
- [273] PORTALES-CASAMAR, E., KIROV, S., LIM, J., LITHWICK, S., SWANSON, M. I., TICOLL, A., SNODDY, J., AND WASSERMAN, W. W. Pazar: a framework for collection and dissemination of cis-regulatory sequence annotation. *Genome Biol* **8**, 10 (2007), R207.
- [274] PUIG, O., MARR, M. T., RUHF, M. L., AND TJIAN, R. Control of cell number by drosophila foxo: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev* **17**, 16 (Aug 2003), 2006–2020.

- [275] QUACKENBUSH, J. Computational analysis of microarray data. *Nat Rev Genet* 2, 6 (Jun 2001), 418–427.
- [276] QUACKENBUSH, J. Microarray data normalization and transformation. *Nat Genet* 32 Suppl (Dec 2002), 496–501.
- [277] QUACKENBUSH, J. Computational approaches to analysis of dna microarray data. *Yearb Med Inform* (2006), 91–103.
- [278] QUESNELLE, K. M., BOEHM, A. L., AND R, G. J. Stat-mediated egfr signaling in cancer. *J Cell Biochem* 102, 2 (Oct 2007), 311–319.
- [279] RAFFETSEDER, U., WERNERT, N., OSTENDORF, T., VAN ROEYEN, C., RAUEN, T., BEHRENS, P., FLOEGE, J., AND MERTENS, P. R. Mesangial cell expression of proto-oncogene ets-1 during progression of mesangioproliferative glomerulonephritis. *Kidney Int* 66, 2 (Aug 2004), 622–632.
- [280] RAFTY, L. A., AND KHACHIGIAN, L. M. Zinc finger transcription factors mediate high constitutive platelet-derived growth factor-b expression in smooth muscle cells derived from aortae of newborn rats. *J Biol Chem* 273, 10 (Mar 1998), 5758–5764.
- [281] RAYCHAUDHURI, S., STUART, J. M., AND ALTMAN, R. B. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput* (2000), 455–466.
- [282] RAYCHAUDHURI, S., SUTPHIN, P. D., CHANG, J. T., AND ALTMAN, R. B. Basic microarray analysis: grouping and feature reduction. *Trends Biotechnol* 19, 5 (May 2001), 189–193.
- [283] R/BIOCONDUCTOR. <http://www.bioconductor.org/>, April 2010.
- [284] REAGAN-SHAW, S., AND AHMAD, N. Rna interference-mediated depletion of phosphoinositide 3-kinase activates forkhead box class o transcription factors and induces cell cycle arrest and apoptosis in breast carcinoma cells. *Cancer Res* 66, 2 (Jan 2006), 1062–1069.
- [285] REAGAN-SHAW, S., AND AHMAD, N. The role of forkhead-box class o (foxo) transcription factors in cancer: a target for the management of cancer. *Toxicol Appl Pharmacol* 224, 3 (Nov 2007), 360–368.
- [286] REINER, A., YEKUTIELI, D., AND BENJAMINI, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19, 3 (Feb 2003), 368–375.
- [287] RINGNÉR, M., AND PETERSON, C. Microarray-based cancer diagnosis with artificial neural networks. *Biotechniques Suppl* (Mar 2003), 30–35.
- [288] ROMASHKOVA, J. A., AND MAKAROV, S. S. Nf-kappab is a target of akt in anti-apoptotic pdgf signalling. *Nature* 401, 6748 (Sep 1999), 86–90.
- [289] ROSA, G. J. M., TEIBE, J. P., AND TEMPELMAN, R. J. Reassessing design and analysis of two-colour microarray experiments using mixed effects models. *Comp Funct Genomics* 6, 3 (2005), 123–131.
- [290] RUPPRECHT, H. D., DANN, P., SUKHATME, V. P., STERZEL, R. B., AND COLEMAN, D. L. Effect of vasoactive agents on induction of egr-1 in rat mesangial cells: correlation with mitogenicity. *Am J Physiol* 263, 4 Pt 2 (Oct 1992), F623–F636.
- [291] RUPPRECHT, H. D., HOFER, G., STERZEL, R. B., AND SCHÖCKLMANN, H. O. Egr-1 transcription factor regulates the growth of glomerular mesangium cells. *Med Klin (Munich)* 92, 2 (Feb 1997), 68–73.
- [292] RUPPRECHT, H. D., HOFFER, G., DE HEER, E., STERZEL, R. B., FALLER, G., AND SCHOECKLMANN, H. O. Expression of the transcriptional regulator egr-1 in experimental glomerulonephritis: requirement for mesangial cell proliferation. *Kidney Int* 51, 3 (Mar 1997), 694–702.
- [293] RUPPRECHT, H. D., SUKHATME, V. P., LACY, J., STERZEL, R. B., AND COLEMAN, D. L. Pdgf-induced egr-1 expression in rat mesangial cells is mediated through upstream serum response elements. *Am J Physiol* 265, 3 Pt 2 (Sep 1993), F351–F360.
- [294] SABUDA-WIDEMANN, D., GRABENSEE, B., SCHWANDT, C., AND BLUME, C. Mycophenolic acid inhibits the autocrine pdgf-b synthesis and pdgf-bb-induced mrna expression of egr-1 in rat mesangial cells. *Nephrol Dial Transplant* 24, 1 (Jan 2009), 52–61.
- [295] SACHINIDIS, A., SCHULTE, K., KO, Y., ZU BRICKWEDDE, M. K. M., HOPPE, V., HOPPE, J., AND VETTER, H. The induction of early response genes in rat smooth muscle cells by pdgf-aa is not sufficient to stimulate dna-synthesis. *FEBS Lett* 319, 3 (Mar 1993), 221–224.
- [296] SACHSENMAIER, C., SADOWSKI, H. B., AND COOPER, J. A. Stat activation by the pdgf receptor requires juxtamembrane phosphorylation sites but not src tyrosine kinase activation. *Oncogene* 18, 24 (Jun 1999), 3583–3592.
- [297] SAEED, A. I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN,

- M., STURN, A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V., AND QUACKENBUSH, J. Tm4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 2 (Feb 2003), 374–378.
- [298] SAND, O., TURATSINZE, J. V., AND VAN HELDEN, J. Evaluating the prediction of cis-acting regulatory elements in genome sequences. In *Modern genome annotation: the BioSapiens network*, D. Frishman and A. Valencia, Eds. Springer, 2008.
- [299] SANDELIN, A., ALKEMA, W., ENGSTRÖM, P., WASSERMAN, W. W., AND LENHARD, B. Jaspar: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 32, Database issue (Jan 2004), D91–D94.
- [300] SANO, H., YOKODE, M., TAKAKURA, N., TAKEMURA, G., DOI, T., KATAOKA, H., SUDO, T., NISHIKAWA, S., FUJIWARA, H., NISHIKAWA, S. I., AND KITA, T. Study on pdgf receptor beta pathway in glomerular formation in neonate mice. *Ann N Y Acad Sci* 947 (Dec 2001), 303–305.
- [301] SAUNDERS, A., CORE, L. J., AND LIS, J. T. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* 7, 8 (Aug 2006), 557–567.
- [302] SCHADT, E. E., LI, C., ELLIS, B., AND WONG, W. H. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J Cell Biochem Suppl* 37 (2001), 120–125.
- [303] SCHENA, M., SHALON, D., DAVIS, R. W., AND BROWN, P. O. Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science* 270, 5235 (Oct 1995), 467–470.
- [304] SCHWACHTGEN, J. L., HOUSTON, P., CAMPBELL, C., SUKHATME, V., AND BRADDOCK, M. Fluid shear stress activation of egr-1 transcription in cultured human endothelial and epithelial cells is mediated via the extracellular signal-related kinase 1/2 mitogen-activated protein kinase pathway. *J Clin Invest* 101, 11 (Jun 1998), 2540–2549.
- [305] SEGAL, E., RAVEH-SADKA, T., SCHROEDER, M., UNNERSTALL, U., AND GAUL, U. Predicting expression patterns from regulatory sequence in drosophila segmentation. *Nature* 451, 7178 (Jan 2008), 535–540.
- [306] SEGAL, E., AND WIDOM, J. What controls nucleosome positions? *Trends Genet* 25, 8 (Aug 2009), 335–343.
- [307] SEIFERT, M., SCHERF, M., EPPLE, A., AND WERNER, T. Multievidence microarray mining. *Trends Genet* 21, 10 (Oct 2005), 553–558.
- [308] SEKI, M., NARUSAKA, M., ABE, H., KASUGA, M., YAMAGUCHI-SHINOZAKI, K., CARNINCI, P., HAYASHIZAKI, Y., AND SHINOZAKI, K. Monitoring the expression pattern of 1300 arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* 13, 1 (Jan 2001), 61–72.
- [309] SEOANE, J., LE, H. V., SHEN, L., ANDERSON, S. A., AND MASSAGUÉ, J. Integration of smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell* 117, 2 (Apr 2004), 211–223.
- [310] SETERNES, O. M., JOHANSEN, B., AND MOENS, U. A dominant role for the raf-mek pathway in forskolin, 12-o-tetradecanoyl-phorbol acetate, and platelet-derived growth factor-induced creb (camp-responsive element-binding protein) activation, uncoupled from serine 133 phosphorylation in nih 3t3 cells. *Mol Endocrinol* 13, 7 (Jul 1999), 1071–1083.
- [311] SHANKAVARAM, U. T., REINHOLD, W. C., NISHIZUKA, S., MAJOR, S., MORITA, D., CHARY, K. K., REIMERS, M. A., SCHERF, U., KAHN, A., DOLGINOW, D., COSSMAN, J., KALDIJIAN, E. P., SCUDIERO, D. A., PETRICCINO, E., LIOTTA, L., LEE, J. K., AND WEINSTEIN, J. N. Transcript and protein expression profiles of the nci-60 cancer cell panel: an integromic microarray study. *Mol Cancer Ther* 6, 3 (2007), 820–832.
- [312] SIMEONE-PENNEY, M. C., SEVERGNINI, M., ROZO, L., TAKAHASHI, S., COCHRAN, B. H., AND SIMON, A. R. Pdgf-induced human airway smooth muscle cell proliferation requires stat3 and the small gtpase rac1. *Am J Physiol Lung Cell Mol Physiol* 294, 4 (Apr 2008), L698–L704.
- [313] SIMON, A. R., RAI, U., FANBURG, B. L., AND COCHRAN, B. H. Activation of the jak-stat pathway by reactive oxygen species. *Am J Physiol* 275, 6 Pt 1 (Dec 1998), C1640–C1652.
- [314] SIMON, A. R., TAKAHASHI, S., SEVERGNINI, M., FANBURG, B. L., AND COCHRAN, B. H. Role of the jak-stat pathway in pdgf-stimulated proliferation of human airway smooth muscle cells. *Am J Physiol Lung Cell Mol Physiol* 282, 6 (Jun 2002), L1296–L1304.
- [315] SIMON, R., RADMACHER, M. D., AND DOBBIN, K. Design of studies using dna microarrays. *Genet Epidemiol* 23, 1 (Jun 2002), 21–36.

