Universiteit Leiden
The Netherlands

Unravelling vascular tumors : combining molecular and computational biology

IJzendoorn, D.G.P. van

Citation
IJzendoorn, D. G. P. van. (2020, January 16). *Unravelling vascular tumors : combining molecular and computational biology*. Retrieved from https://hdl.handle.net/1887/82754

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/82754

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The handle http://hdl.handle.net/1887/82754 holds various files of this Leiden University dissertation.

Author: IJzendoorn, D.G.P. van
Title: Unravelling vascular tumors : combining molecular and computational biology
Issue Date: 2020-01-16

Unravelling Vascular Tumors: Combining Molecular and Computational Biology

Publication of this thesis was financially supported by the Department of Pathology, Leiden University Medical Center.
Cover art: Pollock, J. (ca. 1938-41) *Untitled* [brush and India ink on cardboard]
© 2019. Image copyright The Metropolitan Museum of Art/Art Resource/Scala, Florence.
Printed by: Print Service, Ede, the Netherlands.
Thesis layout: based on a template by Marieke Kuijjer (`https://github.com/mararie/thesis`).
ISBN/EAN: 978-90-830376-1-5

Unravelling Vascular Tumors: Combining Molecular and Computational Biology

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 16 januari 2020
klokke 15.00 uur

door

David Gerardus Pieter van IJzendoorn
geboren te Leiden
in 1990

Promotor:	Prof. dr. J.V.M.G. Bovée
Co-promotor:	Dr. K. Szuhai
Leden promotiecommissie:	Prof. dr. H.J. Tanke Prof. dr. W.T.A. van der Graaf (Netherlands Cancer Institute, Amsterdam) Dr. M.L. Kuijjer (Centre for Molecular Medicine, Norway)

"Stay hungry. Stay foolish."

Steve Jobs

Contents

1 General introduction **9**

2 Vascular tumors of bone: the evolvement of a classification based on molecular developments **25**

I Diagnosis and treatment 51

3 Fusion events lead to truncation of FOS in epithelioid hemangioma of bone **53**

4 Telatinib is an effective targeted therapy for pseudomyogenic hemangioendothelioma **69**

II Model systems 91

5 Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis **93**

6 Pseudomyogenic hemangioendothelioma recapitulated in endothelial cells from human induced pluripotent stem cells engineered to express the *SERPINE1-FOSB* translocation **111**

III Computational biology 145

7 PyPanda: a Python package for gene regulatory network reconstruction **147**

8 Gene regulatory network analysis of translocation driven vascular tumors **153**

9 Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas 167

IV Summary and discussion 191

10 Summary and discussion 193

11 Nederlandse samenvatting 203

V Appendices 209

Curriculum vitae 211

List of publications 213

Acknowledgments 215

Chapter 1

General introduction

Vascular tumors are a group of tumors whose common denominator is that they display endothelial differentiation. It is tempting to speculate that all vascular tumors originate somewhere during the differentiation of mesenchymal stem cells to endothelial cells, and throughout this thesis I propose evidence for this hypothesis. The spectrum of vascular tumors includes frankly benign tumors such as hemangiomas, intermediate entities such as epithelioid hemangioma of bone (which is benign in soft tissue) and pseudomyogenic hemangioendothelioma, and malignant entities including epithelioid hemangioendothelioma and angiosarcoma, as discussed in detail in chapter 2. All vascular tumors except the hemangiomas are very rare, which hampers research; adequate models to study these tumors are lacking and patient material is generally sparse. In this thesis we focused on two of the vascular tumors: epithelioid hemangioma, for which we identified and studied a driving translocation that leads to a truncation of a gene, and pseudomyogenic hemangioendothelioma for which we studied targeted therapy and generated new models using human umbilical vein endothelial cells (HUVECs) as well as human induced pluripotent stem cells (hiPSCs). In the first section of this chapter, I will introduce concepts relating to translocations and tumorigenesis as seen in the vascular tumors. Then, epithelioid hemangioma and pseudomyogenic hemangioendothelioma are introduced. The next section focusses on *in vitro* cell line models and the potential of using lentivirus transductions and CRISPR/Cas9 to alter cells and create biologically relevant models. A common denominator throughout this thesis has been the use of computational biology methods which were used to generate hypotheses that could be tested in the lab, discover new translocations and gain a better understanding of tumorigenesis. Computational biology concepts are introduced in the last section of this chapter where next-generation sequencing, fusion detection, gene regulatory networks and machine learning are discussed.

1.1 Translocations and tumorigenesis

There are a number of different genetic alterations that may produce vascular tumors and tumors in general. The first type of genetic alterations are tumors with numerical and structural chromosomal abnormalities. An example of vascular tumors with numerous chromosomal abnormalities, even though they are known to have some recurrent alterations, are angiosarcomas (1). The second group consists of tumors with specific driver mutations or translocations. Although mutations and gross chromosomal abnormalities are extremely common events in many tumor types, the vascular tumors studied in this thesis are driven by specific gene translocations.

Translocations can occur through a number of different mechanisms. Usually there is a double stranded DNA break at two locations. Because of errors in the DNA double

strand repair mechanisms two strands of DNA, that are originally not attached, become attached to each other (figure 1.1). This can result in a chromosomal translocation, deletion or inversion and two gene parts can fuse together (figure 1.2). Translocations and their resulting chimera genes are capable of driving tumorigenesis through three different mechanisms (2). Firstly, in some cases, such as pseudomyogenic hemangioendothelioma, a promoter of one gene becomes attached to the fusion partner and drives expression. Secondly, a possibility is the generation of a chimeric gene, where both fusion partners contribute domains leading to a protein with a new or altered function. This occurs in epithelioid hemangioendothelioma where *WWTR1* fuses with *CAMTA1* and leads to transport of WWTR1 to the nucleus resulting in activation of the Hippo signaling pathway (3). Lastly, a fusion can lead to loss of a part of the protein, usually resulting in a loss of function. However, in chapter five we describe that this event can also lead to a gain of function.

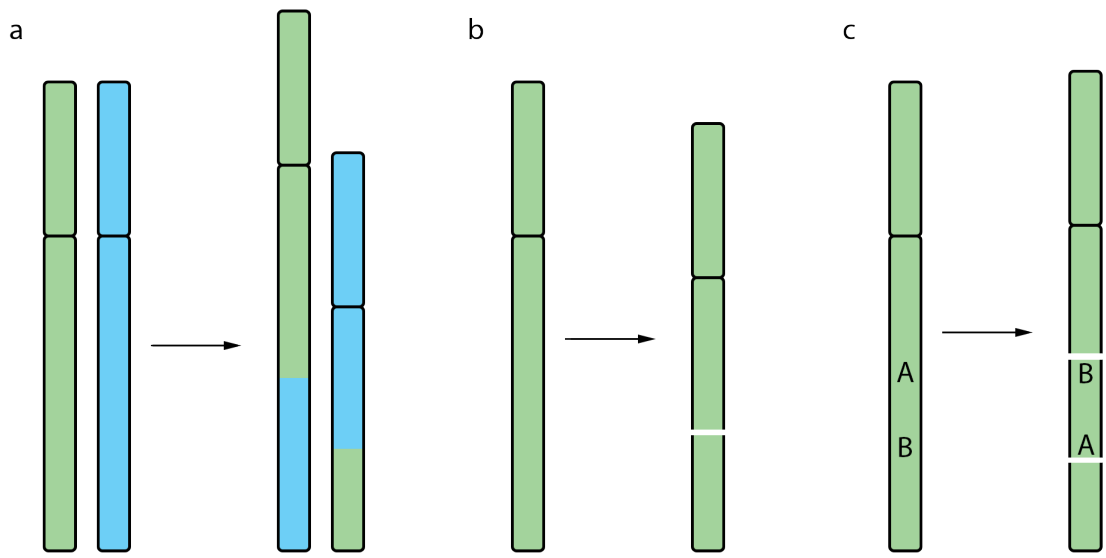


Figure 1.1: Different mechanisms leading to gene translocations. (a) Two double stranded breaks on separate chromosomes can lead to a balanced translocation. (b) Two deletions on one chromosome can lead to the loss of a piece of DNA. (c) Two deletions on one chromosome can also cause the inversion of a fragment of DNA.

1.2 Vascular tumors

Throughout this thesis I have studied epithelioid hemangioma and pseudomyogenic hemangioendothelioma, and their translocations that we hypothesize drive the tumorigenesis. Chapter 2 provides a detailed overview of the vascular tumors. Understanding the

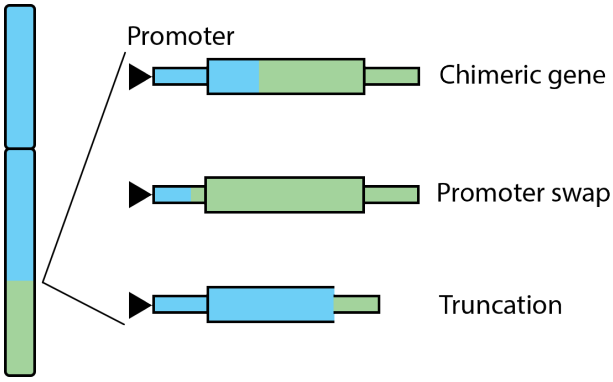


Figure 1.2: Generally, the three effects a translocation can have on the resulting fusion gene are illustrated. On the left a chromosome is depicted, while on the right resulting mRNA is shown. For the mRNA the UTR (slim bars) and coding regions (wide bars) are shown.

Entity	Classification	Prognosis	Immunohistochemistry
Epithelioid hemangioma of bone	Intermediate	100% survival, 2% metastasis, 9% local recurrence	CD31+, CD34+, ERG+
Pseudomyogenic hemangioendothelioma	Intermediate	Limited follow-up	ERG+, FLI1+, Keratin+, CD34-, Desmin-, Retention of INI1, FOSB+

Table 1.1: Summary of the two vascular tumors of bone that were studied most extensively throughout this thesis.

tumorigenesis for the vascular tumors will help develop new targeted therapies or lead to more insights into the pathophysiology.

1.2.1 Epithelioid hemangioma of bone

Epithelioid hemangioma of bone is a very rare intermediate and locally aggressive vascular tumor that can occur at nearly all ages, ranging from 10 to 75 years with a mean of 35, as found in a series of 50 cases (4). Histologically epithelioid hemangioma is recognized by its lobular architecture and well-formed vessels that are lined by the tumor cells. Immunohistochemically the endothelial differentiation of the tumor cells is clear, showing positivity for CD34, CD31 and ERG. The histology and immunohistochemistry therefore give evidence that epithelioid hemangioma displays endothelial differentiation, which was assumed throughout this thesis (table 1.1).

1.2.2 Pseudomyogenic hemangioendothelioma

The naming of pseudomyogenic hemangioendothelioma has been controversial over time. First described as epithelioid sarcoma-like hemangioendothelioma by Billings and colleagues in 2003 (5), Hornick and colleagues described 50 cases of an entity they called pseudomyogenic hemangioendothelioma in 2011 (6). The lack of a concise name led to discussions about who first described this entity, it was found that epithelioid sarcoma-like hemangioendothelioma and pseudomyogenic hemangioendothelioma are indeed the same tumor. In this thesis we used the name pseudomyogenic hemangioendothelioma, in line with the *"WHO Classification of Tumours of Soft Tissue and Bone"* (7). Mean age of occurrence for pseudomyogenic hemangioendothelioma is 31, but ranges from 14 to 80. Most patients present with multifocal disease (7). Histologically the tumor cells show an epithelioid sarcoma-like or rhabdomyoblast-like appearance, with abundant eosinophilic cytoplasm. The cells are positive for keratin AE/AE3 in addition to the vascular marker ERG. CD34 is negative and CD31 is expressed in half of the cases (7). Characteristic for this entity is a balanced translocation between chromosomes 7 and 19, that was first described in 2011 by Trombetta and colleagues (8). This translocation was later found to lead to a fusion between *SERPINE1* and *FOSB* genes by Walther and colleagues in 2014 (9). Recently another recurrent fusion was identified in pseudomyogenic hemangioendothelioma, between *ACTB* and *SERPINE1* (10, 11). It was found that immunohistochemistry for FOSB could be used as a diagnostic marker for pseudomyogenic hemangioendothelioma showing that the fusion leads to an upregulation of FOSB expression (12, 13) (table 1.1).

1.3 Tumor models

To understand neoplasms at a fundamental level, model systems are needed where the variables can be studied in a controlled way. As there are no cell lines available for epithelioid hemangioma and pseudomyogenic hemangioendothelioma a number of different models were used to study these tumors. Here I will introduce the most important aspects relating to *in vitro* cell line models used throughout this thesis.

1.3.1 Cell lines and Lentivirus Vectors

Tumor derived cell lines have been the workhorse in molecular cell biology for many years and are an excellent model which can give insight into the pathways driving neoplastic cells. Ultimately, an understanding of the biology behind tumors can lead to better targeted therapies. Cell lines have been used successfully since the 1960s to study biology

and cancer. With one of the most well-known examples being the HeLa cells. This cell line was derived from the cervical cancer cells of Henrietta Lacks. To date the HeLa cell line is reportedly used in over sixty-five thousand publications (14).

Although it has been extensively tried to culture the intermediate vascular tumors, these efforts have been without success, so far. To our knowledge no cell lines have been established for epithelioid hemangioma and pseudomyogenic hemangioendothelioma. To create a cell line model for the vascular tumors we have used endothelial cells and their precursors (iPS). In chapters four and five of this thesis we have used human umbilical vein endothelial cells (HUVECs) to model epithelioid hemangioma and pseudomyogenic hemangioendothelioma. HUVECs are isolated from the endothelium of veins from the umbilical cord. A large disadvantage of using HUVECs is that they can only be kept in culture for a limited number of passages and doublings (15).

To study tumor biology, we have introduced expression plasmids using a Lentivirus delivery system. Using this delivery system genes or short hairpin RNAs can be efficiently introduced in the genome for expression or repression of genes to mimic the genetic alterations found in human vascular tumors (and thereby create a model system) (16). The downside of using a lentivirus delivery system is that usually genes are introduced without their own promoter and multiple copies of the same gene can become integrated into the genome, therefore gene expression is generally much higher than what would be found in actual tumors.

1.3.2 CRISPR/Cas9 and human induced Pluripotent Stem Cells

The CRISPR/Cas9 gene editing system consists of two components; a Cas9 protein that is guided by a guide RNA to a piece of DNA of interest where it will introduce a double stranded break. Introduction of these breaks will activate either non-homologous end joining or the homologous recombination pathway. Non-homologous end joining is prone to errors as no template is utilized. Often point mutations, deletion or insertions are left at the site that is targeted by the Cas9 protein (17). When homologous recombination is utilized by the cell, a template is used to repair the break. This template can be provided to insert a custom sequence within the break. Because chromosomal translocations are a result of double stranded breaks and non-homologous end joining, CRISPR/Cas9 can be used to introduce chromosomal translocations with reasonable efficiency (figure 1.3) (18). Introducing chromosomal translocations to model gene fusions has large advantages over using an expression system because expression and regulation of the fusion gene remains under control of the original promoter and therefore represents expression and regulation as found in tumor cells.

CRISPR/Cas9 has been used to model fusion driven tumors previously. In mesenchy-

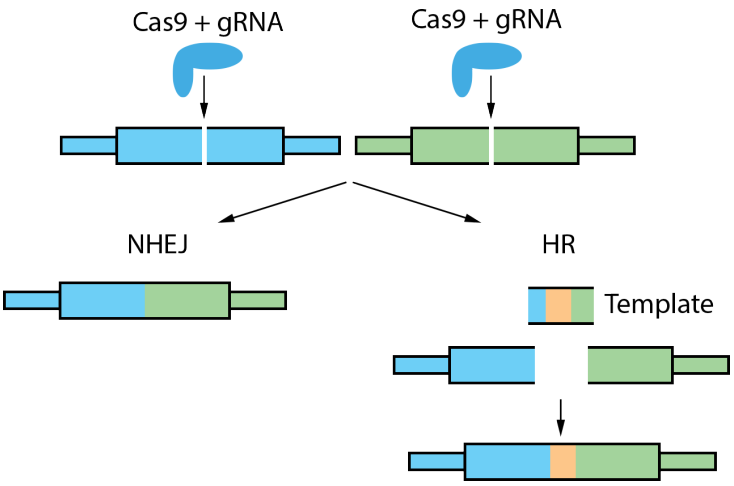


Figure 1.3: Induction of translocations using CRISPR/Cas9. Two break points are introduced and both the NHEJ and HR can lead to the formation of a translocation.

mal stem cells CRISPR/Cas9 was used to introduce the *EWSR1-WT1* fusion, to model desmoplastic small round cell tumors (19). The *EWSR1-FLI1* fusion was introduced in human mesenchymal stem cells to model Ewing sarcoma (20). However, because Ewing sarcoma does not show evident differentiation towards a normal cell type it is not possible to study the effects of the fusion gene on cells with matching differentiation to Ewing sarcoma cells, which influences the observed effects of the fusion gene. In chapter six of this thesis the *SERPINE1-FOSB* fusion is introduced in hiPSCs to overcome the limitations of using HUVECs combined with a lentivirus delivery system as a model to study pseudomyogenic hemangioendothelioma. Because pseudomyogenic hemangioendothelioma shows endothelial differentiation the functional effects of the *SERPINE1-FOSB* fusion were studied in human induced pluripotent stem cells differentiated to endothelial cells.

As indicated before, one of the limitations of using HUVECs is their limited lifespan. To overcome this issue human induced pluripotent stem cells (hiPSCs) have been used. HiPSCs are derived from normal human somatic cells such as fibroblasts that can be reprogrammed to pluripotency. This is done through expression transcription factors such as; Oct4, Sox2, Klf4 and Myc (21). First described in 2008, hiPSCs have been used extensively in recent years, showing large potential as disease models (22, 23). The largest advantage for using hiPSCs is that they can be expanded indefinitely and differentiated to almost any tissue type. To study the effect of the translocation on cells in the endothelial lineage, the hiPSCs are differentiated to the mesoderm lineage, after which CD31 positive endothelial cells are extracted using magnetic beads (24). The endothelial cells derived from hiPSCs express endothelial-specific markers such as VE-Cadherin, von Willebrand factor and LYVE1. Furthermore, they are capable of tube formation when cultured on pericytes or matrigel (25, 26).

1.3.3 Computational biology

Developments in next-generation sequencing have led to affordable sequencing. Moreover, there is a trend to distribute sequencing data in large open available databases. Analyzing this data is done through computational biology. Here the most important next-generation sequencing and computational biology techniques that were used throughout this thesis are introduced.

1.3.4 Next-generation sequencing

Next-generation sequencing (NGS) is a term used to describe massive parallel DNA sequencing techniques succeeding Sanger sequencing. NGS enables rapid and parallel sequencing of many DNA molecules, enabling generation of large datasets (27). NGS is used to sequence DNA, or RNA through generation of cDNA. Therefore, NGS can be used for many purposes including nucleotide-, structural and copy number variant detection but also quantification of gene expression through quantification of the RNA.

There are a number of different platforms used for NGS, in this thesis data was used that is generated by the Illumina platform, a second-generation sequencing technique. Second generation sequencing machines work through sequencing by synthesis, where DNA is amplified and each nucleotide added to the strand generates a signal. In the Illumina platform, nucleotides are added and pyrophosphate is released, used to generate a fluorescent signal that is detected and used to determine the original base in the DNA (28). Although the read length is generally short in the Illumina platform (around 90 bases) it is possible to generate paired-end reads. Paired-end sequencing entails sequencing of both ends of the generated DNA fragments (which is usually around 500 bases long), the two reads from the same DNA fragment are called mate pairs (figure 1.4). Especially for structural variant detection (such as fusion genes) it is essential to generate paired-end reads because this will enable detection of paired-end reads spanning the structural variants (figure 1.4).

1.3.5 Fusion gene detection on transcriptome sequencing data

Most tools to detect fusion genes rely on the detection of spanning read pairs and split reads. Spanning read pairs are read pairs where the two reads align on different locations, with a larger distance between the reads then could be expected based on the fragment size (the size of the generated cDNA fragments). Split reads are reads which partially align on two non-adjacent locations (figure 1.4). False positives are extremely common in fusion gene detection and can occur due to a number of different problems. Firstly, errors in transcription often result in mRNA molecules that are a products of read-through, where

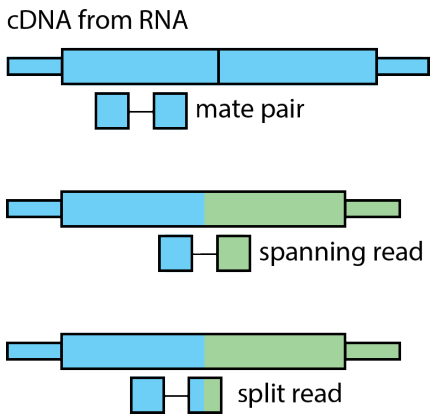


Figure 1.4: Fusion genes can be detected on mRNA (through sequencing of cDNA). Detection of spanning or split reads is generally used to discover fusion gene.

mRNA is not spliced correctly and different genes are attached to each other. Read-throughs can be called as fusion genes. Secondly, there are many sequence homologues in the human genome. These can include paralogs and pseudogenes which can result in calling of false-positive gene fusions. Lastly, SNPs or sequencing errors can result in misalignment of reads and cause calling or missing of fusion-genes.

Through detection of translocations on transcriptome sequencing data, many translocations located in intragenic regions of the genome are eliminated that would be detected on whole-genome sequencing data. Another advantage of translocation detection on transcriptome sequencing data is that translocations that are not expressed, are not detected. Usually translocations that are not expressed do not have clinical consequence (29). A notable exception to this are the translocations affecting tumor suppressors where loss of a gene, due to a translocation, could drive tumorigenesis. Fusion detection tools use a number of different strategies to filter the remaining false-positive reads; including databases with known false positives and filtering fusions involving intra-genic regions. When sufficient computational resources are available, multiple algorithms can be run and consensus fusions are identified in multiple algorithms (30).

1.3.6 Gene regulatory networks

The idea of regulatory networks was first proposed by Butte and colleagues in 1999. They calculated the correlation coefficients for a database with simultaneous laboratory experiments (31). Later they calculated correlation coefficients for expression data and reconstructed part of the regulatory network in *Saccharomyces cerevisiae* (32). Gene regulatory networks are identified based on the correlation coefficient. Genes that are co-expressed show a higher correlation coefficient than would be expected based on chance. A limitation of this first approach is that it does not take protein-protein interactions and

regulation of transcription factors into account that may not necessarily be visible on gene expression level. This approach also does not consider whether genes are transcription factors and are therefore capable of regulating other genes. Among other approaches (33), Glass and colleagues attempted to solve this problem with an algorithm they named Passing Attributes between Networks for Data Assimilation (PANDA) (34). PANDA solves many shortcomings of the regulatory network analysis by integrating multiple data types. Therefore, PANDA can distinguish between direct and indirect interactions in a network analysis.

1.3.7 Machine learning

Machine learning uses algorithms to see patterns and learn concepts without being explicitly programmed to do so. Machine learning algorithms can be subdivided into two categories, unsupervised and supervised algorithms. Unsupervised algorithms are useful on data where a pattern needs to be identified without prior knowledge on the outcome. An example of an unsupervised algorithm is the t-Distributed Stochastic Neighbor Embedding (t-SNE) that reduces dimensions and can be used to identify data clusters such as different cell subtypes within a sample (35). To automatically detect clusters in data, algorithms such as K-means or Density-based spatial clustering of applications with noise (DBSCAN) can be used. These algorithms detect clusters, possibly in data where the dimensions have been reduced using t-SNE or Principal Component Analysis (PCA).

The other category of machine learning algorithms are the supervised algorithms. These algorithms are trained on labeled data and are trained to formulate a hypothesis that captures the relationship between the data and the label. Examples of supervised algorithms are k-nearest neighbor, random forest and support vector machines which can be used for classification problems. Especially the random forest algorithm shows a lot of potential for analysis of expression data as the algorithm can output the weight of each variable, thereby making it possible to find the strongest variables in the decision process (36). One of the most well-known supervised learning algorithm is the neural network. Neural networks use multiple layers of abstraction to form a hypothesis and have led to impressive developments in areas such as speech and image recognition. Although neural networks can be very powerful algorithms they require large datasets of training data to formulate an accurate hypothesis (37).

In genetics unsupervised machine learning algorithms have been used extensively, with Principal Component Analysis and t-SNE used very frequently. There is also large potential for using supervised algorithms. Examples where supervised machine learning has been used include predicting outcome in lung and liver cancer (38–40).

1.4 Aim and outline of the thesis

The aim of this thesis is to study rare tumors, with a focus on the vascular tumors. For these tumors adequate models and knowledge on tumorigenesis are lacking. I aimed to develop new models to study these tumors, discover new genetic alterations and examine their effect on the tumorigenesis through using computational biology.

The research in this thesis is subdivided into three parts. The first part (chapters 2-4) focusses on diagnosis and treatment of vascular tumors. In **chapter 2** our current knowledge on the histopathology, epidemiology and tumorigenesis of the vascular tumors is reviewed. In **chapter 3** we report the discovery of a new fusion gene in epithelioid hemangioma of bone involving *FOS* and different translocation partners. This translocation is present in approximately 60% of epithelioid hemangioma cases. In **chapter 4** a potential treatment for pseudomyogenic hemangioendothelioma is described. A patient with extensive and inoperable pseudomyogenic hemangioendothelioma was treated with telatinib and showed a remarkable response for which we investigated the mechanism of action.

The second part (chapters 5-6) of this thesis covers model systems that were developed and used to study vascular tumors. Initially we used HUVECs overexpressing truncated *FOS* to study epithelioid hemangioma. In **chapter 5** the effect of the truncation of *FOS* on the resulting protein and its function was explored. It was found that the tail of the FOS protein is required for rapid ubiquitin independent degradation. As this part of the protein is lost in the gene fusion that is found in epithelioid hemangioma, it is likely that FOS remains active longer and thereby drives tumorigenesis. In **chapter 6** we developed a new model to study pseudomyogenic hemangioendothelioma. We introduced the *SERPINE1-FOSB* fusion in hiPSCs using CRISPR/Cas9. Thereafter we differentiated the hiPSCs towards CD31 positive endothelial cells. We showed that this model, in part, recapitulates pseudomyogenic hemangioendothelioma and can be used to study tumorigenesis using *in vitro* and *in vivo* assays.

The third part (chapters 7-9) focusses on computational biology. In **chapter 7** we developed a python package to perform gene regulatory network reconstruction. This tool performed much faster than the existing C++ implementation and can therefore help to perform network reconstruction on much larger datasets. We used a cell line based model and network reconstruction methods in **chapter 8** to study epithelioid hemangioma and showed a potential link to the HIPPO signaling pathway. This would explain the similarities in morphology with other vascular tumors as genes involved in the HIPPO signaling pathway are involved in recurrent translocations in other vascular tumors such as epithelioid hemangioendothelioma. **Chapter 9** shows the potential of using machine learning to identify prognostic and diagnostic markers using machine learning algorithms.

For this study we used gene expression data from the Cancer Genome Atlas for soft tissue sarcomas.

This thesis gives insight into the tumorigenesis of the vascular tumors. Many of the findings and models we report can be generalized and therefore could be used to gain insight into other tumors as well.

Bibliography

[1] Verbeke SL, de Jong D, Bertoni F, Sciot R, Antonescu CR, Szuhai K, et al. Array CGH analysis identifies two distinct subgroups of primary angiosarcoma of bone. Genes Chromosomes Cancer. 2015;54(2):72–81. doi:10.1002/gcc.22219.

[2] Roukos V, Misteli T. The biogenesis of chromosome translocations. Nature Cell Biology. 2014;16(4):293–300. doi:10.1038/ncb2941.

[3] Tanas MR, Ma S, Jadaan FO, Ng CK, Weigelt B, Reis-Filho JS, et al. Mechanism of action of a WWTR1(TAZ)-CAMTA1 fusion oncoprotein. Oncogene. 2016;35(7):929–938. doi:10.1038/onc.2015.148.

[4] Nielsen GP, Srivastava A, Kattapuram S, Deshpande V, O'Connell JX, Mangham CD, et al. Epithelioid hemangioma of bone revisited: A study of 50 cases. American Journal of Surgical Pathology. 2009;33(2):270–277. doi:10.1097/PAS.0b013e31817f6d51.

[5] Billings SD, Folpe AL, Weiss SW. Epithelioid Sarcoma-Like Hemangioendothelioma. The American journal of surgical pathology. 2003;27(1):48–57. doi:10.1097/PAS.0b013e31821caf1c.

[6] Hornick JL, Fletcher CD. Pseudomyogenic hemangioendothelioma: a distinctive, often multicentric tumor with indolent behavior. Am J Surg Pathol. 2011;35(2):190–201. doi:10.1097/PAS.0b013e3181ff0901.

[7] Hornick J, Fletcher C, Mertens F. Pseudomyogenic (epithelioid sarcoma-like) haemangioendothelioma. In: Fletcher C, Bridge JA, Hogendoorn P C, Mertens F, editors. WHO Classification of Tumours of Soft Tissue and Bone. Lyon: IARC; 2013. p. 333–4.

[8] Trombetta D, Magnusson L, von Steyern FV, Hornick JL, Fletcher CDM, Mertens F. Translocation t(7;19)(q22;q13)-a recurrent chromosome aberration in pseudomyogenic hemangioendothelioma? Cancer Genetics. 2011;204(4):211–215. doi:10.1016/j.cancergen.2011.01.002.

[9] Walther C, Tayebwa J, Lilljebjörn H, Magnusson L, Nilsson J, Von Steyern FV, et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. Journal of Pathology. 2014;232(5):534–540. doi:10.1002/path.4322.

[10] Agaram NP, Zhang L, Cotzia P, Antonescu CR. Expanding the Spectrum of Genetic Alterations in Pseudomyogenic Hemangioendothelioma With Recurrent Novel ACTB-FOSB Gene Fusions. The American Journal of Surgical Pathology. 2018;42(12):1653–1661. doi:10.1097/PAS.0000000000001147.

[11] Zhu G, Benayed R, Ho C, Mullaney K, Sukhadia P, Rios K, et al. Diagnosis of known sarcoma fusions and novel fusion partners by targeted RNA sequencing with identification of a recurrent ACTB-FOSB fusion in pseudomyogenic hemangioendothelioma. Modern Pathology. 2019;32(5):609–620. doi:10.1038/s41379-018-0175-7.

[12] Hung YP, Fletcher CDM, Hornick JL. FOSB is a useful diagnostic marker for pseudomyogenic hemangioendothelioma. American Journal of Surgical Pathology. 2017;41(5):596–606. doi:10.1097/PAS.0000000000000795.

[13] Sugita S, Hirano H, Kikuchi N, Kubo T, Asanuma H, Aoyama T, et al. Diagnostic utility of FOSB immunohistochemistry in pseudomyogenic hemangioendothelioma and its histological mimics. Diagn Pathol. 2016;11(1):75. doi:10.1186/s13000-016-0530-2.

[14] Landry J, Theodor Pyl P, Rausch T, Zichner T, Tekkedil MM, Stütz AM, et al. The genomic and transcriptomic landscape of a HeLa cell line. ebi. 2013;doi:10.1534g3.113.005777.

[15] Marin V, Kaplanski G, Grès S, Farnarier C, Bongrand P. Endothelial cell culture: Protocol to obtain and cultivate human umbilical endothelial cells. Journal of Immunological Methods. 2001;254(1-2):183–190. doi:10.1016/S0022-1759(01)00408-2.

[16] Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, et al. A third-generation lentivirus vector with a conditional packaging system. Journal of virology. 1998;72(11):8463–71. doi:98440501.

[17] Wang H, La Russa M, Qi LS. CRISPR/Cas9 in Genome Editing and Beyond. Annual Review of Biochemistry. 2016;85(1):227–264. doi:10.1146/annurev-biochem-060815-014607.

[18] Torres R, Martin MC, Garcia A, Cigudosa JC, Ramirez JC, Rodriguez-Perales S. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. Nat Commun. 2014;5:3964. doi:10.1038/ncomms4964.

[19] Vanoli F, Tomishima M, Feng W, Lamribet K, Babin L, Brunet E, et al. CRISPR-Cas9-guided oncogenic chromosomal translocations with conditional fusion protein expression in human mesenchymal cells. Proc Natl Acad Sci U S A. 2017;114(14):3696–3701. doi:10.1073/pnas.1700622114.

[20] Torres-Ruiz R, Martinez-Lage M, Martin MC, Garcia A, Bueno C, Castaño J, et al. Efficient Recreation of t(11;22) EWSR1-FLI1+in Human Stem Cells Using CRISPR/Cas9. Stem Cell Reports. 2017;8(5):1408–1420. doi:10.1016/j.stemcr.2017.04.014.

[21] Park IH, Lerou PH, Zhao R, Huo H, Daley GQ. Generation of human-induced pluripotent stem cells. Nature Protocols. 2008;3(7):1180–1186. doi:10.1038/nprot.2008.92.

[22] Orlova V, Mummery C. SnapShot: Key Advances in hiPSC Disease Modeling. Cell Stem Cell. 2016;18(3):422. doi:10.1016/j.stem.2016.02.013.

[23] Passier R, Orlova V, Mummery C. Complex Tissue and Disease Modeling using hiPSCs. Cell Stem Cell. 2016;18(3):309–321. doi:10.1016/j.stem.2016.02.011.

[24] Orlova VV, Van Den Hil FE, Petrus-Reurer S, Drabsch Y, Ten Dijke P, Mummery CL. Generation, expansion and functional analysis of endothelial cells and pericytes derived from human pluripotent stem cells. Nature Protocols. 2014;9(6):1514–1531. doi:10.1038/nprot.2014.102.

[25] Orlova VV, Drabsch Y, Freund C, Petrus-Reurer S, Van Den Hil FE, Muenthaisong S, et al. Functionality of endothelial cells and pericytes from human pluripotent stem cells demonstrated in cultured vascular plexus and zebrafish xenografts. Arteriosclerosis, Thrombosis, and Vascular Biology. 2014;34(1):177–186. doi:10.1161/ATVBAHA.113.302598.

[26] Halaidych OV, Freund C, van den Hil F, Salvatori DCF, Riminucci M, Mummery CL, et al. Inflammatory Responses and Barrier Function of Endothelial Cells Derived from Human Induced Pluripotent Stem Cells. Stem Cell Reports. 2018;10(5):1642–1656. doi:10.1016/j.stemcr.2018.03.012.

[27] Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA; 2016. Available from: `https://www.sciencedirect.com/science/article/pii/S0888754315300410`.

[28] Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. Science (New York, NY). 1998;281(5375):363, 365.

[29] Liu S, Tsai WH, Ding Y, Chen R, Fang Z, Huo Z, et al. Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. Nucleic Acids Res. 2016;44(5):e47. doi:10.1093/nar/gkv1234.

[30] Hoogstrate Y, Böttcher R, Hiltemann S, Van Der Spek PJ, Jenster G, Stubbs AP. FuMa: Reporting overlap in RNA-seq detected fusion genes. Bioinformatics. 2016;32(8):1226–1228. doi:10.1093/bioinformatics/btv721.

[31] Butte AJ, Kohane IS. Unsupervised knowledge discovery in medical databases using relevance networks. Proceedings / AMIA Annual Symposium AMIA Symposium. 1999;(February):711–715. doi:D005550 [pii].

[32] Butte AJ, Kohane IS. Mutual Information Relevance Networks: Functional Genomic Clustering Using Pairwise Entropy Measurements. In: Biocomputing 2000; 1999. p. 418–429. Available from: `http://psb.stanford.edu/psb-online/proceedings/psb00/butte.pdfhttp://www.worldscientific.com/doi/abs/10.1142/9789814447331{_}0040`.

[33] De Smet R, Marchal K. Advantages and limitations of current network inference methods. Nature Reviews Microbiology. 2010;8(10):717–729. doi:10.1038/nrmicro2419.

[34] Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing Messages between Biological Networks to Refine Predicted Interactions. PLoS ONE. 2013;8(5):e64832. doi:10.1371/journal.pone.0064832.

[35] Van Der Maaten L, Hinton G. Visualizing Data using t-SNE; 2008. Available from: `http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf`.

[36] Tin Kam Ho. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. vol. 1. IEEE Comput. Soc. Press;. p. 278–282. Available from: `http://ieeexplore.ieee.org/document/598994/`.

[37] LeCun YA, Bengio Y, Hinton GE. Deep learning. Nature. 2015;521(7553):436–444. doi:10.1038/nature14539.

[38] Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proceedings of the National Academy of Sciences. 2018;115(13):E2970–E2979. doi:10.1073/pnas.1717139115.

[39] Yu KH, Zhang C, Berry GJ, Altman RB, Re C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun. 2016;7:12474. doi:10.1038/ncomms12474.

[40] Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep Learning based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Research. 2017;doi:10.1101/114892.

Chapter 2

Vascular tumors of bone: the evolvement of a classification based on molecular developments

This chapter is based on the publication: **van IJzendoorn DGP**, Bovée JVMG. Vascular Tumors of Bone: The Evolvement of a Classification Based on Molecular Developments. *Surg Pathol Clin.* 2017;10: 621-635.

2.1 Abstract

The classification of vascular tumours of bone has been under debate over time. Vascular tumors in bone are rare, display highly overlapping morphology, and are therefore considered difficult by pathologists. Compared to their soft tissue counterparts, they are more often multifocal and sometimes behave more aggressively. Over the past decade, with the advent of next generation sequencing, recurrent molecular alterations have been found in some of the entities. The integration of morphology and molecular changes has led to a better characterization of these separate entities.

2.2 Introduction

The common denominator of vascular tumors consist of their endothelial differentiation, with a variable capability of forming mature or immature vessels. Literature on the cell of origin for vascular tumors (other than infantile hemangioma) is scarce, and point to an "endothelial precursor cell" or a haematopoietic precursor cell along its path of endothelial differentiation, for canine and murine hemangioma / angiosarcoma (1, 2). However, the definition of these cells in mice and human is controversial (3, 4).

The classification of vascular tumours of bone has been a matter of discussion over time (5–7). However, with the rapid elucidation of molecular changes in tumors using next generation sequencing, which also included vascular tumours of bone, the classification has evolved and morphology and molecular changes were integrated to better define the separate entities (8), that are sometimes extremely difficult to distinguish based on morphology alone. Like in soft tissue, the entity of "hemangiopericytoma of bone" is no longer recognized, as these lesions are rare presentations of synovial sarcoma, solitary fibrous tumour, and myofibroma primary of bone (9). Moreover, while in the past there has been ample discussion about "haemangioendothelioma of bone" being a separate entity (10), it is now more or less generally accepted that the previously reported cases represent epithelioid hemangioma of bone (8, 9, 11), and with the elucidation of specific genetic alterations in epithelioid hemangioma of bone (12, 13) this discussion may be definitively resolved in the future.

Now that the different vascular tumours have been better characterized, their distinct behaviour in bone as compared to when they are located in the soft tissues is becoming obvious. Vascular tumors of bone are more frequently multifocal, affecting multiple bones (6). Also, although histologically and genetically similar, epithelioid hemangioma in soft tissue is considered benign, while in bone it behaves as a locally aggressive, rarely metastasizing lesion and is therefore considered to be of the intermediate category (8). In addition, atypical epithelioid hemangioma has a preference for bone and penile lo-

Classification	Entity	Prognosis	Treatment
Benign	Hemangioma	100% survival, 0%	Treat symptoms
Intermediate	Epithelioid hemangioma	100% survival, 2% metastases, 9% local recurrence	Curettage or marginal excision
	Pseudomyogenic hemangioendothelioma	Limited follow-up, stable or progressive osseous disease	
Malignant	Epithelioid hemangioendothelioma	85% survival, 25% metastases	Wide resection
	Angiosarcoma	30% survival	Wide resection, consider systemic therapy
	Angiosarcoma	30% survival	Wide resection, consider systemic therapy

Table 2.1: Summary of prognosis and treatment for vascular bone tumors.

cation (14). Moreover, after the morphological and molecular characterization of pseudomyogenic (epithelioid sarcoma-like) haemangioendothelioma (15, 16), cases are reported that are exclusively located in bone, with unique histological findings (17).

Here we will discuss the most common vascular tumors of bone. These tumor entities range from the benign hemangioma of bone, with a good prognosis and no metastasis in all patients, to the intermediate epithelioid hemangioma (including the atypical variant) and the pseudomyogenic hemangioendothelioma whose survival is excellent but with some metastasis and recurrences. Epithelioid hemangioendothelioma is considered low grade malignant, with 85% survival and 25% metastases. Angiosarcoma is high grade malignant with a very poor survival of only 30% over 5 years (table 2.1). This review covers the classic presentations of these tumor entities including the diagnostic pitfalls and immunohistochemistry. We also discuss the recent developments regarding the genetics and tumorigenesis of these vascular tumors of bone (table 2.2).

Entity	Histologic and Molecular Findings	Immunohistochemistry
Hemangioma of bone	-Numerous smaller or larger blood -filled spaces, lined by flat endothelium -Reactive sclerosis of surrounding lamellar bone -No specific genetic alterations	CD31+ CD34+ ERG1+

Entity	Histologic and Molecular Findings	Immunohistochemistry
Epithelioid hemangioma of bone	-Lobular architecture (can be highlightedusing actin immunohistochemistry) -Well-formed vessels lined by epithelioidendothelial cells -Eosinophilic infiltrate -No prominent nuclear atypia or atypical mitoses -Hemorrhagic and spindle cell areas can be prominent, especially in acral lesions -Rearrangement of *FOS*	CD31+ CD34+ ERG+
Atypical epithelioid hemangioma of bone	-Similar to epithelioid hemangioma, with more solid growth, increased cellularity, nuclear pleomorphism, and necrosis -*ZFP36-FOSB* fusion	Similar to epithelioid hemangioma
Pseudomyogenic (epithelioid sarcoma-like) hemangioendothelioma	-Sheets of spindled or epithelioid cells with abundant eosinophilic cytoplasm -Infiltrative growth -Neutrophilic infiltrate -Reactive woven bone and osteoclast-like giant cells can be present -*SERPINE1-FOSB* fusion	ERG+ FLI1+ Keratin+ CD34- Desmin- Retention of INI1 FOSB+
Epithelioid hemangioendothelioma	-Epithelioid endothelial cells in strands and cords embedded in a hyaline or myxoid stroma -Intracytoplasmic vacuoles (blister cells) -No well-formed vessels-Infiltrative growth -Cytologic atypia and mitoses usually limited, but can be prominent -*WWTR1-CAMTA1* fusion	CD31 100% CD34 85% FLI1 100% Keratin 25%-38% D2-40 54% Prox1 54% ERG 98% Claudin-1 88% CAMTA1 86%-88%

Entity	**Histologic and Molecular Findings**	**Immunohistochemistry**
YAP1-TFE3 rearranged epithelioid hemangioendothelioma	-Focally well-formed vasoformative features in addition to solid areas -Voluminous deep eosinophilic or histiocytoid cytoplasm, sometimes feathery -Mild to moderate nuclear atypia -*YAP1-TFE3* fusion	Same as EHE TFE3+
Angiosarcoma	-Vasoformative, with multilayering, or solid -In bone often epithelioid (>90%) -Inflammatory infiltrate -Nuclear atypia (with large nucleoli) -Brisk mitotic atypia, including atypical mitoses -No specific genetic alterations	CD31 95%-100% ERG 96% VWF 60%-75% CD34 39%-40% Actin 61% Keratin 69%-80% D2-40 31%
Intravascular papillary endothelial hyperplasia (Masson tumor)	-Can occur in a blood vessel, a hematoma or in a preexisting vascular lesion -Papillary structures containing fibrin or collagen, covered by a single layer of endothelial cells -No or limited cytologic atypia, no or limited mitotic activity, no multilayering	CD31+ CD34+ ERG+

Table 2.2: Differential diagnosis of vascular tumors of bone.

2.3 Immunohistochemistry

In all vascular tumours endothelial differentiation can be highlighted using a panel of immunohistochemical markers including CD31, CD34 and ERG. ERG positivity can be highly specific for endothelial differentiation although this is dependent on the clone used: antibodies against the N-terminal part of the protein are more specific as compared to antibodies directed against the C-terminal part, which can also be positive in a variety of other mesenchymal tumors (18). Moreover, one should be aware that ∼50% of the prostate carcinomas harbour translocations involving ERG and thereby can be positive (19). FLI1

and Von Willebrand Factor (VWF, factor VIII) can also be used. Smooth muscle actin can highlight the pericytes, while D2-40 (podoplanin) and Prox1 can demonstrate lymphatic differentiation. A notorious pitfall that pathologists should be aware of, especially in bone where vascular tumors are often epithelioid (93-100% (20, 21)), is the expression of keratin in a significant percentage of vascular tumors (20, 22, 23).

2.4 Hemangioma

2.4.1 Definition, epidemiology and clinical features

Hemangiomas are common lesions that rarely ever reach a pathologist. Reported by Mirra et al these tumors are found in ∼10% of all autopsies and they are often seen by radiologists (24). They are usually asymptomatic. The vertebral bodies are most commonly affected (figure 2.1) (25). Kaleem et al analyzed all reported cases of hemangioma affecting the extremities in English literature till 2000 (n=104) and found a mean age of 32 years, and a slight preference for females (60%) (26). When affecting the long bones, the diaphysis or metadiaphysis are the most common location. Medullary origin is most frequent, but 45% of cases are either periosteal (33%) or intracortical (12%) (26). In literature 11 cases have been described with intracortical hemangioma of bone, seven of which were located in the distal tibia (27). Cavernous hemangioma is the most frequent type, although also (areas with) capillary hemangioma can be found. At imaging the lesions are relatively radiolucent due to the lack of bone and abundance of fat on radiological images. Reportedly the tumors give a high MRI signal in T1 and T2 owing to their high fat presence (5). No genetic aberrations have been described thus far.

2.4.2 Histological and immunohistochemistry features

Macroscopically (figure 2.2a, 2.2b and 2.2c) the lesions show trabeculated bone with a dark sponge-like appearance. Histologically, the lesions show numerous blood filled spaces, lined by a thin layer of flat endothelial cells, without atypia. The vascular spaces are surrounded by loose connective tissue, and grow in between the bone trabeculae, that are often thickened (figure 2.2d).

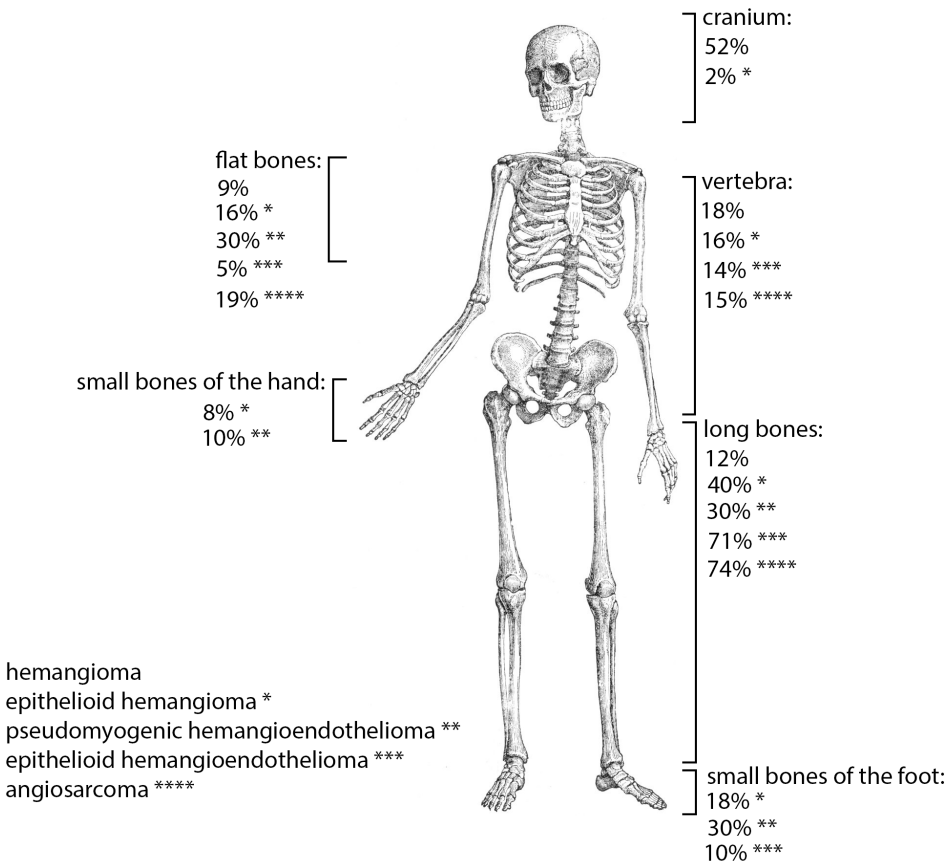


Figure 2.1: Distribution of vascular tumors.

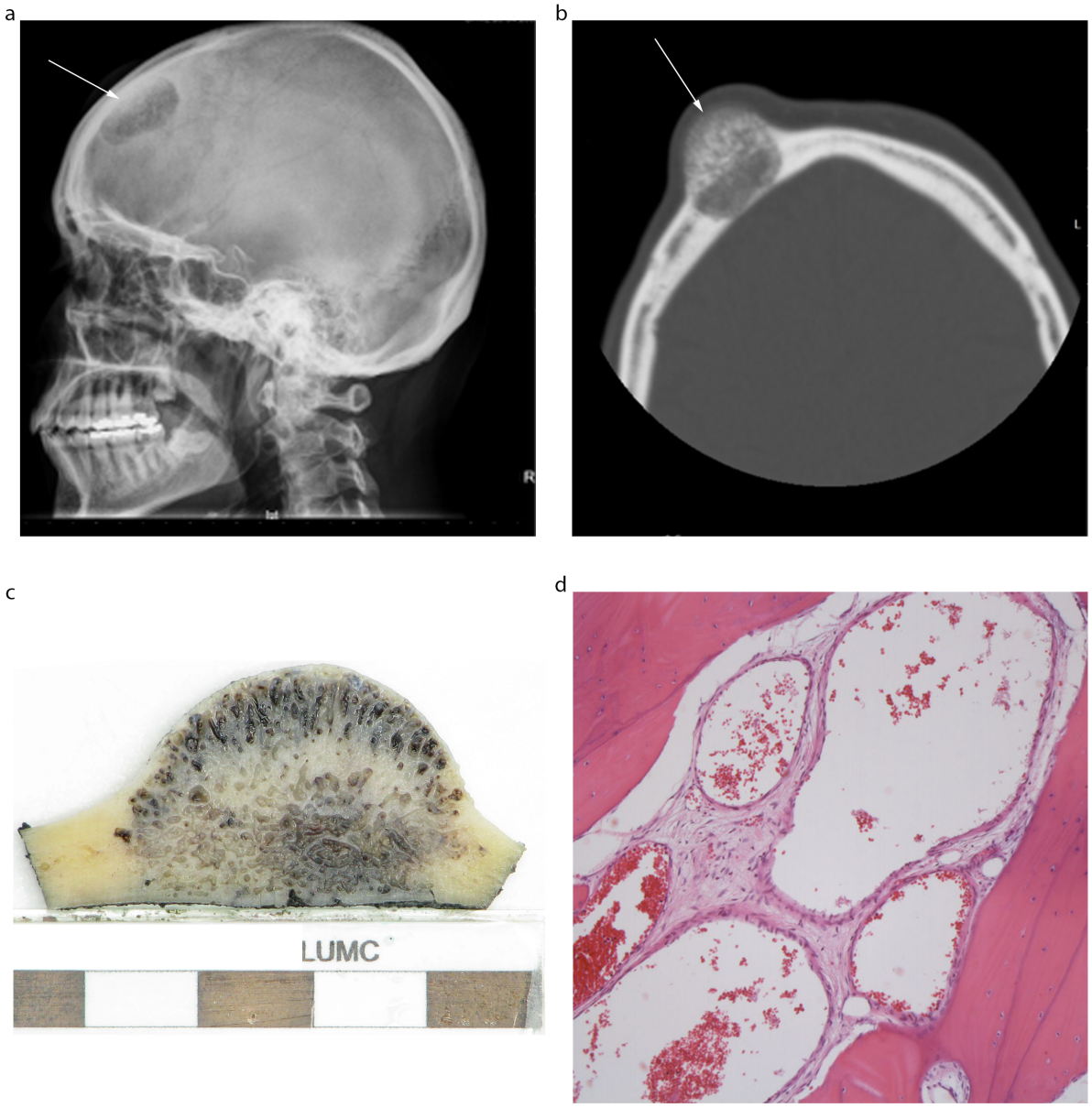


Figure 2.2: Hemangioma of bone. (a) X-ray of the skull with a sharply defined lytic lesion in the cranium. (b) Corresponding CT scan showing large protruding lytic lesion with calcifications. (c) Corresponding gross specimen shows trabeculations of the bone with sponge like appearance. (d) Microscopic image with large cavernous spaces lined by flat endothelium in between bony trabeculae.

2.5 Epithelioid hemangioma

2.5.1 Definition, epidemiology and clinical features

Epithelioid hemangioma of bone is classified as an intermediate, locally aggressive but rarely metastasizing vascular tumor of bone (8, 9). On CT scans a honeycomb pattern can be visible. Concise naming and classification of this tumor entity was only introduced recently by O'Connell et al (28, 29). Previously epithelioid hemangiomas of bone were reported as "haemangioendothelioma of bone" (11) or "haemorrhagic epithelioid and spindle cell hemangioma" (11, 30).

Nielsen et al revisited 50 epithelioid hemangiomas and described the age of occurrence to vary from 10 to 75 with a mean of 35 with a slight preference for males (11). As epithelioid hemangioma is a very rare entity exact prevalence is difficult to determine.

Epithelioid hemangioma has been described to occur in many different locations. Case reports and series of revisited cases seem to show there is a slight preference for the long tubular bones (31), but the spine is also often affected (32–35). Further reports include occurrences in the orbit (36–41). It is also frequently reported to occur in the small tubular bones of the extremities (42–45). Multifocal bone involvement occurs in ∼18% of the cases (11) with one case involving three different bones (46). Involvement of the draining lymph nodes has been described, but is not often confirmed as metastatic (47). Although as described by Nielsen et al the lymph node can contains cells resembling epithelioid hemangioma (11).

2.5.2 Histological and immunohistochemistry features

Epithelioid hemangioma is usually well-defined, with a lobular architecture (figure 2.3c), but can extend into the soft tissue. The vessels are usually well-formed, and lined by epithelioid endothelial cells (figure 2.3d, 2.3e). The cells have an enlarged nuclei with open chromatin, without prominent nuclear atypia. The often loose stroma surrounding the vessels can be infiltrated with eosinophils (figure 2.3e).

2.5.3 Tumorigenesis and genetics

Previously it was believed that epithelioid hemangioma could be a reactive lesion or a low grade variant of epithelioid hemangioendothelioma. This was refuted with the detection of fusion genes involving *FOS* (figure 2.3e) and *FOSB* with various fusion partners (12–14). Discovery of specific fusion genes also showed that tumors with multiple foci are monoclonal suggesting multifocal regional spread instead of multiple primaries (12). Overall, *FOS* rearrangements were found in 29% of epithelioid hemangiomas (13). The frequency

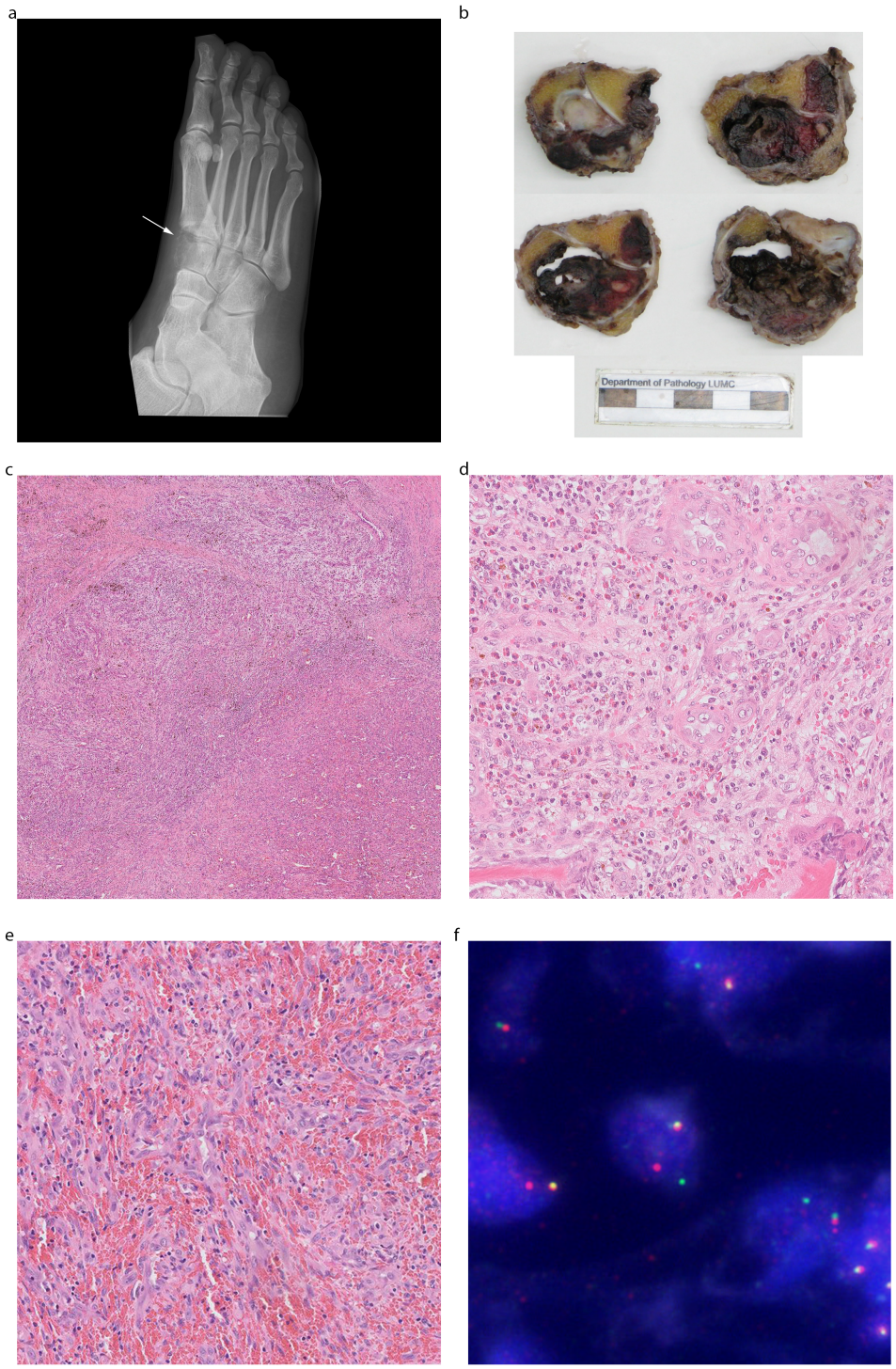


Figure 2.3: Epithelioid hemangioma of bone. (a) X-ray of the foot showing multiple sharply defined lesions. (b) Corresponding gross section. (c) Epithelioid hemangioma typically shows a lobular architecture. (d) Eosinophilic infiltrate surrounding the vessels lined by large epithelioid cells. (e) haemorrhagic and spindled cell appearance can be prominent especially in acral lesions. (f) FISH with break-apart probes surrounding *FOS* shows a split signal.

was higher in epithelioid hemangioma of bone as compared to soft tissue (59-71% versus 19%, respectively) (12, 13). In the *FOS* translocated epithelioid hemangiomas the fusion partners are at the C-terminal end of the protein and lead to a loss of the transactivating domain, regulating FOS turnover. It is speculated that this loss of the transactivation domain of FOS would lead to a reduced turnover, as FOS is normally rapidly degraded. Fusion genes involving *FOSB* are fused at the N-terminal end of the protein and are most likely activating promoter swap events. A specific subset of epithelioid hemangiomas was shown to harbour *ZFP36-FOSB* fusions (14). This subset has atypical histological features including more solid growth, increased cellularity, nuclear pleomorphism and necrosis (14). These cases are predominantly located at the penis and in bone. Both FOS and FOSB are part of the AP-1 transcription factor complex (48).

2.6 Pseudomyogenic (epithelioid-sarcoma like) hemangioendothelioma

2.6.1 Definition, epidemiology and clinical features

Most likely the first description in literature of pseudomyogenic hemangioendothelioma was published in 1992 by Mirra et al who reported a previously undescribed variant of epithelioid sarcoma. They reported five cases of epithelioid sarcoma displaying multicentric involvement of a single limb, osseous involvement, with bland diffuse fibrohistiocytic and rhabdoid cells (49). The first description as a distinct entity was in 2003 when Billings et al reported 7 histologically identical cases under the name epithelioid sarcoma-like hemangioendothelioma (50). In 2011 Hornick and Fletcher presented a large patient series including 29 cases. Twenty four percent of these had concurrent bone involvement. They proposed to designate the tumors as pseudomyogenic hemangioendothelioma. Most of the tumors arise in the extremities (see figure 2.1) with a male predominance (41 vs 9). Mean age was 31 years, ranging from 14-80. Strikingly, 33 of 50 patients presented with multifocal disease in which multiple discontiguous nodules were found in different tissue planes. In 2016 Inyang et al published the largest series of pseudomyogenic hemangioendotheliomas of bone to date, describing 10 cases with a male predominance (9 vs 1) and a mean age of 36 (range 12-74 years). They described the lesions to have intratumoral reactive woven bone and infiltration of osteoclast-like giant cells (17). The tumor is locally aggressive, and rarely metastasizing, and therefore of the intermediate category: one patient out of 50 developed distant metastasis (16). Reportedly the lesions are usually from 0.3 to 5.5 cm in size with ill-defined margins (figure 2.4a).

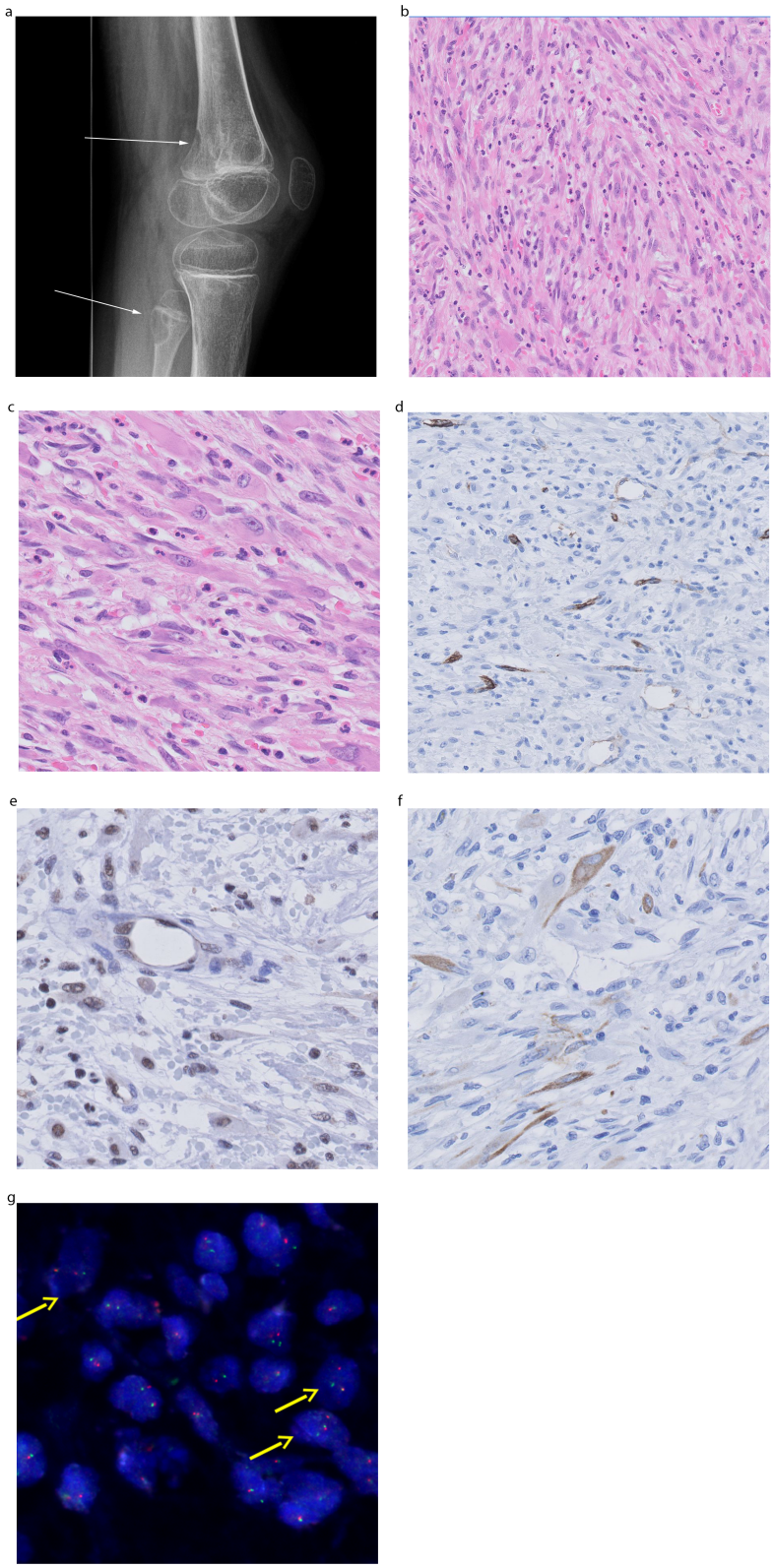


Figure 2.4: Pseudomyogenic hemangioendothelioma of bone. (a) X-ray shows a lesion the the left femur and fibula. (b) Spindle cells are seen admixed with neutrophils (c) the cells display a rhabdomyoblast-like appearance. (d) CD34 is consistently negative. (e) FLI1 shows nuclear staining in the endothelial cells. (f) Keratin staining is positive in the tumor cells. (g) FISH demonstrating split apart using a probe for *FOSB* indicative of *FOSB* rearrangement.

2.6.2 Histological and immunohistochemistry features

The tumor cells characteristically show an epithelioid sarcoma-like or rhabdomyoblast-like appearance, with abundant eosinophilic cytoplasm (figure 42.4b, 2.4c). The lesions can infiltrate in skeletal muscle. Infiltration with neutrophilic granulocytes can be prominent (figure 4b). The tumor cells characteristically express keratin AE1AE3 (figure 2.4f), ERG and FLI1 (figure 2.4e), while CD34 (figure 2.4d) and desmin are negative. CD31 is expressed in ∼50% of the cases and INI1 is retained. FOSB was shown to be an excellent immunohistochemical marker to detect the presence of the *SERPINE1-FOSB* fusion (see below), with 48 out of 50 pseudomyogenic hemangioendothelioma cases showing positive nuclear staining (51, 52).

2.6.3 Tumorigenesis and genetics

Trombetta et al published the first report of a balanced translocation between chromosomes 7 and 19 in pseudomyogenic hemangioendothelioma (53). The exact fusion partners were later on identified as *SERPINE1* and *FOSB* (54). FISH split probes for *FOSB* can be an excellent diagnostic marker (figure 2.4g). Most likely the *SERPINE1-FOSB* fusion leads to up-regulation of FOSB as *FOSB* is retained almost entirely (fused at exon 2) and gains the promoter of *SERPINE1* (fusion occurs in intron 1). Upregulation of FOSB could lead to activation of the AP-1 complex which is a potent transcription factor leading to the tumorigenesis, and thereby the underlying molecular mechanism is very similar to epithelioid hemangioma in which *ZFP36-FOSB*, as well as different types of *FOS* fusions, cause activation of the AP-1 complex.

2.7 Epithelioid hemangioendothelioma

2.7.1 Definition, epidemiology and clinical features

Epithelioid hemangioendothelioma of bone is classified by the WHO as a low grade malignant vascular sarcoma (8). Most are indolent, although 20-30% of the tumors metastasize and mortality is around 15% (55). The first distinction of epithelioid hemangioendothelioma from angiosarcoma was made by Thomas in 1942 who acknowledged that epithelioid hemangioendothelioma resembled epithelium in contrast to angiosarcoma. Moreover, he also described angiosarcoma to have a more malignant clinical course. The first comprehensive description of epithelioid hemangioendothelioma was formulated by Stout in 1943 who described two critical features. First the formation of atypical endothelial cells in greater numbers than are required for the lining of blood vessels. Second the formation of vascular tubes with a delicate framework of reticulin fibers (56). Later the characteristic

blister cells were added to the criteria for epithelioid hemangioendothelioma by Enzinger and Weiss (57). The tumor could easily be mistaken for a carcinoma due to its epithelioid appearance and lack of vasoformation.

The combined literature shows that the age of occurrence is regularly distributed between 10 and 60 years. There is a male predominance (68.9% of patients). The cases with reported metastasis showed that the prefered site for metastasis were the lungs followed by the skeleton, but it remains unclear whether these skeletal metastases should be considered as true metastases or multifocal regional spread. Overall, epithelioid hemangioendothelioma of bone is polyostotic in >50% of the cases (58). It predominantly affects the lower extremity (∼62%), and in up to 18% of the cases concurrent parenchymal tumors are found. Radiologically epithelioid hemangioendothelioma, like the other vascular tumors in bone, presents as a lytic lesion, without a sharp demarcation.

2.7.2 Histological and immunohistochemistry features

Histologically, epithelioid hemangioendothelioma typically consists of epithelioid cells, with abundant eosinophilic cytoplasm, sometimes with intracytoplasmic vacuolization (so-called "blister cells") (figure 2.5a, 2.5b). The cells are organized in short cords or strands and characteristically embedded in hyalinized or myxoid stroma. The tumor has an infiltrative growth pattern, and is lacking a lobular or vasoformative architecture. Marked nuclear atypia and / or necrosis is found in ∼33% of the cases. Inflammatory cells are usually absent. Immunohistochemically, the tumor cells are positive for CD31(100%) (figure 2.5c), CD34 (85%), FLI1 (100%), keratin (25-38%), D2-40 (54%), Prox1 (47%), ERG (98%) and Claudin-1 (88%) (23, 59, 60). Recently, nuclear staining for CAMTA1 was shown to be a highly specific marker for epithelioid hemangioendothelioma positive in 86-88% of the cases (figure 2.5d) (61, 62). TFE3 immunohistochemistry can be used to identify the very specific subset of epithelioid hemangioendotheliomas with *YAP1-TFE3* fusions, although not all TFE3 positive cases carry the translocation (see below) (63). The clinical behaviour of epithelioid hemangioendothelioma is highly variable and difficult to predict based on histological features. Deyrup et al proposed a risk stratification scheme in which tumors larger than 3cm, with >3 mitoses per 50 HPF have a 5 years survival rate of 59% and a metastatic rate of 32% as compared to 100% 5 year survival for patients with tumors smaller than 3 cm with less than 3 mitoses per 50 HPF (55). Whether this risk stratification scheme is also applicable to epithelioid haemangioendotheliomas with primary bone location remains to be established.

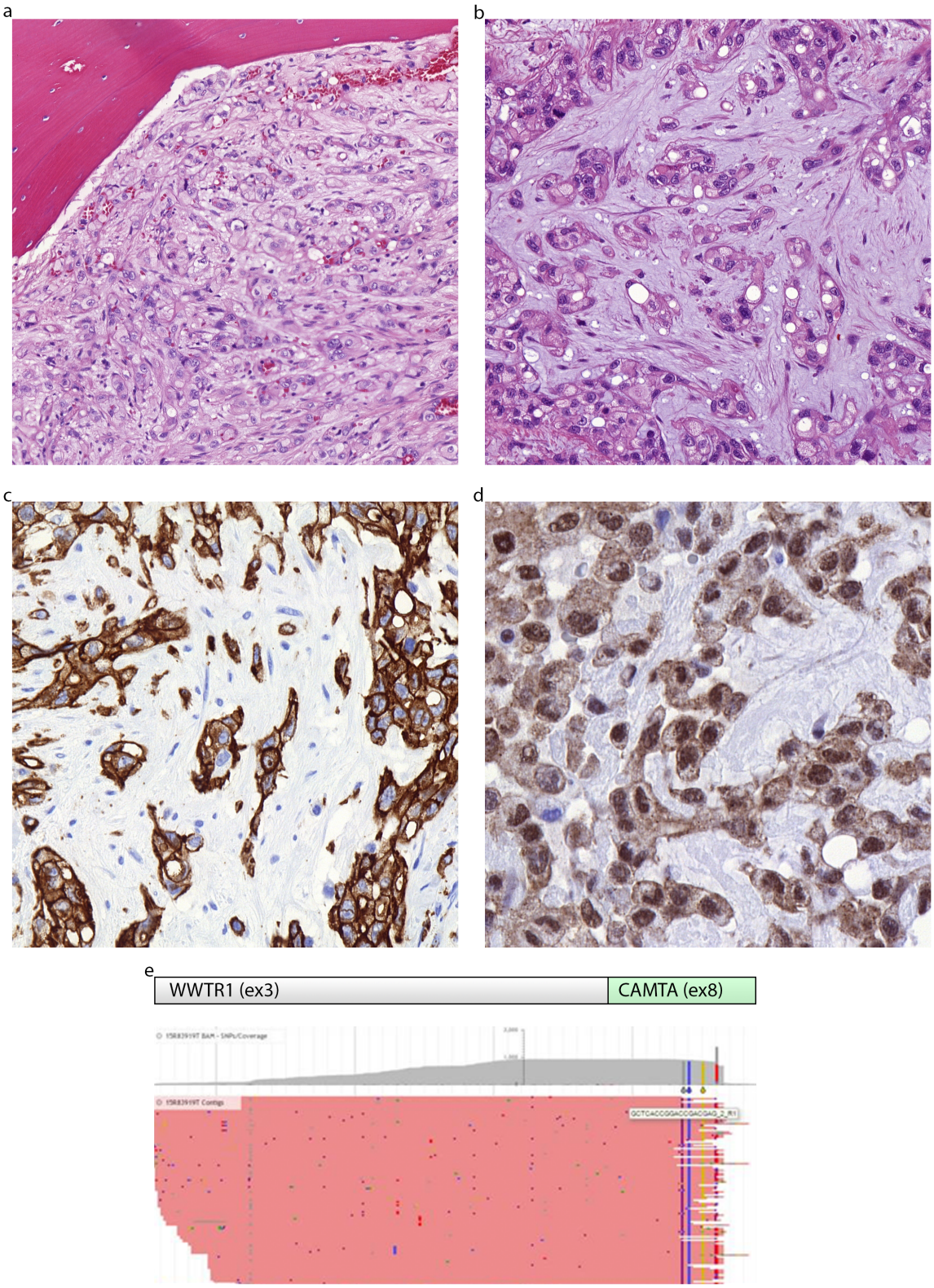


Figure 2.5: Epithelioid hemangioendothelioma of bone. (a) Epithelioid tumor cells in cords and strands embedded in stroma. (b) Intracytoplasmic vacuoles can be seen (Blister cells). (c) CD31 confirms endothelial differentiation. (d) CAMTA1 shows positive nuclear staining. (e) The *WWTR1-CAMTA1* fusion detected using next generation sequencing (Archer sarcoma fusion panel).

2.7.3 Tumorigenesis and genetics

Two fusions have been described for epithelioid hemangioendothelioma. The most common being the *WWTR1-CAMTA1* (figure 2.5e) fusion which was concurrently described (64, 65). Reportedly almost all (89-100%) epithelioid hemangioendothelioma with classic histological features harbor this fusion. Using genetic analysis the monoclonal origin of "multifocal" epithelioid hemangioendothelioma has been established using *WWTR1-CAMTA1* breakpoint analysis. This indicates that multiple lesions arise from local or metastatic spread from a single primary as opposed to multiple independent primaries (64).

In a distinct subset of epithelioid hemangioendotheliomas, which were negative for *WWTR1-CAMTA1*, a *YAP1-TFE3* fusion has been described (66). This specific subset affects predominantly young adults and has a distinct morphology, with vasoformative and vasoinvasive growth, combined with solid areas. The cytoplasm is voluminous, deeply eosinophilic or histiocytoid, sometimes feathery. The nuclei can be mild to moderately atypical. *TFE3* FISH can be used to confirm the diagnosis.

The *WWTR1-CAMTA1* fusion gene has been extensively studied. Interestingly, in contrast to what was speculated when the fusion was first described, the fusion is not a simple promoter swap where the *WWTR1* promoter drives CAMTA1. The *WWTR1-CAMTA1* fusion leads to activation of the Hippo signalling pathway, which is described to be an important regulator of organ size. As the chimeric protein contains the TEAD binding domain from WWTR1 it is able to activate the Hippo signaling pathway. The CAMTA1 part of the fusion protein leads to nuclear localization of the protein (67). Although less well studied, it would seem likely that the *YAP1-TFE3* chimeric protein would also lead to activation of the Hippo signaling pathway.

2.8 Angiosarcoma

2.8.1 Definition, epidemiology and clinical features

Angiosarcomas are highly aggressive sarcomas affecting the cutis, deep soft tissue, bone and viscera. Approximately 4% of all angiosarcomas arise primary in bone, and therefore angiosarcoma of bone is extremely rare. In bone, 30-40% of angiosarcomas are multifocal (21, 68, 69). Angiosarcomas are highly aggressive and predominantly occur in the seventh decade, with a male predominance. Angiosarcomas can be primary, or arise secondary to radiation (70). Angiosarcoma of bone should be treated with wide surgical resection, possibly with adjuvant radiation or chemotherapy. Virtually all patients die within a few years: the 1 year survival is 55% and the 5 year survival is 33% (21, 28). Radiologically angiosarcoma presents as a well-defined, osteolytic lesion, with a geographical

pattern of destruction (69). Cortical destruction is found in 65% of the cases. Because they are often multifocal, it is easily confused by radiologists with metastatic carcinoma. Macroscopically the tumors are hemorrhagic often with prominent necrosis.