- [316] SIMON, R. M., AND DOBBIN, K. Experimental design of dna microarray experiments. *Biotechniques Suppl* (Mar 2003), 16–21.
- [317] SINHA, S., ADLER, A. S., FIELD, Y., CHANG, H. Y., AND SEGAL, E. Systematic functional characterization of cis-regulatory motifs in human core promoters. *Genome Res* 18, 3 (Mar 2008), 477–488.
- [318] SINHA, S., AND TOMPA, M. Ymf: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Res* 31, 13 (Jul 2003), 3586–3588.
- [319] SMITH, D., SHIMAMURA, T., BARBERA, S., AND BEJCEK, B. E. Nf-kappaB controls growth of glioblastomas/astrocytomas. *Mol Cell Biochem* 307, 1-2 (Jan 2008), 141–147.
- [320] STAERK, J., KALLIN, A., DEMOULIN, J. B., VAINCHENKER, W., AND CONSTANTINESCU, S. N. Jak1 and tyk2 activation by the homologous polycythemia vera jak2 v617f mutation: cross-talk with igf1 receptor. *J Biol Chem* 280, 51 (Dec 2005), 41893–41899.
- [321] STEVENSON, A. S., CARTIN, L., WELLMAN, T. L., DICK, M. H., NELSON, M. T., AND LOUNSBURY, K. M. Membrane depolarization mediates phosphorylation and nuclear translocation of creb in vascular smooth muscle cells. *Exp Cell Res* 263, 1 (Feb 2001), 118–130.
- [322] STEWART, S. A., DYKXHOORN, D. M., PALLISER, D., MIZUNO, H., YU, E. Y., AN, D. S., SABATINI, D. M., CHEN, I. S. Y., HAHN, W. C., SHARP, P. A., WEINBERG, R. A., AND NOVINA, C. D. Lentivirus-delivered stable gene silencing by rna in primary cells. *RNA* 9, 4 (Apr 2003), 493–501.
- [323] STILLMAN, B. A., AND TONKINSON, J. L. Fast slides: a novel surface for microarrays. *Biotechniques* 29, 3 (Sep 2000), 630–635.
- [324] STOREY, J. D., AND TIBSHIRANI, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100, 16 (Aug 2003), 9440–9445.
- [325] SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S., AND MESIROV, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102, 43 (Oct 2005), 15545–15550.
- [326] SUI, S. J. H., FULTON, D. L., ARENILLAS, D. J., KWON, A. T., AND WASSERMAN, W. W. opossum: integrated tools for analysis of regulatory motif overrepresentation. *Nucleic Acids Res* 35, Web Server issue (Jul 2007), W245–W252.
- [327] SULTAN, M., SCHULZ, M. H., RICHARD, H., MAGEN, A., KLINGENHOFF, A., SCHERF, M., SEIFERT, M., BORODINA, T., SOLDATOV, A., PARKHOMCHUK, D., SCHMIDT, D., O'KEEFFE, S., HAAS, S., VINGRON, M., LEHRACH, H., AND YASPO, M. L. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 5891 (Aug 2008), 956–960.
- [328] SUN, N., AND ZHAO, H. Reconstructing transcriptional regulatory networks through genomics data. *Stat Methods Med Res* 18, 6 (Dec 2009), 595–617.
- [329] SUN, R., FU, X., GUO, F., MA, Z., GOULBOURNE, C., JIANG, M., LI, Y., XIE, Y., AND MAO, Y. A strategy for meta-analysis of short time series microarray datasets. *Front Biosci* 14 (2009), 4058–4070.
- [330] SZABO, S., DENG, X., KHOMENKO, T., CHEN, L., TOLSTANOVA, G., OSAPAY, K., SANDOR, K., AND XIONG, X. New molecular mechanisms of duodenal ulceration. *Ann N Y Acad Sci* 1113 (Oct 2007), 238–255.
- [331] TABERNERO, A., STEWART, H. J. S., JESSEN, K. R., AND MIRSKY, R. The neuron-glia signal beta neuregulin induces sustained creb phosphorylation on ser-133 in cultured rat schwann cells. *Mol Cell Neurosci* 10, 5/6 (Apr 1998), 309–322.
- [332] TALLQUIST, M., AND KAZLAUSKAS, A. Pdgf signaling in cells and mice. *Cytokine Growth Factor Rev* 15, 4 (Aug 2004), 205–213.
- [333] TAN, Y. D., FORNAGE, M., AND FU, Y. X. Ranking analysis of microarray data: a powerful method for identifying differentially expressed genes. *Genomics* 88, 6 (Dec 2006), 846–854.
- [334] TATSUKAWA, H., FUKAYA, Y., FRAMPTON, G., MARTINEZ-FUENTES, A., SUZUKI, K., KUO, T. F., NAGATSUMA, K., SHIMOKADO, K., OKUNO, M., WU, J., IISMAA, S., MATSUURA, T., TSUKAMOTO, H., ZERN, M. A., GRAHAM, R. M., AND KOJIMA, S. Role of transglutaminase 2 in liver injury via cross-linking and silencing of transcription factor spl. *Gastroenterology* 136, 5 (May 2009), 1783–1795.
- [335] TERRAGNI, J., GRAHAM, J. R., ADAMS, K. W., SCHAFFER, M. E., TULLAI, J. W., AND COOPER, G. M. Phosphatidylinositol 3-kinase signaling in proliferating cells maintains an anti-apoptotic transcriptional program mediated by inhibition of foxo and non-canonical activation of nf- κ b transcription factors. *BMC Cell Biology* 9, 6 (Jan 2008).