2.8.2 Histological and immunohistochemistry features

Microscopically ill-defined blood vessels lined by enlarged endothelial cells with hyperchromatic, pleomorphic nuclei are seen (figure 2.6a, 2.6b). In bone, >90% of the angiosarcomas display epithelioid morphology (20, 21). In addition to the variable presence of vasoformative areas, often with multilayering, solid areas can be found. Mitoses are easily found, sometimes atypical forms (6, 21, 28, 71). Immunohistochemistry shows positivity for CD31 (95-100%) (figure 2.6d), ERG (96%), VWF (60-75%), CD34 (39-40%) and smooth muscle actin (61%) (20, 21, 23, 60). Keratin AE1AE3 is expressed in 69-80% of the angiosarcomas, and in combination with a radiological diagnosis of metastatic carcinoma and epithelioid morphology often causes misdiagnoses (21). D2-40 is expressed in 31% of the angiosarcomas of bone and is associated with a worse outcome (21). Likewise, loss of p16 is associated with a more aggressive clinical behaviour (72). In addition, the presence of a macronucleolus, three or more mitoses per 10 HPF, and less than 5 eosinophilic granulocytes are associated with poor outcome (21).

2.8.3 Tumorigenesis and genetics

Many different genetic alterations have been described for the angiosarcomas. As reported by Verbeke et al two groups of angiosarcomas could be identified; those with complex genetic profiles and a group with few gross genetic alterations (73). Inactivation of the p53 pathway is very common in angiosarcoma. A study on angiosarcoma of the liver reported frequent events leading to inactivation of p53 where p14, p15 and p53 were analyzed for mutations and for methylation. In almost all cases p53 was disabled due to promoter methylation or mutations in the aforementioned genes (74, 75). Antonescu et al described mutations in KDR in 7%, restricted to breast angiosarcomas but later also reported a case in the lumbar spine (76). This mutation could lead to auto-phosphorylation which would provide rationale for treatment with tyrosine kinase inhibitors. Furthermore, unsupervised clustering of gene expression profiles of angiosarcomas and other soft tissue sarcomas revealed that angiosarcomas cluster closely together, indicating they are indeed a highly similar entity in their gene expression pattern even though the specific mutations per case can be different (77). *MYC* amplifications are common events (55-100%) in secondary angiosarcomas (after irradiation or chronic lymphedema) (78). Although *MYC* amplifications were first reported exclusively for secondary angiosarcomas, more

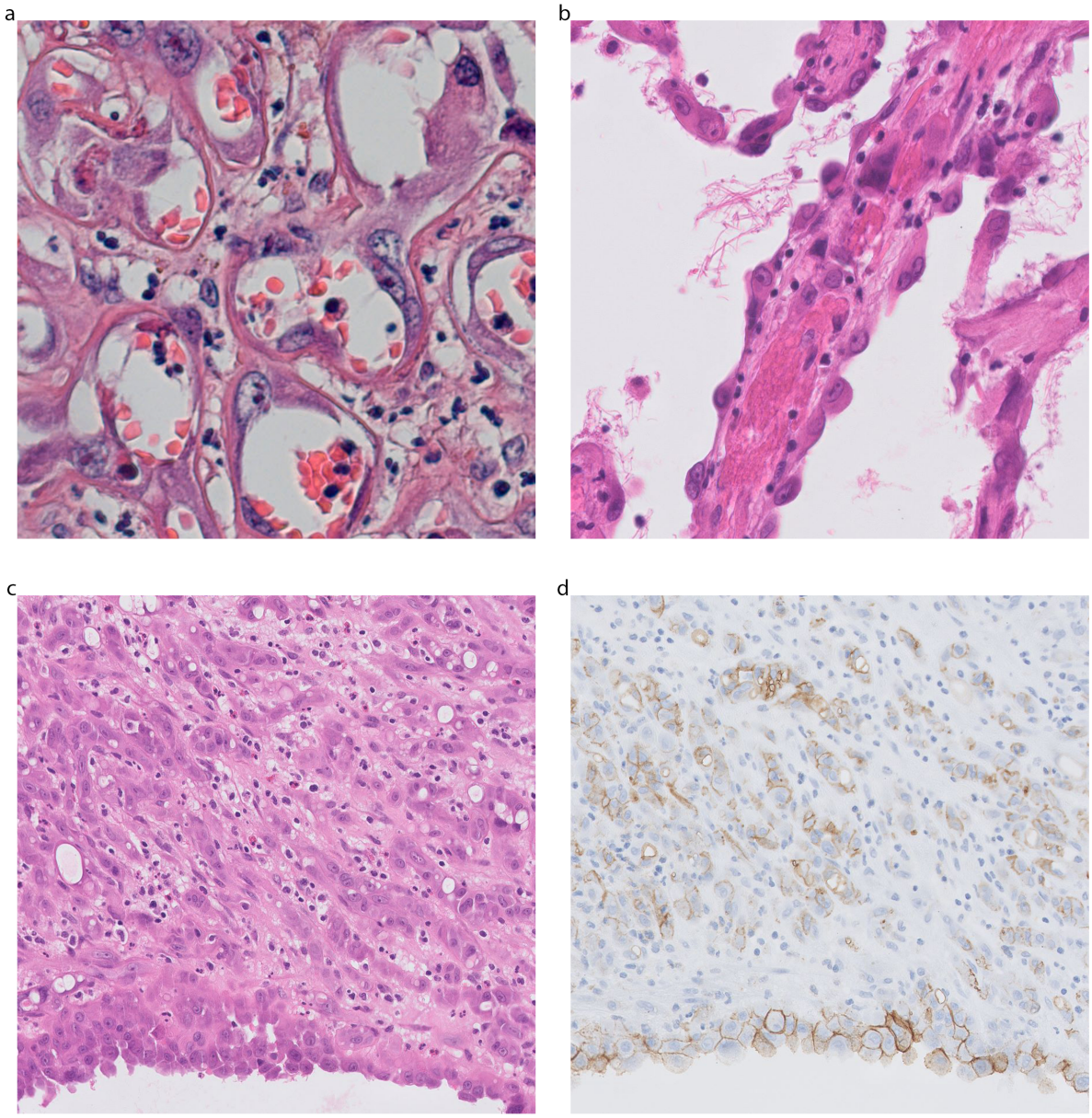


Figure 2.6: Angiosarcoma of bone. (a) Angiosarcoma of bone showing high degree of nuclear atypia, often with prominent nucleoli. (b) Vessels are lined by epithelioid endothelial cells. Cells show nuclear atypia. (c) Diffuse growth can be seen and tumor cells can have cytoplasmic vacuoles. (d) Corresponding CD31 immunohistochemistry.

recently they have been described in primary angiosarcoma cases as well including angiosarcoma of bone (73). Co-amplification of *FLT4*, and additional mutations in PLCG1 and PTPRB can be found in secondary angiosarcoma (70, 79). Furthermore, Huang et al recently reported *CIC* abnormalities occurring in 9% of cases, affecting younger patients with primary AS, with an inferior disease-free survival (80, 81). Although alterations in these genes have not been described in angiosarcoma of bone, it should be noted that genetic information for angiosarcoma of bone is limited.

2.9 Discussion

Vascular tumors consist of endothelial cells, often retaining the capacity to form vessels. Over time there has been much controversy surrounding the classification and naming of the various vascular tumors. But much of the confusion has been clarified with the discovery of a number of fusion genes and genetic alterations that point to distinct tumor entities.

Diagnosis of vascular tumors can be challenging as they are often very similar in appearance but required very different therapeutic approaches. Moreover, the epithelioid variants can show high keratin expression leading to confusion with metastatic carcinoma.

Bibliography

[1] Lamerato-Kozicki AR, Helm KM, Jubala CM, Cutter GC, Modiano JF. Canine hemangiosarcoma originates from hematopoietic precursors with potential for endothelial differentiation. Experimental Hematology. 2006;34(7):870–878. doi:10.1016/J.EXPHEM.2006.04.013.

[2] Kakiuchi-Kiyota S, Crabbs TA, Arnold LL, Pennington KL, Cook JC, Malarkey DE, et al. Evaluation of Expression Profiles of Hematopoietic Stem Cell, Endothelial Cell, and Myeloid Cell Antigens in Spontaneous and Chemically Induced Hemangiosarcomas and Hemangiomas in Mice. Toxicologic Pathology. 2013;41(5):709–721. doi:10.1177/0192623312464309.

[3] Yoder MC, Ingram DA. Endothelial progenitor cell: ongoing controversy for defining these cells and their role in neoangiogenesis in the murine system. Curr Opin Hematol. 2009;16(4):269–273. doi:10.1097/MOH.0b013e32832bbcab.

[4] Kovacic JC, Moore J, Herbert A, Ma D, Boehm M, Graham RM. Endothelial progenitor cells, angioblasts, and angiogenesis–old terms reconsidered from a current perspective. Trends Cardiovasc Med. 2008;18(2):45–51. doi:10.1016/j.tcm.2007.12.002.

[5] Wenger DE, Wold LE. Benign vascular lesions of bone: radiologic and pathologic features. Skeletal Radiology. 2000;29(2):63–74. doi:10.1007/s002560050012.

[6] Wenger DE, Wold LE. Malignant vascular lesions of bone: Radiologic and pathologic features. Skeletal Radiology. 2000;29(11):619–631. doi:10.1007/s002560000261.

[7] Bruder E, Perez-Atayde AR, Jundt G, Alomari AI, Rischewski J, Fishman SJ, et al. Vascular lesions of bone in children, adolescents, and young adults. A clinicopathologic reappraisal and application of the ISSVA classification. Virchows Archiv. 2009;454(2):161–179. doi:10.1007/s00428-008-0709-3.

[8] Fletcher CDM, Bridge JA, Hogendoorn P, Mertens F. Vascular tumours. Lyon: IARC; 2013.

[9] Verbeke SL, Bovee JV. Primary vascular tumors of bone: a spectrum of entities? Int J Clin Exp Pathol. 2011;4(6):541–551.

[10] Volpe R, Mazabraud A. Hemangioendothelioma (angiosarcoma) of bone: A distinct pathologic entity with an unpredictable course? Cancer. 1982;49(4):727–736. doi:10.1002/1097-0142(19820215)49:4<727::AID-CNCR2820490422>3.0.CO;2-V.

[11] Nielsen GP, Srivastava A, Kattapuram S, Deshpande V, O'Connell JX, Mangham CD, et al. Epithelioid hemangioma of bone revisited: A study of 50 cases. American Journal of Surgical Pathology. 2009;33(2):270–277. doi:10.1097/PAS.0b013e31817f6d51.

[12] van IJzendoorn DG, de Jong D, Romagosa C, Picci P, Benassi MS, Gambarotti M, et al. Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. Genes Chromosomes Cancer. 2015;54(9):565–574. doi:10.1002/gcc.22269.

[13] Huang S, Chen H, Zhang L, Sung Y, Dickson B, Krausz T, et al. Frequent FOS Gene Rearrangements in Epithelioid Hemangioma. Lab Invest. 2015;95:18A–19A.

[14] Antonescu CR, Chen HW, Zhang L, Sung YS, Panicek D, Agaram NP, et al. ZFP36-FOSB fusion defines a subset of epithelioid hemangioma with atypical features. Genes Chromosomes and Cancer. 2014;53(11):951–959. doi:10.1002/gcc.22206.

[15] Walter JW, North PE, Waner M, Mizeracki A, Blei F, Walker JWT, et al. Somatic mutation of vascular endothelial growth factor receptors in juvenile hemangioma. Genes Chromosomes and Cancer. 2002;33(3):295–303. doi:10.1002/gcc.10028.

[16] Hornick JL, Fletcher CD. Pseudomyogenic hemangioendothelioma: a distinctive, often multicentric tumor with indolent behavior. Am J Surg Pathol. 2011;35(2):190–201. doi:10.1097/PAS.0b013e3181ff0901.

[17] Inyang A, Mertens F, Puls F, Sumathi V, Inwards C, Folpe A, et al. Primary pseudomyogenic hemangioendothelioma of bone. American Journal of Surgical Pathology. 2016;40(5):587–598. doi:10.1097/PAS.0000000000000613.

[18] Machado I, Mayordomo-Aranda E, Scotlandi K, Picci P, Llombart-Bosch A. Immunoreactivity using anti-ERG monoclonal antibodies in sarcomas is influenced by clone selection. Pathol Res Pract. 2014;210(8):508–513. doi:10.1016/j.prp.2014.04.005.

[19] Perner S, Mosquera JM, Demichelis F, Hofer MD, Paris PL, Simko J, et al. TMPRSS2-ERG fusion prostate cancer: An early molecular event associated with invasion. American Journal of Surgical Pathology. 2007;31(6):882–888. doi:10.1097/01.pas.0000213424.38503.aa.

[20] Deshpande V, Rosenberg AE, O'Connell JX, Nielsen GP. Epithelioid angiosarcoma of the bone: A series of 10 cases. American Journal of Surgical Pathology. 2003;27(6):709–716. doi:10.1097/00000478-200306000-00001.

[21] Verbeke SL, Bertoni F, Bacchini P, Sciot R, Fletcher CD, Kroon HM, et al. Distinct histological features characterize primary angiosarcoma of bone. Histopathology. 2011;58(2):254–264. doi:10.1111/j.1365-2559.2011.03750.x.

[22] Barak S, Wang Z, Miettinen M. Immunoreactivity for calretinin and keratins in desmoid fibromatosis and other myofibroblastic tumors: A diagnostic pitfall. American Journal of Surgical Pathology. 2012;36(9):1404–1409. doi:10.1097/PAS.0b013e3182556def.

[23] Miettinen M, Wang ZF, Paetau A, Tan SH, Dobi A, Srivastava S, et al. ERG transcription factor as an immunohistochemical marker for vascular endothelial tumors and prostatic carcinoma. American Journal of Surgical Pathology. 2011;35(3):432–441. doi:10.1097/PAS.0b013e318206b67b.

[24] Mirra JM, Gold RH, Marcove RC. Bone Tumors, Diagnosis and Treatment: Diagnosis and Treatment. Lippincott Williams & Wilkins; 1980.

[25] Picci P. Atlas of Musculoskeletal Tumors and Tumorlike Lesions. Springer Science & Business Media; 2014. Available from: `http://link.springer.com/10.1007/978-3-319-01748-8`.

[26] Kaleem Z, Kyriakos M, Totty WG. Solitary skeletal hemangioma of the extremities. Skeletal Radiol. 2000;29(9):502–513. doi:10.1007/s002560000290502.256.

[27] López-Barea F, Hardisson D, Rodríguez-Peralto JL, Sánchez-Herrera S, Lamas M. Intracortical hemangioma of bone. Report of two cases and review of the literature. The Journal of bone and joint surgery American volume. 1998;80(11):1673–8.

[28] O'Connell JX, Nielsen GP, Rosenberg aE. Epithelioid vascular tumors of bone: a review and proposal of a classification scheme. Advances in anatomic pathology. 2001;8(2):74–82.

[29] O'Connell JX, Kattapuram SV, Mankin HJ, Bhan AK, Rosenberg AE. Epithelioid hemangioma of bone: A tumor often mistaken for low-grade angiosarcoma or malignant hemangioendothelioma. American Journal of Surgical Pathology. 1993;17(6):610–617. doi:10.1097/00000478-199306000-00009.

[30] Keel SB, Rosenberg AE. Hemorrhagic epithelioid and spindle cell hemangioma: a newly recognized, unique vascular tumor of bone. Cancer. 1999;85(9):1966–1972.

[31] Mridha AR, Kinra P, Sable M, Sharma MC, Rastogi S, Khan SA, et al. Epithelioid hemangioma of distal femoral epiphysis in a patient with congenital talipes equinovarus. Malays J Pathol. 2014;36(1):63–66.

[32] Boyaci B, Hornicek FJ, Nielsen GP, Delaney TF, Pedlow FX, Mansfield FL, et al. Epithelioid hemangioma of the spine: A case series of six patients and review of the literature. Spine Journal. 2013;13(12):e7–13. doi:10.1016/j.spinee.2013.06.070.

[33] Calderaro J, Guedj N, Dauzac C, Wassef M, Guigui P, Bedossa P. Un cas d'hémangiome épithélioïde du rachis. Annales de Pathologie. 2011;31(4):312–315. doi:10.1016/j.annpat.2011.05.008.

[34] Weaver SM, Kumar AB. Epithelioid hemangioma of the spine: an uncommon cause of spinal cord compression. Acta Neurol Belg. 2015;115(4):843–845. doi:10.1007/s13760-015-0437-9.

[35] Akgun B, Ozturk S, Ucer O, Erol FS. Epithelioid hemangioma of the thoracic spine. Neurol India. 2015;63(4):610–611. doi:10.4103/0028-3886.162082.

[36] Cummings TJ, Proia AD, Woodward J. Epithelioid hemangioma of the orbit. Clinical and Surgical Ophthalmology. 2004;22(8):224–226. doi:10.1016/S0161-6420(00)00011-7.

[37] Budimir I, Demirovic A, Ivekovic R, Pazanin L. Epithelioid hemangioma of the orbit: case report. Acta Clin Croat. 2015;54(1):92–95.

[38] Sanchez-Orgaz M, Insausti-Garcia A, Gregorio LY, Duralde AA, Romero-Martin R. Epithelioid hemangioma of the orbit or angiolymphoid hyperplasia with eosinophilia. Ophthalmic Plastic and Reconstructive Surgery. 2014;30(3):e70–2. doi:10.1097/IOP.0b013e31829c4216.

[39] Alder B, Proia A, Liss J. Distinct, bilateral epithelioid hemangioma of the orbit. Orbit (Amsterdam, Netherlands). 2013;32(1):51–53. doi:10.3109/01676830.2012.739674.

[40] Fernandes BF, Al-Mujaini A, Petrogiannis-Haliotis T, Al-Kandari A, Arthurs B, Burnier MN. Epithelioid hemangioma (angiolymphoid hyperplasia with eosinophilia) of the orbit: a case report. Journal of Medical Case Reports. 2007;1(1):30. doi:10.1186/1752-1947-1-30.

[41] Baili L, Cheour M, Hachicha F, Aydi Z, Zoubeidi H, Ben Brahim E, et al. Hémangiome épithélioïde de l'orbite: à propos d'un cas. Journal Francais d'Ophtalmologie. 2014;37(9):e133–e136. doi:10.1016/j.jfo.2014.01.017.

[42] Svajdler M, Bohus P, Baumohlova H, Sokol L, Bielek J. [Epithelioid hemangioma of the foot]. Cesk Patol. 2006;42(2):86–90.

[43] Luna JTP, DeGroot H. Five-Year Follow-up of Structural Allograft Reconstruction for Epithelioid Hemangioma of the Talus and Navicular: A Case Report and Review of the Literature. Foot & Ankle International. 2007;28(3):379–384. doi:10.3113/FAI.2007.0379.

[44] Werhahn C, Lang M, Merkel KH, Pickartz H. [Epithelioid hemangioma–a rare tumor of the hand]. Handchir Mikrochir Plast Chir. 1990;22(4):214–218.

[45] El Harroudi T, Moumen M, Tijami F, El Otmany A, Jalil A. Hémangiome épithélioïde géant de la main. Chirurgie de la Main. 2008;27(5):240–242. doi:10.1016/j.main.2008.08.006.

[46] Sirikulchayanonta V, Arthit J, Suphaneewan J. Epithelioid hemangioma involving three contiguous bones: A case report with a review of the literature. Korean Journal of Radiology. 2010;11(6):692–696. doi:10.3348/kjr.2010.11.6.692.

[47] Zhou Q, Lu L, Fu Y, Xiang K, Xu L. Epithelioid hemangioma of bone: a report of two special cases and a literature review. Skeletal radiology. 2016;45(12):1723–1727. doi:10.1007/s00256-016-2482-8.

[48] Eferl R, Wagner EF. AP-1: A double-edged sword in tumorigenesis. Nature Reviews Cancer. 2003;3(11):859–868. doi:10.1038/nrc1209.

[49] Mirra JM, Kessler S, Bhuta S, Eckardt J. The fibroma-like variant of epithelioid sarcoma. A fibrohistiocytic/myoid cell lesion often confused with benign and malignant spindle cell tumors. Cancer. 1992;69(6):1382–1395. doi:10.1002/1097-0142(19920315)69:6<1382::AID-CNCR2820690614>3.0.CO;2-Y.

[50] Billings SD, Folpe AL, Weiss SW. Epithelioid Sarcoma-Like Hemangioendothelioma. The American journal of surgical pathology. 2003;27(1):48–57. doi:10.1097/PAS.0b013e31821caf1c.

[51] Hung YP, Fletcher CDM, Hornick JL. FOSB is a useful diagnostic marker for pseudomyogenic hemangioendothelioma. American Journal of Surgical Pathology. 2017;41(5):596–606. doi:10.1097/PAS.0000000000000795.

[52] Sugita S, Hirano H, Kikuchi N, Kubo T, Asanuma H, Aoyama T, et al. Diagnostic utility of FOSB immunohistochemistry in pseudomyogenic hemangioendothelioma and its histological mimics. Diagn Pathol. 2016;11(1):75. doi:10.1186/s13000-016-0530-2.

[53] Trombetta D, Magnusson L, von Steyern FV, Hornick JL, Fletcher CDM, Mertens F. Translocation t(7;19)(q22;q13)-a recurrent chromosome aberration in pseudomyogenic hemangioendothelioma? Cancer Genetics. 2011;204(4):211–215. doi:10.1016/j.cancergen.2011.01.002.

[54] Walther C, Tayebwa J, Lilljebjörn H, Magnusson L, Nilsson J, Von Steyern FV, et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. Journal of Pathology. 2014;232(5):534–540. doi:10.1002/path.4322.

[55] Deyrup AT, Tighiouart M, Montag AG, Weiss SW. Epithelioid hemangioendothelioma of soft tissue: a proposal for risk stratification based on 49 cases. Am J Surg Pathol. 2008;32(6):924–927.

[56] Thomas A. Vascular tumors of bone: pathological and clinical study of 27 cases. Surg Gynecol Obstet. 1942;777–95.

[57] Stout AP. Hemangio-Endothelioma: a Tumor of Blood Vessels Featuring Vascular Endothelial Cells. Ann Surg. 1943;118(3):445–464.

[58] Weiss SW, Enzinger FM. Epithelioid Hemangioendothelioma - a Vascular Tumor Often Mistaken for a Carcinoma. Cancer. 1982;50(5):970–981. doi:Doi 10.1002/1097-0142(19820901)50:5<970::Aid-Cncr2820500527>3.0.Co;2-Z.

[59] Campanacci M, Boriani S, Giunti A. Hemangioendothelioma of bone: A study of 29 cases. Cancer. 1980;46(4):804–814. doi:10.1002/1097-0142(19800815)46:4<804::AID-CNCR2820460427>3.0.CO;2-1.

[60] Gill R, O'Donnell RJ, Horvai A. Utility of immunohistochemistry for endothelial markers in distinguishing epithelioid hemangioendothelioma from carcinoma metastatic to bone. Arch Pathol Lab Med. 2009;133(6):967–972. doi:10.1043/1543-2165-133.6.967.

[61] Miettinen M, Rikala MS, Rys J, Lasota J, Wang ZF. Vascular endothelial growth factor receptor 2 as a marker for malignant vascular tumors and mesothelioma: an immunohistochemical study of 262 vascular endothelial and 1640 nonvascular tumors. Am J Surg Pathol. 2012;36(4):629–639. doi:10.1097/PAS.0b013e318243555b.

[62] Shibuya R, Matsuyama A, Shiba E, Harada H, Yabuki K, Hisaoka M. CAMTA1 is a useful immunohistochemical marker for diagnosing epithelioid haemangioendothelioma. Histopathology. 2015;67(6):827–835. doi:10.1111/his.12713.

[63] Doyle LA, Fletcher CDM, Hornick JL. Nuclear expression of CAMTA1 distinguishes epithelioid hemangioendothelioma from histologic mimics. American Journal of Surgical Pathology. 2016;40(1):94–102. doi:10.1097/PAS.0000000000000511.

[64] Flucke U, Vogels RJ, de Saint Aubain Somerhausen N, Creytens DH, Riedl RG, van Gorp JM, et al. Epithelioid Hemangioendothelioma: clinicopathologic, immunhistochemical, and molecular genetic analysis of 39 cases. Diagn Pathol. 2014;9:131. doi:10.1186/1746-1596-9-131.

[65] Errani C, Zhang L, Sung YS, Hajdu M, Singer S, Maki RG, et al. A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. Genes Chromosomes and Cancer. 2011;50(8):644–653. doi:10.1002/gcc.20886.

[66] Tanas MR, Sboner A, Oliveira AM, Erickson-Johnson MR, Hespelt J, Hanwright PJ, et al. Identification of a disease-defining gene fusion in epithelioid hemangioendothelioma. Sci Transl Med. 2011;3(98):98ra82. doi:10.1126/scitranslmed.3002409.

[67] Antonescu CR, Le Loarer F, Mosquera JM, Sboner A, Zhang L, Chen CL, et al. Novel YAP1-TFE3 fusion defines a distinct subset of epithelioid hemangioendothelioma. Genes Chromosomes Cancer. 2013;52(8):775–784. doi:10.1002/gcc.22073.

[68] Patel NR, Salim AA, Sayeed H, Sarabia SF, Hollingsworth F, Warren M, et al. Molecular characterization of epithelioid haemangioendotheliomas identifies novel WWTR1-CAMTA1 fusion variants. Histopathology. 2015;67(5):699–708. doi:10.1111/his.12697.

[69] Young RJ, Brown NJ, Reed MW, Hughes D, Woll PJ. Angiosarcoma. The Lancet Oncology. 2010;11(10):983–991. doi:10.1016/S1470-2045(10)70023-1.

[70] Vermaat M, Vanel D, Kroon HM, Verbeke SLJ, Alberghini M, Bovee JVMG, et al. Vascular tumors of bone: Imaging findings. European Journal of Radiology. 2011;77(1):13–18. doi:10.1016/j.ejrad.2010.06.052.

[71] Mentzel T, Schildhaus HU, Palmedo G, Buttner R, Kutzner H. Postradiation cutaneous angiosarcoma after treatment of breast carcinoma is characterized by MYC amplification in contrast to atypical vascular lesions after radiotherapy and control cases: clinicopathological, immunohistochemical and molecular analysis o. Mod Pathol. 2012;25(1):75–85. doi:10.1038/modpathol.2011.134.

[72] Evans HL, Raymond AK, Ayala AG. Vascular tumors of bone: A study of 17 cases other than ordinary hemangioma, with an evaluation of the relationship of hemangioendothelioma of bone to epithelioid hemangioma, epithelioid hemangioendothelioma, and high-grade angiosarcoma. Human Pathology. 2003;34(7):680–689. doi:10.1016/S0046-8177(03)00249-1.

[73] Verbeke SL, Bertoni F, Bacchini P, Oosting J, Sciot R, Krenacs T, et al. Active TGF-beta signaling and decreased expression of PTEN separates angiosarcoma of bone from its soft tissue counterpart. Mod Pathol. 2013;26(9):1211–1221. doi:10.1038/modpathol.2013.56.

[74] Verbeke SL, de Jong D, Bertoni F, Sciot R, Antonescu CR, Szuhai K, et al. Array CGH analysis identifies two distinct subgroups of primary angiosarcoma of bone. Genes Chromosomes Cancer. 2015;54(2):72–81. doi:10.1002/gcc.22219.

[75] Weihrauch M, Markwarth A, Lehnert G, Wittekind C, Wrbitzky R, Tannapfel A. Abnormalities of the ARF-p53 pathway in primary angiosarcomas of the liver. Human Pathology. 2002;33(9):884–892. doi:10.1053/hupa.2002.126880.

[76] Garcia JM, Gonzalez R, Silva JM, Dominguez G, Vegazo IS, Gamallo C, et al. Mutational status of K-ras and TP53 genes in primary sarcomas of the heart. Br J Cancer. 2000;82(6):1183–1185. doi:10.1054/bjoc.1999.1060.

[77] Antonescu CR, Yoshida A, Guo T, Chang NE, Zhang L, Agaram NP, et al. KDR activating mutations in human angiosarcomas are sensitive to specific kinase inhibitors. Cancer Res. 2009;69(18):7175–7179. doi:10.1158/0008-5472.CAN-09-2068.

[78] Segal NH, Pavlidis P, Antonescu CR, Maki RG, Noble WS, DeSantis D, et al. Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. Am J Pathol. 2003;163(2):691–700. doi:10.1016/S0002-9440(10)63696-6.

[79] Manner J, Radlwimmer B, Hohenberger P, Mossinger K, Kuffer S, Sauer C, et al. MYC high level gene amplification is a distinctive feature of angiosarcomas after irradiation or chronic lymphedema. Am J Pathol. 2010;176(1):34–39. doi:10.2353/ajpath.2010.090637.

[80] Behjati S, Tarpey PS, Sheldon H, Martincorena I, Van Loo P, Gundem G, et al. Recurrent PTPRB and PLCG1 mutations in angiosarcoma. Nature Genetics. 2014;46(4):376–379. doi:10.1038/ng.2921.

[81] Huang SC, Zhang L, Sung YS, Chen CL, Kao YC, Agaram NP, et al. Recurrent CIC gene abnormalities in angiosarcomas: A molecular study of 120 cases with concurrent investigation of PLCG1, KDR, MYC, and FLT4 gene alterations. American Journal of Surgical Pathology. 2016;40(5):645–655. doi:10.1097/PAS.0000000000000582.

Part I

Diagnosis and treatment

Chapter 3

Fusion events lead to truncation of FOS in epithelioid hemangioma of bone

This chapter is based on the publication: **van IJzendoorn DGP**, de Jong D, Romagosa C, Picci P, Benassi MS, Gambarotti M, Daugaard S, van de Sande M, Szuhai K, Bovée JVMG. Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. *Genes Chromosom Cancer*. 2015;54: 565-574.

3.1 Abstract

Epithelioid hemangioma of bone is a locally aggressive vascular neoplasm. It can be challenging to diagnose because of the wide histological spectrum, which can make it difficult to differentiate from other vascular neoplasms such as epithelioid hemangioendothelioma or epithelioid angiosarcoma. COBRA-FISH karyotyping identified a balanced t(3;14) translocation. Transcriptome sequencing of the index case and two other epithelioid hemangiomas revealed a recurrent translocation breakpoint involving the FOS gene, which was fused to different partners in all three cases. The break was observed in exon 4 of the *FOS* gene and the fusion event led to the introduction of a stop codon. In all instances, the truncation of the *FOS* gene would result in the loss of the transactivation domain (TAD). Using FISH probes, we found a break in the *FOS* gene in two additional cases, in none of these cases a recurrent fusion partner could be identified. In total, *FOS* was split in 5/7 evaluable samples. We did not observe point mutations leading to early stop codons in any of the 10 cases where RNA was available. Detection of *FOS* rearrangement may be a useful diagnostic tool to assist in the often-difficult differential diagnosis of vascular tumors of bone. Our data suggest that the translocation causes truncation of the FOS protein, with loss of the TAD, which is thereby a novel mechanism involved in tumorigenesis.

3.2 Introduction

Epithelioid hemangioma of bone (previously known as angiolymphoid hyperplasia with eosinophilia or histiocytoid hemangioma) is a locally aggressive neoplasm composed of cells that have an endothelial phenotype and epithelioid morphology (1). Tumors usually involve long tubular bones (40%), distal lower extremities (18%), flat bones, (18%), vertebrae (16%), and small bones of the hands (8%) (1, 2). It is slightly more common in males and occurs at an average age of 35 years (3). Approximately 18-25% of the tumors display multifocal regional spread (2, 4). The tumors have a lobular architecture, and consist of epithelioid endothelial cells that form vascular lumina or grow in solid sheets (1). The diagnosis of epithelioid hemangioma is challenging, and distinction from epithelioid angiosarcoma can sometimes be difficult (4). Treatment with curettage or marginal en bloc excision is sufficient in most cases, although the tumor can show locally aggressive growth and rare lymph node involvement (2).

The classification of vascular tumors of bone has been controversial, and different classification systems have been proposed over the years (5). They include a heterogeneous group ranging from benign to malignant tumors, including hemangioma, epithelioid hemangioma, epithelioid hemangioendothelioma, and epithelioid angiosarcoma. Part of the

controversy in literature is caused by the fact that cases reported in the past as "hemangioendothelioma of bone" (6–10) probably reflect epithelioid hemangiomas (2, 5, 11), a term which is now generally accepted (1). The new classification of vascular tumors of bone as proposed in the 2013 WHO is supported by the rapid elucidation of novel, characteristic translocations in the different entities. Epithelioid hemangioendothelioma was shown to contain *WWTR1-CAMTA1* or, more rarely, *YAP1-TFE3* fusions (12–14). Recently, a *ZFP36-FOSB* fusion was identified in a subset of epithelioid hemangiomas with atypical features (15). In this study, we used molecular karyotyping to characterize classic epithelioid hemangioma of bone, followed by next generation sequencing analysis. Using COBRA-FISH a balanced translocation t(3;14) was found and by breakpoint mapping using Bacterial Artificial Chromosomes (BAC) clones the approximate locations of the translocation breakpoints were identified. Using transcriptome sequencing, a chimeric sequence was found resulting in a *FOS-MBNL1* fusion gene, as well as fusions involving FOS rearranged with different partner genes in two additional cases. In all three cases, the fusion led to a truncated form of FOS due to early termination of translation. Additional cases were analyzed by interphase FISH probes flanking the *FOS* locus.

3.3 Materials and methods

3.3.1 Patient samples

Eleven epithelioid hemangiomas of bone were available (table 3.1). Cases were acquired from the archives of the Leiden University Medical Center (LUMC), Leiden, The Netherlands, Rizzoli Institute Laboratory of Oncologic Research, Bologna, Italy, and the Department of Pathology, University of Copenhagen, Denmark. Two of the cases from the LUMC showed multifocal regional spread and for all these cases frozen and FFPE material was available from multiple foci (figure 3.1).

For all cases histology was reviewed by two pathologists (JVMGB, CR) and the diagnosis of epithelioid hemangioma was confirmed (table 3.1). All samples were handled according to the Dutch code of proper secondary use of human material as accorded by the Dutch society of pathology (`www.federa.org`). The samples were handled in a coded (pseudonymised) fashion according to the procedures as accorded by the LUMC ethical board.

Case ID	Age/Sex	Location	Multifocal regional spread	Treatment	Follow-up (month)	Break-Apart FOS (FISH)	RT-PCR
L3933[a,b]	F/42	Medial cuneiform bone, intermediate cuneiform bone, base metatarsal 1 and 2	Yes	Amputation	23, NED	Positive	*FOS-MBNL1*
L3141[b]	F/60	Metatarsal 4, metatarsal 2	Yes	Curettage, followed by amputation after relapse. Radiotherapy	73, NED	Positive	*FOS-VIM*
L4065[b]	M/50	Navicular bone, medial cuneiform bone, intermediate cuneiform bone, metatarsal 1	Yes	Amputation	14, NED	Not scorable	*FOS-lincRNA*
L3140	F/45	Vertebra T7	No	Vertebrectomy	63, NED	Not scorable	No
L3142	M/42	Femur diaphysis	No	Resection, radiotherapy	20, NED	Positive	No
L3151	M/33	Distal tibia, cuboid, cuneiform 3	Yes	Curettage	19, NED	Not scorable	No
L3153	M/83	Proximal tibia, distal tibia, fibula	Yes	Curettage, radiotherapy	72, NED	Not scorable	No
L3154	M/41	Metacarpal 2	No	Curettage	175, NED	Positive	No
L3157	M/40	Tibia diaphysis, proximal tibia, femur diaphysis	Yes	Knee disarticulation curettage, radiotherapy	114, NED	Not scorable	No
L3160	M/28	Proximal tibia	No	Curettage	36, NED	Negative	No
L3336	F/29	Metatarsal 1	No	Curettage	15, NED	Negative	No

Table 3.1: Clinical Features of Epithelioid Hemangioma Cases. [a]Index case that was used for the COBRA-FISH. [b]Cases that were transcriptome sequenced. NED: no evidence of disease; NA: not available.

3.3.2 Combined binary ratio labeling with multiple Fluorescence In Situ Hybridization (COBRA-FISH)

The tumor sample L3933 was obtained immediately after surgery and metaphase slides were prepared as described before (16). COBRA-FISH was performed as described by our group (16, 17).

3.3.3 Fluorescence In Situ Hybridization

FISH mapping using BAC clones selected from the 1MB clone set from the Welcome Trust, Sanger Institute (supplementary data 1 available online) were used to narrow down the translocation breakpoints in case L3933. Later, clones flanking the FOS locus were selected. Proximal and distal to *FOS*, BAC clones RP11-173A8 pooled with RP11-316E14 and RP11-361H10 pooled with RP11-368K8 were selected to detect rearrangements, respectively. DNA was extracted from the BAC clones using the High Pure Plasmid isolation kit (Roche, Woerden, The Netherlands). The two-color FISH was performed by labeling the probes with either biotin-11-dUTP or digoxigenin-11-dUTP (Roche, Woerden, The Netherlands) using a nick translation labeling reaction (18).

A two-color FISH for *FOS* break-apart (translocation) detection was performed on all cases. Four micrometer paraffin slides were prepared as described before and the labeled probe and the slides were denatured simultaneously at 80°C for 10 min. Thereafter slides were hybridized in a moist chamber overnight at 37°C. Posthybridization washing steps and fluorescent detection of probes were performed as described earlier (18). Slides were scanned using the Pannoramic MIDI scanner (3DHistech, Budapest, Hungary). Tumor areas were selected based on the hematoxylin and eosin stained slides and scored using Pannoramic Viewer software (3DHistech, Budapest, Hungary) by counting 100 nuclei. Cases were considered positive for a break when >20% of the nuclei showed a break-apart signal. Slides were scored by one observer (DGPIJ) and the conclusion was confirmed by a pathologist (JVMGB).

3.3.4 RNA sequencing using transcriptome sequencing

RNA was extracted from samples L3141, L3393, and L4065 for RNA sequencing. The samples were collected after surgery and stored at -80°C. From each sample 15, 20 μm slides were cut for RNA isolation. TRIzol (Life technologies, Carlsbad) was added to the cut tissue sections. The nucleotides were extracted from the mixture using chloroform and the content was precipitated using 2-propanol. The total nucleotides were washed using 75% ethanol and resuspended in Milli-Q water. An additional RNA purification step was included using the RNeasy Mini kit (Qiagen, Venlo, The Netherlands) according to the

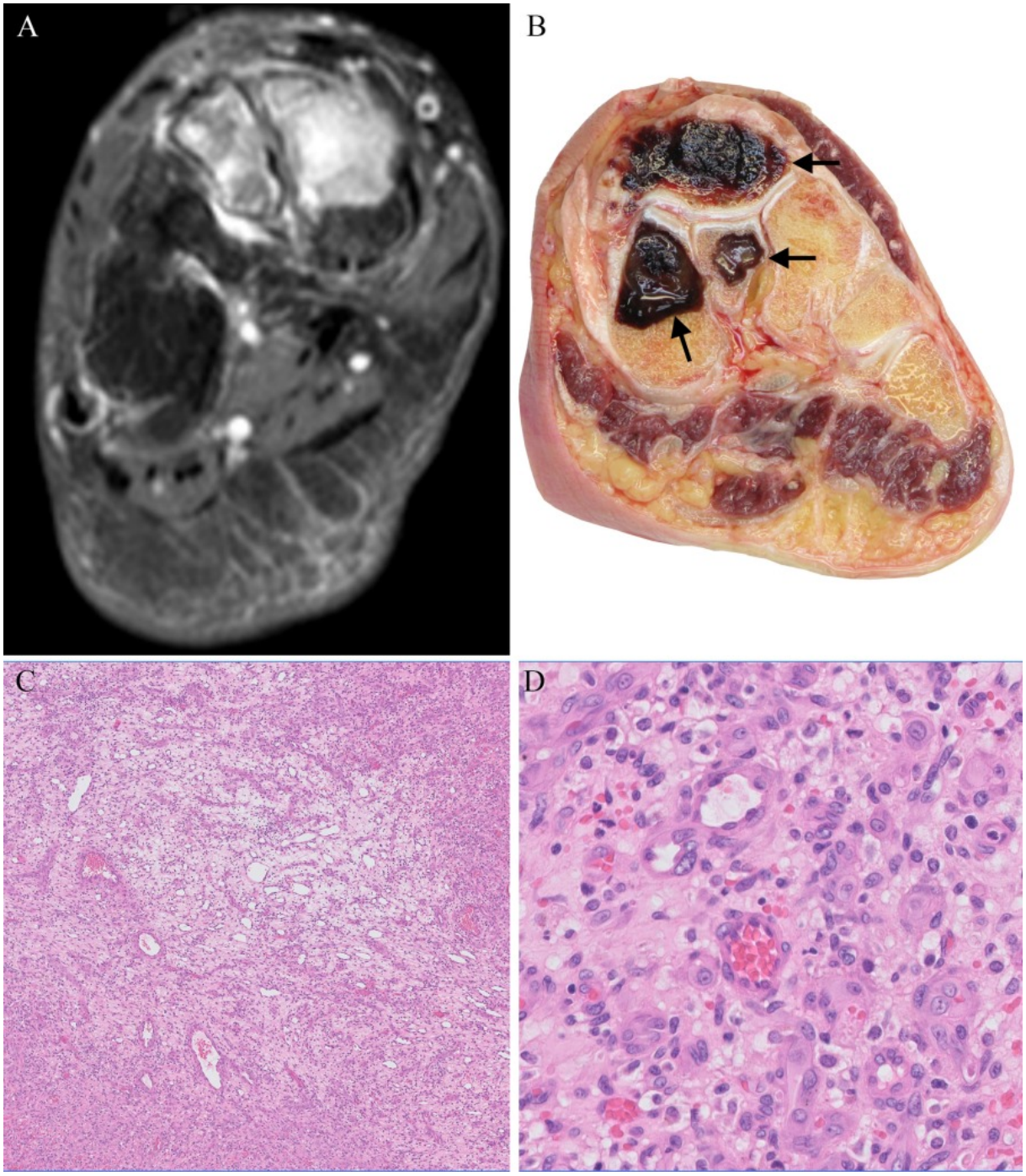


Figure 3.1: Epithelioid hemangioma case L4065 displaying multifocal regional spread revealing identical translocations in all separate tumors. (a) MRI image of multiple lesions in the foot. (b) Macroscopical image, at the same plane as the MRI, of the foot showing involvement of three separate tarsal bones (indicated by arrows). (c) H&E staining demonstrating vasoformative areas admixed with spindled and hemorrhagic areas at low power view. (d) High power view shows epithelioid cells lining vascular channels.

manufacturers protocol. The total RNA was send to BGI (Hong Kong) for sequencing with the HiSeq2000 platform (Illumina, San Diego).

3.3.5 RNA sequencing analysis with Defuse

An in-house pipeline was used to align the sequenced reads to the hg19 reference genome (`http://genome.ucsc.edu`) and analyze the data for SNVs. Sequence alignment was performed with TopHat2 (v2.0.13) aligning to the hg19 (`http://genome.ucsc.edu/`) and hg38 (`http://genome.ucsc.edu/`). SNVs were called with VarScan (v2.3.7). Annovar was used to annotate and filter the detected variants using the genomicSuperDups (`http://genome.ucsc.edu/`), snp138 (`http://genome.ucsc.edu/`) 1000 Genomes (`http://www.1000genomes.org/`), Exome Sequencing Project (`http://evs.gs.washington.edu/EVS/`), and the LJB23 nonsynonymous variant annotations (v2.3). Fusion detection was performed with Defuse (v0.6.2) aligning to the hg19 (`http://genome.ucsc.edu/`) and filtering with the repeats library from UCSC (`http://genome.ucsc.edu/`). Detected fusions were sorted according to fusion spanning reads.

3.3.6 RT-PCR

Total RNA was isolated from 10 frozen epithelioid hemangioma samples according to the protocol described above. cDNA was made using a mixture of RNasin, 5x RT-buffer, oligodT, random primer, dNTP's, and AMV-RT enzyme (Promega, Madison), which was added to the total RNA after denaturing the RNA for 15 min at 60°C and placing it on ice. The reverse transcriptase reaction was performed at 42°C for 1 hr. Then the enzyme was inactivated by heating the mix to 65°C for 15 min. Primers were designed for the fusions that were identified with Defuse. Primers for *FOS-MBNL1*, *FOS-VIM*, and *FOS-linc*RNA(RP11-326N17.1), were designed using Primer3 (v0.4.0). For *FOS-VIM* the forward primer 5'-GAGAAAAGGAGAATCCGAAGG-3' and reverse primer 5'-ATCTTCCGCTAGCAAGATGC-3' were used. For *FOS-MBNL1* the forward primer 5'-GAGAAAAGGAGAATCCGAAGG-3' and reverse primer 5'-TCCATAAGACGTGTGGG-TGT-3' were used. For *FOS-lincRNA*(RP11-326N17.1) the forward primer 5'-GAGAAAA-GGAGAATCCGAAGG-3' and the reverse primer 5'- GAAAATCTGAGCTGTAACCAAGC-3' were used. To compare wild-type *FOS* expression with the expression of the fusion genes primers were also made for the normal exon 4 of *FOS*. The forward primer 5'-GAGAAAAGGAGAATCCGAAGG-3' and reverse primer 5'-GTCAGAGGAAGGCTCAT-TGC-3' were used. RT-PCR was performed using the CFX touch 96 (BIO-RAD, Hercules). Primers were added to SYBR Green I (Life technologies, Carlsbad) and cDNA. RT-PCR ran at 55°C for 40 cycles with a 40 sec elongation time at 72°C. Sanger sequenc-

ing of the PCR product was performed by the LGTC (Leiden, The Netherlands) and the sequences were analyzed using Chromas (v2.1.1).

3.4 Results

3.4.1 A balanced t(3;14) and the approximate location for a translocation is determined

Using COBRA-FISH, we identified a balanced translocation 46,XX,t(3;14) in L3933 (figure 3.2). The approximate location of the translocation breakpoints was determined using BAC clones. Annealing locations for the BAC clones on the L3933 case are shown in supplementary data 1 available online. The break on chromosome 3 was between nt 151,839,044 and 152,021,576. The break on chromosome 14 was between nt 75,518,664 and 76,145,087.

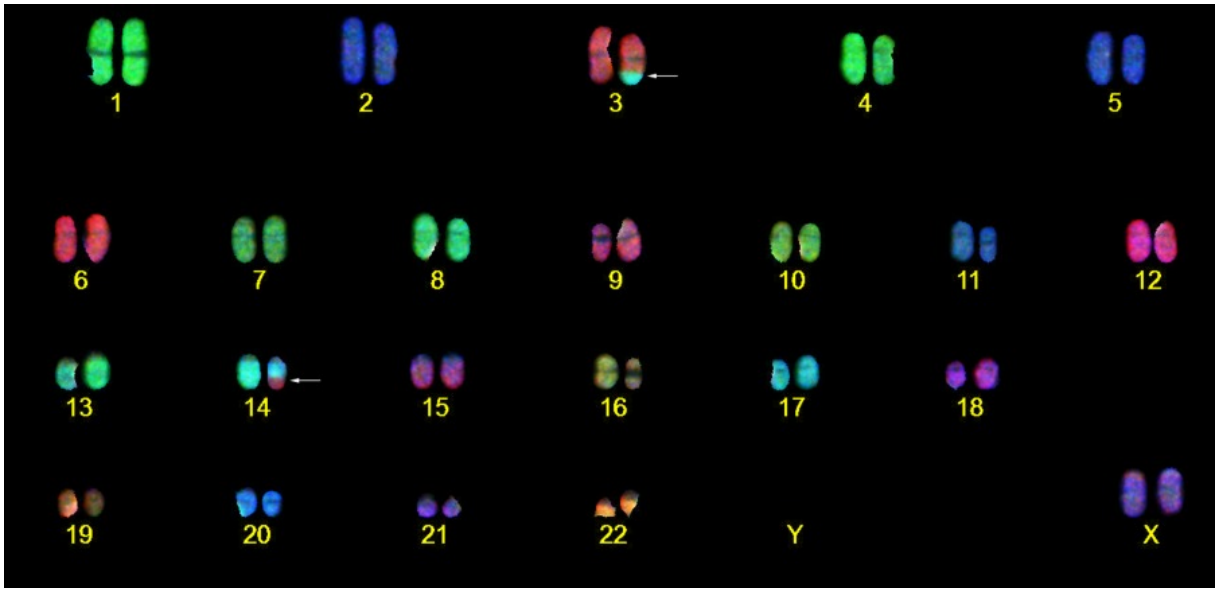


Figure 3.2: COBRA-FISH identifies a balanced 46,XX,t(3;14) translocation. Arrowheads are indicating the translocated chromosomes.

3.4.2 Three fusions involving *FOS* were identified

In index case L3933, Defuse identified a *FOS-MBNL1* translocation. *FOS* starts at position 75,745,477 on chromosome 14, matching to the fusion location that was identified using COBRA-FISH. *MBNL1* is located on chromosome 3 and starts at position 151,961,617 which matched to the expected location from the COBRA-FISH. In addition, Defuse identified fusions involving *FOS* in the two additional cases: L3141 had a *FOS-VIM* fusion

and L4065 had a *FOS-lincRNA* (RP11-326N17.1). All fusions were validated by RT-PCR (figure 3.3). In all three cases, the fusion-breakpoint in *FOS* is in exon 4. In the index case (L3933) nt position c.858 in exon 4 of *FOS* was fused to intron 2 of *MBNL1*, leading immediately to a stop codon at translation (figure 3.3a and GenBank accession number: KP790137). At the protein level this would lead to a truncation of the last 95 C terminal amino acids of the FOS protein. In case L3141 the break in *FOS* occurred at c.828 in exon 4 fused to the reverse complementary strand downstream of *VIM*. In the predicted fusion protein FOS would gain an additional (scrambled) 13 amino acids at the C terminal end before there is a stop codon (GenBank accession number: KP790138). In the case of L4065 the fusion occurred at c.806 in exon 4, where *FOS* was fused to a long non-coding RNA (RP11-326N17.1). FOS would gain an additional ten scrambled amino acids before a stop codon is encountered (GenBank accession number: KP790139). Therefore, in all three cases approximately 100 amino acids would be lost at the C terminal of the FOS protein, with loss of the transactivation domain (TAD) of *FOS*.

We sought for recurrent translocation events in our panel of epithelioid hemangioma samples performing an RT-PCR in the remaining 7 cases using all three translocation combinations using the index cases as positive controls; however, a recurrent fusion product was not detected in any of these cases.

Since truncation could also be caused by point mutation of the *FOS* gene, we examined exon 4 of *FOS* using RT-PCR and Sanger sequencing. No mutations were found in any of the 11 cases for which frozen material was available.

3.4.3 Cases with multifocal regional spread display identical fusion products in different Foci

For index case L3933 RNA was available from six separate tumors and for case L4065 there was RNA from three separate tumors. RT-PCR was performed on all tumor locations for the three identified fusions and the products were Sanger sequenced. None of the cases were positive for the *FOS-VIM* fusion. In the index case all separate tumors were positive for identical *FOS-MBNL1* fusions and in L4065 all separate tumors were positive for identical *FOS-lincRNA* (RP11-326N17.1), suggesting that the separately located tumors within one patient are derived from one single clone. Using FISH it was clear that the normal cells surrounding the tumor did not carry the translocation, excluding the possibility of a germline translocation.

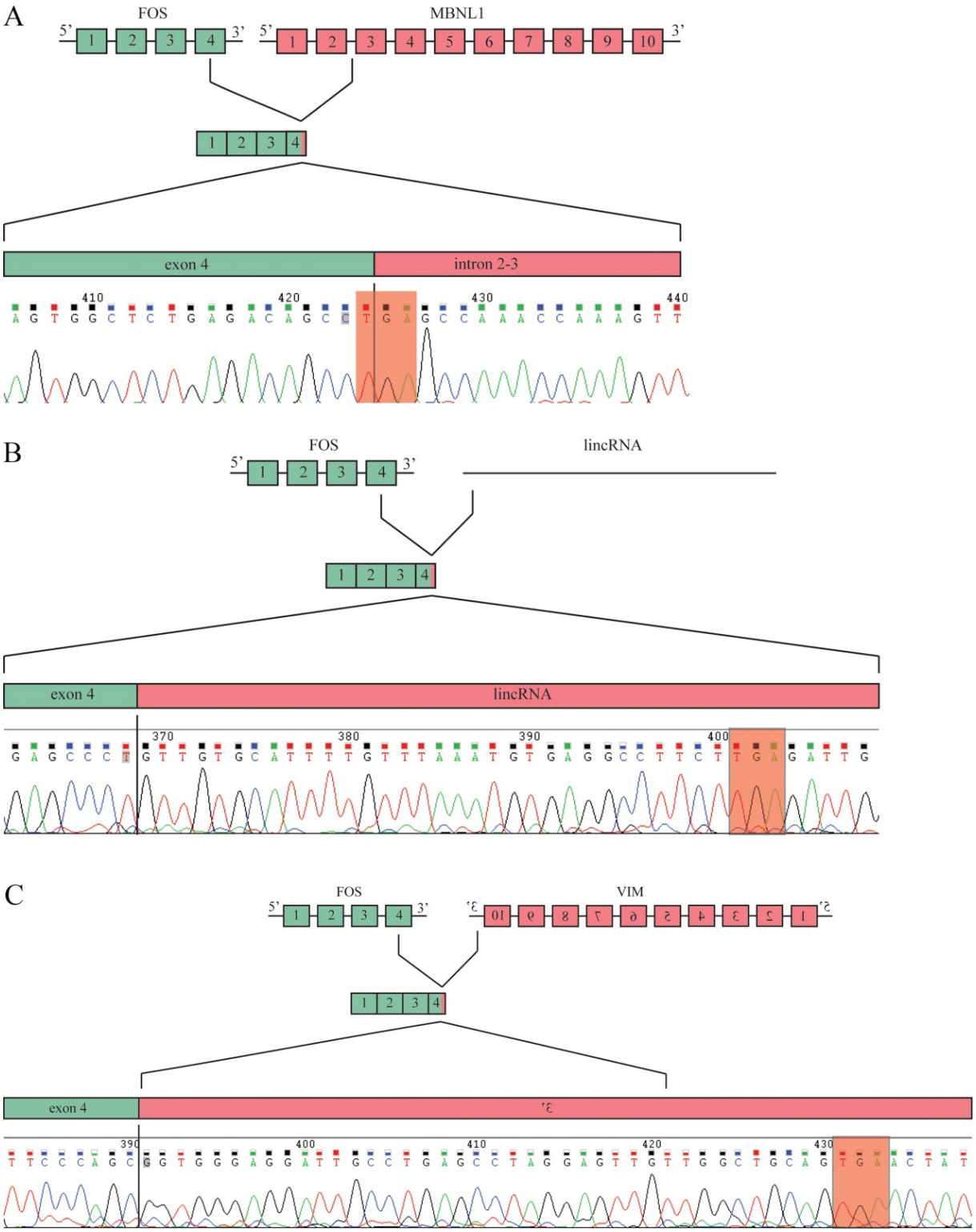


Figure 3.3: (a) The fusion product between *FOS* and *MBNL1* in case L3933. *FOS* gains a stop codon directly at the site of the translocation truncating the gene. (b) The fusion between *FOS* and *lincRNA* (RP11-326N17.1) in case L4065 resulting in a stop codon. (c) The *FOS-VIM* fusion in case L3141. *FOS* fuses to the reverse strand of *VIM* but gains a stop codon. Stop codons are highlighted in red.

3.4.4 A recurrent FOS break is present in 5/7 samples

As no recurrent translocations were detected using RT-PCR, interphase FISH analysis was performed to identify *FOS* rearrangement in the other cases. Five out of seven cases had a break-apart of *FOS* of which four out of six showed a scorable break-apart signal in the *FOS* locus (figure 3.4). Five cases could not be scored, including case L4065 that was found to have a fusion by transcriptome sequencing and RT-PCR.

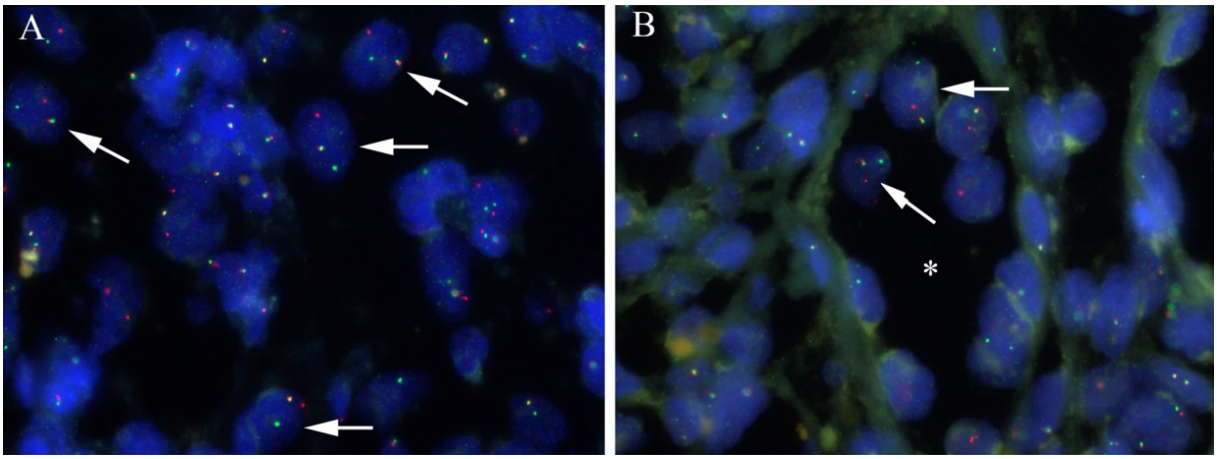


Figure 3.4: *FOS* break events. (a) Case L3142 showing four cells with a split signal (arrows). (b) Case L3154 shows two cells with a split signal (arrows). Both cases show a split and a colocalized signal in cells lining the vascular channel (star). The surrounding cells show only colocalized signals.

3.5 Discussion

While usually translocations in mesenchymal tumors lead to tumorigenesis by the creation of a chimeric transcription factor (deregulating transcription), or lead to upregulation of expression of a specific (onco) gene by promotor swap (deregulating transcription control) (19), we here report an uncommon mechanism in which a balanced translocation leads to truncation of a specific protein. We observed translocations involving the *FOS* gene with various partners in three epithelioid hemangiomas using transcriptome sequencing. Involvement of the *FOS* gene in bone and soft tissue epithelioid hemangioma was also very recently presented by Antonescu et al in an abstract (20). Intriguingly, based on the sequence it can be predicted that none of the fusion genes lead to an extended translation and formation of a chimeric protein. Instead, the translocation leads to a transcribed chimeric RNA with an early stop codon that may result in truncation of the FOS protein at translation.

Involvement of *FOS* by fusion events seems to be a highly specific and driving event in

epithelioid hemangioma. Complex genomic rearrangements were not detected in our index case using COBRA FISH based karyotyping, indicating that the balanced translocation might be a sole event. Also, transcriptome analysis of the three cases revealed no other recurrent mutations. Moreover, we observed *FOS* rearrangement by using split-apart FISH probe sets in two additional cases. Thus, in total, *FOS* rearrangement was found in five out of seven epithelioid hemangiomas of bone. Interestingly, none of the detected fusion events were found to be recurrent, indicating that in these two additional cases, yet other, novel fusion partners are involved. Since the introduction of a stop codon could also occur due to point mutations or small in/dels we sequenced exon 4 of *FOS* in 11 cases including the two cases without *FOS* rearrangements, without finding any mutations.

The diagnosis of vascular tumors of bone has been difficult, as there is considerable histomorphological overlap and immunohistochemical markers that can help in the distinction are lacking. The recent elucidation of *WWTR1-CAMTA1* and *YAP1-TFE3* fusions in epithelioid hemangioendothelioma has provided novel molecular tools that can help in the classification (14, 21, 22). Recently, a new translocation was identified involving a fusion between *ZFP36* and *FOSB* in a subset of epithelioid hemangiomas with atypical histological features (15). Also, for pseudomyogenic hemangioendothelioma, a recently described rare vascular tumor that also can occur in bone, a *SERPINE1-FOSB* fusion was described leading to upregulation of *FOSB* (23).

Approximately 18-25% of epithelioid hemangiomas demonstrate multifocal regional spread (7, 14), which was also seen in six of our cases. In the two cases for which we elucidated the exact fusion product using transcriptome sequencing, we could demonstrate using RT-PCR that in separate tumors, affecting different bones, an identical fusion product was present. Similar to our observations, in multifocal epithelioid hemangioendothelioma of the liver, identical *WWTR1-CAMTA1* fusions could be identified in the different foci (12). Our results indicate that also in epithelioid hemangioma the separate tumours affecting multiple bones as well as soft tissue are of monoclonal origin. Moreover, our FISH results excluded that the translocation occurs in the germline. Therefore, we may conclude that germline predisposition or a "field-effect" predisposing to the development of multiple independent neoplastic lesions in multiple adjacent bones is highly unlikely. Instead of the generally used "metastasis", we prefer the term "multifocal regional spread" as the multiple foci represent local spread of a single neoplastic clone to adjacent bones and soft tissue.

The predicted loss of approximately 100 amino acids at the C terminal region of the FOS protein results in the loss of the TAD. With primers covering the exon 4 of *FOS* we performed RT-PCR showing a substantial expression of the wild-type *FOS* (with average Ct value of normal *FOS* = 30, for fusions = 28, data not shown) in all cases,

indicating that the translocation does not lead to strong decreased expression of the intact allele. The in silico predicted, truncated FOS protein shows a high structural similarity to FOSL1, which also lacks the TAD (figure 3.5). FOS and FOSB are part of the FOS family of proteins, which also includes FOSL1 (also known as FRA-1) and FOSL2 (also now as FRA-2). All the members of the FOS family are involved in forming the Activator Protein-1 (AP-1) protein. The canonical AP-1 complex consists of FOS and JUN protein heterodimer, but other variants involving members of the FOS and JUN family of transcription factors have been described to be involved in the heterodimer as well, because of their structural similarities (24). FOSL1 knockdown in endothelial cells resulted in a decrease in angiogenesis through upregulation of the aVβ3 receptor, which is an important regulator of angiogenesis (25). It was shown that FOSL1 correlated with upregulation of adhesion proteins such as CD44 and integrin α5 resulting in an increase in adhesion (26). This is in contrast to endothelial cells where FOSL1 downregulates integrins, illustrating the different tissue-specific functions of FOSL1.

In models where *FOS* was knocked down, it was shown that many of the functions of FOS were taken over by FOSL1, although for some target genes FOSL1 fails to induce expression, indicating FOSL1 is not a full replacement for FOS (27). Furthermore, the activation pattern of FOSL1 and FOS is different. It was shown that by stimulation of fibroblasts by serum FOS and FOSB are rapidly and transiently induced whereas FOSL1 and FOSL2 are expressed in a more delayed and stable pattern (28, 29). In ER-positive breast cancer, patients with high FOSL1 expression showed significantly shorter survival and higher rates of lung metastasis. Furthermore, it was shown that fibroblastoid cells, which did not express FOSL1, transfection with *FOSL1* led to a strong enhancement of mobility in the cell line (30). It is tempting to speculate that these functions of FOSL1 may underlie the multifocal regional spread of epithelioid hemangioma of bone.

In conclusion, we identified rearrangement of FOS with various non-recurrent fusion partners in five out of seven classic epithelioid hemangiomas of bone. Detection of FOS rearrangement may therefore be a useful diagnostic tool to assist in the often-difficult differential diagnosis of vascular tumors of bone. In silico prediction shows that the translocations cause truncation of the FOS protein, with loss of the TAD, which is thereby a novel mechanism involved in tumorigenesis, which requires further studies.

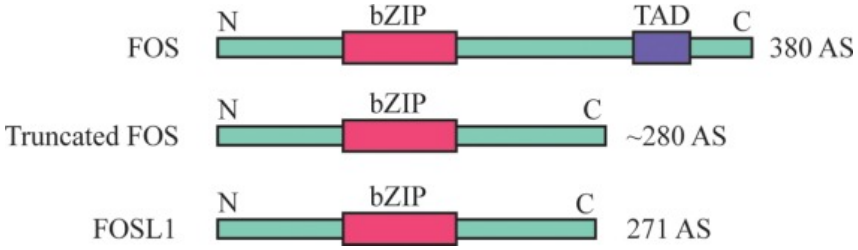


Figure 3.5: The fusions are predicted to result in the loss of 100 amino acids from FOS. This results in the loss of the TAD. The truncated FOS shows a large similarity in structure with FOSL1, which is also part of the FOS family of proteins.

Bibliography

[1] Rosenberg Ae BJ. Epithelioid haemangioma. In: Fletcher Cd BJAHPCWMF, editor. WHO Classification of Tumours of Soft Tissue and Bone. Lyon: IARC; 2013. p. 333–4.

[2] Nielsen GP, Srivastava A, Kattapuram S, Deshpande V, O'Connell JX, Mangham CD, et al. Epithelioid hemangioma of bone revisited: A study of 50 cases. American Journal of Surgical Pathology. 2009;33(2):270–277. doi:10.1097/PAS.0b013e31817f6d51.

[3] Fletcher CDM, Bridge JA, Hogendoorn P, Mertens F. Vascular tumours. Lyon: IARC; 2013.

[4] Errani C, Zhang L, Panicek DM, Healey JH, Antonescu CR. Epithelioid hemangioma of bone and soft tissue: A reappraisal of a controversial entity. Clinical Orthopaedics and Related Research. 2012;470(5):1498–1506. doi:10.1007/s11999-011-2070-0.

[5] Verbeke SLJ, Bovée JVMG. Primary vascular tumors of bone: A spectrum of entities? International Journal of Clinical and Experimental Pathology. 2011;4(6):541–551.

[6] Unni KK, Ivins JC, Beabout JW, Dahlin DC. Hemangioma, hemangiopericytoma, and hemangioendothelioma (angiosarcoma) of bone. Cancer. 1971;27(6):1403–1414. doi:10.1002/1097-0142(197106)27:6<1403::AID-CNCR2820270621>3.0.CO;2-6.

[7] Campanacci M, Boriani S, Giunti A. Hemangioendothelioma of bone: A study of 29 cases. Cancer. 1980;46(4):804–814. doi:10.1002/1097-0142(19800815)46:4<804::AID-CNCR2820460427>3.0.CO;2-1.

[8] Dorfman H, Czerniak B. Vascular lesions. Bone Tumors. 1998; p. 1:729–814.

[9] Wenger DE, Wold LE. Benign vascular lesions of bone: radiologic and pathologic features. Skeletal Radiol. 2000;29(2):63–74.

[10] Evans HL, Raymond AK, Ayala AG. Vascular tumors of bone: A study of 17 cases other than ordinary hemangioma, with an evaluation of the relationship of hemangioendothelioma of bone to epithelioid hemangioma, epithelioid hemangioendothelioma, and high-grade angiosarcoma. Human Pathology. 2003;34(7):680–689. doi:10.1016/S0046-8177(03)00249-1.

[11] Bruder E, Perez-Atayde AR, Jundt G, Alomari AI, Rischewski J, Fishman SJ, et al. Vascular lesions of bone in children, adolescents, and young adults. A clinicopathologic reappraisal and application of the ISSVA classification. Virchows Archiv. 2009;454(2):161–179. doi:10.1007/s00428-008-0709-3.

[12] Errani C, Zhang L, Sung YS, Hajdu M, Singer S, Maki RG, et al. A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. Genes Chromosomes and Cancer. 2011;50(8):644–653. doi:10.1002/gcc.20886.

[13] Tanas MR, Ma S, Jadaan FO, Ng CK, Weigelt B, Reis-Filho JS, et al. Mechanism of action of a WWTR1(TAZ)-CAMTA1 fusion oncoprotein. Oncogene. 2016;35(7):929–938. doi:10.1038/onc.2015.148.

[14] Antonescu CR, Le Loarer F, Mosquera JM, Sboner A, Zhang L, Chen CL, et al. Novel YAP1-TFE3 fusion defines a distinct subset of epithelioid hemangioendothelioma. Genes Chromosomes and Cancer. 2013;52(8):775–784. doi:10.1002/gcc.22073.

[15] Antonescu CR, Chen HW, Zhang L, Sung YS, Panicek D, Agaram NP, et al. ZFP36-FOSB fusion defines a subset of epithelioid hemangioma with atypical features. Genes Chromosomes and Cancer. 2014;53(11):951–959. doi:10.1002/gcc.22206.

[16] Szuhai K, Tanke HJ. COBRA: Combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping. Nature Protocols. 2006;1(1):264–275. doi:10.1038/nprot.2006.41.

[17] Szuhai K, Bezrookove V, Wiegant J, Vrolijk J, Dirks RW, Rosenberg C, et al. Simultaneous molecular karyotyping and mapping of viral DNA integration sites by 25-color COBRA-FISH. Genes Chromosomes and Cancer. 2000;28(1):92–97. doi:10.1002/(SICI)1098-2264(200005)28:1<92::AID-GCC11>3.0.CO;2-2.

[18] Rossi S, Szuhai K, Ijszenga M, Tanke HJ, Zanatta L, Sciot R, et al. EWSR1-CREB1 and EWSR1-ATF1 fusion genes in angiomatoid fibrous histiocytoma. Clinical Cancer Research. 2007;13(24):7322–7328. doi:10.1158/1078-0432.CCR-07-1744.

[19] Roukos V, Misteli T. The biogenesis of chromosome translocations. Nature Cell Biology. 2014;16(4):293–300. doi:10.1038/ncb2941.

[20] Huang S, Chen H, Zhang L, Sung Y, Dickson B, Krausz T, et al. Frequent FOS Gene Rearrangements in Epithelioid Hemangioma. Lab Invest. 2015;95:18A–19A.

[21] Tanas MR, Sboner A, Oliveira AM, Erickson-Johnson MR, Hespelt J, Hanwright PJ, et al. Identification of a disease-defining gene fusion in epithelioid hemangioendothelioma. Science Translational Medicine. 2011;3(98):98ra82–98ra82. doi:10.1126/scitranslmed.3002409.

[22] Errani C, Sung YS, Zhang L, Healey JH, Antonescu CR. Monoclonality of multifocal epithelioid hemangioendothelioma of the liver by analysis of WWTR1-CAMTA1 breakpoints. Cancer Genetics. 2012;205(1-2):12–17. doi:10.1016/j.cancergen.2011.10.008.

[23] Walther C, Tayebwa J, Lilljebjörn H, Magnusson L, Nilsson J, Von Steyern FV, et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. Journal of Pathology. 2014;232(5):534–540. doi:10.1002/path.4322.

[24] Hess J. AP-1 subunits: quarrel and harmony among siblings. Journal of Cell Science. 2004;117(25):5965–5973. doi:10.1242/jcs.01589.

[25] Galvagni F, Orlandini M, Oliviero S. Role of the AP-1 transcription factor FOSL1 in endothelial cell adhesion and migration. Cell Adhesion and Migration. 2013;7(5):408–411. doi:10.4161/cam.25894.

[26] Oliveira-Ferrer L, K??rschner M, Labitzky V, Wicklein D, M??ller V, L??ers G, et al. Prognostic impact of transcription factor Fra-1 in ER-positive breast cancer: contribution to a metastatic phenotype through modulation of tumor cell adhesive properties. Journal of Cancer Research and Clinical Oncology. 2015;141(10):1715–1726. doi:10.1007/s00432-015-1925-2.

[27] Fleischmann A, Hafezi F, Elliott C, Remé CE, Rüther U, Wagner EF. Fra-1 replaces c-Fos-dependent functions in mice. Genes and Development. 2000;14(21):2695–2700. doi:10.1101/gad.187900.

[28] Milde-Langosch K. The Fos family of transcription factors and their role in tumourigenesis. European Journal of Cancer. 2005;41(16):2449–2461. doi:10.1016/j.ejca.2005.08.008.

[29] Chalmers CJ, Gilley R, March HN, Balmanno K, Cook SJ. The duration of ERK1/2 activity determines the activation of c-Fos and Fra-1 and the composition and quantitative transcriptional output of AP-1. Cellular Signalling. 2007;19(4):695–704. doi:10.1016/j.cellsig.2006.09.001.

[30] Tkach V, Tulchinsky E, Lukanidin E, Vinson C, Bock E, Berezin V. Role of the Fos family members, c-Fos, Fra-1 and Fra-2, in the regulation of cell motility. Oncogene. 2003;22(32):5045–5054. doi:10.1038/sj.onc.1206570.

Chapter 4

Telatinib is an effective targeted therapy for pseudomyogenic hemangioendothelioma

This chapter is based on the publication: **van IJzendoorn DGP**, Sleijfer S, Gelderblom H, Eskens FALM, van Leenders GJLH, Szuhai K, Bovée JVMG. Telatinib Is an Effective Targeted Therapy for Pseudomyogenic Hemangioendothelioma. *Clin Cancer Res.* 2018;24: 2678-2687.

4.1 Abstract

Pseudomyogenic hemangioendothelioma (PHE) is an extremely rare locally aggressive neoplasm with endothelial differentiation, which often presents with multiple lesions. These tumors have characteristic *SERPINE1-FOSB* fusions. We report a 17 years old patient with advanced unresectable PHE with a durable complete remission to the multi-tyrosine kinase inhibitor telatinib. The aim of this study was to generate an in vitro model for PHE, to study the functional consequences of *SERPINE1-FOSB* in endothelial cells, and its interaction with telatinib, to biologically substantiate the complete response to telatinib.

As the fusion results in overexpression of a truncated form of *FOSB*, we overexpressed truncated FOSB in normal endothelial cells.

Truncated FOSB significantly affected tumor growth in three-dimensional (3D) on matrigel with increased and sustained sprouting. Moreover, truncated FOSB acted as an active transcription factor capable to regulate its own transcription, as well as to upregulate *PDGFRA* and *FLT1* expression (four-fold). Telatinib decreased proliferation and tumor growth in 3D and induced apoptosis. As expected, telatinib blocked VEGF signaling as phosphorylation of ERK was abolished. Interestingly, in FOSB overexpressing cells, telatinib specifically affected PDGFRA, FLT1, and FLT4 signaling and downregulated SERPINE1, thereby affecting the self-regulation of the fusion gene.

We provide a biological substantiation of a complete clinical remission that was seen in a patient with PHE, showing that telatinib indirectly interferes with the self-regulated expression of the fusion product. Thus, telatinib or any other currently available VEGFR1-4/PDGFRA inhibitor could be a highly specific treatment option for patients with multifocal unresectable PHE.

4.2 Introduction

Pseudomyogenic hemangioendothelioma (PHE) is a locally aggressive and rarely metastasizing tumor, predominantly affecting young adults. It is an extremely rare entity that occurs more frequently in males than in females (41 vs. 9). Most PHE patients present with multifocal disease (33 of 50 patients) with multiple discontiguous nodules present in different tissue planes (1, 2). This multicentric appearance combined with its locally aggressive behavior can make PHE difficult to treat.

PHE was first described as "epithelioid sarcoma-like hemangioendothelioma" in 2003, which was based on the presence of large cells with abundant eosinophilic cytoplasm at microscopy with keratin positivity (3). In 2011, based on a series of 50 patients, the proposed terminology was changed into PHE based on the myogenic appearance combined

with the evidence of endothelial differentiation by IHC, as well as the lack of a relation with epithelioid sarcoma (2).

Characteristic for PHE is a translocation between chromosomes 7 and 19. This translocation was later found to involve *SERPINE1* and *FOSB*. The translocation leads to a swap of the 5'-UTR regions and a loss of the first 48 amino acids of the FOSB protein (4, 5). FOSB is a member of the Fos family of proteins.

Fos family members can bind to Jun proteins thereby forming heterodimers, which make up the Activator Protein 1 (AP-1) transcription factor. By regulating the function of several genes such as FLI1 and FAS, this transcription factor is involved in many cellular functions including proliferation, differentiation, and transformation (6, 7). Among vascular tumors there is often involvement of the Fos family members. In epithelioid hemangioma, another locally aggressive vascular tumor, FOS and FOSB were also described to be involved in a translocation (8–10).

Based on the observation of a durable complete response following prolonged exposure to telatinib, an orally available tyrosine kinase inhibitor targeting VEGFR, PDGFR, and KIT, in a 17 years old male who presented with advanced, unresectable PHE, we here aimed to elucidate the underlying molecular mechanism of response to telatinib in PHE. As there are no PHE cell lines available we first needed to establish a model for functional analysis of PHE. We opted to use normal endothelial cells, human umbilical vein endothelial cells (HUVEC), in which we overexpressed the truncated FOSB protein, that is, lacking the first 48 amino acids of the FOSB protein that are lost in PHE due to the *SERPINE1-FOSB* fusion. We subsequently use this model to investigate the effect of telatinib on PHE. Based on the complete response we observed in the patient, we hypothesized that telatinib has a direct effect on the function of the *SERPINE1-FOSB* fusion product, the tumor-driving event in this particular tumor type.

4.3 Patient and Methods

4.3.1 Patient

A 17 years old male, with no prior history of disease, presented with multiple skin lesions on the head and neck. Excisional biopsy of one of the lesions revealed a cellular proliferation of spindled and epithelioid cells, with abundant eosinophilic cytoplasm, that occupied the dermis and extended into the subcutis (figure 4.1a). The nuclei were round to oval with one or more prominent nucleoli. IHC revealed that the tumor cells were positive for CD31, vimentin, and keratin AE1AE3, whereas CD34, S100, CD68, HHV8, podoplanin (D2-40), melan A, CD56, CD57, EMA, smooth muscle actin, KL1, CAM5.2, CD45, keratin 20, CD30, actin (HHF35), and desmin were negative. The differential diagnosis

included epithelioid sarcoma and epithelioid sarcoma-like hemangioendothelioma/PHE, an entity that was just described at that time (2, 3). Based on the strong expression of CD31 the diagnosis of PHE was favored. As the patient also had enlarged lymph nodes in the cervical area, a fine needle aspiration was performed and tumor involvement of locoregional lymph nodes in the neck was confirmed. CT of the chest and abdomen showed no distant metastasis. Initially, the patient was treated with six rounds of docetaxel, which was chosen given the then recently described activity of taxanes in vascular sarcomas of the scalp (11). Docetaxel yielded a partial, although short-lasting response, followed by rapid progression of the disease in the head and neck area after three months. The disease was too extensive for an operation (figure 4.1e, left). An excisional biopsy of one of the lesions confirmed the lack of response to chemotherapy. Two months later, the patient was included in a multicentre phase I dose escalation study for telatinib (BAY 57-9352), an orally available, small-molecule multi-tyrosine kinase inhibitor (12), which was open for patients with advanced or metastatic solid tumors and for whom no standard therapy is available. The patient has given written informed consent for the use of material, and publication of clinical data including photographs.