- [336] THEILGAARD-MONCH, K., KNUDSEN, S., FOLLIN, P., AND BORREGAARD, N. The transcriptional activation program of human neutrophils in skin lesions supports their important role in wound healing. The Journal of Immunology 172, 12 (Jun 2004), 7684–7693.
- [337] THEISEN, C. S., WAHL, J. K., JOHNSON, K. R., AND WHELOCK, M. J. Nherf links the n-cadherin/catenin complex to the platelet-derived growth factor receptor to modulate the actin cytoskeleton and regulate cell motility. Mol Biol Cell 18, 4 (Apr 2007), 1220–1232.
- [338] THOMAS, P. D., CAMPBELL, M. J., KEJARIWAL, A., MI, H., KARLAK, B., DAVERMAN, R., DIEMER, K., MURUGANUJAN, A., AND NARECHANIA, A. Panther: a library of protein families and subfamilies indexed by function. Genome Res 13, 9 (Sep 2003), 2129–2141.
- [339] THOMAS-CHOLLIER, M., SAND, O., TURATSINZE, J. V., JANKY, R., DEFRANCE, M., VERVISCH, E., BROHÉE, S., AND VAN HELDEN, J. Rsat: regulatory sequence analysis tools. Nucleic Acids Res 36, Web Server issue (Jul 2008), W119–W127.
- [340] TIBSHIRANI, R., HASTIE, T., NARASIMHAN, B., AND CHU, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99, 10 (May 2002), 6567–6572.
- [341] TIBSHIRANI, R., HASTIE, T., NARASIMHAN, B., AND CHU, G. Class prediction by nearest shrunken centroids, with applications to dna microarrays. Statistical Science 18, 1 (2003), 104–117.
- [342] TIBSHIRANI, R., WALTHER, G., AND HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 63, 2 (2001), 411–423.
- [343] TOFFALINI, F., KALLIN, A., VANDENBERGHE, P., PIERRE, P., MICHAUX, L., COOLS, J., AND DEMOULIN, J. B. The fusion proteins tel-pdgfr β and fip111-pdgfr α escape ubiquitination and degradation. Hematologica 94, 8 (2009). doi:10.332/hematol.2008.001149.
- [344] TRAPNELL, C., AND SALZBERG, S. L. How to map billions of short reads onto genomes. Nat Biotechnol 27, 5 (May 2009), 455–457.
- [345] TSENG, G. C., OH, M. K., ROHLIN, L., LIAO, J. C., AND WONG, W. H. Issues in cdna microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic Acids Res 29, 12 (Jun 2001), 2549–2557.
- [346] TUSHER, V. G., TIBSHIRANI, R., AND CHU, G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98, 9 (Apr 2001), 5116–5121.
- [347] VALGEIRSDÓTTIR, S., PAUKKU, K., SILVENOINEN, O., HELDIN, C. H., AND CLAESSON-WELSH, L. Activation of stat5 by platelet-derived growth factor (pdgf) is dependent on phosphorylation sites in pdgf beta-receptor juxtamembrane and kinase insert domains. Oncogene 16, 4 (Jan 1998), 505–515.
- [348] VAN DE SLUIS, B., MULLER, P., DURAN, K., CHEN, A., GROOT, A. J., KLOMP, L. W., LIU, P. P., AND WIJMEGA, C. Increased activity of hypoxia-inducible factor 1 is associated with early embryonic lethality in commd1 null mice. Mol Cell Biol 27, 11 (Jun 2007), 4142–4156.
- [349] VAN DER HEIDE, L. P., HOEKMAN, M. F. M., AND SMIDT, M. P. The ins and outs of foxo shuttling: mechanisms of foxo translocation and transcriptional regulation. Biochem J 380, Pt 2 (Jun 2004), 297–309.
- [350] VAN ROEYEN, C. R. C., OSTENDORF, T., DENECHE, B., BOKEMEYER, D., BEHRMANN, I., STRUTZ, F., LICHENSTEIN, H. S., LAROCHELLE, W. J., PENA, C. E., CHAUDHURI, A., AND FLOEGE, J. Biological responses to pdgf-bb versus pdgf-dd in human mesangial cells. Kidney Int 69, 8 (Apr 2006), 1393–1402.
- [351] VANDEPOELE, K., CASNEUF, T., AND VAN DE PEER, Y. Identification of novel regulatory modules in dicotyledonous plants using expression data and comparative genomics. Genome Biol 7, 11 (2006), R103.
- [352] VANTLER, M., CAGLAYAN, E., ZIMMERMANN, W. H., BÄUMER, A. T., AND ROSENKRANZ, S. Systematic evaluation of anti-apoptotic growth factor signaling in vascular smooth muscle cells. only phosphatidylinositol 3'-kinase is important. J Biol Chem 280, 14 (Apr 2005), 14168–14176.
- [353] VANTLER, M., HUNTGEBURTH, M., CAGLAYAN, E., FREYHAUS, H. T., SCHNABEL, P., AND ROSENKRANZ, S. Pi3-kinase/akt-dependent anti-apoptotic signaling by the pdgf alpha receptor is negatively regulated by src family kinases. FEBS Lett 580, 30 (Dec 2006), 6769–6776.
- [354] VEERLA, S., AND HÖGLUND, M. Analysis of promoter regions of co-expressed genes identified by microarray analysis. BMC Bioinformatics 7 (2006), 384.
- [355] VERNELL, R., HELIN, K., AND MÜLLER, H. Identification of target genes of the p16ink4a-prb-e2f

- p pathway.
- J Biol Chem*
- 278, 46 (Nov 2003), 46124–46137.
- [356] VOGT, P. K., JIANG, H., AND AOKI, M. Triple layer control: phosphorylation, acetylation and ubiquitination of foxo proteins. *Cell Cycle* 4, 7 (Jul 2005), 908–913.
- [357] WANG, D., HUANG, J., XIE, H., MANZELLA, L., AND SOARES, M. B. A robust two-way semi-linear model for normalization of cDNA microarray data. *BMC Bioinformatics* 6 (2005), 14.
- [358] WANG, L., FENG, F., WANG, X., WANG, X., AND ZHANG, X. Degseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 1 (Jan 2010), 136–138.
- [359] WANG, Y. Z., WHARTON, W., GARCIA, R., KRAKER, A., JOVE, R., AND PLEDGER, W. J. Activation of stat3 preassembled with platelet-derived growth factor beta receptors requires src kinase activity. *Oncogene* 19, 17 (Apr 2000), 2075–2085.
- [360] WANG, Z., GERSTEIN, M., AND SNYDER, M. RNA-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 1 (Jan 2009), 57–63.
- [361] WARBURTON, D., WUENSCHALL, C., FLORES-DELGADO, G., AND ANDERSON, K. Commitment and differentiation of lung cell lineages. *Biochem Cell Biol* 76, 6 (1998), 971–995.
- [362] WASSERMAN, W. W., AND FICKETT, J. W. Identification of regulatory regions which confer muscle-specific gene expression. *J Mol Biol* 278, 1 (Apr 1998), 167–181.
- [363] WATSON, M. Coxpress: differential co-expression in gene expression data. *BMC Bioinformatics* 7 (2006), 509.
- [364] WEI, D., WILLIAMS, D., AND BROWDER, W. Activation of ap-1 and sp1 correlates with wound growth factor gene expression in glucan-treated human fibroblasts. *Int Immunopharmacol* 2, 8 (Jul 2002), 1163–1172.
- [365] WERNER, T. Bioinformatics applications for pathway analysis of microarray data. *Curr Opin Biotechnol* 19, 1 (Feb 2008), 50–54.
- [366] WHITE, M. K., AND STRAYER, D. S. Survival signaling in type II pneumocytes activated by surfactant protein-A. *Exp Cell Res* 280, 2 (Nov 2002), 270–279.
- [367] WIERENGA, A. T. J., VELLENGA, E., AND SCHURINGA, J. J. Maximal stat5-induced proliferation and self-renewal at intermediate stat5 activity levels. *Mol Cell Biol* 28, 21 (Nov 2008), 6668–6680.
- [368] WILHELM, B. T., MARGUERAT, S., GOODHEAD, I., AND BÄHLER, J. Defining transcribed regions using RNA-seq. *Nat Protoc* 5, 2 (2010), 255–266.
- [369] WILHELM, B. T., MARGUERAT, S., WATT, S., SCHUBERT, F., WOOD, V., GOODHEAD, I., PENKETT, P. J., ROGERS, J., AND BÄHLER, J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 7199 (Jun 2008), 1239–1243.
- [370] WILSON, D., CHAROENSAWAN, V., KUMMERFELD, S. K., AND TEICHMANN, S. A. Dbd-taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res* 36, Database issue (Jan 2008), D88–D92.
- [371] WINGENDER, E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinform* 9, 4 (July 2008), 326–332.
- [372] WINGENDER, E., KEL, A. E., KEL, O. V., KARAS, H., HEINEMEYER, T., DIETZE, P., KNUPPEL, R., ROMASCHENKO, A. G., AND KOLCHANOV, N. A. TRANSFAC, TRRD and COMPIL: towards a federated database system on transcriptional regulation. *Nucleic Acids Res* 25, 01 (January 1997), 265–268.
- [373] WOLFINGER, R. D., GIBSON, G., WOLFINGER, E. D., BENNETT, L., HAMADEH, H., BUSHEL, P., AFSHARI, C., AND PAULES, R. S. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol* 8, 6 (2001), 625–637.
- [374] WORKMAN, C., JENSEN, L. J., JARMER, H., BERKA, R., GAUTIER, L., NIELSER, H. B., SAXILD, H. H., NIELSEN, C., BRUNAK, S., AND KNUDSEN, S. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* 3, 9 (Aug 2002), research0048.
- [375] WU, Z., IRIZARRY, R. A., GENTLEMAN, R., MURILLO, F., AND SPENCER, F. A model based background adjustment for oligonucleotide expression arrays. *Technical Report, John Hopkins University, Department of Biostatistics Working Papers, Baltimore, MD* (2003).
- [376] YANG, J. Y., ZONG, C. S., XIA, W., YAMAGUCHI, H., DING, Q., XIE, X., LANG, J. Y., LAI, C. C., CHANG, C. J., HUANG, W. C., HUANG, H., KUO, H. P., LEE, D. F., LI, L. Y., LIEN, H. C., CHENG, X., CHANG, K. J., HSIAO, C. D., TSAI, F. J., TSAI, C. H., SAHIN, A. A., MULLER, W. J., MILLS, G. B., YU, D., HORTOBAGYI, G. N., AND HUNG, M. C. Erk promotes tumorigenesis by inhibiting

- foxo3a via mdm2-mediated degradation. *Nat Cell Biol* 10, 2 (Feb 2008), 138–148.
- [377] YANG, L., LIN, C., AND LIU, Z. R. P68 rna helicase mediates pdgf-induced epithelial mesenchymal transition by displacing axin from beta-catenin. *Cell* 127, 1 (Oct 2006), 139–155.
- [378] YANG, L., LIN, C., ZHAO, S., WANG, H., AND LIU, Z. R. Phosphorylation of p68 rna helicase plays a role in platelet-derived growth factor-induced cell proliferation by up-regulating cyclin d1 and c-myc expression. *J Biol Chem* 282, 23 (Jun 2007), 16811–16819.
- [379] YANG, L., LIN, C., ZHAO, S., WANG, H., AND LIU, Z. R. Phosphorylation of p68 rna helicase plays a role in platelet-derived growth factor-induced cell proliferation by up-regulating cyclin d1 and c-myc expression. *J Biol Chem* 282, 23 (Jun 2007), 16811–16819.
- [380] YANG, X., QIAO, D., MEYER, K., AND FRIEDL, A. Signal transducers and activators of transcription mediate fibroblast growth factor-induced vascular endothelial morphogenesis. *Cancer Res* 69, 4 (Feb 2009), 1668–1677.
- [381] YANG, Y. H., BUCKLEY, M. J., AND SPEED, T. P. Analysis of cDNA microarray images. *Brief Bioinform* 2, 4 (Dec 2001), 341–349.
- [382] YANG, Y. H., DUDOIT, S., LUU, P., LIN, D. M., PENG, V., NGAI, J., AND SPEED, T. P. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30, 4 (Feb 2002), e15.
- [383] YANG, Y. H., AND SPEED, T. P. Design issues for cDNA microarray experiments. *Nat Rev Genet* 3, 8 (Aug 2002), 579–588.
- [384] YAO, H., PENG, F., FAN, Y., ZHU, X., HU, G., AND BUCH, S. J. Trpc channel-mediated neuroprotection by pdgf involves pyk2/erk/creb pathway. *Cell Death Differ* 16, 12 (Dec 2009), 1681–1693.
- [385] YELLATURU, C. R., BHANOORI, M., NEELI, I., AND RAO, G. N. N-ethylmaleimide inhibits platelet-derived growth factor bb-stimulated akt phosphorylation via activation of protein phosphatase 2a. *J Biol Chem* 277, 42 (Oct 2002), 40148–40155.
- [386] YEUNG, K. Y., HAYNOR, D. R., AND RUZZO, W. L. Validating clustering for gene expression data. *Bioinformatics* 17, 4 (Apr 2001), 309–318.
- [387] ZACHARIA, E., AND MAROULIS, D. An original genetic approach to the fully automatic gridding of microarray images. *IEEE Trans Med Imaging* 27, 6 (Jun 2008), 805–813.
- [388] ZHANG, G. S., LI, B., PEI, M. F., DAI, C. W., ZHENG, W. L., AND SHEN, J. K. Identification of fip111-pdgfra fusion, and expression of signal transducer and activator of transcription 5 in hypereosinophilic syndrome. *Zhonghua Yi Xue Za Zhi* 84, 18 (Sep 2004), 1541–1544.
- [389] ZHANG, J. Y., ADAMS, A. E., RIDKY, T. W., TAO, S., AND KHAVARI, P. A. Tumor necrosis factor receptor 1/c-jun-nh2-kinase signaling promotes human neoplasia. *Cancer Res* 67, 8 (Apr 2007), 3827–3834.
- [390] ZHAO, F., XUAN, Z., LIU, L., AND ZHANG, M. Q. Tred: a transcriptional regulatory element database and a platform for in silico gene regulation studies. *Nucleic Acids Research* 35, Database issue (Jan 2007), 137–140.
- [391] ZHAO, Z., PEYTAVI, R., DIAZ-QUIJADA, G. A., PICARD, F. J., HULETSKY, A., LEBLANC, E., FRENETTE, J., BOIVIN, G., VERES, T., DU-MOULIN, M. M., AND BERGERON, M. G. Plastic polymers for efficient dna microarray hybridization: application to microbiological diagnostics. *J Clin Microbiol* 46, 11 (Nov 2008), 3752–3758.
- [392] ZHENG, W. H., KAR, S., AND QUIRION, R. Insulin-like growth factor-1-induced phosphorylation of transcription factor fkhrl1 is mediated by phosphatidylinositol 3-kinase/akt kinase and role of this pathway in insulin-like growth factor-1-induced survival of cultured hippocampal neurons. *Mol Pharmacol* 62, 2 (Aug 2002), 225–233.
- [393] ZHOU, B., BAGRI, A., AND BECKENDORF, S. K. Salivary gland determination in drosophila: a salivary-specific, fork head enhancer integrates spatial pattern and allows fork head autoregulation. *Dev Biol* 237, 1 (Sep 2001), 54–67.
- [394] ZHU, J., AND HASTIE, T. Classification of gene microarrays by penalized logistic regression. *Biostatistics* 5, 3 (Jul 2004), 427–443.

Annexes

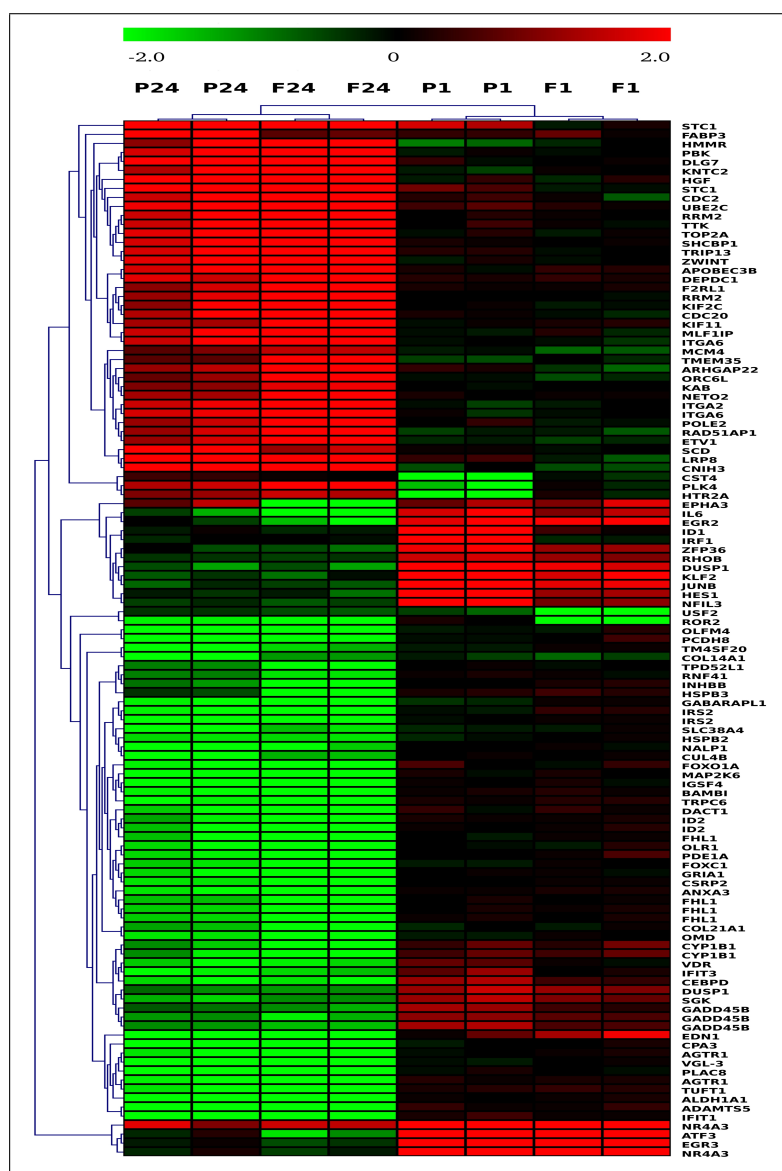
SUPPLEMENTARY TABLE 1

Supplementary Table 1: Oligonucleotides used for QPCR, ChIP and cloning, plus shRNA targeting FOXO1 gene.