4.3.2 Immunohistochemistry

FOSB and KIT IHC was performed using 4-μm-thick tissue sections. Paraffin was removed with xylene and sections were rehydrated in a gradient of ethanol. Endogenous peroxidase was blocked using 0.3% H_2O_2. Microwave antigen retrieval was performed in Tris-EDTA (pH 9.0). FOSB antibody (Cell Signaling) or KIT antibody (Dako) was incubated with the cells overnight at 4°C. Rabbit secondary antibody was used and detected with DAB (3,3'-diaminobenzidine). Counterstaining was performed with hematoxylin.

4.3.3 Fluorescence in situ hybridization

FISH has previously been described by our group (8). BAC probes were selected surrounding *FOSB* (CTB-14D10 and RP11-84C16) and *SERPINE1* (RP11-44M6 and RP5-1059M17). Four-micrometer-thick tissue sections were cut from the paraffin embedded excisional biopsy before chemotherapy. BAC probes were tested on metaphase control slides.

4.3.4 Polymerase chain reaction

RNA was isolated from fresh frozen tissue and cultured cells using the Direct-zol RNA Isolation Kit (Zymo research). cDNA was made using the iScript cDNA Synthesis Kit (Bio-Rad). Real-time PCR was performed with SybrGreen (Bio-Rad) on a Bio-Rad CFX384

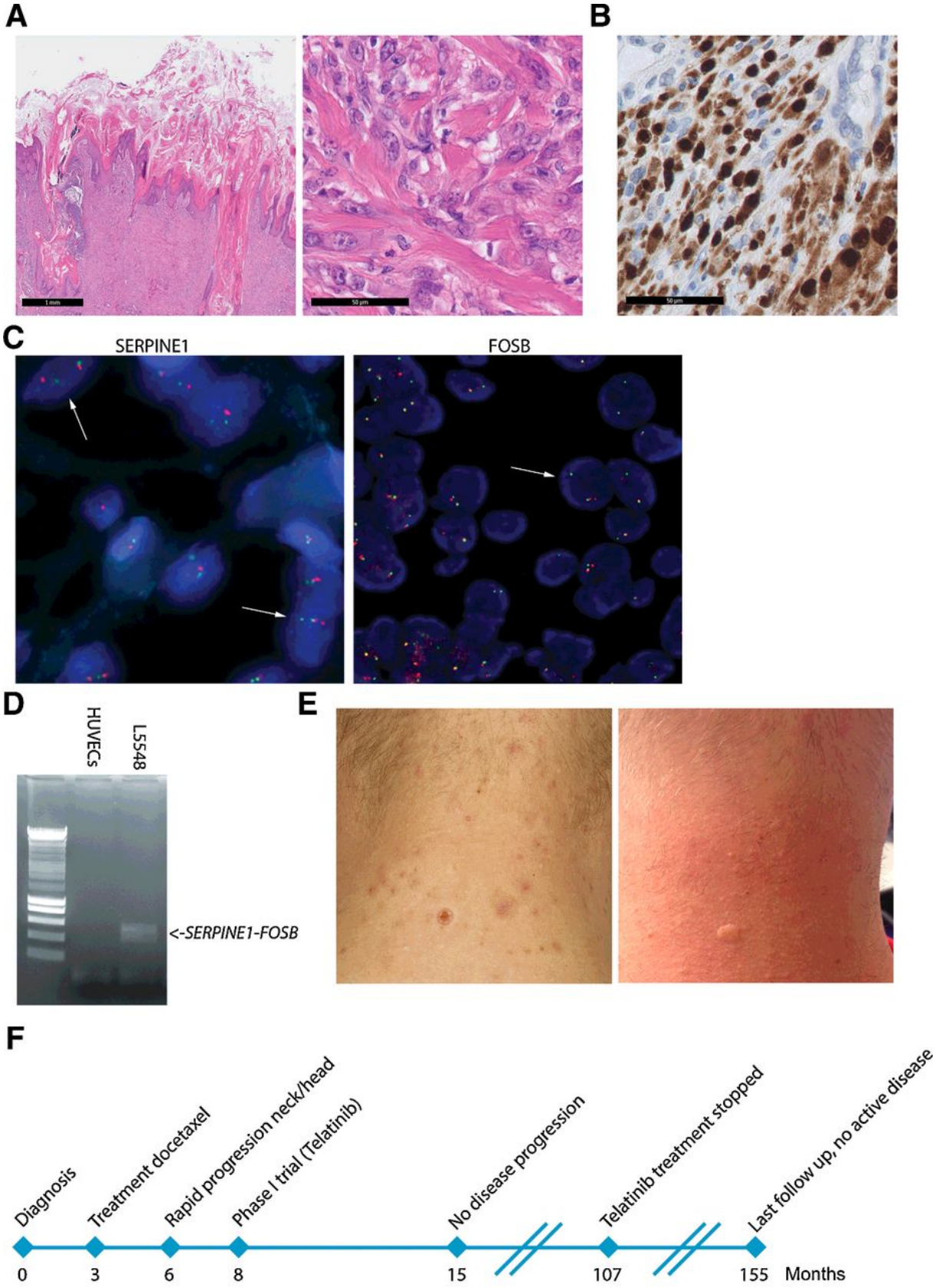


Figure 4.1: Caption on next page.

Figure 4.1: Diagnosis of PHE was confirmed by detection of the *SERPINE1-FOSB* gene fusion in a tumor from a 17 years old patient with advanced inoperable PHE who was treated with telatinib and showed a complete remission. (a) Left panel shows a low power H&E image of the lesion showing that the lesion infiltrates the epidermis. The scale bar indicates 1 mm. Right panel shows a high power H&E showing tumor cells with a rhabdomyoblast-like appearance. The scale bar indicates 50 μm. (b) IHC confirms high expression of FOSB in the nuclei of the tumor cells. The scale bar indicates 50 μm. (c) Using FISH, the involvement of *SERPINE1* locus (left) and *FOSB* locus (right) was confirmed using split apart FISH probes (arrows indicate cells with split apart signals). (d) PCR on cDNA made from RNA of frozen patient tumor tissue (L5548) shows a distinct band indicating expression on RNA level of the fusion gene at the expected size. (e) Left panel shows a picture of the skin lesion as seen at the time of diagnosis. Multiple nodules with hyperkerastosis and central ulcerations were seen. Right panel is the same area 156 months after first diagnosis showing complete remission of the lesions. The picture shows that the hyperkeratotic lesions have disappeared. (f) Timeline shows the course of the disease, from the initial diagnosis to rapid progression while treated with docetaxel to the complete remission nine years after starting telatinib treatment.

Touch (Bio-Rad). For regular PCR Phusion proofreading enzyme was used (NEB). To visualize amplified DNA fragments were size separated on a 1% agarose gel. Used primers are listed (Supplementary Table S1 available online). All real-time PCR experiments were performed in triplicates.

4.3.5 Cell culture

Primary pooled HUVECs (Lonza) were cultured in EGM-2 medium (Lonza) on 0.2% gelatin coated culture dishes. Cells were serum starved with M199 medium (Gibco) for 6 hours before introducing EGM-2 or M199 medium supplemented with VEGF. Cells were checked for mycoplasma with the MycoAlert Kit (Lonza).

4.3.6 Plasmid and shRNA

Human *FOSB* and truncated *FOSB* were cloned in frame with a Flag tag into pLV lentiviral expression vector. ShRNA was selected from the Sigma Mission shRNA library. Knockdown efficiency of the selected shRNAs was tested with real-time qPCR (Supplementary Table S2 available online).

4.3.7 HUVEC tube formation assay, proliferation assay, and analysis

Tube formation assays were performed on 96-well plates coated with 60 μL matrigel (Lonza). Cells were seeded at a density of 20,000 cells per well in 200 μL EGM-2 medium. To induce tube formation, 50 ng/mL VEGF was added. Tube formation was analyzed with Stacks (in house developed software tool, Molecular Cell Biology, LUMC). To measure relative proliferation 11 μL PrestoBlue Cell Viability reagent (Thermo Fisher Scientific) was added to each well after which the cells were incubated for 30 minutes. Measurements were performed with the VICTOR Multilabel Plate Reader (PerkinElmer). As vehicle control, DMSO was added to the control cells at corresponding concentrations to exclude effects from the DMSO used to dissolve telatinib. All tube formation assays were performed in triplicates.

4.3.8 Chromatin immunoprecipitation

In short, chromatin immunoprecipitation (ChIP) was performed by crosslinking protein and DNA with a final concentration of 1% formaldehyde. Crosslinking was stopped with 1.25 M glycine. Chromatin was sonicated 15 minutes using a Bioruptor at 30 second intervals to generate 500 to 1,000 bp DNA fragments. Immunoprecipitation was performed with ProtA beads and monoclonal FOSB rabbit antibody (Cell Signaling). Thereafter DNA was eluted and quantitative real-time PCR was performed. Primers were designed to detect three different AP-1 binding sites near the *SERPINE1* promoter (Supplementary Table S3 available online). Targets for ChIP were identified with the DECODE database (SABioscience).

4.3.9 Antibodies, growth factors, and drugs

Western blotting and IHC have previously been described by our group (13). Antibodies were obtained from the following sources: Flag monoclonal rabbit antibody (F7425; Sigma); FOS monoclonal rabbit antibody (HPA018531; Protein Atlas); FOSB monoclonal rabbit antibody (#2251; Cell Signaling); phosphorylated ERK mouse monoclonal antibody (M8159; Sigma); JUN monoclonal rabbit (#9165; Cell Signaling); USP7 monoclonal rabbit (A300-033A; Bethyl); KIT polyclonal rabbit (A4502; Dako). Tube formation was stimulated with 50 ng/mL VEGF 165 (R&D systems). Cycloheximide was used to block translation at a concentration of 50 μg/mL (Sigma). Telatinib (Selleckchem) dissolved in DMSO was used at different concentrations.

4.3.10 Apoptosis and cell-cycle analysis

The Nucleocounter NC-250 (Chemometec) was used for apoptosis and cell-cycle analysis. For apoptosis, the cells were resuspended in PBS and stained with propidium iodide (PI) and VitaBright-48 (VB-48). For cell-cycle analysis the cells were fixed, permeabilized, and thereafter stained with DAPI. Analysis was performed with FlowJo (FlowJo, v10). Apoptotic cells were selected by gating for the PI and VB-48 negative cells.

4.4 Results

4.4.1 Detection of the *SERPINE1-FOSB* fusion confirms the diagnosis of PHE

We used IHC for FOSB, which was recently reported as a very specific marker for PHE (14), on the initial excisional biopsy and confirmed overexpression of FOSB in the nuclei of the tumor cells (figure 4.1b). Using FISH with split apart probes flanking *SERPINE1* and *FOSB* genes, we confirmed that both genes were involved in the fusion (figure 4.1c). Expression of the *SERPINE1-FOSB* fusion transcript was confirmed using RT-PCR performed on RNA isolated from fresh frozen tissue from the excisional biopsy (figure 4.1d).

4.4.2 Complete response after long-term telatinib treatment in a patient with pseudomyogenic hemangioendothelioma

Six weeks after starting telatinib treatment, one of the larger skin lesions was shed from the skin and disease progression had stopped. The treatment seemed to have a gradual, cyclic effect, with regular shedding of the skin lesions. The lymph nodes were small and non-palpable and therefore initially followed by CT but remained stable over the years. Therefore, it was decided to use visual inspection of the skin lesions to monitor disease activity. The patient had no side effects of the telatinib other than headache and treatment was continued. Another 4 years later, all skin lesions had disappeared and the disease was considered in complete remission (figure 4.1e). Nine years after first diagnosis, telatinib treatment was stopped (as it was no longer available) and the patient is still tumor free 4 years later after treatment cessation (figure 4.1f).

4.4.3 Overexpression of truncated *FOSB* in HUVECs as a model to study PHE

We overexpressed truncated *FOSB* (*FOSB*49-338) in HUVECs to generate a model to study PHE (figure 4.2a). We grew the HUVECs on matrigel to evaluate their behavior in 3D. We found that the HUVECs overexpressing truncated *FOSB* not only showed more tube formation compared to control HUVECs (Supplementary figure S1 available online), but also retained tube formation for more than 48 hours, where the pLV control cell network collapsed after 24 hours, which is normal for this cell type (figure 4.2b). 2D cell culture also showed a small difference in proliferation between truncated *FOSB* and pLV control HUVECs after 48 hours (P = 0.04; figure 4.2c).

4.4.4 Truncated FOSB acts as an active transcription factor capable of self-regulation

Next, we studied if the fusion affects the function, lifetime, and localization of FOSB. With immunoprecipitation we show that truncated FOSB forms a complex with JUN (figure 4.3a), showing that similar to normal FOSB, truncated FOSB still forms an AP-1 complex. Immunofluorescence showed that truncated FOSB, similar to wild-type FOSB, localizes to the nucleus (figure 4.3b). To see if the loss of the first 48 amino acids of FOSB affected its stability we treated the cells with cycloheximide, effectively blocking all translation. After blocking translation both the truncated and full length FOSB showed to be highly stable (figure 4.3c). To evaluate differences in downstream signaling, we evaluated expression of previously identified AP-1 targets as downstream markers of upregulated AP-1 signaling (13). With real-time qPCR we found that truncated *FOSB* upregulates *HEY1*, *JAG1*, *FAS* (respectively 3.18, 0.78, 0.98 log2 fold) and downregulates *VWF* and *ADAMTS13* (respectively -4.08, -1.53 log2 fold) compared with control HUVECs (figure 4.3d). To check the validity of our model we also investigated expression in the patient tumor tissue (which contains ∼60% tumor cells). *FOSB* was upregulated in both the patient biopsy and the HUVECs overexpressing truncated *FOSB* (respectively 8.68 and 13.99 log2 fold; Supplementary figure S2A available online). Similarly, in the tumor tissue from the patient, *HEY1*, *JAG1*, and *FAS* were upregulated (respectively 3.57, 1.51, and 3.31 log2 fold) and *VWF* was downregulated (-5.61 log2 fold) compared with normal HUVECs (Supplementary figure S2B available online). *ADAMTS13* was not detectable in the patient tumor biopsy. Thus, truncated *FOSB* functions as an active nuclear transcription factor in HUVECs regulating known AP-1 target genes. In the *SERPINE1* promoter we found three AP-1 consensus sites by searching in the DECODE (SABioscience) database (Chr7:100760761, Chr7:100761757, and Chr7:100766067 in hg18), indicating the AP-1

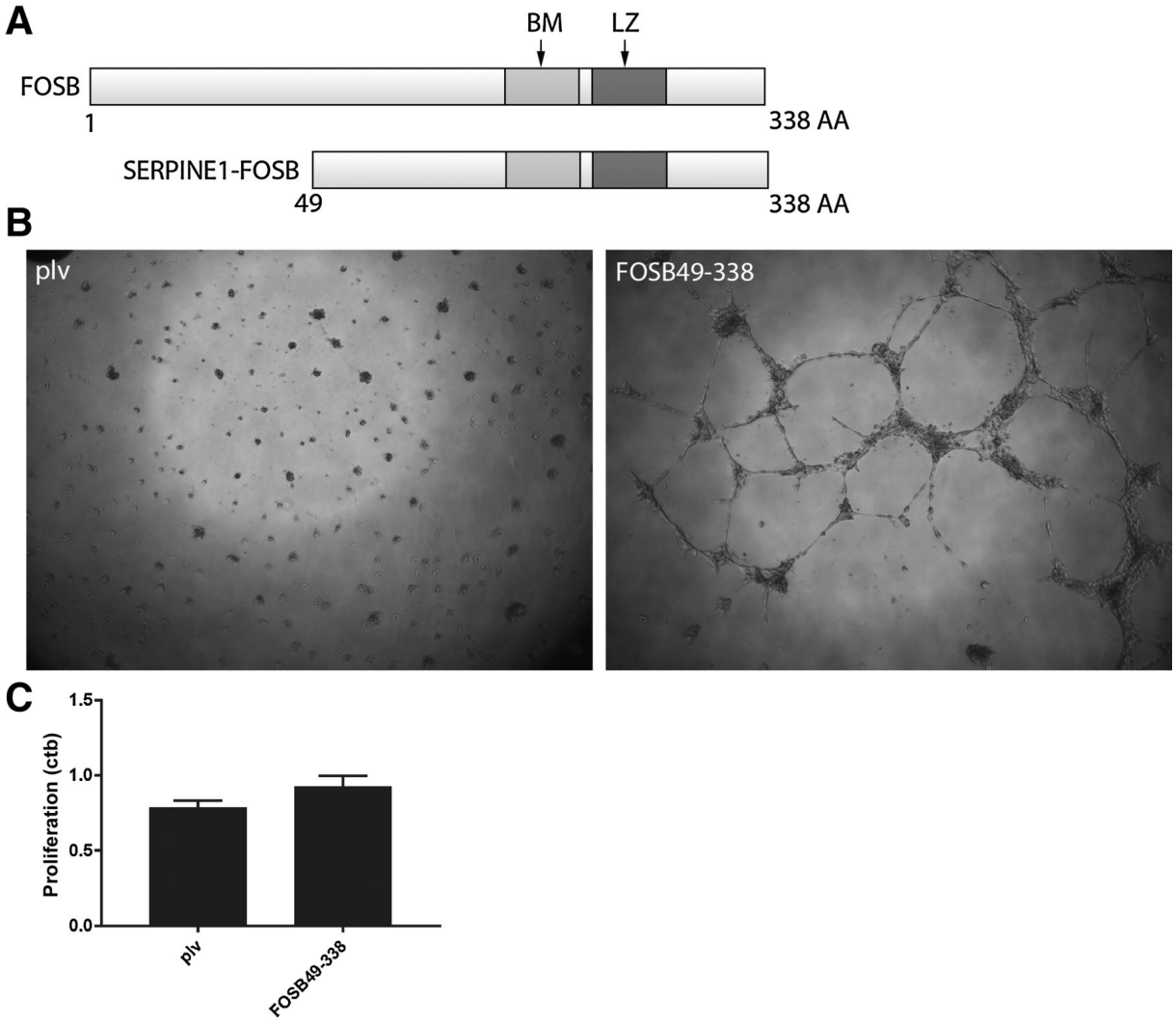


Figure 4.2: Truncated *FOSB* is representative of *SERPINE1-FOSB* and affects HUVEC growth in 2D and 3D with increased and sustained sprouting. (a) Schematic view of truncated FOSB protein. *SERPINE1* only retains its 5'-UTR whereas *FOSB* loses the coding region for the first 48 amino acids. The truncated FOSB retains the DNA Binding Motif (BM) and Leucine Zipper (LZ). (b) Tube formation after 48 hours. Control tube network (pLV) has collapsed while truncated FOSB overexpressing HUVECs retain the tube network showing that overexpression of truncated FOSB leads to increased and prolonged tube formation. (c) Presto Blue assay shows overexpression of truncated FOSB compared to pLV controls cells leads to an increase in proliferation (P = 0.0429).

transcription factor is an important regulator of *SERPINE1*. To investigate if truncated FOSB directly regulates *SERPINE1*, we first performed a chromatin pull-down with a FOSB antibody. With real-time qPCR for the AP-1 consensus sites in the *SERPINE1* promoter we detected an average enrichment of 2.18 log2 fold compared to the control cells (figure 4.3e). Furthermore, real-time qPCR shows an increase in *SERPINE1* expression (0.34 log2 fold) when truncated *FOSB* is expressed (figure 4.3f). Thus, it is likely that the *SERPINE1-FOSB* fusion is able to function as its own promoter and thereby regulates its own expression.

4.4.5 Telatinib blocks VEGFR downstream signaling and induces apoptosis in truncated FOSB overexpressing HUVECs

We assessed the response of HUVECs overexpressing truncated FOSB to the inhibitor. When growing HUVECs overexpressing truncated FOSB on matrigel, telatinib inhibits most tube formation at 5 μmol/L and completely abolishes tube formation at 10 μmol/L (figure 4.4a). To investigate if telatinib blocks signaling downstream of the VEGF receptors, we investigated the effect of telatinib on MAPK/ERK signaling. We found that telatinib reduces phosphorylation of KDR (VEGFR2) at 100 nmol/L and completely blocks phosphorylation at 2 μmol/L (figure 4.4b). Most phosphorylation of ERK is absent at 1 μmol/L and higher concentrations show no additional effect (figure 4.4c).

To see if telatinib leads to apoptosis (and not senescence or quiescence), we measured the cell cycle and the apoptotic fraction using an automatic fluorescence cell counter. We found no difference between the untreated and telatinib treated cells in the G2 fraction (from 8.97% in the untreated to 9.44% in the 10 μmol/L telatinib treated cells; supplementary figure S3 available online). The sub-G1 fraction however increases from 4.7% in the untreated to 10.7% in the 10 μmol/L telatinib treated cells. Furthermore, we found the fraction of apoptotic cells (negative for PI and VB-48) increases from 16.1% in the untreated to 26.5% in the 10 μmol/L telatinib treated cells (figure 4.4d).

4.4.6 Truncated FOSB sensitizes HUVECs to inhibition of the surface receptors FLT1 and PDGFRA by telatinib, leading to a reduced expression of the fusion product

Reportedly telatinib inhibits signaling through KIT, VEGFR, and PDGFR (15). To investigate which receptors are essential for growth of truncated FOSB overexpressing HUVECs in 3D, we selected shRNA targeting the aforementioned receptors. Normal HUVECs were sensitive to knockdown of *FLT4* (*VEGFR3*) and *KIT*. HUVECs overex-

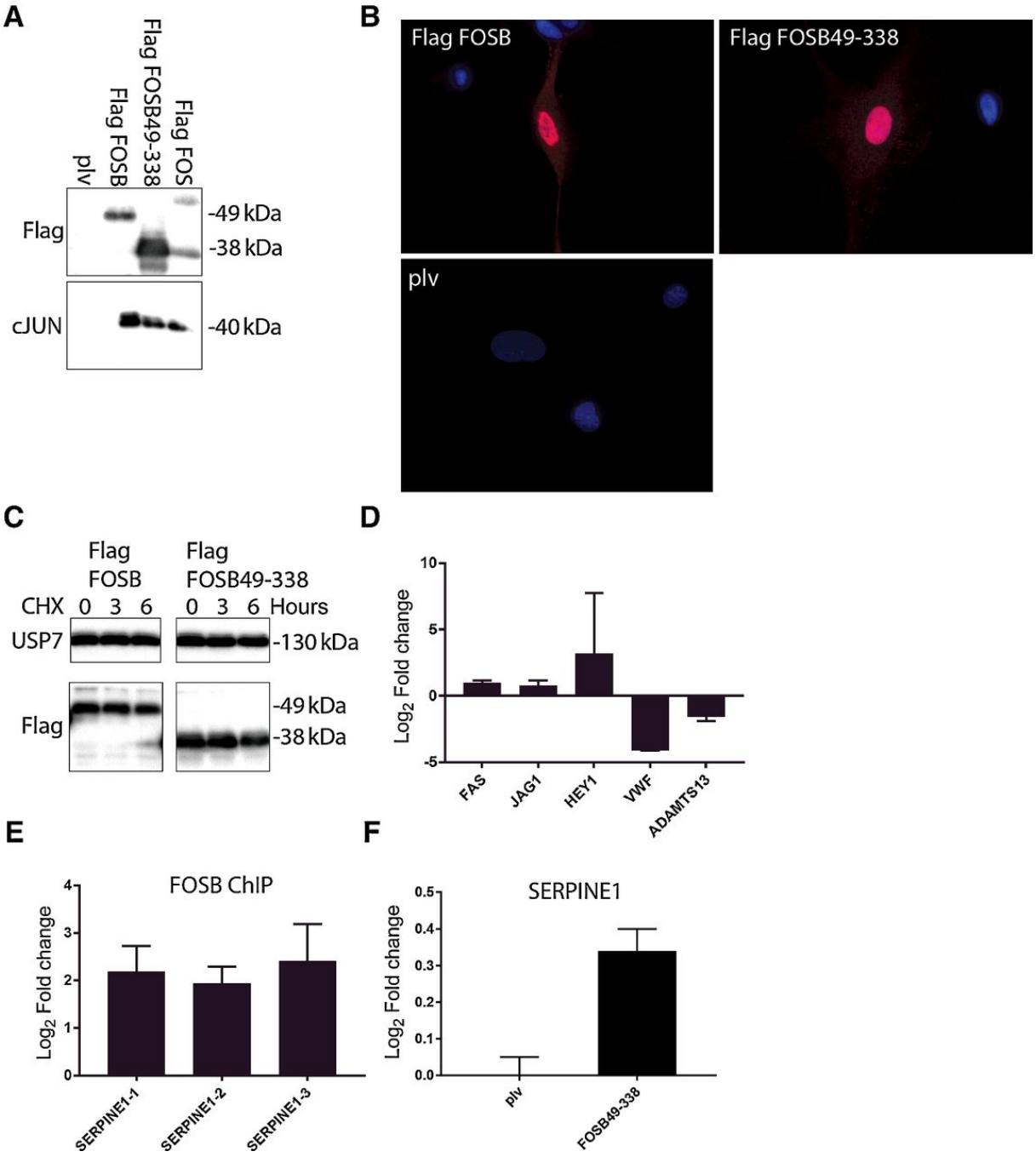


Figure 4.3: Caption on next page.

Figure 4.3: Truncated FOSB still functions as a transcription factor in the AP-1 complex, localizes to the nucleus, and can regulate its own expression by acting on the *SERPINE1* promoter. (a) FOSB and truncated FOSB form heterodimers with JUN as seen when immunoprecipitating with Flag beads and subsequent detection of JUN suggesting they can form an AP-1 transcription factor. Flag tagged FOS was used as a positive control. (b) Immunofluorescence using anti-Flag antibody shows both FOSB and truncated FOSB (red) localize to the nucleus (blue) whereas no signal is detected in the pLV control cells. (c) HUVECs overexpressing FOSB and truncated FOSB were treated with Cycloheximide to see if the loss of the first 48 amino acids influences the turnover of the protein. No difference in lifetime was detected between the variants showing that the truncation does not affect the half-life of the protein and that both forms of FOSB proteins are active for over 6 hours. (d) Expression levels of the indicated transcripts in HUVECs overexpressing truncated FOSB compared to normal control HUVECs were determined by real-time qPCR showing upregulation of *FAS*, *JAG1*, and *HEY1*. *VWF* and *ADAMTS13* are downregulated. Truncated FOSB is therefore still capable of regulating downstream target genes. Expression was normalized against *HPRT1*. (e) All three AP-1 consensus sites in *SERPINE1* promoter were retained in the *SERPINE1-FOSB* fusion. With qPCR after ChIP we showed an enrichment for the promoter binding site when truncated FOSB was overexpressed. Immunoprecipitation was performed with a FOSB antibody. This indicates that *SERPINE1* expression is under direct control of truncated FOSB. (f) Overexpression of truncated FOSB upregulates SERPINE1 compared to control HUVECs as found with real-time qPCR for *SERPINE1* transcript. Expression was normalized against *HPRT1*.

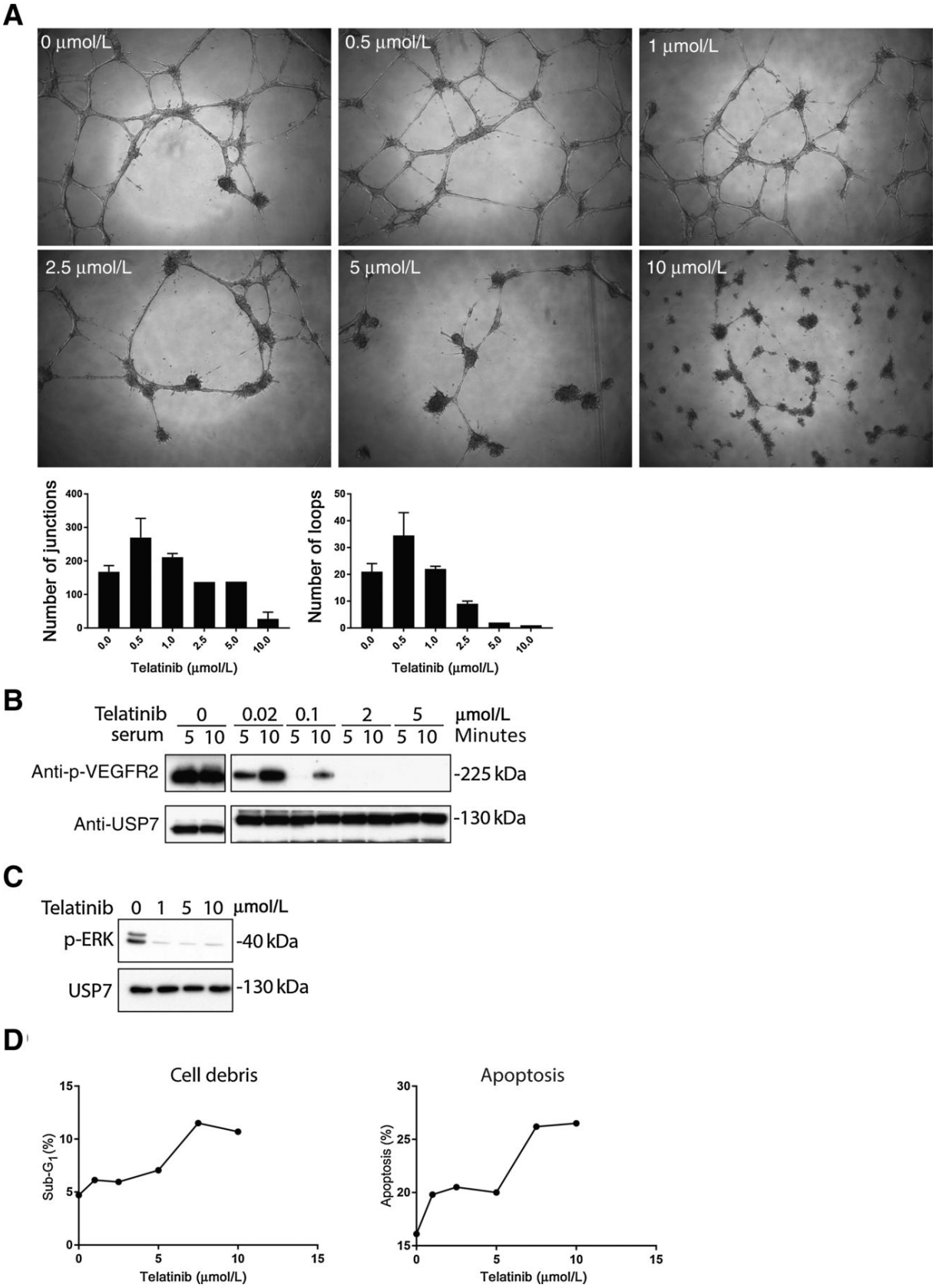


Figure 4.4: Caption on next page.

Figure 4.4: Telatinib inhibits proliferation and tube formation through blocking VEGF receptor and its downstream signaling. (a) 3D assay on matrigel with HUVECs overexpressing truncated FOSB. Most tube formation is blocked at 5 μmol/L telatinib, whereas no tube formation is present at 10 μmol/L. Bottom part of the panel shows the decrease in calculated number of loops and junctions of the tubes with increasing doses of telatinib. (b) Serum starved and stimulated HUVECs overexpressing truncated FOSB were harvested 5 and 10 minutes after addition of serum. The cells were treated with telatinib at the indicated concentrations for 6 hours prior to addition of serum. At 100 nmol/L most phosphorylated VEGFR2 (KDR) is absent, whereas at 2 μmol/L it is completely absent. (c) HUVECs overexpressing truncated FOSB were treated with the indicated concentrations of telatinib 6 hours prior to addition of VEGF. Already at 1 μmol/L most phosphorylation of ERK is absent as detected with a phospho-ERK antibody detecting bands at 42 and 44 kDa, the 44 kDa band completely disappears whereas there is still a faint band visible at 42 kDa, which shows no further reduction at higher concentrations. (d) The left panel shows the sub-G1 cell fraction as found with the NC-250 nucleocounter, showing an increase in sub-G1 fraction with an increase in telatinib concentration. The right panel shows cell cycle and apoptosis as analyzed with the Nucleocounter NC-250. HUVECs overexpressing truncated FOSB were treated with telatinib at the indicated concentrations. An approximate two-fold increase of apoptosis was detected with the highest telatinib dose.

pressing truncated FOSB were also sensitive to knockdown of *FLT1* and *PDGFRA* as well as knockdown of *FLT4* (figure 4.5a; supplementary figure S4A available online). With real-time qPCR, we demonstrated that overexpression of truncated FOSB upregulates *FLT1* and *PDGFRA* (3.53 and 4.35 log2 fold, respectively; figure 4.5b). Thus, overexpression of truncated FOSB results in high expression of *FLT1* and *PDGFRA*, whereas its inhibition alters tumor growth in 3D, suggesting dependence. With the overexpression of truncated FOSB, *KIT* expression was reduced which is in line with the lack of KIT expression in the patient's tumor (supplementary figure S4B). The expression of *FLT1* and *PDGFRA* in the patients tumor, compared with normal HUVECs was 0.13 and 4.92 log2 fold, respectively (supplementary figure S2B).

As AP-1 is reportedly a downstream transcription factor of VEGF signaling (16), we hypothesized that inhibition of the VEGF receptors with telatinib would indirectly lead to a reduction in *SERPINE1* expression. We investigated this by serum starving and stimulating HUVECs which leads to an upregulation of SERPINE1. This upregulation could be blocked by treatment with telatinib (figure 4.5c), suggesting that blocking FLT1, FLT4, and PDGFRA with telatinib could lead to a reduction in expression of the *SERPINE1-FOSB* fusion protein.

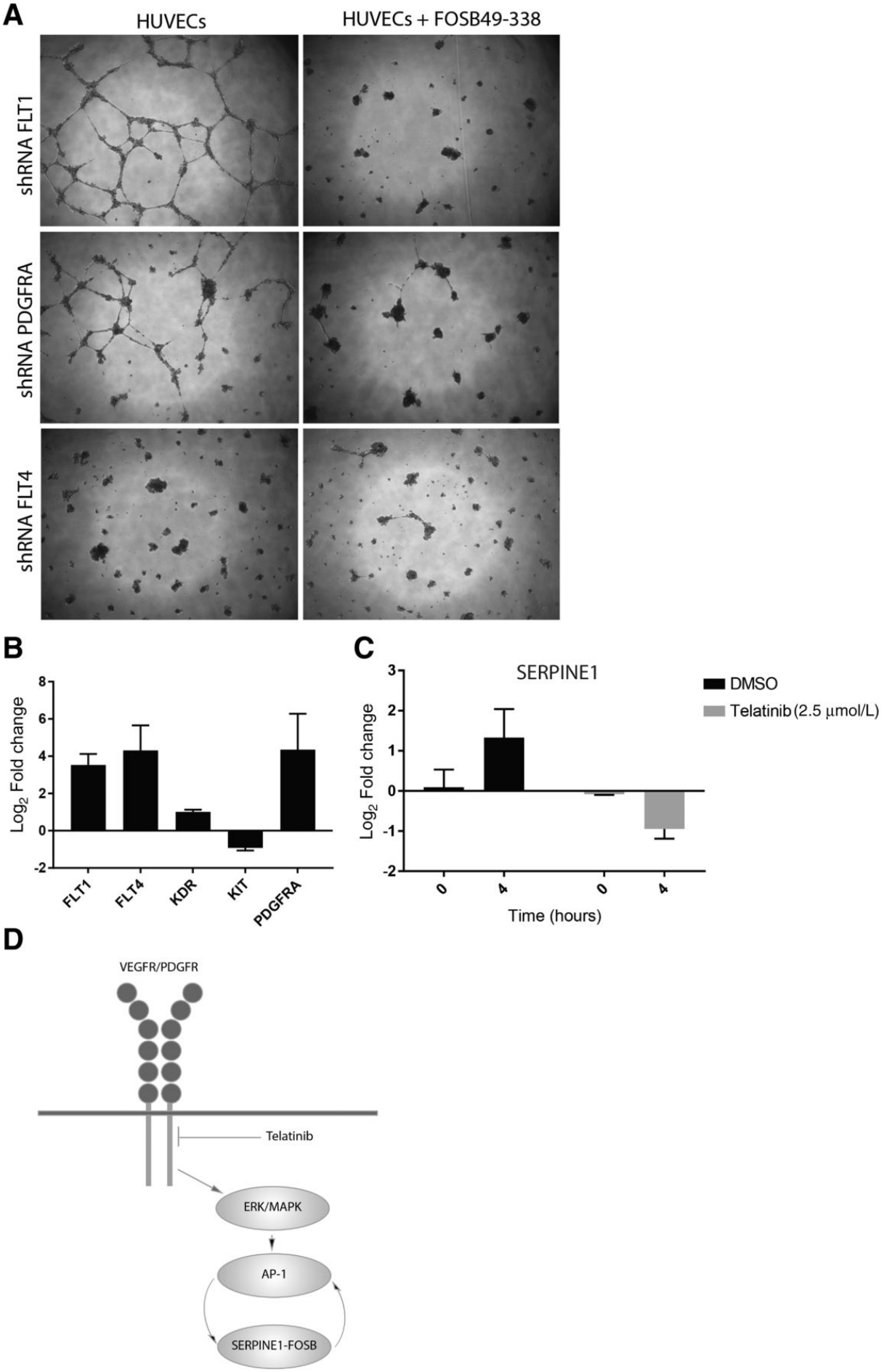


Figure 4.5: Caption on next page.

Figure 4.5: Telatinib targets FLT1, FLT4, and PDGFRA, which are shown to be essential for tube formation in truncated FOSB expressing HUVECs. (a) 3D assay on matrigel with normal HUVECs and with HUVECs overexpressing truncated FOSB. ShRNA for *FLT4* efficiently blocks tube formation in both conditions whereas *FLT1* and *PDGFRA* knockdown only blocks tube formation when truncated FOSB is overexpressed. This suggests that truncated FOSB sensitizes the cells to *PDGFRA* knockdown. (b) Expression levels of the indicated transcripts in HUVECs overexpressing truncated FOSB compared with normal HUVECs were determined by real-time qPCR. *KIT* expression was reduced while *FLT1*, *FLT4*, and *PDGFRA* expression is increased by overexpression of truncated FOSB. Expression was normalized against *HPRT1*. (c) Serum starving and stimulating cells show an upregulation of the *SERPINE1* transcript as found with real-time qPCR. This upregulation is blocked when the HUVECs are treated with telatinib. Expression was normalized against *HPRT1*. (d) We propose a model for the modus of action for telatinib treatment in PHE. *SERPINE1-FOSB* can regulate its own expression while also being under control from the downstream signaling cascade of VEGFR and PDGFR surface receptors. Inhibition of these receptors reduces *SERPINE1-FOSB* expression, and expression is further diminished because self-regulation is reduced.