Ref #	Gene	Application	Forward primer	Reverse primer
JB408/409	FOXO1 exons 2-3 (3' UTR)	qPCR	GCTTCCACACAGTGTCAAGAC	CCTGCTGTCAGACAATCTGAAGTAC
JB456/457	FOXO1 exons 1-2 & ectopic ¹	qPCR	CGCAGATCTACGAGTGGATGGT	GGACAGATTGTGGCGAATTGAA
JB517/518	FOXO3 exons 2-3 (3' UTR)	qPCR	GGGGAACTTCACTGGTGCTA	TGTCCACTTGCTGAGAGCAG
JB388/389	FOXO3 exons 1-2 & ectopic ¹	qPCR	TGAGAAGTTCCCAGCGACTT	TCCAACCCATCAGCATCCA
JB410/411	FOXO4 exons 2-3	qPCR	TTGGAGAACCTGGAGTGTGACA	AAGCTTCCAGGCATGACTCAG
JB414/415	FOXO6 exons 1-2	qPCR	CCGTTACGTGCCCTACTTCAAG	TTCTGCACGCGGATGAAAC
OPC1845/1846	p27 kip1 exons 1-2-3	qPCR	CCTGCAACCGACGATTCTTC	GGGCGTCTGCTCCACAGA
JB458/459	RPLP0 exons 6-7	qPCR	TCTATCATCAACGGGTACAAACGA	GGCCTTGACCTTTTCAGCAA
JB553/554	FOXO1 promoter (108 bp)	ChIP	AGTACTCGGCTCTGCTGCTC	TGTTTCTGTGATTCTCTCGCC
JB441/442	FOXO1 promoter (1238 bp)	Cloning	GACAGATCGCAGCGAAAGAAAA	GCCGCGAGGAATTCTACTCAA
FOXO1 shRNA sequences:				
TRCN0000039578	FOXO1 shRNA	FOXO1 gene silencing	CCGGGCCTGTTATCAATCTGCTAAACTCGAGTTTAGCAGATTGATAACAGGCTTTTGG	
TRCN0000039581	FOXO1 shRNA	FOXO1 gene silencing	CCGGCAGGACAATAAGTCGAGTTATCTCGAGATAACTCGACTTATTGTCCTGTTTTG	

¹These oligonucleotides amplify FOXO constructs expressed from plasmids and retroviral vectors.

SUPPLEMENTARY FIGURE 1



Supplementary figure 1: Hierarchical clustering of genes regulated by PDGF-BB and b-FGF.
 SAM analysis was performed between conditions and the significant probe sets were hierarchical clustered. The intensities are in log₂-ratios. F1: b-FGF(1h); F24: b-FGF(24h); P1: PDGF-BB(1h); P24: PDGF-BB(24h).

**SUPPLEMENTARY TABLE 2: TFACTS
PREDICTIONS IN EOL-1 EXPERIMENT.**

Activated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
STAT	1.00E-005	2.50E-004	8.95E-005	2.00E-003	9	89	0
STAT1	2.00E-005	5.00E-004	8.95E-005	4.00E-003	5	23	0
STAT3	9.00E-005	2.25E-003	2.69E-004	6.00E-003	6	48	0
STAT5	4.50E-004	1.13E-002	1.01E-003	8.00E-003	5	42	0
NFKB1	3.19E-002	7.97E-001	5.70E-002	1.00E-002	1	2	0

Inhibited Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
LEF1	1.61E-002	4.02E-001	4.02E-001	2.00E-003	1	1	0

Regulated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
STAT	0.00E+000	0.00E+000	0.00E+000	6.94E-004	10	89	3
STAT1	0.00E+000	0.00E+000	0.00E+000	1.39E-003	8	61	1
STAT5	0.00E+000	0.00E+000	0.00E+000	2.08E-003	6	42	1
STAT3	0.00E+000	0.00E+000	0.00E+000	2.78E-003	7	67	2
NFKB1	0.00E+000	0.00E+000	0.00E+000	3.47E-003	9	119	1
TCF/BETA-CATENIN	2.00E-005	1.44E-003	2.16E-005	4.17E-003	11	306	15
RELA	2.00E-005	1.44E-003	2.16E-005	4.86E-003	6	70	1
BCL6	5.00E-005	3.60E-003	4.72E-005	5.56E-003	3	10	0
NFKB	2.80E-004	2.02E-002	2.35E-004	6.25E-003	5	74	0
STAT5B	3.60E-004	2.59E-002	2.72E-004	6.94E-003	3	18	0
FOXO	4.20E-004	3.02E-002	2.89E-004	7.64E-003	7	175	5
REL	5.70E-004	4.10E-002	3.59E-004	8.33E-003	3	21	2
STAT5A	6.60E-004	4.75E-002	3.84E-004	9.03E-003	3	22	0
FOXO1	9.50E-004	6.84E-002	4.89E-004	9.72E-003	6	145	3
E2F1	9.70E-004	6.98E-002	4.89E-004	1.04E-002	3	25	0
JUND	1.09E-003	7.85E-002	5.15E-004	1.11E-002	3	26	0
GLI	1.22E-003	8.78E-002	5.42E-004	1.18E-002	7	209	5
LEF1	1.66E-003	1.20E-001	6.97E-004	1.25E-002	3	30	0
MYC	4.12E-003	2.97E-001	1.64E-003	1.32E-002	11	564	19
P53	5.34E-003	3.85E-001	2.02E-003	1.39E-002	5	143	0
ELK1	6.12E-003	4.41E-001	2.20E-003	1.46E-002	2	15	0
EGR1	6.43E-003	4.63E-001	2.21E-003	1.53E-002	4	94	1
GAS	7.97E-003	5.74E-001	2.52E-003	1.60E-002	1	1	0
JUN	8.00E-003	5.76E-001	2.52E-003	1.67E-002	3	52	0
ATF4	8.78E-003	6.32E-001	2.65E-003	1.74E-002	2	18	0
ETS1	9.15E-003	6.59E-001	2.65E-003	1.81E-002	4	104	1
GLI2	9.46E-003	6.81E-001	2.65E-003	1.88E-002	4	105	2
NFIX	1.59E-002	1.14E+000	4.24E-003	1.94E-002	1	2	0
SP1	1.63E-002	1.17E+000	4.24E-003	2.01E-002	8	417	16
MAZ	2.37E-002	1.71E+000	5.60E-003	2.08E-002	1	3	0
FOSL1	2.37E-002	1.71E+000	5.60E-003	2.15E-002	1	3	0
ESR2	2.37E-002	1.71E+000	5.60E-003	2.22E-002	1	3	0
EGR3	3.15E-002	2.27E+000	7.00E-003	2.29E-002	1	4	0
FOXO4	3.15E-002	2.27E+000	7.00E-003	2.36E-002	1	4	0
TFAP2A	3.55E-002	2.55E+000	7.65E-003	2.43E-002	3	91	0
E2F	3.65E-002	2.63E+000	7.65E-003	2.50E-002	2	38	0
CREB1	4.48E-002	3.23E+000	9.08E-003	2.57E-002	4	169	5
ARNT	4.69E-002	3.38E+000	9.08E-003	2.64E-002	1	6	0
ATF6	4.69E-002	3.38E+000	9.08E-003	2.71E-002	1	6	0

SUPPLEMENTARY TABLE 3: TFACTS PREDICTIONS IN HUMAN FIBROBLAST EXPERIMENT.

TFactS predictions based on significant gene lists regulated more than 2-fold by PDGF-BB and b-FGF after 24h of stimulation.

Activated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
SREBP	1.00E-005	3.00E-004	3.00E-004	1.67E-003	9	46	0
E2F	7.70E-003	2.31E-001	1.16E-001	3.33E-003	5	38	0
PPARA	3.40E-002	1.02E+000	3.40E-001	5.00E-003	2	9	0

Inhibited Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
FOXO	1.00E-005	3.00E-004	3.00E-004	1.67E-003	18	175	0
FOXO1	2.00E-004	6.00E-003	3.00E-003	3.33E-003	14	144	0
FOXO3	1.29E-003	3.87E-002	1.29E-002	5.00E-003	8	66	0
RUNX2	3.34E-002	1.00E+000	2.50E-001	6.67E-003	1	1	0

Regulated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
FOXO1	0.00E+000	0.00E+000	0.00E+000	5.26E-004	15	145	7
SREBP	0.00E+000	0.00E+000	0.00E+000	1.05E-003	9	46	0
FOXO	0.00E+000	0.00E+000	0.00E+000	1.58E-003	19	175	10
GLI	0.00E+000	0.00E+000	0.00E+000	2.11E-003	16	209	15
TCF/BETA-CATENIN	0.00E+000	0.00E+000	0.00E+000	2.63E-003	24	306	35
FOXO3	2.00E-005	1.90E-003	2.43E-005	3.16E-003	8	68	4
SP1	5.00E-005	4.75E-003	5.21E-005	3.68E-003	19	417	56
GLI2	6.00E-005	5.70E-003	5.47E-005	4.21E-003	9	105	5
AP1	1.00E-004	9.50E-003	8.11E-005	4.74E-003	8	88	4
CREB	1.90E-004	1.81E-002	1.39E-004	5.26E-003	6	51	1
E2F	3.80E-004	3.61E-002	2.52E-004	5.79E-003	5	38	3
E2F4	4.30E-004	4.09E-002	2.58E-004	6.32E-003	4	22	1
ATF1	4.60E-004	4.37E-002	2.58E-004	6.84E-003	6	60	7
STAT	6.70E-004	6.37E-002	3.49E-004	7.37E-003	7	89	2
STAT3	8.30E-004	7.89E-002	4.04E-004	7.90E-003	6	67	0
FLI1	1.11E-003	1.06E-001	5.06E-004	8.42E-003	4	28	2
ESR1	1.65E-003	1.57E-001	7.08E-004	8.95E-003	5	52	1
CREB1	1.97E-003	1.87E-001	7.99E-004	9.47E-003	9	169	8
ZBTB16	2.68E-003	2.55E-001	1.03E-003	1.00E-002	2	5	0
ATF2	3.19E-003	3.03E-001	1.16E-003	1.05E-002	4	37	3
TFAP2A	4.00E-003	3.80E-001	1.39E-003	1.11E-002	6	91	1
MYC	5.33E-003	5.06E-001	1.75E-003	1.16E-002	18	564	66
SP4	5.50E-003	5.23E-001	1.75E-003	1.21E-002	2	7	0
RARA	7.58E-003	7.20E-001	2.24E-003	1.26E-002	4	47	1
NFKB	7.68E-003	7.30E-001	2.24E-003	1.32E-002	5	74	3
ETS2	7.99E-003	7.59E-001	2.24E-003	1.37E-002	3	25	0
JUN	1.08E-002	1.03E+000	2.92E-003	1.42E-002	4	52	0
ETV4	1.33E-002	1.26E+000	3.46E-003	1.47E-002	3	30	1
RUNX2	1.67E-002	1.59E+000	3.70E-003	1.53E-002	1	1	0
GAS	1.67E-002	1.59E+000	3.70E-003	1.58E-002	1	1	0
FRA1	1.67E-002	1.59E+000	3.70E-003	1.63E-002	1	1	0
AP2G	1.67E-002	1.59E+000	3.70E-003	1.68E-002	1	1	0
ELK4	1.67E-002	1.59E+000	3.70E-003	1.74E-002	1	1	0
CEBPB	1.86E-002	1.77E+000	3.99E-003	1.79E-002	4	61	2
ETS	2.29E-002	2.18E+000	4.78E-003	1.84E-002	4	65	4
PPARG	2.87E-002	2.73E+000	5.75E-003	1.90E-002	3	40	1
RELA	2.91E-002	2.77E+000	5.75E-003	1.95E-002	4	70	2
AP2B	3.32E-002	3.15E+000	6.20E-003	2.00E-002	1	2	0
HOXA10	3.32E-002	3.15E+000	6.20E-003	2.05E-002	1	2	0
CEBPA	3.53E-002	3.35E+000	6.44E-003	2.11E-002	5	109	14
PPARA	4.33E-002	4.11E+000	7.50E-003	2.16E-002	2	20	2
SP3	4.71E-002	4.47E+000	7.50E-003	2.21E-002	5	118	3
REL	4.73E-002	4.49E+000	7.50E-003	2.26E-002	2	21	0
HLF	4.93E-002	4.69E+000	7.50E-003	2.32E-002	1	3	0
NR2E1	4.93E-002	4.69E+000	7.50E-003	2.37E-002	1	3	0
ARX	4.93E-002	4.69E+000	7.50E-003	2.42E-002	1	3	0
HNF1	4.93E-002	4.69E+000	7.50E-003	2.47E-002	1	3	0
FOSL1	4.93E-002	4.69E+000	7.50E-003	2.53E-002	1	3	0

SUPPLEMENTARY TABLE 4: TFACTS PREDICTIONS IN HUMAN FIBROBLAST EXPERIMENT.

TFactS predictions based on significant gene lists regulated more than 2-fold by PDGF-BB and b-FGF after 1h of stimulation.

Activated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Detection(%)
CREB	4.03E-003	8.06E-002	8.06E-002	2.50E-003	3	51	0
NFKB	9.42E-003	1.88E-001	9.42E-002	5.00E-003	3	69	0
AP1	1.51E-002	3.02E-001	1.01E-001	7.50E-003	3	82	1
SMAD	3.56E-002	7.11E-001	1.45E-001	1.00E-002	2	46	0
GLI	4.32E-002	8.63E-001	1.45E-001	1.25E-002	4	210	0
TCF/BETA-CATENIN	4.34E-002	8.68E-001	1.45E-001	1.50E-002	5	310	0

Regulated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Detection(%)
NFKB	3.00E-005	1.74E-003	1.50E-004	8.62E-004	4	74	0
RELA	3.00E-005	1.74E-003	1.50E-004	1.72E-003	4	70	1
CEBPA	1.60E-004	9.28E-003	5.34E-004	2.59E-003	4	109	0
NFKB1	2.20E-004	1.28E-002	5.50E-004	3.45E-003	4	119	2
CREB	3.00E-004	1.74E-002	6.00E-004	4.31E-003	3	51	0
P53	4.50E-004	2.61E-002	7.15E-004	5.17E-003	4	143	1
CEBPB	5.00E-004	2.90E-002	7.15E-004	6.03E-003	3	61	0
ELK1	6.80E-004	3.94E-002	8.50E-004	6.90E-003	2	15	0
CREB1	8.40E-004	4.87E-002	9.20E-004	7.76E-003	4	169	4
TCF/BETA-CATENIN	9.20E-004	5.34E-002	9.20E-004	8.62E-003	5	306	1
REL	1.34E-003	7.77E-002	1.08E-003	9.48E-003	2	21	0
CEBPD	1.34E-003	7.77E-002	1.08E-003	1.03E-002	2	21	0
AP1	1.47E-003	8.53E-002	1.08E-003	1.12E-002	3	88	0
TFAP2A	1.62E-003	9.40E-002	1.08E-003	1.21E-002	3	91	0
JUND	1.62E-003	9.40E-002	1.08E-003	1.29E-002	2	23	0
EGR1	1.78E-003	1.03E-001	1.10E-003	1.38E-002	3	94	0
GLI	1.86E-003	1.08E-001	1.10E-003	1.47E-002	4	209	0
USF1	2.58E-003	1.50E-001	1.43E-003	1.55E-002	3	107	0
ATF2	4.16E-003	2.41E-001	2.19E-003	1.64E-002	2	37	0
SMAD	6.38E-003	3.70E-001	3.19E-003	1.72E-002	2	46	0
ESR2	7.92E-003	4.59E-001	3.68E-003	1.81E-002	1	3	0
ESR1	8.09E-003	4.69E-001	3.68E-003	1.90E-002	2	52	0
FOXO	1.02E-002	5.92E-001	4.27E-003	1.98E-002	3	175	0
STAT6	1.06E-002	6.12E-001	4.27E-003	2.07E-002	1	4	0
ATF1	1.07E-002	6.19E-001	4.27E-003	2.16E-002	2	60	0
ETS	1.24E-002	7.22E-001	4.79E-003	2.24E-002	2	65	2
STAT3	1.32E-002	7.64E-001	4.85E-003	2.33E-002	2	67	2
FOXO3	1.36E-002	7.87E-001	4.85E-003	2.41E-002	2	68	0
NFKB2	1.58E-002	9.15E-001	5.09E-003	2.50E-002	1	6	0
CEBPG	1.58E-002	9.15E-001	5.09E-003	2.59E-002	1	6	0
JUND1	1.58E-002	9.15E-001	5.09E-003	2.67E-002	1	6	0
SP1	2.12E-002	1.23E+000	6.56E-003	2.76E-002	4	417	2
USF2	2.16E-002	1.25E+000	6.56E-003	2.85E-002	2	87	0
SRF	2.36E-002	1.37E+000	6.74E-003	2.93E-002	1	9	0
NFAT2	2.36E-002	1.37E+000	6.74E-003	3.02E-002	1	9	0
ETS1	3.02E-002	1.75E+000	8.25E-003	3.10E-002	2	104	1
GLI2	3.07E-002	1.78E+000	8.25E-003	3.19E-002	2	105	0
NFAT1	3.13E-002	1.82E+000	8.25E-003	3.28E-002	1	12	0
ATF4	4.67E-002	2.71E+000	1.20E-002	3.36E-002	1	18	0

SUPPLEMENTARY TABLE 5: TFACTS PREDICTIONS IN HUMAN FIBROBLAST EXPERIMENT.