4.5 Discussion

PHE is a locally aggressive tumor of the so-called intermediate category in the WHO 2013 classification (17), which often presents with multiple discontiguous nodules. In this study we elucidated the biological effect seen in a patient who presented with advanced inoperable PHE, who was treated with telatinib and showed a durable complete response. We here present an in vitro model for PHE in which we show that truncated FOSB is capable of regulating *SERPINE1*, as well as being under control of AP-1 through signaling from VEGF- and PDGF-receptors yielding a positive feedback loop. We propose a model where telatinib inhibits FLT1, FLT4, and PDGFRA signaling, reducing initial expression of *SERPINE1-FOSB*. As there would be less *SERPINE1-FOSB* available, self-regulation is reduced which would further diminish the levels of *SERPINE1-FOSB* in the tumor cells. This effect, combined with essential signaling being blocked by inhibition of FLT1, FLT4, and PDGFRA, could give a biological substantiation to the complete remission as seen in the PHE patient (figure 4.5d).

To model PHE we overexpressed truncated FOSB in HUVECs. As the cell of origin is still enigmatic, we chose primary endothelial cells as a model because these cells are genetically normal (in contrast to cancer cell lines and transformed cells such as 293T). Histologically PHE does not form vessels making it difficult to formally prove endothelial cells are the precursor cells. We think however, given the fact that the tumor cells are known to express CD31, ERG, and FLI1, that endothelial cells, or their precursors, are the most likely precursors for these tumor cells (1, 2). Other candidates would have been

mesenchymal stem cells (MSC), which cannot be completely excluded as potential precursors for PHE. To overexpress truncated FOSB we used a lentiviral system. Truncated FOSB was under control of a constitutively active CMV promoter. We showed that truncated FOSB is able to regulate the *SERPINE1* promoter and that *FLT1* and *PDGFRA* are upregulated. We confirmed that the pattern of expression of these downstream target genes showed a similar trend in the patient's tumor tissue, both using qPCR (for *FAS*, *JAG1*, *HEY1*, *VWF*, *ADAMTS13*, *FLT1*, and *PDGFRA*) compared to normal HUVECs and using IHC (for KIT). However, a formal comparison with the patient tumor biopsy is difficult as the tumor biopsy also contains 40% normal cells and there was no normal tissue from the patient available. As truncated FOSB is constitutively active due to the CMV promotor, our model is not suited to compare the effect of telatinib on proliferation in normal HUVECs with HUVECs overexpressing truncated FOSB. Given the fact however that patients tolerate a high dose of telatinib (comparable to the concentration we used in vitro) for an extended period of time (12), it is likely that truncated FOSB expressing cells are more sensitive to telatinib than normal endothelial cells. We cannot completely exclude that the complete response seen in our patient was due to a spontaneous remission of the disease irrespective of the telatinib treatment. A literature search revealed one described case with a spontaneous remission of PHE as seen with PET/CT (18). Given the fact that the tumor progression halted shortly after starting telatinib treatment and that we showed in our in vitro model that telatinib could specifically interfere with the expression of its driver *SERPINE1-FOSB*, we think it is very likely that the remission observed here can be attributed to the telatinib treatment.

Based on our proposed mechanism of action for telatinib in PHE it is tempting to speculate that telatinib could also function as a targeted therapy in other vascular tumors. In atypical epithelioid hemangioma, a *ZFP36-FOSB* fusion is the most likely driving event (10). In this fusion *ZFP36* exon 1 fuses with exon 2 of *FOSB*. This fusion would result in a truncated FOSB protein driven by the *ZFP36* promoter, while retaining a slightly larger part of FOSB compared to the *SERPINE1-FOSB* fusion. Moreover, according to the DECODE database there are 14 AP-1 consensus binding sites in the *ZFP36* promoter making it most probable that the signaling cascade from the VEGF and PDGF receptors and self-regulation are important in driving expression of this fusion product. In both cases inhibition of the VEGF and PDGF receptors would likely lead to a reduction in expression of the fusion products as in both cases the fusion genes are downstream in the signaling cascade.

Three case reports described the mTOR inhibitors sirolimus, everolimus, and rapamycin as efficient treatment options for PHE (19–21). Reportedly, the AKT/mTOR signaling is, among other surface receptors, also downstream of VEGF-receptor signal-

ing (22). Moreover, it is also reported that AKT/mTOR signaling can lead to upregulation of AP-1 family members (23). Therefore, it is possible that treatment with the mTOR inhibitors would also result in a reduction in expression of the SERPINE-FOSB fusion protein leading to a similar inhibitory effect as telatinib treatment.

Based on our results, it is likely that other tyrosine kinase inhibitors could have a similar effect as telatinib in the treatment of PHE. Pazopanib is one of these tyrosine kinase inhibitors that target VEGF-receptors 1 to 3 and the PDGF-receptors, therefore showing a large overlap with telatinib. Pazopanib was approved as a treatment for soft tissue sarcomas, including angiosarcoma and epithelioid hemangioendothelioma, where standard therapy has failed (24).

In conclusion, we show that truncated FOSB protein, resulting from *SERPINE1-FOSB* fusion, is downstream of VEGF- and PDGF- receptor signaling, and that inhibition of signaling from these surface receptors reduces expression of the fusion protein. Moreover, we showed that *SERPINE1-FOSB* is capable of regulating its own promoter. Consequently, an initial reduction in expression of the *SERPINE1-FOSB* fusion would result in a further reduction as self-regulation is diminished.

Bibliography

[1] van IJzendoorn DGP, Bovée JVMG. Vascular Tumors of Bone: The Evolvement of a Classification Based on Molecular Developments. Surgical Pathology Clinics. 2017;10(3):621–635. doi:10.1016/j.path.2017.04.003.

[2] Hornick JL, Fletcher CD. Pseudomyogenic hemangioendothelioma: a distinctive, often multicentric tumor with indolent behavior. Am J Surg Pathol. 2011;35(2):190–201. doi:10.1097/PAS.0b013e3181ff0901.

[3] Billings SD, Folpe AL, Weiss SW. Epithelioid Sarcoma-Like Hemangioendothelioma. The American journal of surgical pathology. 2003;27(1):48–57. doi:10.1097/PAS.0b013e31821caf1c.

[4] Trombetta D, Magnusson L, von Steyern FV, Hornick JL, Fletcher CDM, Mertens F. Translocation t(7;19)(q22;q13)-a recurrent chromosome aberration in pseudomyogenic hemangioendothelioma? Cancer Genetics. 2011;204(4):211–215. doi:10.1016/j.cancergen.2011.01.002.

[5] Walther C, Tayebwa J, Lilljebjörn H, Magnusson L, Nilsson J, Von Steyern FV, et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. Journal of Pathology. 2014;232(5):534–540. doi:10.1002/path.4322.

[6] Eferl R, Wagner EF. AP-1: A double-edged sword in tumorigenesis. Nature Reviews Cancer. 2003;3(11):859–868. doi:10.1038/nrc1209.

[7] Shaulian E, Karin M. AP-1 as a regulator of cell life and death. Nat Cell Biol. 2002;4(5):E131–6. doi:10.1038/ncb0502-e131.

[8] van IJzendoorn DG, de Jong D, Romagosa C, Picci P, Benassi MS, Gambarotti M, et al. Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. Genes Chromosomes Cancer. 2015;54(9):565–574. doi:10.1002/gcc.22269.

[9] Huang S, Chen H, Zhang L, Sung Y, Dickson B, Krausz T, et al. Frequent FOS Gene Rearrangements in Epithelioid Hemangioma. Lab Invest. 2015;95:18A–19A.

[10] Antonescu CR, Chen HW, Zhang L, Sung YS, Panicek D, Agaram NP, et al. ZFP36-FOSB fusion defines a subset of epithelioid hemangioma with atypical features. Genes Chromosomes and Cancer. 2014;53(11):951–959. doi:10.1002/gcc.22206.

[11] Nagano T, Yamada Y, Ikeda T, Kanki H, Kamo T, Nishigori C. Docetaxel: a therapeutic option in the treatment of cutaneous angiosarcoma: report of 9 patients. Cancer. 2007;110(3):648–651. doi:10.1002/cncr.22822.

[12] Eskens FA, Steeghs N, Verweij J, Bloem JL, Christensen O, van Doorn L, et al. Phase I dose escalation study of telatinib, a tyrosine kinase inhibitor of vascular endothelial growth factor receptor 2 and 3, platelet-derived growth factor receptor beta, and c-Kit, in patients with advanced or metastatic solid tumors. J Clin Oncol. 2009;27(25):4169–4176. doi:10.1200/JCO.2008.18.8193.

[13] Van IJzendoorn DGP, Forghany Z, Liebelt F, Vertegaal AC, Jochemsen AG, Bovée JVMG, et al. Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis. Journal of Biological Chemistry. 2017;292(52):21282–21290. doi:10.1074/jbc.C117.815845.

[14] Hung YP, Fletcher CDM, Hornick JL. FOSB is a useful diagnostic marker for pseudomyogenic hemangioendothelioma. American Journal of Surgical Pathology. 2017;41(5):596–606. doi:10.1097/PAS.0000000000000795.

[15] Strumberg D, Schultheis B, Adamietz IA, Christensen O, Buechert M, Kraetzschmar J, et al. Phase I dose escalation study of telatinib (BAY 57-9352) in patients with advanced solid tumours. British Journal of Cancer. 2008;99(10):1579–1585. doi:10.1038/sj.bjc.6604724.

[16] Jia J, Ye T, Cui P, Hua Q, Zeng H, Zhao D. AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation. Microvascular Research. 2016;105:103–108. doi:10.1016/j.mvr.2016.02.004.

[17] Hornick J, Fletcher C, Mertens F. Pseudomyogenic (epithelioid sarcoma-like) haemangioendothelioma. In: Fletcher C, Bridge JA, Hogendoorn P C, Mertens F, editors. WHO Classification of Tumours of Soft Tissue and Bone. Lyon: IARC; 2013. p. 333–4.

[18] Bryanton M, Makis W. Pseudomyogenic hemangioendothelioma mimicking multiple myeloma on 18F-FDG PET/CT, followed by spontaneous regression. Clinical Nuclear Medicine. 2015;40(7):579–581. doi:10.1097/RLU.0000000000000800.

[19] Ozeki M, Nozawa A, Kanda K, Hori T, Nagano A, Shimada A, et al. Everolimus for Treatment of Pseudomyogenic Hemangioendothelioma. J Pediatr Hematol Oncol. 2017;39(6):e328–e331. doi:10.1097/MPH.0000000000000778.

[20] Gabor KM, Sapi Z, Tiszlavicz LG, Fige A, Bereczki C, Bartyik K. Sirolimus therapy in the treatment of pseudomyogenic hemangioendothelioma. Pediatr Blood Cancer. 2018;65(2). doi:10.1002/pbc.26781.

[21] Joseph J, Wang WL, Patnana M, Ramesh N, Benjamin R, Patel S, et al. Cytotoxic and targeted therapy for treatment of pseudomyogenic hemangioendothelioma. Clin Sarcoma Res. 2015;5:22. doi:10.1186/s13569-015-0037-8.

[22] Abid MR, Guo S, Minami T, Spokes KC, Ueki K, Skurk C, et al. Vascular endothelial growth factor activates PI3K/Akt/forkhead signaling in endothelial cells. Arterioscler Thromb Vasc Biol. 2004;24(2):294–300. doi:10.1161/01.ATV.0000110502.10593.06.

[23] Bishnupuri KS, Luo Q, Murmu N, Houchen CW, Anant S, Dieckgraefe BK. Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. Gastroenterology. 2006;130(1):137–149. doi:10.1053/j.gastro.2005.10.001.

[24] Kollar A, Jones RL, Stacchiotti S, Gelderblom H, Guida M, Grignani G, et al. Pazopanib in advanced vascular sarcomas: an EORTC Soft Tissue and Bone Sarcoma Group (STBSG) retrospective analysis. Acta Oncol. 2017;56(1):88–92. doi:10.1080/0284186X.2016.1234068.

Part II

Model systems

Chapter 5

Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis

This chapter is based on the publication: **van IJzendoorn DGP**, Forghany Z, Liebelt F, Vertegaal AC, Jochemsen AG, Bovée JVMG, Szuhai K, Baker DA. Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis. *J Biol Chem.* 2017;292: 21282-21290.

5.1 Abstract

Epithelioid hemangioma is a locally aggressive vascular neoplasm, found in bones and soft tissue, whose cause is currently unknown, but may involve oncogene activation. FOS is one of the earliest viral oncogenes to be characterized, and normal cellular FOS forms part of the activator protein 1 (AP-1) transcription factor complex, which plays a pivotal role in cell growth, differentiation, and survival as well as the DNA damage response. Despite this, a causal link between aberrant FOS function and naturally occurring tumors has not yet been established. Here, we describe a thorough molecular and biochemical analysis of a mutant FOS protein we identified in these vascular tumors. The mutant protein lacks a highly conserved helix consisting of the C-terminal four amino acids of FOS, which we show is indispensable for fast, ubiquitin-independent FOS degradation via the 20S proteasome. Our work reveals that FOS stimulates endothelial sprouting and that perturbation of normal FOS degradation could account for the abnormal vessel growth typical of epithelioid hemangioma. To the best of our knowledge, this is the first functional characterization of mutant FOS proteins found in tumors.

5.2 Introduction

Epithelioid hemangioma is a neoplasm composed of cells that are phenotypically endothelial, which form vascular lumina or grow as solid sheets (figure 5.1a) (1). Until now, the molecular underpinnings of this disease have yet to be deciphered. A recent cytogenetic and karyotypic survey of the disease, by us (2) and others (3), aimed at refining diagnoses and tumor classification, unearthed a significant number of *FOS* translocations raising the possibility that disruption of FOS function could promote tumorigenesis. The immediate-early FOS proto-oncogene is activated rapidly and transiently in response to a wide spectrum of cell stimuli (4–6), including serum, growth factors, cytokines, tumor-promoting agents, and DNA damage (7). The encoded FOS protein is a component of the crucial AP-1 transcription factor complex whose normal activity is regulated by controlled proteasome degradation (8, 9), and corruption of this process can lead to cell transformation (10–12). In this study, we have investigated the role of a novel mutant FOS protein we discovered in epithelioid hemangioma. We provide evidence that sustained expression of mutant FOS, due to loss of the C terminus, might drive the formation of vascular neoplasms by perturbing matrix metalloproteinase (MMP)4 production and the Notch signaling pathway that are known to facilitate both physiological and pathological angiogenesis. Operationally, we found that the extreme C terminus of FOS renders it intrinsically susceptible to ubiquitin-independent degradation by the 20S proteasome, an essential mechanism bypassed by tumor FOS proteins. This is the first report of a module

that directly mediates ubiquitin-independent proteasomal degradation (UIPD) and emphasizes the importance of UIPD in normal as well as tumor cells. Our work establishes the first demonstrable connection between mutations of FOS and the development of a naturally occurring tumor and unveils a potential, novel approach to treating epithelioid hemangioma by targeted inhibition of FOS or proteins whose expression is activated by FOS.

5.3 Materials and Methods

5.3.1 Patient samples

Epithelioid hemangioma case L3933 was acquired from the archives of the Leiden University Medical Center (LUMC), Leiden, The Netherlands. The diagnosis of epithelioid hemangioma was established by a bone and soft tissue pathologist (J. V. M. G. B.). The study was approved by the LUMC Medical Ethical Commission under protocol B17006.

5.3.2 Cell culture, biochemistry, and molecular biology

Primary HUVECs (Lonza) were cultured in EGM2 medium (Lonza). Chondrosarcoma HT1080 and human embryonic kidney 293T cells were cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (Gibco). Transfections, lentivirus production and cell infections, Western blotting, and co-immunoprecipitations have been described previously (13). FOS stability assays were performed by incubating cells in the presence or absence of cycloheximide for a defined time course (hours). Protein levels were determined by Western blotting.

5.3.3 Plasmid and shRNA construction

Human *FOS* cDNAs fused in-frame with a *FLAG* or an *HA* epitope tag were cloned into the pLV lentiviral vector and pCS2 expression plasmid. Gene-specific shRNA-expressing lentiviruses were generated using the TRC2-pLKO lentiviral vector system.

5.3.4 Transcriptome profiling

RNA was isolated from HUVECs stably expressing FOS or FOSΔ by treatment with TRIzol (Invitrogen) column purification (Direct-zol RNA isolation kit-Zymo Research). RNA quality was verified with a Bioanalyzer (Agilent), and sequencing was performed on the Illumina HiSeq 2500 (Genome Scan). HUVEC transcript sequencing data have been deposited under GenBankTM accession no. PRJNA390521.

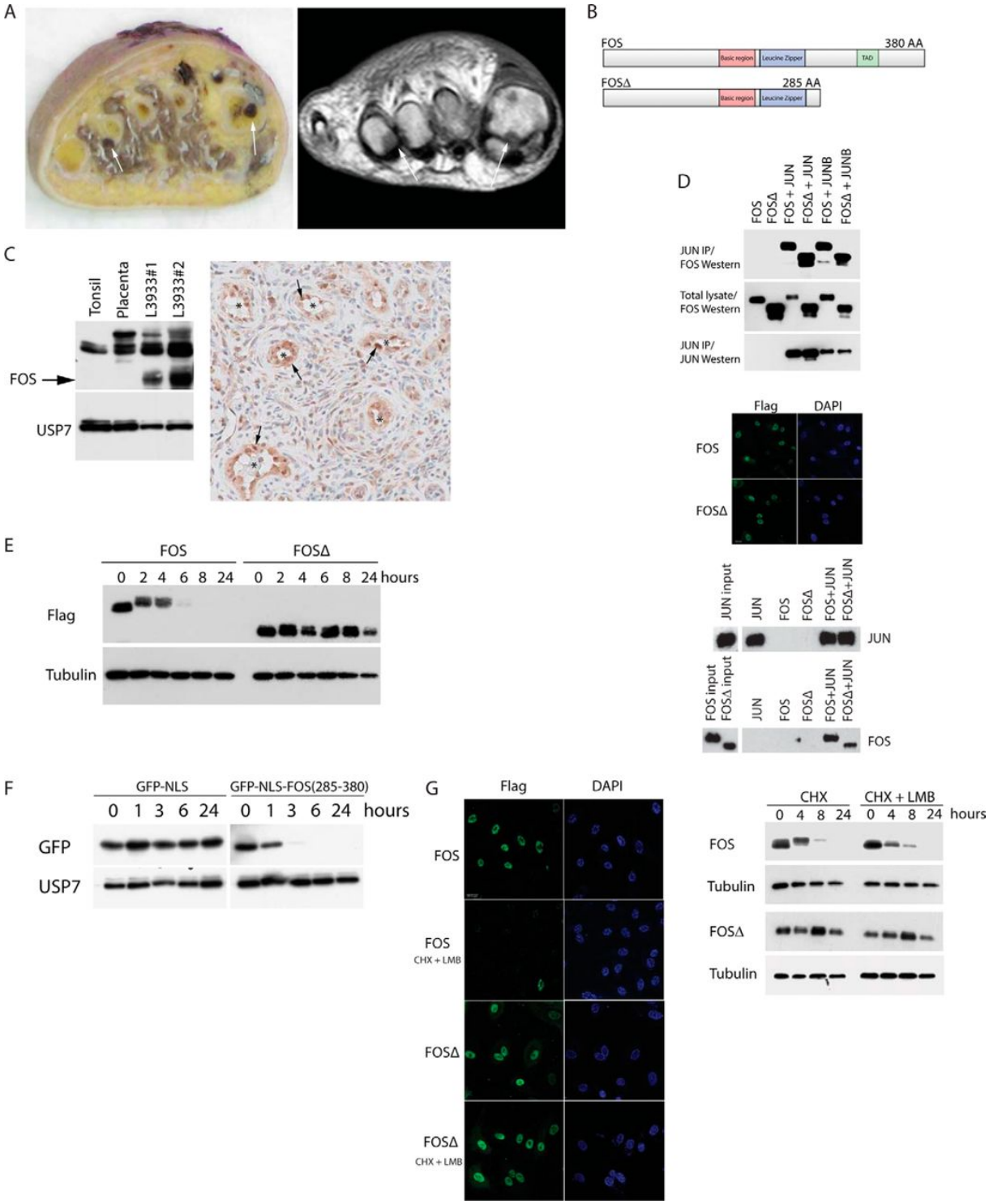


Figure 5.1: Caption on next page.

Figure 5.1: (a) Epithelioid hemangioma case L3933. Left panel, gross specimen with polyostotic localization of a hemorrhagic tumor in the 1st and 4th metatarsal bones of the foot (arrows). Right panel, corresponding T1 weighted MR image. (b) Tumor FOSΔ lacks the C-terminal 95 amino acids (including the C-terminal TAD). IP, immunoprecipitation. (c) Left panel, Western blot of endogenous FOS proteins in control tonsil and placenta cell lysates compared with epithelioid hemangioma tumor cell lysates. Mutant FOSΔ protein is highlighted with an arrow. Right panel, high FOS expression (arrows) is indicated in the endothelial cells of epithelioid hemangioma tumor blood vessels (*). (d) AP-1 heterodimers were immunopurified from cells transfected with the indicated constructs (top panel). Immunofluorescence shows both FOS and FOSΔ localize to the nucleus (middle panel). FOS (and FOSΔ), JUN heterodimers bind to consensus AP-1 DNA-binding sites (bottom panel). (e) FOS stability assay on HUVECs stably expressing FOS or FOSΔ. (f) Protein stability assay on HUVECs stably expressing either GFP or a GFP-FOS fusion (encompassing the C-terminal 95 amino acids of FOS). (g) HUVECs expressing the indicated proteins were incubated with or without leptomycin B (LMB) in the presence of cycloheximide (CHX). Left panel, immunofluorescence. Right panel, Western blots.

5.3.5 Analysis of mRNA expression

RNA isolation, first strand cDNA synthesis, and analysis of expression of transcripts by quantitative PCR were performed as described previously (14).

5.3.6 Ubiquitination assay

293T cells were transfected with the appropriate plasmids. Proteasome degradation was blocked for 8 h with 10 μm MG132 (Sigma). HIS pulldowns were performed as described previously (15).

5.3.7 HUVEC sprouting assay

96-well plates were coated with 60 μl of Matrigel/well 30 min prior to seeding HUVECs. EGM-2 medium was supplemented with 50 ng/ml recombinant human VEGF 165 (R & D Systems). Images were taken at multiple time points. Analysis of the sprouting was performed with Stacks (in-house software, Department of Molecular Cell Biology, LUMC).

5.3.8 Immunohistochemistry/Immunofluorescence

Staining was performed on 4-μm tissue sections. Paraffin was removed with xylene, and sections were rehydrated in a gradient of ethanol. Exogenous peroxidase was blocked using 0.3% H_2O_2. Microwave antigen retrieval was performed in Tris-EDTA (pH 9.0).

FOS antibody was used at a 1:400 concentration. Antibody was detected with 3,3'-diaminobenzidine, and counterstaining was performed with hematoxylin. Immunostaining was performed as described previously (15).

5.3.9 Proteasome purification and in vitro degradation assay

HT1080 cells, stably expressing GFP-PSMD12, were lysed in buffer containing 40 mm Tris (pH 7.5), 40 mm NaCl, 2 mm β-mercaptoethanol, 5 mm $MgCl_2$, 2 mm ATP, 10% glycerol, and 0.5% Nonidet P-40. Lysates were cleared by ultracentrifugation at 36,000 rpm for 45 min at 4 °C. Cleared lysates were incubated for 3 h at 4 °C with prewashed Chromotek GFP-Trap bead slurry. Beads were washed four times in wash buffer containing 40 mm Tris (pH 7.5), 40 mm NaCl, 2 mm β-mercaptoethanol, 5 mm $MgCl_2$, 2 mm ATP, and 10% glycerol. Activity of purified 26S proteasome and 20S proteasome (Enzo LifeSciences) was measured using 100 μm suc-LLVY-AMC substrate (Bachem) in a buffer containing 50 mm Tris (pH 7.5), 40 mm KCl, 5 mm $MgCl_2$, 1 mm DTT (0.5 mm ATP for the 26S proteasome) (absorbance/emission = 353/442 nm). In vitro-translated FOS proteins were prepared using the TnT-coupled reticulocyte in vitro translation system (Promega). Cell-free degradation assays were performed as described previously (16).

5.3.10 Protein-DNA interaction assays

In vitro-translated protein was made as above. 50 pmol of biotinylated double-stranded oligonucleotides harboring three contiguous AP-1 DNA-binding sites were coupled to My-One streptavidin C1 beads (Invitrogen). Reactions were incubated at 4 °C with vigorous shaking for 30 min in the presence of 1 μg of poly(dI/dC), 4 mm spermidine, 50 mm KCl, 10 mm HEPES (pH 7.6), 5 mm $MgCl_2$, 10 mm Tris (pH 8), 0.05 mm EDTA (pH 8), 0.1% Triton X-100, and 20% glycerol. Beads were successively washed three times with the aforementioned buffer. Associated proteins were eluted in Laemmli buffer, and protein-DNA interactions were determined by Western blotting.

5.3.11 ChIP

ChIP analyses were performed on confluent 10-cm tissue culture dishes of HUVECs as described previously (13).

5.3.12 Antibodies, growth factor, and drugs

Antibodies were obtained from the following sources: FLAG mouse M2 monoclonal (Sigma); anti-HA.11 mouse monoclonal (Covance); anti-FOS rabbit (Cell Signaling); anti-

HA rabbit polyclonal (Abcam); anti-FOS rabbit (Sigma); anti-FLAG rabbit (Sigma); anti-USP7 rabbit (Bethyl); anti-γ-tubulin (Sigma); anti-GFP (GeneTex); anti-His (Sigma); and anti-PSMA1 (Sigma). Drugs were used at the following concentrations: MG132 (Sigma), 10 μm; cycloheximide (Sigma), 50 μg/ml; epoxomicin (Sigma), 10 μm; leptomycin B (Sigma) 35 nm; MLN-7243 (Active Biochem), 10 μm; Batimastat (Calbiochem), 10 μm; DAPT (Tocris Bioscience), 10 μm.

5.3.13 Bioinformatics

Rosetta (RosettaCommons) was used for structure prediction of the FOS C terminus (17). Secondary structure was predicted using Psipred (version 4.01, UCL). Degree of disorder was predicted using Disopred (version 3.16, UCL).

5.4 Results and discussion

5.4.1 C-terminally truncated FOS mutant is expressed in epithelioid hemangioma

To determine whether mutations that disrupt the normal function of FOS might promote tumorigenesis, we investigated the role of a novel mutant FOS protein in epithelioid hemangioma (figure 5.1a). Figure 5.1b depicts schematically a FOS deletion mutant (hereafter termed FOSΔ) that resulted from a *FOS-MBNL1* translocation (2). The mutant transcript is predicted to encode a FOS isoform lacking the C-terminal 95 amino acids but including the bZIP domain. Western blot analysis of lysates prepared from patient tumor tissue revealed a truncated FOS protein of the expected size demonstrating that the mutant *FOS* gene is translated in vivo (figure 5.1c). Moreover, immunohistochemistry of tumor sections showed a significant enrichment of FOS in tumor blood vessel endothelial cells (figure 5.1c). In common with wild-type FOS, FOSΔ is localized to the nucleus, can heterodimerize both with JUN and JUNB, and is efficiently associated with a consensus AP-1 DNA-binding site (figure 5.1d). However, FOSΔ protein levels appear to be significantly higher than wild-type FOS protein levels in patient cells (see figure 5.1c) suggesting that the mutant protein may be aberrantly stable. To elucidate the mechanistic consequences of this deletion, we first assessed FOS protein stability in primary endothelial cells. Figure 5.1e shows that wild-type FOS, as expected, has a relatively short half-life of $\sim$1-2 h. By contrast, the deleted version of FOS is highly stable (half-life in excess of 8 h) suggesting that wild-type FOS harbors a destabilizing element in its C terminus, which is absent in the patient FOSΔ protein. In support of this view, tethering the FOS C terminus to a GFP reporter construct, led to a striking destabilization of the

GFP protein (figure 5.1f), whereas a truncated FOS C terminus did not substantially alter the stability of the GFP reporter (see figure 5.1g). This observation is consistent with previous reports relating to FOS stability (8, 18, 19). Additionally, blocking nuclear export had no effect either on rapid FOS degradation or the stability of FOSΔ indicating that FOS is degraded in the nucleus and that FOSΔ is resistant to this process (figure 5.1g). The above findings were confirmed in diploid HT1080 cell and HEK293T cells indicating that this mechanism is likely to be generic.

5.4.2 Mutant FOS is resistant to proteasomal degradation

To precisely delineate how FOS is degraded (and why FOSΔ is not), we monitored FOS protein degradation by the proteasome. Figure 5.2a shows that pharmacological inhibition of the proteasome, using either the specific inhibitor epoxomicin or MG132, markedly stabilized the wild-type FOS protein such that its half-life was comparable with that of the mutant FOSΔ protein. The half-life of FOSΔ was refractory to proteasome inhibition indicating that this deletion essentially lacks the motif(s) responsible for this degradative process (figure 5.2a). It is established that degradation by the proteasome is either ubiquitin-dependent (20) or ubiquitin-independent (21–23), and multiple different mechanisms have been reported to regulate FOS stability (24–26). In agreement with others (23, 25, 27), figure 5.2b shows that wild-type FOS can be ubiquitinated and subsequently processed by the 26S proteasome (see also figure 5.3c). We found that patient FOSΔ protein was not detectably ubiquitinated (figure 5.2b and figure 5.3c), and it fails to bind the E3 ligase, KDM2b, which has been demonstrated to stimulate FOS ubiquitination (supplementary figure S1 available online) (25). These observations could suggest that patient FOSΔ stabilization results from the absence of FOSΔ ubiquitin-dependent degradation.

However, several lines of evidence support the view that the tumor FOSΔ protein is intrinsically resistant to ubiquitin-independent proteasome degradation and that this is the principal cause of its substantially increased stability. First, pharmacological inhibition of ubiquitin-activating enzymes ablated FOS ubiquitination but had no detectable impact on FOS degradation (figure 5.2c). In the same experiment, ubiquitin-dependent degradation of an established substrate of the proteasome, MDM2, was completely abrogated (see figure 5.2c). Second, wild-type non-ubiquitinated FOS but strikingly not FOSΔ was efficiently degraded by the 20S proteasome in a cell-free in vitro system (figure 5.2d). By contrast, FOS was completely resistant to degradation by the 26S proteasome under identical conditions (figure 5.2e). Consistent with these findings, selective inhibition of the 26S proteasome, but not the 20S proteasome (through shRNA-mediated abolition of the 19S subunits PSMD2 and PSMD14), stabilized the ubiquitinated fraction of FOS but

failed to conspicuously inhibit FOS protein degradation, in sharp contrast to inhibiting both the 20S and 26S proteasomes (see figure 5.2a), indicating that degradation is principally via the 20S and not the 26S proteasome (supplementary figure S2 available online). These results show that the FOS C terminus is vital for both ubiquitin-independent and ubiquitin-dependent proteasomal degradation and that both of these processes are lost by the mutant FOSΔ protein expressed in epithelioid hemangioma. These data show that the normal process of FOS degradation is severely corrupted in the tumor FOSΔ mutant protein and substantiate previous studies (26, 28) which suggest that FOS stability is governed chiefly by ubiquitin-independent proteasome degradation.

5.4.3 Mutant FOS lacks a conserved motif essential for ubiquitin-independent degradation by the 20S proteasome

To identify the motif(s) in the C terminus of FOS, which mediates FOS degradation (and is absent in FOSΔ), we performed a thorough mutational analysis of the FOS C terminus. Figure 5.3a shows that deleting the C-terminal four amino acids was sufficient to strongly stabilize FOS and that lack of these amino acids might therefore cause the aberrant stability of the FOSΔ tumor protein. Ab initio modeling (17) of the FOS tail revealed that the C terminus is composed of an intrinsically unstructured region terminating in a helix composed of the C-terminal four amino acids (LLAL), which is conserved in all metazoans sequenced to date (figure 5.3b). Unlike FOSΔ, eliminating this helical region, either through point mutation or deletion, had no effect upon FOS ubiquitination (figure 5.3c). The same mutations did, however, efficiently block FOS degradation to the same degree as the tumor FOSΔ protein (see figure 5.3d and figure 5.3e). The integrity of the four C-terminal amino acids, but not residues immediately adjacent to this motif, is absolutely required for priming FOS labileness (figure 5.3f). Deletions or point mutations of adjacent amino acids, which include consensus phosphorylation sites for ERK and GSK, failed to augment FOS stability. Indeed, a subset of these, in agreement with others (29), served to enhance FOS instability suggesting they play a role in stabilizing but not destabilizing the FOS protein (correspondingly, chemical inhibitors of MEK or deletion of the consensus ERK-docking site had a comparable effect; data not shown). Three additional experiments further validated the importance of the C-terminal motif. One, deletion of the C-terminal four amino acids strongly attenuated the capacity of the FOS C terminus to destabilize GFP (figure 5.3g). Two, a deletion mutant lacking the intrinsically disordered region (IDR) (which is highly stable) but retaining the C-terminal four amino acids was as unstable as wild-type FOS (figure 5.3h). Three, in common with tumor FOSΔ, a mutant FOS lacking the C-terminal four amino acids, was highly resistant to 20S proteasomal degradation in a cell-free in vitro assay (figure

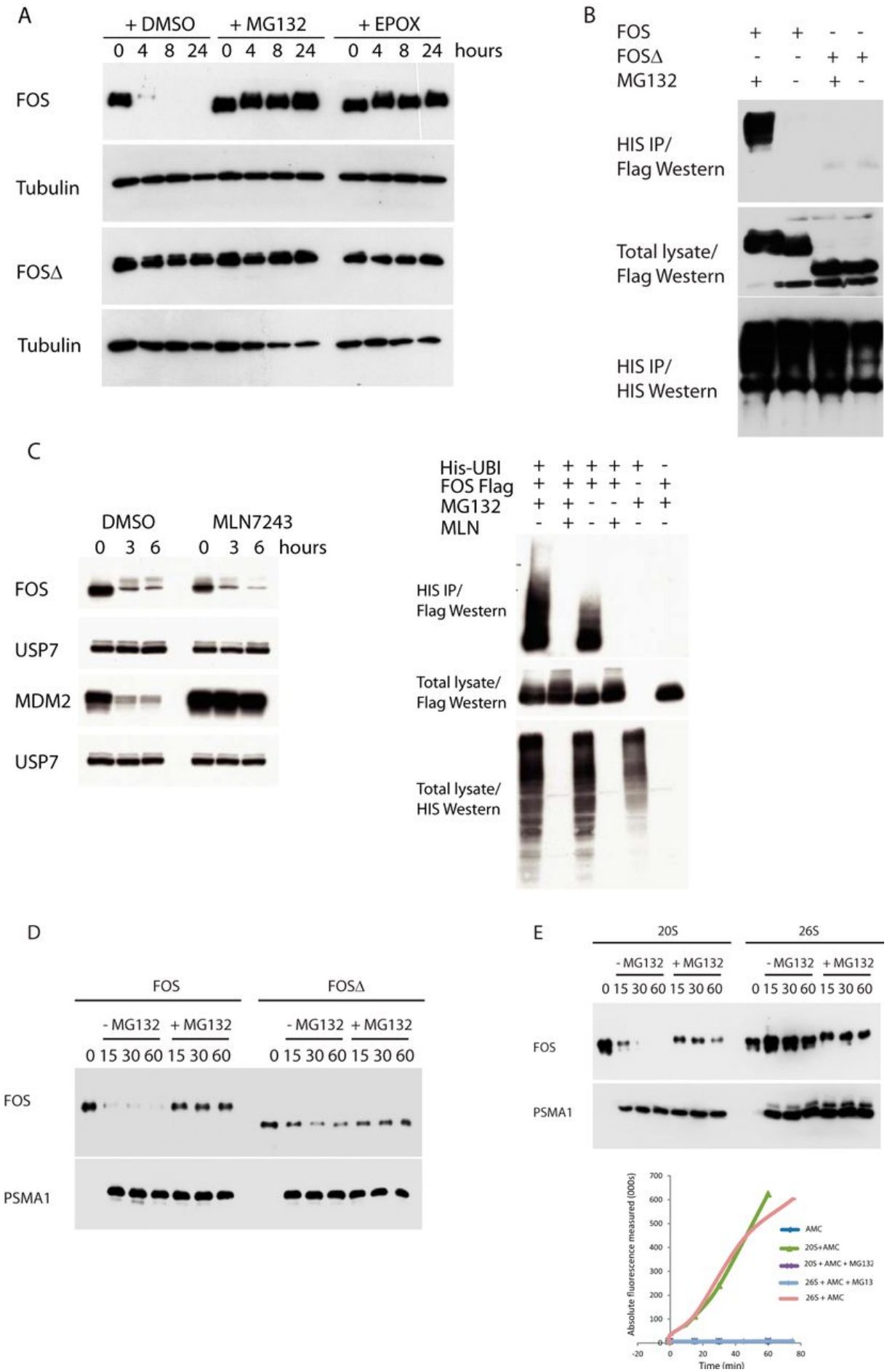


Figure 5.2: Caption on next page.

Figure 5.2: (a) FOS stability assay on HUVECs stably expressing FOS or FOSΔ. (b) Ubiquitin assay of cells transfected with the indicated constructs together with 10$\times$ HIS epitope-tagged ubiquitin. (c) Left panel, FOS stability assay on HUVECs stably expressing FOS in the presence or absence of MLN7243. Right panel, ubiquitin assay of cells transfected with the indicated constructs and cultured in the presence or absence of MG132 and MLN7243. IP, immunoprecipitation. (d) In vitro translated FOS proteins were incubated with purified 20S proteasomes for the shown time course (minutes). 20S protein levels were determined by Western blotting using an antibody directed against PSMA1. 20S proteasome activity was independently quantified using the suc-Leu-Leu-Val-Tyr-AMC peptide (as shown in e). (e) Experiment performed as in d.

5.3i). These data highlight a short helical region at the extreme C terminus of FOS as a crucial determinant of FOS stability and that perturbation of this motif leads to pronounced FOS stabilization. A block in ubiquitin-independent degradation, due to loss of the extreme C terminus, is sufficient to explain mutant FOSΔ stability. Experiments in cell-free systems indicate that this motif can orchestrate direct proteasomal degradation of FOS independently of accessory proteins. IDRs have been reported to strongly influence proteasomal degradation (30), including the IDR found in the C terminus of FOS (31). Our data show that the FOS IDR, by itself, does not stimulate FOS degradation. Rather, a highly conserved helical motif at the extreme C terminus of FOS is essential for triggering ubiquitin-independent degradation.

5.4.4 FOS potently stimulates endothelial sprouting

Vascular neoplasms result from the dysregulated growth of endothelial cells or their precursors (1) and represent a unique model for gaining insights into pathological as well as normal angiogenesis. To recapitulate the cell biological consequences of the mutant FOS stabilization observed in epithelioid hemangioma, we ectopically expressed wild-type and mutant FOS proteins in primary HUVECs and assessed their ability to sprout. Figure 5.4a shows that whereas loss of FOS abolished sprouting, sustained expression of FOS strongly promoted endothelial sprouting and the formation of stable endothelial cell networks. Similarly, expression of tumor FOSΔ or FOS lacking an intact C-terminal four amino acids strongly stimulated endothelial sprouting of HUVECs (figure 5.4b). This phenomenon was independent of marked changes in cell proliferation (figure 5.4b). Supplementary figure S3 available online shows that FOS and FOSΔ also stimulated sprouting of human lung microvascular endothelial cells. The endothelial cell networks produced by cells expressing FOS and FOSΔ were stable and persistent. In this assay, ordinarily the sprouting network is relatively short-lived and collapses after $\sim$24 h. By contrast, endothelial cell networks expressing elevated levels of FOS and FOSΔ were sustained for

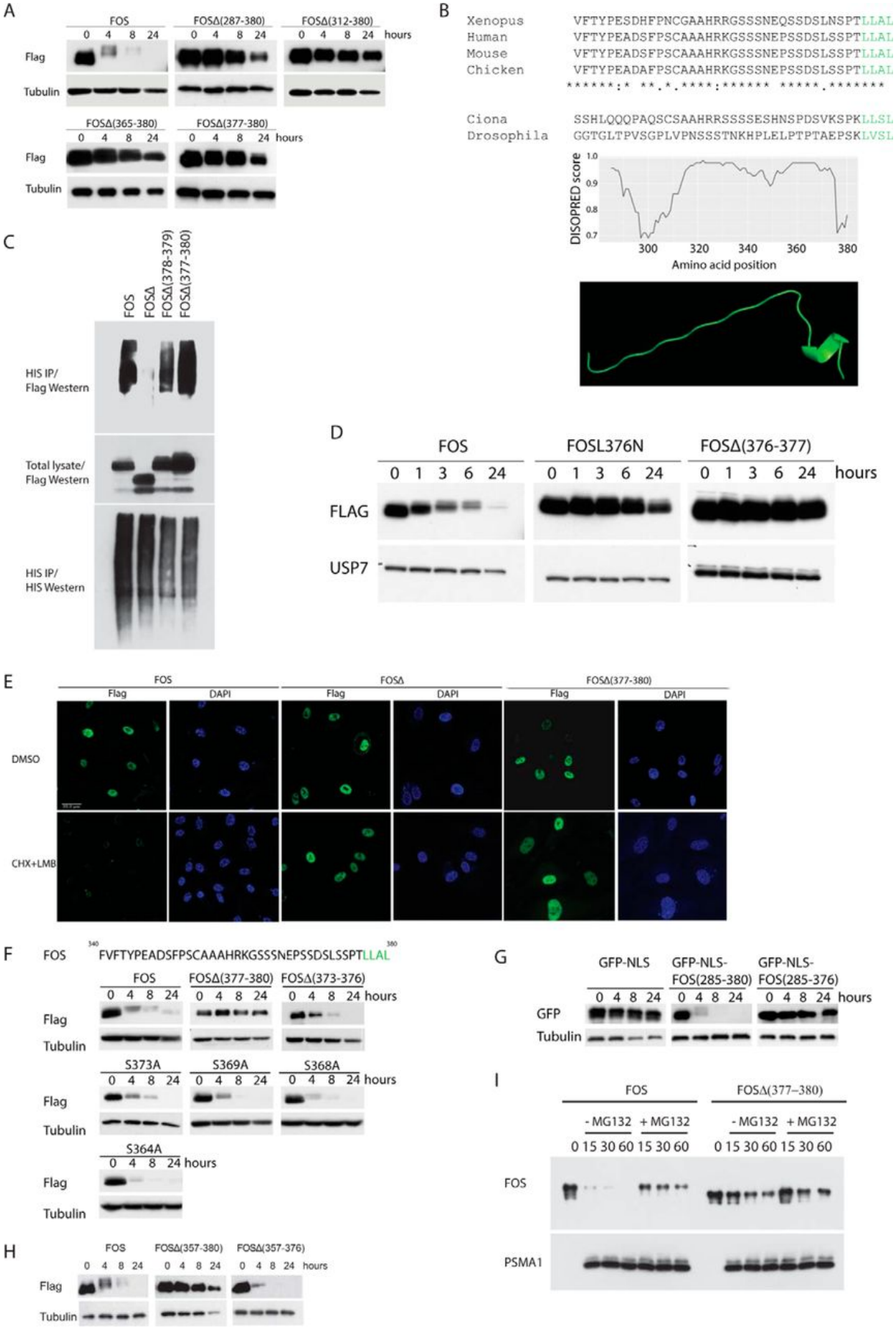


Figure 5.3: Caption on next page.

Figure 5.3: (a) FOS stability assay on HUVECs stably expressing the indicated FOS deletion mutants. (b) Ab initio modeling of the FOS C terminus. (c) Ubiquitin assay performed on cells transfected with the indicated constructs together with 10x HIS epitope-tagged ubiquitin. Cells were cultured in the presence of MG132. IP, immunoprecipitation. (d) FOS stability assay on HUVECs stably expressing the indicated FOS deletion mutants. (e) HUVECs expressing the indicated proteins were incubated with or without leptomycin B (LMB) in the presence of cycloheximide (CHX). FOS was visualized by immunofluorescence. (f) FOS stability assay on HUVECs stably expressing the indicated FOS deletion mutants. (g) Protein stability assay on HUVECs stably expressing either GFP, a GFP-FOS fusion (encompassing the C-terminal 95 amino acids of FOS), or the same fusion lacking the last four amino acids of FOS. (h) FOS stability assay on HUVECs stably expressing the indicated FOS deletion mutants. FOSΔ(357-380) lacks the C-terminal 23 amino acids (the IDR). FOSΔ(357-376) lacks the IDR but retains the C-terminal four amino acids. (i) In vitro FOS stability assay as described in figure 5.2d and figure 5.2e.

at least 2 weeks of culture (sprouting networks expressing FOSΔ were noticeably more robust than the wild-type FOS-expressing networks), which resembles the illicit vessel growth observed in human epithelioid hemangioma.

5.4.5 Mutant FOS-driven sprouting is dependent on MMPs and Notch signalling

To understand the mechanistic basis of FOS-driven sprouting, we performed global transcriptome analyses of sprouts formed by FOS or patient FOSΔ-expressing primary endothelial cells. Figure 5.4c shows a confirmatory qPCR of a selection of angiogenesis-control genes, which were up-regulated, including MMPs and components of the Notch-signaling pathway that are known to facilitate both physiological and pathological angiogenesis (14, 32–36). ChIP analyses showed that endogenous FOS bound to these promoters (supplementary figure S4 available online), and complementary ChIP studies showed that FOSΔ directly interacts with these promoters (see figure 5.4c). Our experiments uncover a previously unreported role for FOS as an activator of endothelial sprouting and show that patient FOSΔ could stimulate illicit endothelial sprouting by activating the Notch signaling pathway and increasing the production of MMPs. In this regard, it is notable that inhibitors of either MMPs or Notch signaling significantly inhibited the sprouting of FOSΔ-expressing endothelial cells (figure 5.4d). The same inhibitors had relatively little effect on cells expressing wild-type FOS under these assay conditions. This could reflect the fact that both MMP production and Notch signaling (as well as other FOS target pathways, see under "Transcriptome profiling") were significantly more augmented in cells expressing wild-type FOS compared with cells expressing FOSΔ

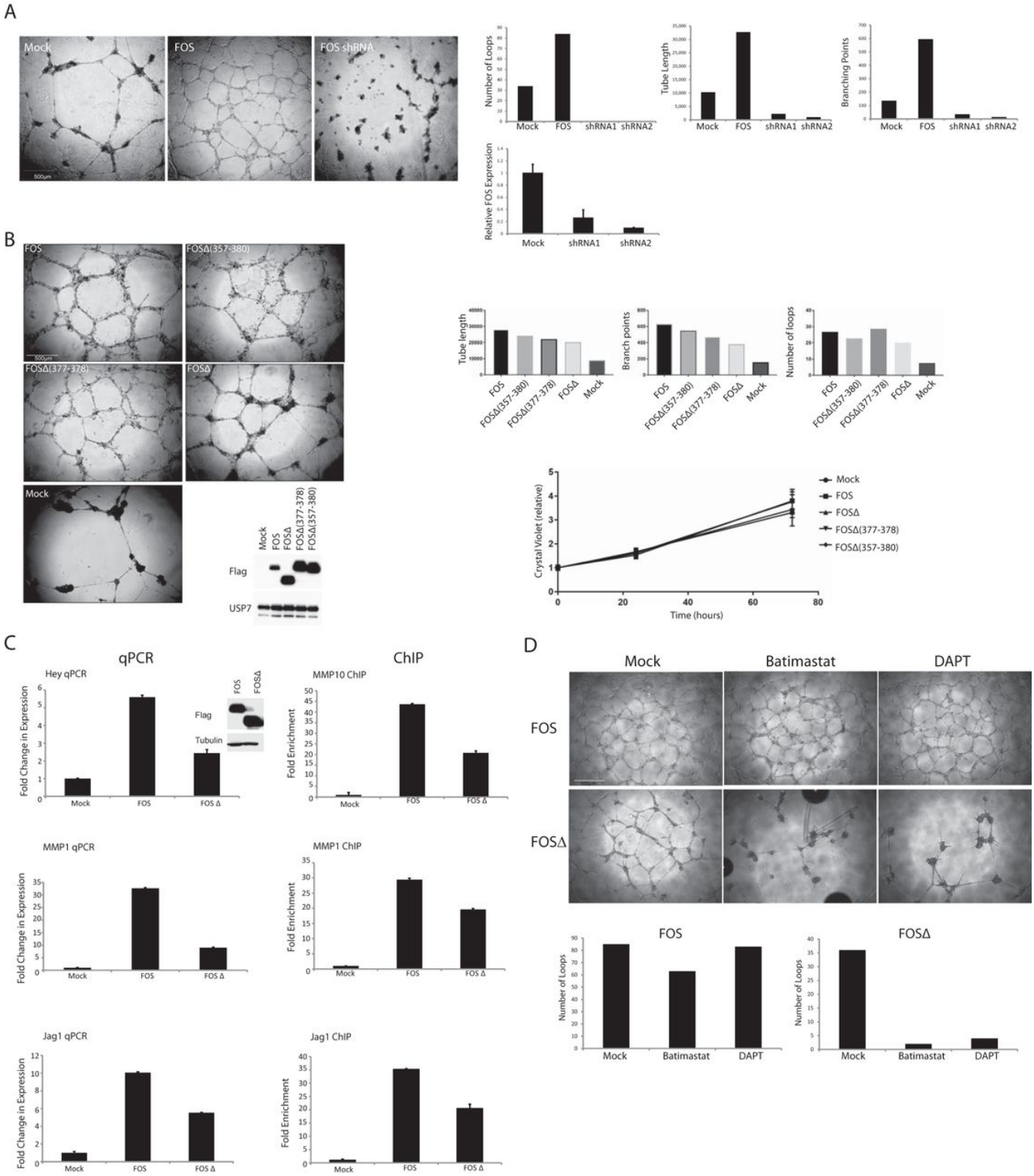


Figure 5.4: Caption on next page.

Figure 5.4: (a) HUVECs lacking endogenous FOS or ectopically expressing wild-type FOS were grown on Matrigel. A representative of several independent experiments is shown. Sprouting was quantified after 24 h using in-house computer software. Loss of *FOS* was determined by qPCR (lowermost graph). (b) Matrigel sprouting assay (see A) on HUVECs stably expressing the indicated FOS proteins. Lower graph, cell proliferation assay of the same HUVECs lines. Triplicate measurements were made at each time point. Values are means ± S.E. of the mean. (c) Left panel, expression levels of the indicated transcripts in HUVECs were determined by real-time qPCR. All values were averaged relative to TATA-binding protein (TBP), signal recognition particle receptor (SRPR), and calcium-activated neutral proteinase 1 (CAPNS1). Values were normalized against mock-treated cells. Values represent ± S.D. (n = 3). Right panel, a ChIP analysis of FOS association with the indicated promoters in HUVECs stably expressing FOS or tumor FOSΔ. Three different primer sets were used for each promoter region. A single representative is shown (all three gave similar results). Results are presented as mean fold changes in recovery (as a fraction of input) relative to the Mock infected cells. Error bars represent the standard deviation (n = 3). Relative FOS and FOSΔ protein levels were determined by Western blotting. (d) HUVECs stably expressing the indicated FOS proteins were grown on Matrigel in the presence or absence of the MMP inhibitor, batimastat (10 μM), or the γ-secretase inhibitor, DAPT (10 μM). Sprouting was quantified after 48 h.

(presumably because FOSΔ lacks the C-terminal TAD). Accordingly, a recently reported small molecule inhibitor of FOS (37), which has advanced to human Phase II clinical trials for the treatment of rheumatoid arthritis, efficiently inhibited FOS-driven endothelial sprouting (supplementary figure S5 available online).

In summary, our data have uncovered a previously unreported role for FOS in stimulating endothelial cell sprouting. We show that sustained expression of FOS, due to loss of the C terminus, could drive the formation of vascular neoplasms. By analyzing the C-terminal region of FOS, which is deleted in epithelioid hemangioma, we have discovered a highly conserved motif at the extreme C terminus of FOS that is critical for controlling its stability by rendering it intrinsically susceptible to ubiquitin-independent degradation by the 20S proteasome. Our work suggests that targeted inhibition of FOS or proteins whose expression is activated by FOS might represent a legitimate novel approach to treating these locally aggressive tumors.

Bibliography

[1] Rosenberg Ae BJ. Epithelioid haemangioma. In: Fletcher Cd BJAHPCWMF, editor. WHO Classification of Tumours of Soft Tissue and Bone. Lyon: IARC; 2013. p. 333–4.

[2] van IJzendoorn DG, de Jong D, Romagosa C, Picci P, Benassi MS, Gambarotti M, et al.

Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. Genes Chromosomes Cancer. 2015;54(9):565–574. doi:10.1002/gcc.22269.

[3] Huang SC, Zhang L, Sung YS, Chen CL, Krausz T, Dickson BC, et al. Frequent FOS gene rearrangements in epithelioid hemangioma: A molecular study of 58 cases with morphologic reappraisal. American Journal of Surgical Pathology. 2015;39(10):1313–1321. doi:10.1097/PAS.0000000000000469.

[4] Jochum W, Passegue E, Wagner EF. AP-1 in mouse development and tumorigenesis. Oncogene. 2001;20(19):2401–2412. doi:10.1038/sj.onc.1204389.

[5] Johnson RS, Spiegelman BM, Papaioannou V. Pleiotropic effects of a null mutation in the c-fos proto-oncogene. Cell. 1992;71(4):577–86.

[6] Shaulian E, Karin M. AP-1 as a regulator of cell life and death. Nat Cell Biol. 2002;4(5):E131–6. doi:10.1038/ncb0502-e131.

[7] Haas S, Kaina B. c-Fos is involved in the cellular defence against the genotoxic effect of UV radiation. Carcinogenesis. 1995;16(5):985–91.

[8] Wilson T, Treisman R. Fos C-terminal mutations block down-regulation of c-fos transcription following serum stimulation. The EMBO journal. 1988;7(13):4193–202.

[9] Gomard T, Jariel-Encontre I, Basbous J, Bossis G, Mocquet-Torcy G, Piechaczyk M. Fos family protein degradation by the proteasome. Biochemical Society Transactions. 2008;36(5):858–863. doi:10.1042/BST0360858.

[10] Van Beveren C, van Straaten F, Curran T, Müller R, Verma IM. Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular fos gene products have different carboxy termini. Cell. 1983;32(4):1241–55.

[11] Eferl R, Wagner EF. AP-1: A double-edged sword in tumorigenesis. Nature Reviews Cancer. 2003;3(11):859–868. doi:10.1038/nrc1209.

[12] Monje P, Marinissen MJ, Gutkind JS. 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. Molecular and cellular biology. 2003;23(19):7030–43. doi:10.1128/MCB.23.19.7030-7043.2003.

[13] Roukens MG, Alloul-Ramdhani M, Baan B, Kobayashi K, Peterson-Maduro J, Van Dam H, et al. Control of endothelial sprouting by a Tel-CtBP complex. Nature Cell Biology. 2010;12(10):933–942. doi:10.1038/ncb2096.

[14] Adams RH, Alitalo K. Molecular regulation of angiogenesis and lymphangiogenesis. Nature Reviews Molecular Cell Biology. 2007;8(6):464–478. doi:10.1038/nrm2183.

[15] Roukens MG, Alloul-Ramdhani M, Moghadasi S, Op den Brouw M, Baker DA. Down-regulation of vertebrate Tel (ETV6) and Drosophila Yan is facilitated by an evolutionarily conserved mechanism of F-box-mediated ubiquitination. Mol Cell Biol. 2008;28(13):4394–4406. doi:10.1128/MCB.01914-07.

[16] Asher G, Tsvetkov P, Kahana C, Shaul Y. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. Genes & development. 2005;19(3):316–21. doi:10.1101/gad.319905.

[17] Baker D. Centenary Award and Sir Frederick Gowland Hopkins Memorial Lecture. Protein folding, structure prediction and design. Biochemical Society transactions. 2014;42(2):225–9. doi:10.1042/BST20130055.

[18] Ferrara P, Andermarcher E, Bossis G, Acquaviva C, Brockly F, Jariel-Encontre I, et al. The structural determinants responsible for c-Fos protein proteasomal degradation differ according to the conditions of expression. Oncogene. 2003;22(10):1461–1474. doi:10.1038/sj.onc.1206266.

[19] Acquaviva C, Brockly F, Ferrara P, Bossis G, Salvat C, Jariel-Encontre I, et al. Identification of a C-terminal tripeptide motif involved in the control of rapid proteasomal degradation of c-Fos proto-oncoprotein during the G(0)-to-S phase transition. Oncogene. 2001;20(51):7563–7572. doi:10.1038/sj.onc.1204880.

[20] Collins GA, Goldberg AL. The Logic of the 26S Proteasome. Cell. 2017;169(5):792–806. doi:10.1016/J.CELL.2017.04.023.

[21] Erales J, Coffino P. Ubiquitin-independent proteasomal degradation. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2014;1843(1):216–221. doi:10.1016/J.BBAMCR.2013.05.008.

[22] Jariel-Encontre I, Bossis G, Piechaczyk M. Ubiquitin-independent degradation of proteins by the proteasome. Biochim Biophys Acta. 2008;1786(2):153–177. doi:10.1016/j.bbcan.2008.05.004.

[23] Ben-Nissan G, Sharon M. Regulating the 20S proteasome ubiquitin-independent degradation pathway. Biomolecules. 2014;4(3):862–884. doi:10.3390/biom4030862.

[24] Stancovski I, Gonen H, Orian A, Schwartz AL, Ciechanover A. Degradation of the proto-oncogene product c-Fos by the ubiquitin proteolytic system in vivo and in vitro: identification and characterization of the conjugating enzymes. Molecular and cellular biology. 1995;15(12):7106–16. doi:10.1128/MCB.15.12.7106.

[25] Han XR, Zha Z, Yuan HX, Feng X, Xia YK, Lei QY, et al. KDM2B/FBXL10 targets c-Fos for ubiquitylation and degradation in response to mitogenic stimulation. Oncogene. 2016;35(32):4179–4190. doi:10.1038/onc.2015.482.

[26] Adler J, Reuven N, Kahana C, Shaul Y. c-Fos Proteasomal Degradation Is Activated by a Default Mechanism, and Its Regulation by NAD(P)H:Quinone Oxidoreductase 1 Determines c-Fos Serum Response Kinetics. Molecular and Cellular Biology. 2010;30(15):3767–3778. doi:10.1128/MCB.00899-09.

[27] Basbous J, Chalbos D, Hipskind R, Jariel-Encontre I, Piechaczyk M. Ubiquitin-independent proteasomal degradation of Fra-1 is antagonized by Erk1/2 pathway-mediated phosphorylation of a unique C-terminal destabilizer. Mol Cell Biol. 2007;27(11):3936–3950. doi:10.1128/MCB.01776-06.

[28] Sasaki T, Kojima H, Kishimoto R, Ikeda A, Kunimoto H, Nakajima K. Spatiotemporal Regulation of c-Fos by ERK5 and the E3 Ubiquitin Ligase UBR1, and Its Biological Role. Molecular Cell. 2006;24(1):63–75. doi:10.1016/j.molcel.2006.08.005.

[29] Okazaki K, Sagata N. The Mos/MAP kinase pathway stabilizes c-Fos by phosphorylation and augments its transforming activity in NIH 3T3 cells. The EMBO journal. 1995;14(20):5048–59.

[30] Prakash S, Tian L, Ratliff KS, Lehotzky RE, Matouschek A. An unstructured initiation site is required for efficient proteasome-mediated degradation. Nature Structural & Molecular Biology. 2004;11(9):830–837. doi:10.1038/nsmb814.

[31] Campbell KM, Terrell AR, Laybourn PJ, Lumb KJ. Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. Biochemistry. 2000;39(10):2708–2713. doi:10.1021/bi9923555.

[32] Heo SH, Choi YJ, Ryoo HM, Cho JY. Expression profiling of ETS and MMP factors in VEGF-activated endothelial cells: Role of MMP-10 in VEGF-induced angiogenesis. Journal of Cellular Physiology. 2010;224(3):734–742. doi:10.1002/jcp.22175.

[33] Lee S, Jilani SM, Nikolova GV, Carpizo D, Iruela-Arispe ML. Processing of VEGF-A by matrix metalloproteinases regulates bioavailability and vascular patterning in tumors. The Journal of Cell Biology. 2005;169(4):681–691. doi:10.1083/jcb.200409115.

[34] Kopan R, Ilagan MXG. The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. Cell. 2009;137(2):216–233. doi:10.1016/j.cell.2009.03.045.

[35] Herbert SP, Stainier DYR. Molecular control of endothelial cell behaviour during blood vessel morphogenesis. Nature Reviews Molecular Cell Biology. 2011;12(9):551–564. doi:10.1038/nrm3176.

[36] Juncker-Jensen A, Deryugina EI, Rimann I, Zajac E, Kupriyanova TA, Engelholm LH, et al. Tumor MMP-1 Activates Endothelial PAR1 to Facilitate Vascular Intravasation and Metastatic Dissemination. Cancer Research. 2013;73(14):4196–4211. doi:10.1158/0008-5472.CAN-12-4495.

[37] Ye N, Ding Y, Wild C, Shen Q, Zhou J. Small molecule inhibitors targeting activator protein 1 (AP-1). Journal of Medicinal Chemistry. 2014;57(16):6930–6948. doi:10.1021/jm5004733.

Chapter 6

Pseudomyogenic hemangioendothelioma recapitulated in endothelial cells from human induced pluripotent stem cells engineered to express the *SERPINE1-FOSB* translocation

This chapter is based on the manuscript: **van IJzendoorn DGP**, Salvatori DCF, Cao X, van den Hil F, Briaire-de Bruijn IH, de Jong D, Mei H, Mummery CL, Szuhai K, Bovée JVMG, Orlova VV. Pseudomyogenic hemangioendothelioma recapitulated in endothelial cells from human induced pluripotent stem cells engineered to express the *SERPINE1-FOSB* translocation.

6.1 Abstract

Chromosomal translocations are prevalent among soft tissue tumors including those of the vasculature. Pseudomyogenic hemangioendothelioma (PHE) is one such tumor. It has features of endothelial cells (ECs) and a tumor-specific t(7;19)(q22;q13) *SERPINE1-FOSB* translocation, but has been difficult to study since to date no cell lines have been derived from the tumor. To address this, we engineered the PHE chromosomal translocation into human induced pluripotent stem cells (hiPSCs) using CRISPR/Cas9 and differentiated these into ECs (hiPSC-ECs). Comparison of parental and modified (PHE) lines with the t(7;19)(q22;q13) *SERPINE1-FOSB* translocation showed (i) elevated expression of *FOSB* specifically in hiPSC-ECs *in vitro* and *in vivo* (ii) increased proliferation and tube formation but decreased endothelial barrier function (iii) invasive growth and abnormal vessel formation in mouse after transplantation of the mutated cells (iv) transcriptome alterations specific for hiPSC-ECs that reflect the PHE phenotype and elucidate pathways regulated by the fusion that can be targeted for treatment (PI3K-Akt and MAPK signaling). hiPSC-ECs carrying the *SERPINE1-FOSB* translocation thus recapitulated functional features of PHE and demonstrated that this approach can yield models of translocation-driven tumors for identification of therapeutic targets and deeper understanding of underlying tumorigenic mechanisms.

6.2 Introduction

Chromosomal translocations and their corresponding gene fusions are common in neoplasia and are important in the initiation of tumorigenesis (1). These gene fusions are especially prevalent in soft tissue tumors, of which ∼15-20% carry a recurrent chromosomal translocation with no, or few, additional genomic alterations (2). Moreover, translocations are usually specific for each subtype. The identification of specific fusion genes has significantly increased the understanding of the pathogenesis of these (often-rare) tumor types and are used as an auxiliary diagnostic tool.

Pseudomyogenic hemangioendothelioma (PHE) is a rare soft tissue tumor characterized by a specific recurrent balanced translocation, t(7;19)(q22;q13), which fuses *SERPINE1* to *FOSB* (3, 4). The translocation leads to the loss of the first exon of *FOSB* containing the start codon, resulting in a novel start codon in exon 2 of *FOSB*. The translocation therefore causes loss of 48 amino acids at the start of the FOSB protein which then falls consequently under control of the *SERPINE1* promoter (4). PHE is locally aggressive, rarely metastasizing and often affecting young adults, especially men between 20-50 years of age. The disease most often presents as multiple discontiguous lesions in different tissue planes (5). Approximately 60% of the patients show relapse after surgical removal

or develop additional nodules which can necessitate limb amputation. The tumors display loose spindle-shaped cells with abundant eosinophilic cytoplasm, that invade in the surrounding soft tissues, expressing vascular (CD31, ERG) and epithelial (keratin) markers. Moreover, the translocation results in overexpression of FOSB protein in patient tumor samples (6). Although PHE does not form functional blood vessels, vascular markers are expressed suggesting that PHE arises from endothelial cells (ECs) or their precursors. The tumor is therefore classified among the group of vascular tumors (5, 7, 8).

Like many other soft tissue tumors with translocations, PHE is rare and no cell lines have yet been derived from the tumor, confounding understanding of tumorigenesis and the identification of potential therapeutic targets. A possible approach to model translocation driven tumors is to engineer the complete chromosomal translocation in human pluripotent stem cells and examine the effects on appropriately differentiated derivatives (9). Engineered nucleases were recently shown to be useful in generating chromosomal translocations in human cells. Clustered regularly interspaced short palindromic repeats (CRISPR) and Cas9 nucleases have been used to introduce chromosomal translocations in human umbilical cord-derived mesenchymal stromal cells (hMSCs), umbilical cord blood-derived CD34+ cells, and more recently human induced pluripotent stem cells (hiPSCs) (10–13). hiPSCs in particular are increasingly used as human disease models, as they can be propagated indefinitely *in vitro* and differentiated into most cell types of the body (14), including ECs (15–17). They are thus a renewable source of cells to study human physiology and disease. We hypothesized that hiPSC-derived ECs (hiPSC-ECs) could be valuable for modeling rare tumors such as those of the vasculature and demonstrated in the study described here that this is indeed the case for PHE.

We introduced the t(7;19)(q22;q13) *SERPINE1-FOSB* translocation into hiPSCs and thus generated control and modified isogenic hiPSC pairs. We carried out functional analysis of hiPSC-ECs and whole genome and transcriptome sequencing of isogenic pairs of hiPSC and hiPSC-EC with and without translocation. We showed that hiPSC-ECs with the *SERPINE1-FOSB* fusion were distinct from their isogenic controls and exhibited phenotypic and transcriptomic characteristics very similar to PHE. More importantly, in mice mutant hiPSC-ECs became invasive and formed abnormal vessels. Our hiPSC model thus mimics PHE, but in more general terms, the approach can serve as a blueprint for using CRISPR/Cas9 in hiPSCs to explore the role of fusion genes in the development of specific rare cancer subtypes for which cell lines are lacking, providing deeper understanding of tumorigenesis resulting from gene fusions.

6.3 Materials and Methods

6.3.1 hiPSC lines and culture

The SeV reprogrammed hiPSC line LUMC0054iCTRL was used (additional information available in public databases: `http://hpscreg.eu/cell-line/LUMCi001-A` and `http://hpscreg.eu/cell-line/LUMCi001-A-1`) (15). hiPSCs were cultured on recombinant vitronectin (VN)-coated plates in TeSR-E8 all from STEMCELL Technologies (SCT), according to the manufacturer's instructions. For targeting experiments, hiPSCs were adapted to single cell passaging on mouse embryonic fibroblasts (MEFs) in Dulbecco's modified Eagle's medium/Ham's F-12 medium (DMEM/F12) supplemented with 20% knockout serum replacement (Invitrogen), 1 mM L-glutamine, 0.1 mM nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 8 ng/ml recombinant human basic fibroblast growth factor (bFGF; Milteny). Single cell adapted hiPSC were passaged using 1X TrypLE Select with additional supplementation with 1X RevitaCell (Invitrogen).

6.3.2 Construction of dual-guide Cas9-encoding plasmids and repair template

A dual sgRNA and Cas9-expressing plasmid was generated by introducing a second gRNA scaffold in the SpCas9-2A-Puro V2.0 (Addgene, Feng Zang) plasmid using Gibson ligation as described (18). The final plasmid contains *FOSB* sgRNA TCCACTACACCGTGACGCAG and *SERPINE1* sgRNA TGAACACTAGGGCAAGGTGC. The repair template was generated by blunt ligation of *FOSB* and *SERPINE1* homology arms (around 1kb each) into a P15 backbone containing a Neomycin resistance cassette surrounded by two flippase recognition target (FRT) sequences (kindly provided by Dr. Konstantinos Anastassiadis, Technical University Dresden). The CAGGs-Flpo-IRES-puro vector which expresses codon-optimized Flp recombinase was used for transient transfection to recombine FRT sites (19) (kind gift of Dr Konstantinos Anastassiadis). The U6 vector used for the Gibson ligation was a kind gift from Dr. Andrea Ventura (Addgene plasmid # 69312).

6.3.3 Transfection

hiPSCs were transfected at 60-70% confluence the day after seeding in a 60 mm dish. Transfection was carried out using Lipofectamine 2000 (Invitrogen). First, 20 μl Lipofectamine 2000 was diluted in 300 μl Opti-MEM Medium and incubated at RT for 5 min. In parallel, 8 μg of both the repair template and double guide RNA/Cas9 was diluted in 300 μl Opti-MEM Medium. Diluted plasmid DNA was added to diluted Lipofectamine 2000 in

a 1:1 ratio and incubated another 5 min at RT before the DNA-lipid complex was added to the cells in a drop-wise manner. Cells were allowed to grow in the incubator for ∼18 hours before the medium was changed. Antibiotic selection with 50 μg/ml G-418 was performed 24 hours post transfection and was continued for 7 days to select for targeted cells. Once recovered, cells were passaged into 6-well plates and transfected the next day with 4 μg Flp recombinase expression vector to excise the neomycin cassette (using Lipofectamine 2000, according to the manufacturer's protocol). At 24h post transfection the medium was supplemented with 0.5 μg/ml Puromycin for 48h to enrich for transfected cells. At 80% confluence, the cells were passaged for clonal expansion on 96-well plates using limited dilution.

6.3.4 Fluorescence In Situ Hybridization

Three-color Fluorescence In Situ Hybridization (FISH) was performed using BAC clones (BACPAC Resource Center). Proximal to *SERPINE1* BAC clone RP11-395B7 was selected. Proximal and distal to *FOSB* respectively BAC clone RP11-84C16 and RP11-902P17 were selected. BAC DNA was extracted using the High Pure plasmid isolation kit (Roche). The RP11-395B7, RP11-84C16 and RP11-902P17 were respectively labeled with Cy5-dUTP, Fluorescein-12-dCTP and Cy3-dUTP using a nick translation labeling reaction (20). FISH was performed as previously described by our group (21). Representative images were taken using a fluorescence microscope (Leica).

6.3.5 Identification of targeted hiPSC clones by PCR

PCR screening was performed to determine the presence of both the 5' homology arm of *SERPINE1* (primers SF and FR), the 3' homology arm of *FOSB* (primer F2 and R2), the wild-type *SERPINE1* (primer SF and SR) and wild-type *FOSB* (primer FF and FR) in clonal lines (table 6.1). Colonies were picked in maximum 2 μl hESC-food and added to 20 μl QuickExtract Solution (Epicentre) in 0.5 mL tubes. The tubes were vortexed for 15s and DNA was extracted by heating the samples to 65 °C for 15 minutes, 68 °C for 15 minutes and 98 °C for 10 minutes in a thermocycler. 2-Step PCR was performed with Terra PCR Direct Polymerase (TaKaRa) according to the manufacturer's protocol. Sanger sequencing was performed (BaseClear) to confirm the *SERPINE1-FOSB* fusion and to screen the *SERPINE1* and *FOSB* wild-type allele for on-target mutations due to NHEJ.

Name	Sequence
SERPINE1 (SF)	ACACAGGCAGAGGGCAGAAAGGTCAA
SERPINE1 (SR)	CCTGCGCCACCTGCTGAAACAC
FOSB (FF)	GCCTTCAGAGCAGTTCCAGGAGTCCATTTA
FOSB (FR)	ACCGACACACACACACCCAACACACATAA
F2	TGGGCTGCAAAGGCAGAGAGTGGTAAT
R2	AAGCGATCCTCCCACTAAAGCCTCCATAGT

Table 6.1: PCR primers used to screen targeted clones.

6.3.6 COBRA-FISH

COmbined Binary RAtio Fluorescence in Situ Hybridization (COBRA-FISH) was performed on metaphase cells as previously described in detail (22, 23).

6.3.7 Differentiation and characterization of hiPSCs to ECs

hiPSCs were differentiated to hiPSC-ECs and characterized as previously described (15–17).

6.3.8 Real-Time qPCR

RNA isolation was performed with the Direct-zol RNA isolation kit (Zymo-research) according to the manufacturer's protocol. cDNA was synthesized using M-MLV with oligo dT primers (Promega) according to the manufacturer's protocol. Real-Time qPCR was performed with Sybr Green (Bio-Rad) on a CFX384 thermocycler (Bio-Rad). All real time PCR experiments were performed in triplicate. Primers are listed in table 6.2.

Name	Sequence
HPRT_f	TGACACTGGCAAAACAATGCA
HPRT_r	GGTCCTTTTCACCAGCAAGCT
FOSB_f	AGCAGCAGCTAAATGCAGGA
FOSB_r	CCAACTGATCTGTCTCCGCC

Table 6.2: qPCR primers.

6.3.9 Western blotting

Western blotting was performed as previously described (24) using FOSB monoclonal rabbit antibody (#2251; Cell Signaling) and USP7 monoclonal rabbit antibody (A300-033A; Bethyl).

6.3.10 Assessment of hiPSC-EC proliferation

To quantify proliferation, cells were cultured in a 96-well plate for 24 hours. Presto Blue (ThermoFisher) was subsequently added to the medium and cells were incubated at 37 °C for 30 minutes before determining the Relative Fluorescence Units (RFU) using a plate reader (Perkin Elmer).

6.3.11 Matrigel tube formation assay

Tube formation assays were performed in 96-well plates coated with 50 μl Matrigel (Corning). hiPSC-ECs were seeded at a density of 15,000 cells per well in 150 μl EC-SFM supplemented with 1% BSA and 50 ng/μl VEGF. Tube formation was analyzed with ImageJ (NHI, v1.51s). Tube formation was imaged with the EVOS Cell Imaging System (ThermoFisher). To quantify tube formation a custom plugin in ImageJ (NIH, v1.51s) was used. Analysis scripts are available on GitHub (`github.com/davidvi`) for analysis of tube formation.

6.3.12 Endothelial barrier function and analysis

Endothelial barrier function was determined as previously described (15). Briefly, hiPSC-ECs were plated on FN-coated ECIS arrays (8W10E PET, Applied Biophysics) at a density of 50,000 cells/cm2. Wounding of the cells grown on the electrodes was performed by applying a 10 sec pulse of 5V at 60 kHz. Barrier function was estimated by applying a current to the electrodes at 4 kHz and measuring the R [ohm]. Barrier function was measured for over 8 hours. Quantification was performed over a period of 5 hours, when the barrier had stabilized.

6.3.13 Immunohistochemistry

Immunofluorescence was performed as previously described (15, 16). Briefly, hiPSC-ECs were fixed with 4% PFA and permeabilized with 0.1% Triton-X100. The following primary antibodies were used: anti-ZO1 (61-7300; ThermoFisher), VEC (53-1449-42; CellSignaling), CD31 (M082301; Dako) and cells were counterstained with A488 conjugated Phalloidin (ThermoFisher). Incubation with primary antibodies was overnight at +4 °C and secondary antibodies for 30 minutes at room temperature. Immunohistochemistry with human-specific CD31 (huCD31) and FOSB was performed as previously described (15). Images were acquired using EVOS FL AUTO2 Imaging System (ThermoFisher) or with the WLL1 confocal microscope (Leica), using 40x DRY objective and 0.75 Zoom factor. Used Antibodies are listed (table 6.3).

Antibody	Cat#	Manufacturer	Clone	Dilutions
VE-cadherin-A488	53-1449-42	eBiosciences	16B1	1:100
KDR-PE	FAB357P	R&D systems	89106	1:50
VEGFR3-PE	FAB3492P	R&D systems	54733	1:50
CD31-APC	17-0319-42	eBiosciences	WM59	1:200
CD34- PerCP-Cy5.5	347222	BD Pharmingen	8G12	1:100
CD105-VioBlue	130-099-666	Miltenyi Biotec	43A4E1	1:50
Phalloidin-A488	A12379	ThermoFisher	N/A	1:20
VE-Cadherin	2158S	Cell Signaling	Polyclonal	1:200
ZO-1	61-7300	ThermoFisher	Polyclonal	1:200
CD31	M0823	Dako	JC70A	1:200 IF 1:30 IHC
FOSB	2251S	Cell Signaling	5G4	1:200 IF 1:30000 WB
USP7	A300-033A	Bethyl Laboratories	Polyclonal	1:10000

Table 6.3: Antibodies and their dilutions.

6.3.14 Vasculogenesis *in vivo* in mice

All animal experiments were performed in accordance with legal regulations with approved protocols by the Central Commissie voor Dierproeven (CCD, Central Commission for Animal Experiments). Mice were maintained at the animal facility of Leiden University Medical Center (LUMC). Teratoma and Matrigel plug assays (figure 6.5a) were performed in eight week old male NSG mice (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ, Charles River).

6.3.15 Teratoma assay

The teratoma assay was performed on the parental hiPSC^{WT} and hiPSC$^{SERPINE\text{-}FOSB(D3)}$ as reported before (25). On the same day, three animals per cell-line were injected using the same batch of cells for each mouse.

6.3.16 *In vivo* mouse Matrigel plug assay

The Matrigel plug assay was performed as described previously (15, 26). Plugs were removed after 4 and 16 weeks. For each time points three mice were injected with hiPSC^{WT} and three mice with hiPSC$^{SERPINE\text{-}FOSB(D3)}$ (figure 6.5a). Each mouse was subcutaneously injected in the right and left flank with a mixture of hiPSC-ECs, human bone marrow-derived stromal cells (BMSCs) (PromoCell) and Matrigel (Corning). Vessel density was estimated by quantifying the human CD31+ area in serial sections as described previously (15).