TFactS predictions based on significant gene lists regulated specifically by PDGF-BB after 1h of stimulation.

Activated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
STAT3	0.00E+000	0.00E+000	0.00E+000	2.38E-003	8	48	2
STAT	0.00E+000	0.00E+000	0.00E+000	4.76E-003	10	89	2
STAT1	5.00E-005	1.05E-003	3.47E-004	7.14E-003	4	23	1
STAT5	5.40E-004	1.13E-002	2.81E-003	9.52E-003	4	42	2
SMAD	8.64E-003	1.81E-001	3.60E-002	1.19E-002	3	46	0
NFKB1	1.89E-002	3.96E-001	6.54E-002	1.43E-002	1	2	0

Regulated Transcription Factors:

Transcription Factor	P.value	E.value	Q.value	FDR control (B-H)	Intersection	Target genes	Random Control(%)
STAT3	0.00E+000	0.00E+000	0.00E+000	7.35E-004	8	67	1
STAT	0.00E+000	0.00E+000	0.00E+000	1.47E-003	10	89	0
JUN	1.00E-005	6.80E-004	2.37E-005	2.21E-003	5	52	0
JUND	1.00E-005	6.80E-004	2.37E-005	2.94E-003	4	26	0
RELA	3.00E-005	2.04E-003	5.68E-005	3.68E-003	5	70	0
STAT5	6.00E-005	4.08E-003	9.47E-005	4.41E-003	4	42	0
SMAD3	1.60E-004	1.09E-002	2.13E-004	5.15E-003	3	21	0
SP1	1.80E-004	1.22E-002	2.13E-004	5.88E-003	9	417	8
STAT1	2.50E-004	1.70E-002	2.63E-004	6.62E-003	4	61	0
NFKB1	3.20E-004	2.18E-002	3.03E-004	7.35E-003	5	119	0
NFKB	5.20E-004	3.54E-002	4.48E-004	8.09E-003	4	74	3
SMAD	1.63E-003	1.11E-001	1.26E-003	8.82E-003	3	46	1
FOS	1.73E-003	1.18E-001	1.26E-003	9.56E-003	3	47	0
TCF/BETA-CATENIN	4.21E-003	2.86E-001	2.44E-003	1.03E-002	6	306	2
JUNB	4.62E-003	3.14E-001	2.44E-003	1.10E-002	2	20	0
SMAD4	4.62E-003	3.14E-001	2.44E-003	1.18E-002	2	20	1
GAS	5.16E-003	3.51E-001	2.44E-003	1.25E-002	1	1	0
AP2G	5.16E-003	3.51E-001	2.44E-003	1.32E-002	1	1	0
FRA1	5.16E-003	3.51E-001	2.44E-003	1.40E-002	1	1	0
PITX2	5.16E-003	3.51E-001	2.44E-003	1.47E-002	1	1	0
AP2B	1.03E-002	7.00E-001	4.43E-003	1.54E-002	1	2	0
RELB	1.03E-002	7.00E-001	4.43E-003	1.62E-002	1	2	0
HIF1A	1.46E-002	9.90E-001	5.83E-003	1.69E-002	2	36	0
MSX2	1.54E-002	1.05E+000	5.83E-003	1.77E-002	1	3	0
FOSL1	1.54E-002	1.05E+000	5.83E-003	1.84E-002	1	3	1
E2F	1.61E-002	1.10E+000	5.88E-003	1.91E-002	2	38	1
MYC	2.25E-002	1.53E+000	7.88E-003	1.99E-002	7	564	16
RARA	2.41E-002	1.64E+000	8.15E-003	2.06E-002	2	47	2
AP2A	2.55E-002	1.74E+000	8.33E-003	2.13E-002	1	5	0
ESR1	2.91E-002	1.98E+000	9.19E-003	2.21E-002	2	52	0
ARNT	3.06E-002	2.08E+000	9.33E-003	2.28E-002	1	6	1
MEF2A	3.56E-002	2.42E+000	1.02E-002	2.35E-002	1	7	2
SP4	3.56E-002	2.42E+000	1.02E-002	2.43E-002	1	7	1
P53	3.66E-002	2.49E+000	1.02E-002	2.50E-002	3	143	0
CDX2	4.05E-002	2.76E+000	1.10E-002	2.57E-002	1	8	0
HNF1B	4.55E-002	3.09E+000	1.16E-002	2.65E-002	1	9	0
GATA1	4.55E-002	3.09E+000	1.16E-002	2.72E-002	1	9	0
FOXO3	4.75E-002	3.23E+000	1.18E-002	2.79E-002	2	68	1

SUMMARY

During this work, we developed TFactS (www.tfacts.org), a bioinformatics tool that predicts the activation (or inhibition) of transcription factors based on a list of regulated genes generated by expression microarrays. Using Fisher's statistics, TFactS compares a list of regulated genes with a database of transcription factors associated with well-established target genes. This approach is an alternative to the systematic analysis of gene promoters using consensus transcription factor binding sites. We showed that it can be used to analyze published cancer microarray data. TFactS was validated experimentally by analyzing gene regulation by platelet-derived growth factor (PDGF) receptors.

PDGF stimulates cell migration and proliferation in wound healing, angiogenesis and embryogenesis, as well as in different types of cancers. We performed microarrays to identify genes regulated by PDGF receptors in fibroblasts and leukemic cells. TFactS suggested that PDGF activates transcription factors of the STAT and SREBP families, in line with our experimental data. TFactS has also predicted an inhibition of Forkhead box O (FOXO) transcription factors. It was already known that PDGF inhibits these factors by stimulating their phosphorylation by AKT. Our data demonstrate that PDGF also represses the expression of FOXO genes at the mRNA level. We showed that FOXO3 and FOXO1 can bind to FOXO1 gene promoter and stimulate FOXO1 mRNA transcription in a positive feed-back loop that is disrupted by growth factors. This process contributes to the growth promoting activity of PDGF. The data also suggest a mechanism by which growth factors may down-regulate FOXO genes in tumors.

In conclusion, this work describes a new approach for transcription factor mining in microarray expression data. This allowed us to decipher the transcriptional regulations triggered by PDGF receptor signaling. This tool may contribute to the global analysis of transcriptional networks in cancer cells.

RÉSUMÉ

Nous avons développé TFactS (www.tfacts.org), un outil qui prédit l'état de régulation (activation ou inhibition) des facteurs de transcription à partir des données des microarrays. TFactS utilise le test de Fisher pour comparer les listes de gènes régulés à une base de données liant les facteurs de transcription à leur gènes cibles. Cette approche représente une alternative aux méthodes dites *in silico*, qui se basent sur la recherche de motifs de fixation de facteurs de transcription dans les promoteurs de gènes régulés. TFactS a été validé expérimentalement sur les gènes régulés par le PDGF.

Le PDGF est un facteur de croissance qui induit la migration et la prolifération des cellules lors de la cicatrisation, l'angiogenèse et l'embryogenèse, ainsi que dans différents types de cancers. Nous avons réalisé des expériences de microarrays pour identifier les gènes régulés par le récepteur du PDGF dans les fibroblastes normaux et dans une lignée cancéreuse de leucémie. TFactS, appliqué sur ces données, a suggéré une régulation de plusieurs facteurs de transcription dont STAT et SREBP, validés expérimentalement. TFactS a aussi prédit l'inhibition par le PDGF d'un groupe de facteurs de transcription FOXO, suppresseurs de tumeurs, impliqués dans l'apoptose et l'arrêt du cycle cellulaire. Le récepteur du PDGF induit la phosphorylation de FOXO par AKT provoquant son inactivation et son exclusion du noyau. Nous avons montré que ce récepteur, en plus, inhibe l'expression des gènes FOXO. En effet, nos résultats montrent que FOXO3 et FOXO1 activent l'expression du gène FOXO1 en se fixant directement sur son promoteur. Ceci suggère un maintien de l'activité de FOXO, en absence d'un signal de croissance, par une boucle d'activation positive dans laquelle l'activation de FOXO induit sa propre expression. Lors d'une stimulation par le PDGF, cette boucle est rompue et la cellule retrouve les conditions nécessaires pour la prolifération. Ce mécanisme pourrait être impliqué dans différents cancers.

Ce travail présente une nouvelle approche pour la détection des facteurs de transcription régulés à partir de données d'expression. Son application au PDGF a contribué à la compréhension des régulations induites par ce facteur. TFactS pourrait permettre l'élaboration et l'analyse des réseaux de transcription dans les tumeurs.

ACKNOWLEDGEMENTS

Écrire une thèse est un exercice qui n'est pas facile et demande de la patience. Cependant, je pense que la partie la plus difficile est celle-ci. Pourquoi? Tout d'abord par ce que, sincèrement, je ne trouve pas des mots assez pesants pour exprimer ma gratitude envers celles et ceux qui m'ont aidé à mener ce travail jusqu'au bout. Ensuite par ce que j'ai peur d'oublier quelqu'un qui m'a aidé (et ils sont nombreux) dans un moment ou un autre et que sans son aide je n'aurais pas pu avancé. Ainsi, j'en profite, tout de suite, pour dire un grand merci à tous les membres de l'*Institut de Duve*, sans exception. Un grand merci à l'UCL pour avoir accepté de m'inscrire au doctorat et de me compter parmi ses membres de personnel. Et sachez que les mots que j'écris ne sont que des signes qui n'expriment qu'une infime partie de ma reconnaissance.

C'est avec honneur et fierté que j'annonce que cette thèse a été menée sous la direction du *Pr. Jean-Baptiste Demoulin* qui par ses qualités humaines, ses connaissances et sa rigueur scientifiques n'a pas hésité un moment à m'orienter quand je perds le Nord, à m'encourager quand je perds de la combativité et même à m'épauler dans les manipulations au tout début de mes balbutiements sur la paillasse. Que dire d'autre que mille merci pour le destin qui nous croise avec des gens comme lui, avec qui on apprend l'essence même de la recherche scientifique qu'est le partage du savoir à la fois pratique et théorique.

Je remercie aussi les membres du jury qui ont accepté de m'accompagner tout au long de ce travail. Par leurs pertinentes suggestions et critiques, leurs humbles encouragements et conseils, j'ai su avancer avec confiance tout au long de ce travail.

Ainsi, je remercie humblement *Pr. Fred Oppendoes* de m'avoir honoré par sa présidence du jury, son charisme personnel et scientifique et sa modestie me marqueront à jamais.

De même, je remercie fortement *Pr. Stefan N Constantinescu* d'avoir assisté mon travail en tant que membre de jury et de contribuer à son avancement et sa cohérence à travers ses suggestions et commentaires, mais aussi pour les diverses conversations et opportunités scientifiques qu'on a échangé en dehors de ma thèse à travers lesquelles j'ai reconnu sa grande maîtrise des détails et sa grande personnalité scientifique. J'en profite aussi pour remercier les membres de son laboratoire pour les discussions, les entre-aides, les réactifs et les protocoles qu'on a échangés.

I would also like to thank *Pr. Lodewijk Wessels*, who reviewed this thesis. Thank you for coming from the Netherlands to assist at my defense. It is always a pleasure to talk to great Bioinformaticians like you and to discuss this work with you. I have really enjoyed your suggestions and appreciated your comments during my thesis defense and during the BBC 2009 conference in Liège, in which you were the chairman of the session where my talk was scheduled. What a coincidence!

Aussi vaste que le domaine des microarrays soit-il, il nous permet au moins de rencontrer des gens comme le *Pr. Laurent Knoops*, que je remercie tout d'abord d'avoir accepté d'assister mon travail en tant que membre de jury et ensuite pour toutes les discussions scientifiques qui ont largement aidé dans le projet. Par ses qualités humaines et ses encouragements, il ne peut qu'être énormément remercié et apprécié. J'en profite, ici, pour remercier à travers lui le *Pr. Jean-Christophe Renauld* et son unité ainsi le *Pr. Laure Dumoutier* pour leur aide matériel et scientifique.

Dans la lignée de ces personnalités "mastodontes" se situe *Pr. Jacques van Helden*, avec qui j'ai eu le plaisir et l'honneur de travailler dans le cadre de mon DEA en Bioinformatique. Avec lui j'ai appris un déterminisme scientifique qui n'a pas peur d'approvoiser les modèles probabilistes. Sa méthodologie de travail presque logarithmique m'a beaucoup aidé dans mon travail de thèse. Et enfin, je reconnais en lui une personnalité humaine chaleureuse et généreuse qui n'hésite pas à soutenir les idées neuves et à sacrifier de son temps précieux. Merci aussi à son ancienne équipe transMaze pour son support et toutes les interactions qu'on a eu.

Je voudrais remercier aussi le *Pr. Yves Deville* pour son support et son aide et à travers lui remercier toute son équipe, principalement *Jean-Noël* et *Pierre*.

Un grand merci et profonde gratitude à ceux qui ont contribué au financement de ce projet : l'organisme Wallonie-Bruxelles international, la bourse du patrimoine de la faculté de médecine (UCL) et la région Wallonne.

I would also like to thank Professors *Valeria Giandomenico*, *Nathalie Delzenne* and *Paul Coffey* for the fruitful collaborations we had during this PhD work.

Chaque grand professeur à ses mousquetaires de la recherche. Celles (et ceux) du laboratoire du *Pr. Demoulin* sont d'une espèce très particulière. A la fois bosseuses et énergétiques dans leur domaine, elles ont aussi l'art de semer la vie et la joie autour d'elles rendant ainsi l'environnement du travail agréable et paisible. Par ces mots je remercie *Amélie*, *Alexandra*, *Axelle*, *Carmen*, *Sandrine* et *Seima* et je leur souhaite une bonne continuation dans leurs vies professionnelles et privées. Merci de m'avoir encouragé et soutenu. Je remercie aussi *Catherine* et *Nicolas* pour leur aide précieuse dans le projet Foxo. Sans votre aide, je serai toujours entrain de faire des manip pour faire aboutir ce travail. Special thanks to *Federica*, the neighbor on my right-side. Since the first hours of my work in the lab you have supported, with *Anders*, my teasing and my jokes. As soon as time goes, during the coffee drinks, for instance :) I discovered on you a great person with a highly sociable and friendly soul. And of course no one would argue about your molecular biology experience and knowledge, which you didn't hesitate to share and make others like me benefit from it. Thank you so much for your help and friendship. Your last mobib picture has no artifact, you're really shiny! :) In this line, I would like to thank also the great man from the snowy North, *Anders*, for his help, his kindness and his human qualities. Thank you for your friendship. Un grand merci à tous les membres du laboratoire, techniciens, postdocs, doctorants et mémorants pour toutes les discussions scientifiques, culturelles, sociales, politiques, philosophiques et religieuses durant lesquelles on a échangé nos idées et nos points de vue.

A *mes parents*, je ne saurais pas exprimer ma gratitude et ma profonde reconnaissance envers vous. Depuis mon enfance, vous avez été porteurs de conseils et de réconfort. Vous m'avez appris que le travail couplé à la patience est la recette magique de toute réussite. Je m'incline devant vos enseignements. Et que Dieu vous garde pour moi, mon frère et mes sœurs, que je remercie pour leur fraternité et amitié irremplaçable.

Merci aussi aux fidèles amis *Hamdi*, *Said*, *Said*, *Samir* et *Mohamed* du Maroc et à mes amis dans le pays d'accueil (Belgique) : *Steve*, *Ben*, *Mathieu*, *Pauline*, *Fabrice*, *Nicolas* et *Omar*. Votre amitié n'a pas de prix!

Enfin, chacun ses ressources et sa tendresse pour mener à bien son travail. Les miennes je les retrouve chez moi, à côté de vous *Hanane* et *Anir*, chaque jour et chaque moment qu'on passe ensemble sont irremplaçables. Grâce à vous, les moments les plus durs passent inaperçus et les bons moments restent à jamais. Merci aussi à *Amal*, *Tarik* et tous les autres ...