6.3.17 Phosphotungstic acid-haematoxylin staining and analysis

PTAH staining was performed on 4 μm FFPE sections. Paraffin was removed with xylene and sections were rehydrated in an ethanol gradient. Sections were incubated for 15 minutes in 0.25% potassium permanganate then 5 minutes in 5% oxalic acid. Last, the sections were incubated for 24 hours in PTAH solution. To analyze thrombus formation in vessels all vessels with fibrin were counted in an area of 5.7 mm^2.

6.3.18 Whole genome and transcriptome sequencing and analysis

DNA was isolated for whole genome sequencing using the Wizard Genomic DNA Purification Kit (Promega) and sequenced on the BGISEQ-500 platform (BGI). Reads were aligned to the GRCh38/hg19 reference genome using the Burrow-Wheeler Aligner (v0.7.12) and further processed according to the GATK (Broad institute) best practice pipeline. Copy Number Analysis (CNA) was performed using VarScan (v2.2.4) and analyzed using DNAcopy R package (v3.6). Off-target sites for the used gRNAs were determined using an online tool (`crispr.cos.uni-heidelberg.de`). Data was visualized with the circlize R package (v0.4.3). RNA for transcriptome sequencing was isolated using Direct-zol RNA miniprep kit (Zymo Research). After library preparation, sequencing was performed on the BGISEQ-500 platform (BGI). Raw data was processed using the LUMC BIOPET Gentrap pipeline (`https://github.com/biopet/biopet`), which comprises FASTQ preprocessing, alignment and read quantification. Sickle (v1.2) was used to trim low-quality read ends15. Cutadapt (v1.1) was used for adapter clipping16, reads were aligned to the human reference genome GRCh38 using GSNAP (gmap-2014-12-23) (27, 28) and gene read quantification with htseq-count (v0.6.1p1) against the Ensembl v94. Gene length and GC content bias were normalized using the R package cqn (v1.28.1) (29). Genes were excluded if the number of reads was below 5 reads in $\leq$90% of the samples. The final dataset comprised gene expression levels of 6 samples and 16,510 genes. Differentially expressed genes were identified using generalized linear models as implemented in (30). P-values were adjusted using the Benjamini-Hochberg procedure and PFDR ≤ 0.05 was considered significant. Normalized RPKM values were log2 transformed and standardized across each gene using z-scores and heatmap was produced with the R package ggplot2 (v2.2.1). KEGG pathway enrichment analysis was carried out using Enrichr (31, 32) computational tool and $q < 0.05$ was used as the cutoff for significant pathways. Gene ontology (GO) enrichment analysis and cnetplot of selected GOs were done with R package clusterProfiler (v3.10.1) (33), $q < 0.05$ was used as the cutoff for significant GOs. Interaction networks of input genes were predicated using Interaction network analysis function of Ingenuity Pathway Analysis (IPA) software. Then, interac-

tions between specific genes and selected networks were generated using the Build function of IPA.

6.3.19 Statistical Analysis

Statistics and graphs for real-time PCR, proliferation, tube formation and barrier function were generated with GraphPad Prism (GraphPad Software). One-way ANOVA with Tukey's multiple comparison for the analysis of three or more groups or Mann-Whitney test for analysis of two groups were used. The data are reported as mean ±SD.

6.3.20 Data availability

Whole-genome and transcriptome sequencing data was deposited to the Sequence Read Archive under accession PRJNA448372.

6.4 Results

6.4.1 Introduction of t(7;19)(q22;q13) *SERPINE1-FOSB* translocation in hiPSCs

We used CRISPR/Cas9-facilitated gene targeting to introduce the t(7;19)(q22;q13) translocation in hiPSCs. We generated a fusion between intron 1 of *SERPINE1* and intron 1 of *FOSB*, which leads to the same novel start codon as found in PHE tumors from patients (figure 6.1a). Two double stranded DNA breaks were introduced in the genome guided by two gRNAs targeting *SERPINE1* intron 1 and *FOSB* intron 1. A repair template was provided for homologous directed recombination (HDR) containing two 1000 bp homology arms for *SERPINE1* and *FOSB*, separated by an excisable neomycin resistance cassette flanked by Flp-recombinase sequences (FRT) (figure 6.1a). A wild-type hiPSC line generated from an anonymous "healthy" donor using non-integrating Sendai virus (SeV) was used for targeting (15). hiPSCs were simultaneously transfected with vectors containing Cas9, gRNAs and HDR template (schematic overview of the targeting strategy in hiPSCs is shown in supplementary figure 6.1a). Neomycin selection allowed enrichment of hiPSCs with integration of the targeting template. The neomycin cassette was next removed by transient transfection of Flp-recombinase. Three color fluorescence in situ hybridization (FISH) revealed that translocations occurred relatively frequently, with 20 of 100 screened cells harboring a split of the *FOSB* bracketing probes (chromosome 19) and a colocalization of the distal *FOSB* probe to the *SERPINE1* (chromosome 7) (supplementary figure 6.1b). hiPSC clones derived from single cells were screened by PCR

and the presence of the *SERPINE1-FOSB* gene fusion was confirmed in 2 out of 73 (2.7% of targeted cells, clones D3 and G6) (figure 6.1b). Sanger sequencing of PCR products confirmed the correct translocation (figure 6.1a and supplementary figure 6.2a). This shows that although translocations between chromosomes 7 and 19 were relatively common events (20% of targeted cells showed translocation detected by FISH), most of these translocations likely occur via non-homologous end joining (NHEJ) and possibly contain large deletions/insertions. They were therefore not detected during PCR screening, resulting in only 2 correctly targeted clones (2.7% targeting efficiency). The targeted allele of hiPSC clone D3 was found to have an FRT remaining between *SERPINE1* and *FOSB* as expected (figure 6.1a), while this insert was absent in the targeted allele of hiPSC clone G6 due to the translocation occurring via NHEJ (supplementary figure 6.2a). The non-targeted wild-type alleles of *SERPINE1* and *FOSB* were also Sanger sequenced. In clone D3 a single nucleotide insertion was found in both the non-targeted wild-type *SERPINE1* intron 1 and the non-targeted wild-type *FOSB* intron 1 (supplementary figure 6.2b,c). In clone G6, a 9 base-pair deletion was found in the non-targeted wild-type *FOSB* intron 1 (supplementary figure 6.2b). In addition, clone G6 contained an insertion of ∼1220 bp of the repair template in the non-targeted wild-type *SERPINE1* intron 1, which was evident on the DNA gel and Sanger sequencing (figure 6.1b and supplementary figure 6.2c). Analysis of the corresponding cDNA showed presence of fused *SERPINE1* 5' UTR and *FOSB* exon 2 in both clones D3 and G6 (supplementary figure 6.2d) identical to that found in PHE patients, and the presence of correctly spliced wild-type *SERPINE1* (supplementary figure 6.2e). Neither clone D3 or G6 had other karyotypic abnormalities, other than the balanced t(7;19)(q22;q13) translocation, as seen using COmbined Binary RAtio Labeling Fluorescence in Situ Hybridization (COBRA-FISH) (figure 6.1c).

To verify that the targeting with CRISPR/Cas9 did not result in deleterious off-target effects, whole-genome sequencing was performed. No additional copy number variations, insertions or deletions, structural variants or single nucleotide variants (SNVs) were found in the coding genome of the targeted hiPSC clones D3 and G6 (hiPSC$^{SERPINE1\text{-}FOSB}$) compared to the parental control hiPSCs (hiPSC^{WT}) (figure 6.1d). Furthermore, the in silico predicted off target sites for the guide-RNAs used showed no additional alterations (synonymous or non-synonymous) compared with the untargeted parental control hiPSCs (figure 6.1d). To verify pluripotency of the targeted and parental control hiPSC lines, teratoma assay was performed in mice. Targeted and parental control hiPSCs retained the ability to form tissues derived from all three germ layers (endoderm, mesoderm and ectoderm), showing that CRISPR-Cas9 targeting has not affected pluripotency (figure 6.1e).

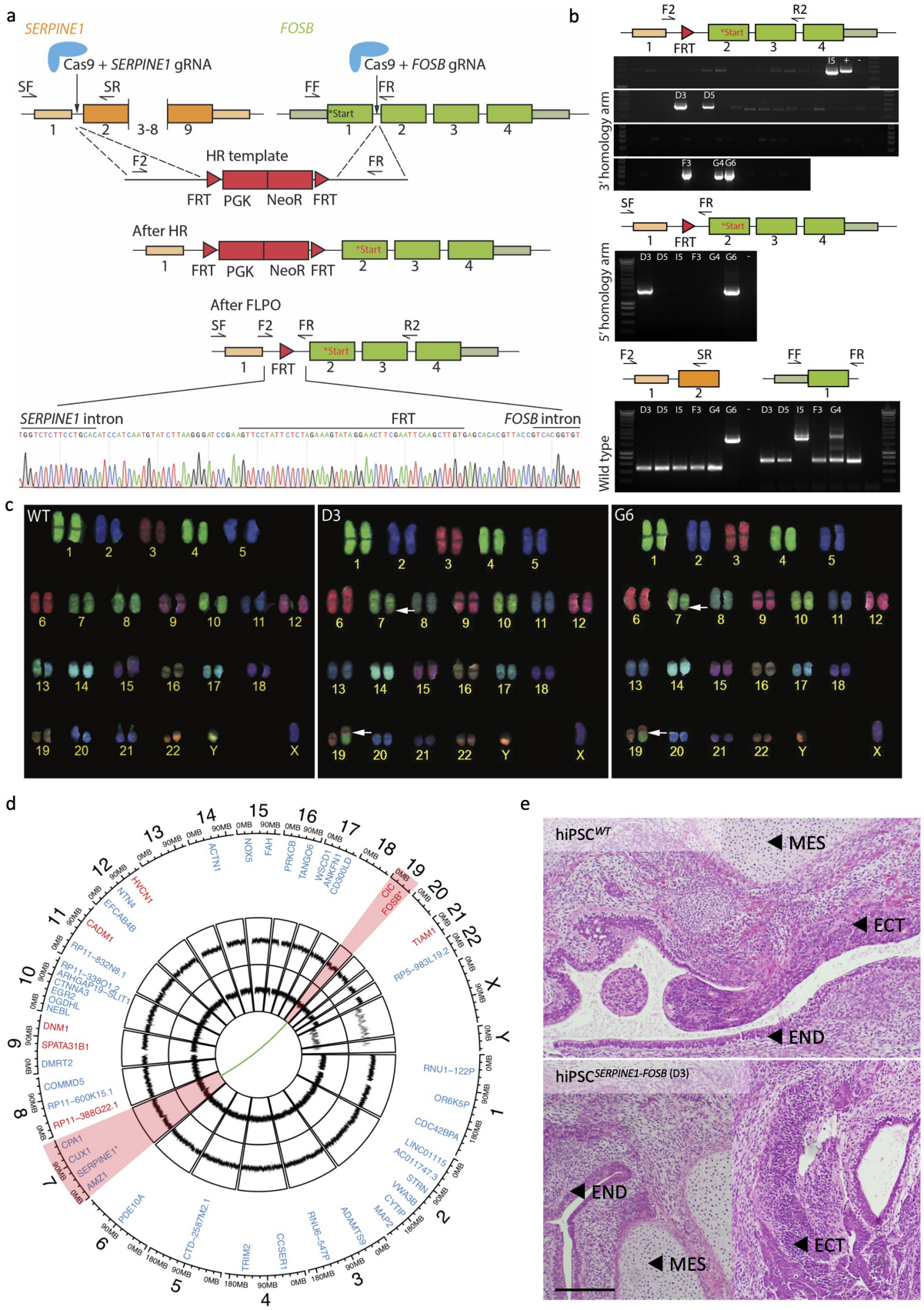


Figure 6.1: Caption on next page.

Figure 6.1: Generation and characterization of hiPSCs with a *SERPINE-FOSB* fusion. (a) Schematic overview of the targeting strategy for generation of a *SERPINE1-FOSB* gene fusion. Filled boxes are exons, lines introns. *FOSB* start codons are labeled in the figure, in black text represents the original start codon while the new start codon after the fusion is shown in red. Two double stranded breaks were introduced in the genome guided by two gRNAs in *SERPINE1* intron 1 and *FOSB* intron 1. A repair template used for homologous recombination (HR template) with neomycin resistance cassette flanked by Flp-recombinase sequences (FRT), as well targeted genomic locus prior (After HR) and after FLP-mediated neomycin removal (after FLPO). Bottom panel shows Sanger sequencing of PCR products from the clone with translocation validating HDR recombination of *SERPINE1* and *FOSB* with remaining FRT sequence left from the repair template (D3 clone). (b) Representative results of PCR screen on single cell-derived hiPSC clones using primers (F2, R2 and SF, SR; F2, SR and FF and FR) shown in panel above the PCR screen results. Two targeted clones (D3 and G6) were identified out of 73 screened clones. PCR shows clone G6 has a large insert in the *SERPINE1* wild-type allele. (c) COBRA-FISH on colony metaphase cells of WT, D3 and G6 hiPSC clones shows a balanced translocation t(7;19)(q22;q13); furthermore no additional chromosomal abnormalities were evident in any of the screened cells. (d) Whole genome sequencing was performed and the results are summarized in a Circos plot. The first layer shows all genes that are potential off-target sites for the gRNA for *FOSB* (red) and *SERPINE1* (blue). No mutations were found in the off-target sites and the surrounding 100 bases. The second and third layers show Copy Number Analysis (CNA) for clones D3 and G6 respectively compared to the isogenic control. No Copy Number Variations (CNVs) are detected. The green connection line shows the detected *SERPINE1-FOSB* fusion, as detected in both clones D3 and G6. Chromosomes 7 and 19, involved in the translocation, are highlighted in red. (e) Teratoma formation in mice. Top panel shows teratomas formed from the hiPSC^{WT}, the bottom panel from the hiPSC$^{SERPINE1\text{-}FOSB(D3)}$; two sections of each are shown. Cellular derivatives of the three germ lineages are indicated: mesoderm (MES), ectoderm (ECT) and endoderm (END). Scale bar indicates 200 μm.

6.4.2 Differentiation and characterization of ECs from hiPSC^{WT} and hiPSC$^{SERPINE1\text{-}FOSB}$ lines

hiPSCs with and without *SERPINE1-FOSB* fusion were differentiated into ECs using a protocol previously described (15–17) (figure 6.2a). hiPSC-ECs were purified on day 10 of differentiation by CD31+ cell selection, expanded and cryopreserved for further characterization. hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ differentiated from two targeted clones (D3 and G6) exhibited typical EC morphology (figure 6.2b and data not shown) and showed cell surface expression of known EC markers, such as vascular endothelial (VE)-cadherin (VEC), CD31, CD34, VEGFR2, VEGFR3 and CD105, as expected and in accordance with our previous findings (16) (figure 6.2c,d and supplementary figure 6.3a,b). Interestingly, hiPSC-ECs derived from both the D3 and G6 targeted clones displayed increased expression of CD105 (figure 6.2d and supplementary figure 6.3b), which is known to be upregulated in tumor endothelial cells (34) as well as in vascular tumors (35). Moreover, expression of *FOSB* mRNA showed a 4.9 log2 fold increase in hiPSC-ECs from clone D3 and a 5.9 log2 fold increase in hiPSC-ECs from clone G6 compared to the isogenic hiPSC-ECs derived from parental non-targeted hiPSC line (figure 6.2e and supplementary figure 6.3c). The increase in *FOSB* expression at the mRNA level was also evident as an increase in protein expression by Western blot, where FOSB was detected in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ but not in hiPSC-ECs^{WT}(figure 6.2f and supplementary figure 6.3d).

6.4.3 Functionality of ECs from hiPSC^{WT} and hiPSC$^{SERPINE1\text{-}FOSB}$ lines

To investigate the effect of the *SERPINE1-FOSB* fusion on hiPSC-ECs functionality, we next preformed assessment of proliferation and Matrigel tube formation assay. *SERPINE1-FOSB* fusion caused increased EC proliferation. The effect measured after 24 hours was most prominent in basal EC growth medium supplemented with 1% platelet-poor serum (PPS) (1.9-fold increase), followed by basal EC growth medium supplemented with both 1% PPS and VEGF (1.58 vs 2.27-fold) (figure 6.2g). No significant differences in EC proliferation were observed using complete EC growth medium (full) that in addition to VEGF also contained bFGF, indicating that FOSB overexpression caused by the *SERPINE1-FOSB* fusion may result in a VEGF-independent growth advantage for ECs. Matrigel tube formation assays showed significant increase in number of junctions (147 vs 218, $p<0.001$) and meshes (53 vs 85, $p<0.005$) in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ compared to isogenic control hiPSCs-ECs^{WT} after 48 hours (figure 6.2h).

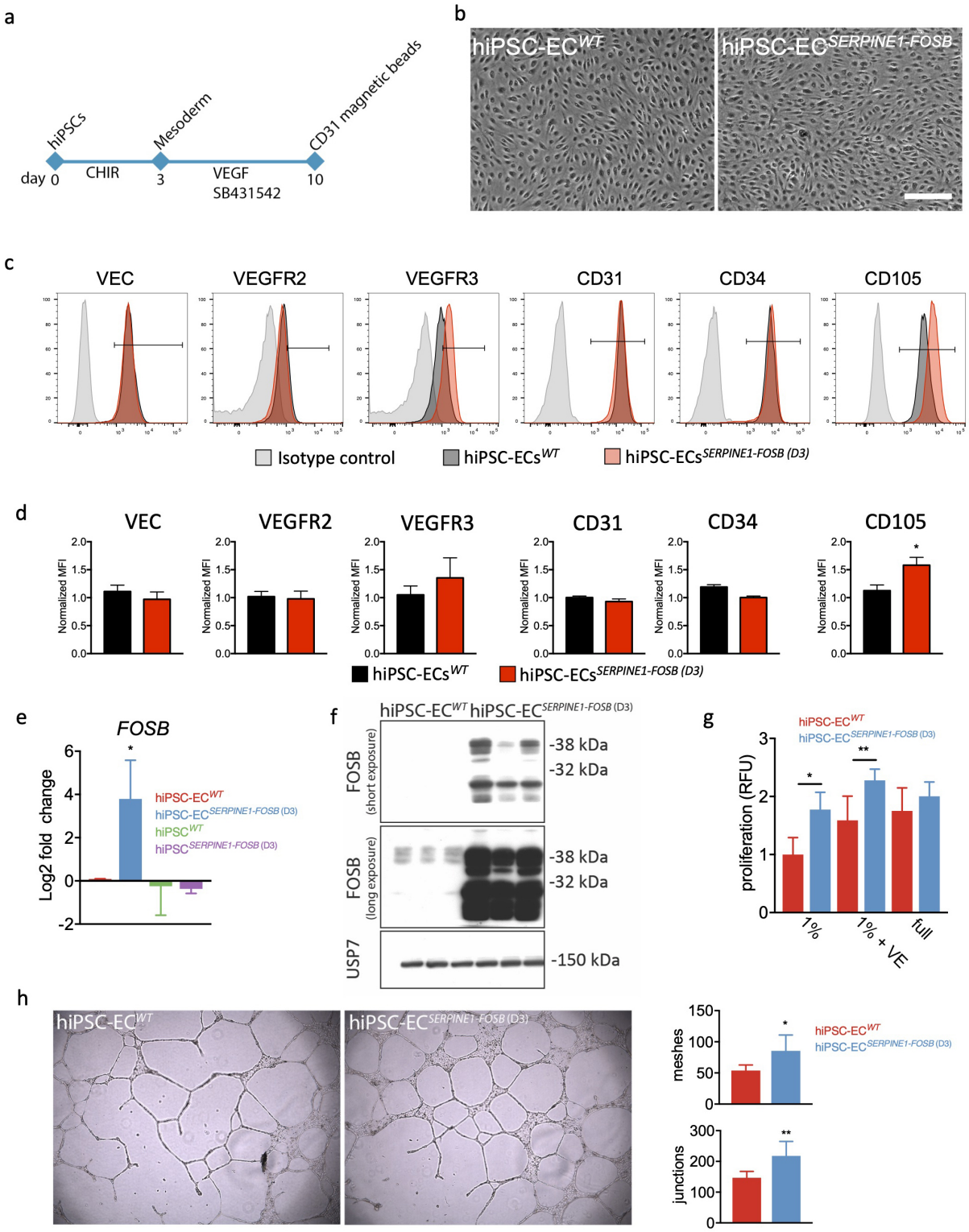


Figure 6.2: Caption on next page.

Figure 6.2: Differentiation and characterization of hiPSC-ECs with *SERPINE1-FOSB* fusion. (a) Schematic overview of the differentiation protocol and purification of ECs from hiPSCs. (b) Bright field images showing typical EC morphology of hiPSC-ECs. Scale bar represents 500 μm. (c) FACS analysis of EC markers expression on isolated ECs at passage 3 (P3) from hiPSC-ECs^{WT}(black filled histogram) and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ (red filled histogram), and relevant isotype control (gray filled histogram). (d) Quantification of normalized relative surface expression levels (MFI) of VEC, VEGFR2, VEGFR3, CD31, CD34 and CD105. Error bars are SD, * $p<0.005$. (e) Real-time qPCR analysis of FOSB expression in hiPSCs^{WT}, hiPSCs$^{SERPINE1\text{-}FOSB(D3)}$, hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$. Expression is determined relative to hiPSC-ECs^{WT}, shown as log2 fold change. Error bars are SD, * $p<0.05$. (f) Western blot of FOSB expression in hiPSC-ECs^{WT} and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$. Short and long exposure of the gel is shown. USP7 was used as a housekeeping control. (g) Analysis of hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ proliferation rates when cultured in basal endothelial cell growth medium supplemented with 1%PPS (1%), 1%PPS supplemented with 50 ng/ml VEGF (1% VE) or complete EC growth medium (full) for 24 hours. Proliferation was determined by using a Presto Blue assay. Error bars are shown as SD, * $p<0.0001$, ** $p<0.0005$. (h) Representative images of Matrigel tube formation assay using hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ at the 48h time point. Right panel shows quantification of the number of junctions and meshes. Error bars are SD, * $p<0.005$, ** $p<0.001$. All experiments were performed in triplicate using three independent batches of hiPSC-ECs.

6.4.4 Barrier function of ECs from hiPSC^{WT} and hiPSC$^{SERPINE1\text{-}FOSB}$ lines

Barrier function of hiPSC-ECs with and without the *SERPINE1-FOSB* fusion was next examined by real-time impedance spectroscopy with an integrated assay of electric wound healing, as demonstrated previously (15). The *SERPINE1-FOSB* fusion resulted in a significant decrease in barrier function of hiPSC-ECs (figure 6.3a,b and supplementary figure 6.4a,b). Barrier function depends on the integrity of cell junction complexes that form tight- and adherence junctions. Therefore, we also investigated junctional integrity in hiPSC-ECs with *SERPINE1-FOSB* fusion using the tight junctional marker zonula occludens (ZO)-1, the adherence junctional marker VEC and counterstained for F-actin (figure 6.3d). Presence of less organized "zig-zag" patterns of ZO1 and VEC was evident for hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ compared to the hiPSC-ECs^{WT}(figure 6.3d). Which is in line with reduces barrier function of hiPSC-ECs with *SERPINE1-FOSB* fusion.

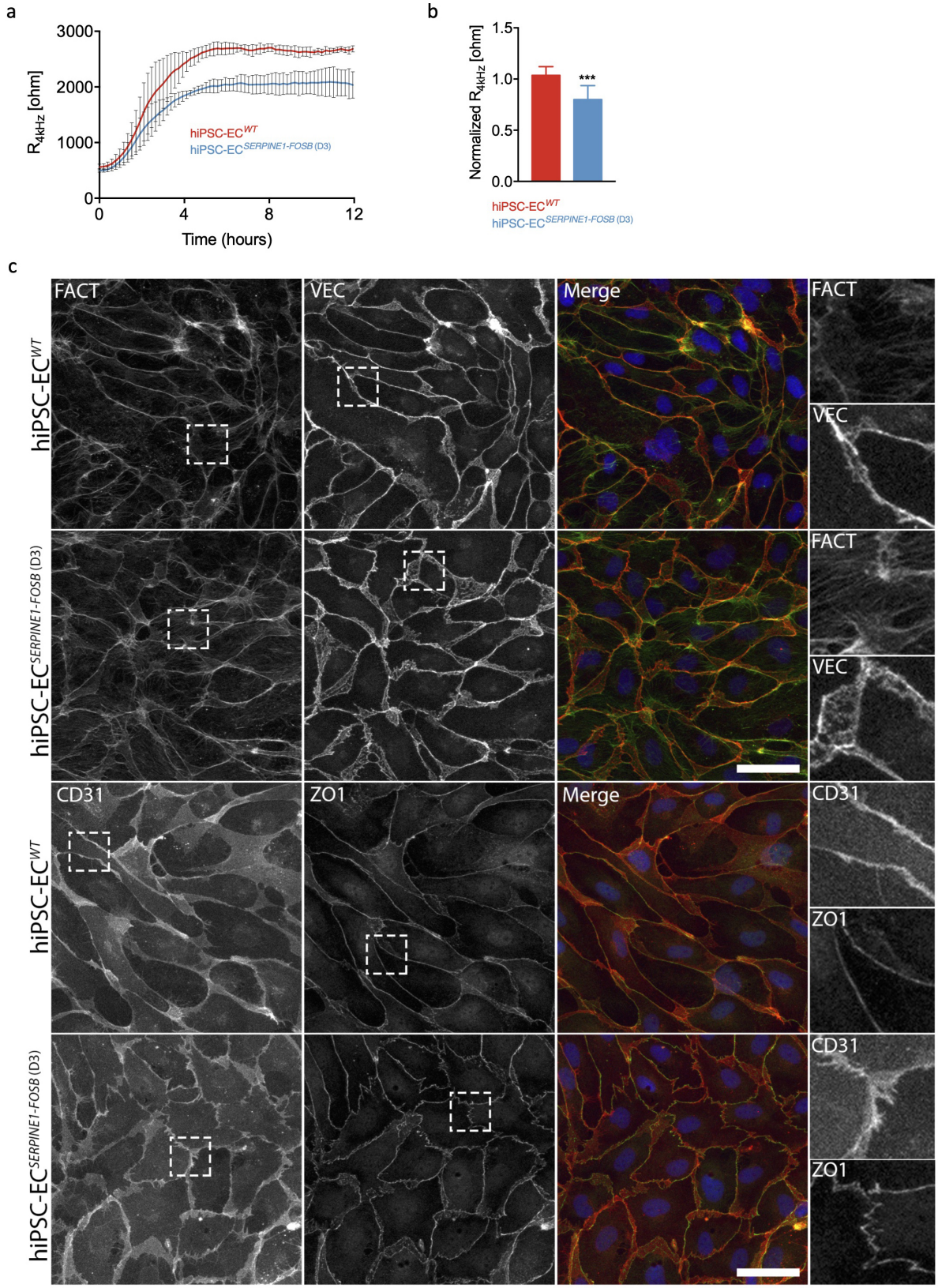


Figure 6.3: Caption on next page.

Figure 6.3: Assessment of barrier function of hiPSC-ECs with *SERPINE1-FOSB* fusion. (a) Representative absolute resistance of the EC monolayer in complete EC growth medium. Errors bars are shown as ±SD. (b) Normalized resistance [4 kHz] of the EC monolayer in complete EC growth medium. Error bars are shown as ±SD of six independent biological experiments. *** $p<0.001$. (c) Representative immunofluorescent images of FACT, VEC, CD31 and ZO1 to analyze the cell adherence and tight junctions. The top panels with merged images show FACT in green, VEC in red and DAPI in blue. The bottom panels with merged images show CD31 in red, ZO1 in green and DAPI in blue. The right panels show further enlarged areas selected from the shown images (dashed squares). Scale bar represents 50 μm. Experiments were performed in triplicate using three independent batches of hiPSC-ECs.

6.4.5 Functionality of hiPSC-ECs from hiPSC^{WT} and hiPSC$^{SERPINE1\text{-}FOSB}$ lines *in vivo* in mice

In order to test the functionality and the ability to form functional perfused blood vessels, hiPSC-ECs with and without *SERPINE-FOSB* translocation were injected in mice in a Matrigel Plug Assay that allows assessment of vasculogenesis, as described previously (15). Matrigel plugs were excised and analyzed 4- and 16-weeks post-transplantation. Both hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ formed stable vessels *in vivo* composed of human ECs evident at the 4-weeks (figure 6.4a) and 16-weeks post-transplantation (figure 6.4c). Quantification of the vessel density showed comparable areas covered by human vessels, demonstrating that both hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ had similar abilities to form vessels *in vivo* (figure 6.4b,d). The vessels were perfused (as indicated by the presence of red blood cells) (figure 6.4a,c). Moreover, FOSB positive ECs were evident in the Matrigel plugs with hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$, but not the Matrigel plugs with hiPSC-ECs^{WT}(figure 6.4e). Furthermore, FOSB positive hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ also invaded the surrounding mouse soft tissues (the striated muscle) at 16-weeks post-transplantation in two of the three mice whereas this was not observed in any of the mice with hiPSC-ECs^{WT}transplants (figure 6.4f, and 6.4g).

The hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ vessels at 16 weeks were disorganized and disarrayed and often contained thrombi (conglomeration of fibrin and platelets, containing red and white blood cells) (figure 6.4h,i). Thrombi were quantified using phosphotungstic acid-haematoxylin (PTAH) staining (thrombus positive vessels 20.67 vs 81.33 counted in 5.7 mm^2, n=3, p=0.1) (figure 6.5j).

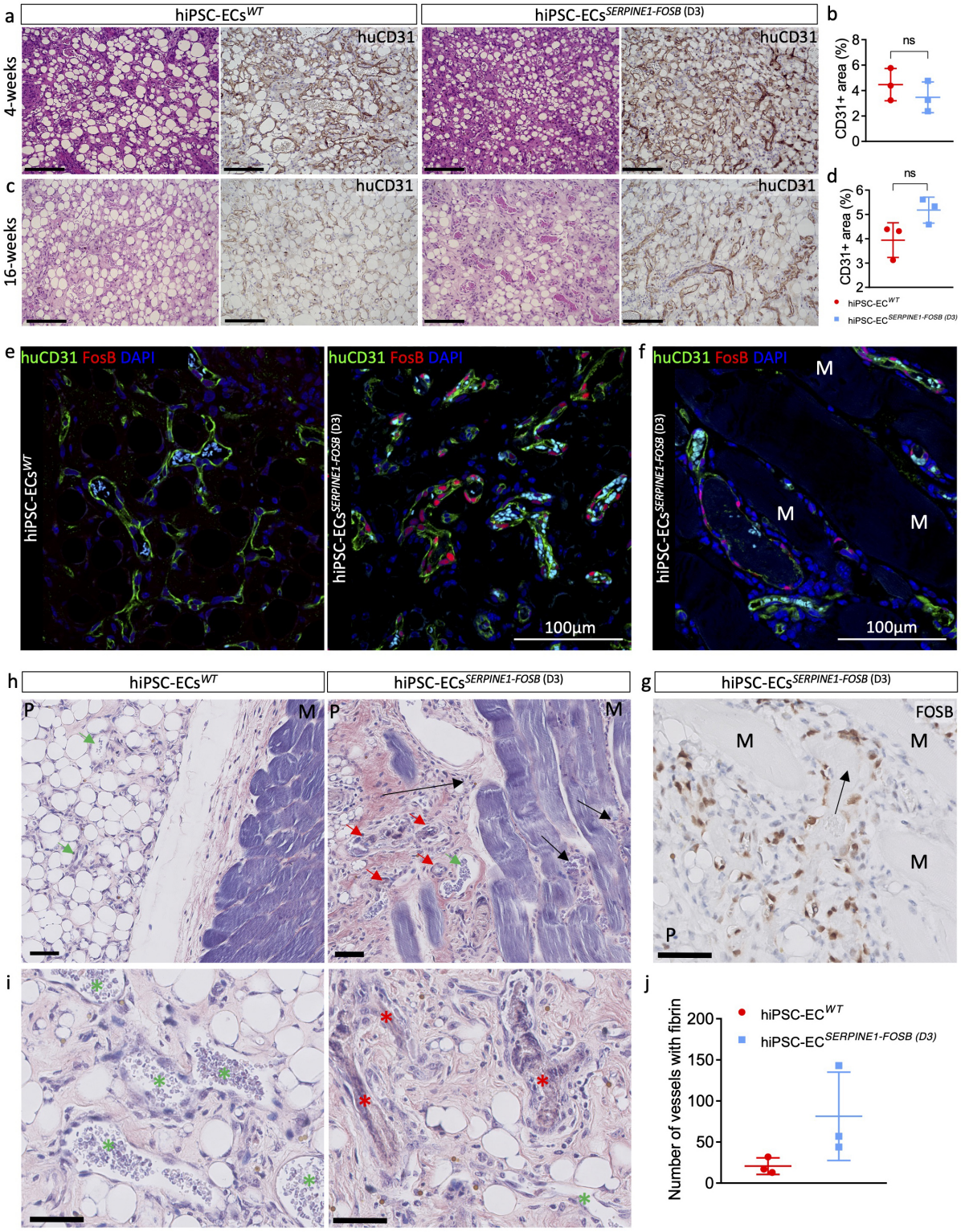


Figure 6.4: Caption on next page.

Figure 6.4: *In vivo* vasculogenesis assay for hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$. (a,c) H&E and CD31 staining of FFPE tissue from the Matrigel plug harvested after 4- (shown in panel (a)) and 16- (shown in panel (c)) weeks. Both hiPSC^{WT} and hiPSC$^{SERPINE\text{-}FOSB(D3)}$ show vessel formation. Scale bar indicates 100 μm. (b,d) Vessel density was estimated by quantification of the human CD31+ area at 4- and 16-weeks. The 4- and 16-week timepoints showed no significant difference in human CD31+ area. (e,f) Double immunofluorescent staining with FOSB and human CD31 antibodies counterstained with DAPI on cryosections from Matrigel plug containing hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE\text{-}FOSB(D3)}$. FOSB is shown in red, CD31 in green and DAPI is blue. The left panel shows the hiPSC-ECs^{WT}experiment, the right panel the hiPSC-ECs$^{SERPINE\text{-}FOSB(D3)}$ experiment. (g) FOSB IHC on FFPE tissue from Matrigel plug with hiPSCSERPINE-FOSB (D30) ECs. Scale bar 50 μm. Panels (f,g) show invasion of FOSB positive hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ into the striated muscle at 16-weeks post-transplantation. Surrounding mouse muscle (indicated by M) and the matrigel plug (indicated by a P). (h,i) PTAH stained sections from the *in vivo* vasculogenesis assay. hiPSC-ECs^{WT}(left panel) and hiPSC-ECs$^{SERPINE\text{-}FOSB(D3)}$ (right panel) are shown. Both images show the surrounding mouse muscle (indicated by M) and the Matrigel plug (indicated by a P). Vessels with and without thrombi are indicated by red and green arrows, respectively (h) and red and green stars (i). The black arrows indicate areas with infiltration in the mouse muscle. The scale bar indicates 50 μm. (j) Quantification of vessels containing PTAH positive thrombi in hiPSC-ECs^{WT}and hiPSC-ECs$^{SERPINE\text{-}FOSB(D3)}$, in an area of 5.7 mm2, n=3 and p=0.1.

6.4.6 Transcriptome analysis of ECs from hiPSC^{WT} and hiPSC$^{SERPINE1\text{-}FOSB}$ lines

The transcriptomes of hiPSC-ECs with and without *SERPINE1-FOSB* fusion were compared. 630 and 592 differentially expressed genes (DEGs)(PFDR$\leq$0.05) were upregulated and downregulated respectively in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ compared to hiPSC-ECs^{WT}(figure 6.5a). Both *FOSB* and *SERPINE1* were significantly upregulated in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ compared to hiPSC-ECs^{WT}. Enrichment analysis using the KEGGs (Kyoto Encyclopedia of Genes and Genomes) pathway database revealed several signaling pathways significantly enriched in DEGs upregulated in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$. These included focal adhesion, ECM-receptor interaction, Pathways in cancer, PI3K-Akt, MAPK, TGF-beta and HIF-1 signaling pathways, and Glycolysis/Gluconeogenesis (figure 6.5b and supplementary figure 6.5). Upregulation of glycolytic genes in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ (supplementary figure 6.5h) indicates possible changes in the metabolic state of ECs, as previously demonstrated for tumor ECs (36). No signaling pathways were significantly enriched in DEGs upregulated in hiPSC-EC^{WT}. Gene Ontology (GO) enrichment analysis revealed alterations in the following biological processes in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$: extracellular matrix organization, angiogenesis, cell-matrix adhesion,

inflammatory response, cell junction organization, regulation of TGF-beta receptor signaling pathway and others (figure 6.5c). By contrast, response to interferon-gamma was the only biological process significantly enriched in DEGs upregulated in hiPSC-EC^{WT} (figure 6.5c). To demonstrate the relationship between the genes and identified GOs, a gene network map was next constructed using DEGs upregulated in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ (total of 182 genes) (figure 6.5d). Gene interaction analysis was next performed using Ingenuity pathway analysis (IPA). Gene interaction networks related to cancer, cellular movement and growth, and TGF-β signaling pathway were used to demonstrate interaction between the identified dysregulated genes and *FOSB* in hiPSC-ECs$^{SERPINE1\text{-}FOSB}$ (figure 6.5e). FOSB regulates *SERPINE1* directly, which is in line with our previous finding that truncated FOSB was able to regulate its own transcription (22), as well as via SMAD3. Both exhibit a self-regulatory mechanism, which could further activate many genes in the network of cellular growth and proliferation and cancer processes directly or indirectly through activation of the TGF-β signaling pathway (figure 6.5e).

6.5 Discussion

There is an urgent need for *in vitro* models to study rare translocation-driven tumors, both to identify the functional consequences of the translocation, and to identify potential therapeutic targets. We used CRISPR/Cas9 to induce a tumor-associated translocation in hiPSCs, specifically the *SERPINE1-FOSB* translocation in hiPSCs to model the rare vascular tumor PHE. Two hiPSC clones among 73 clones screened contained the translocation. In one of the clones (D3), the translocation was introduced via HDR using the donor DNA template, while in the second clone (G6) the translocation occurred via NHEJ. As the breakpoints were in intronic regions of the two involved genes, in both clones the transcribed and spliced RNA resulted in the typical *SERPINE1-FOSB* chimeric RNA. Other groups have also shown that CRISPR/Cas9 can be used to introduce chromosomal translocations in other cells, notably hMSCs via both NHEJ, but also via HDR using donor DNA templates and additional exposure to low doses of DNA-PKC inhibitor (NU7441) to block NHEJ (13). Although blocking NHEJ to increase HDR proved efficient in mouse embryonic stem cells (mESCs) and human cell lines (37), the efficiency of this approach remains debated in hiPSCs (38). Our results showed that both NHEJ and HDR pathways can be used to introduce translocations in hiPSCs. It would be interesting to investigate further whether the use of NHEJ inhibitors can indeed enhance the occurrence of translocations via HDR.

Whole genome sequencing of both parental and targeted hiPSC lines showed no deleterious structural variations, copy number variations or mutations at the predicted off-target

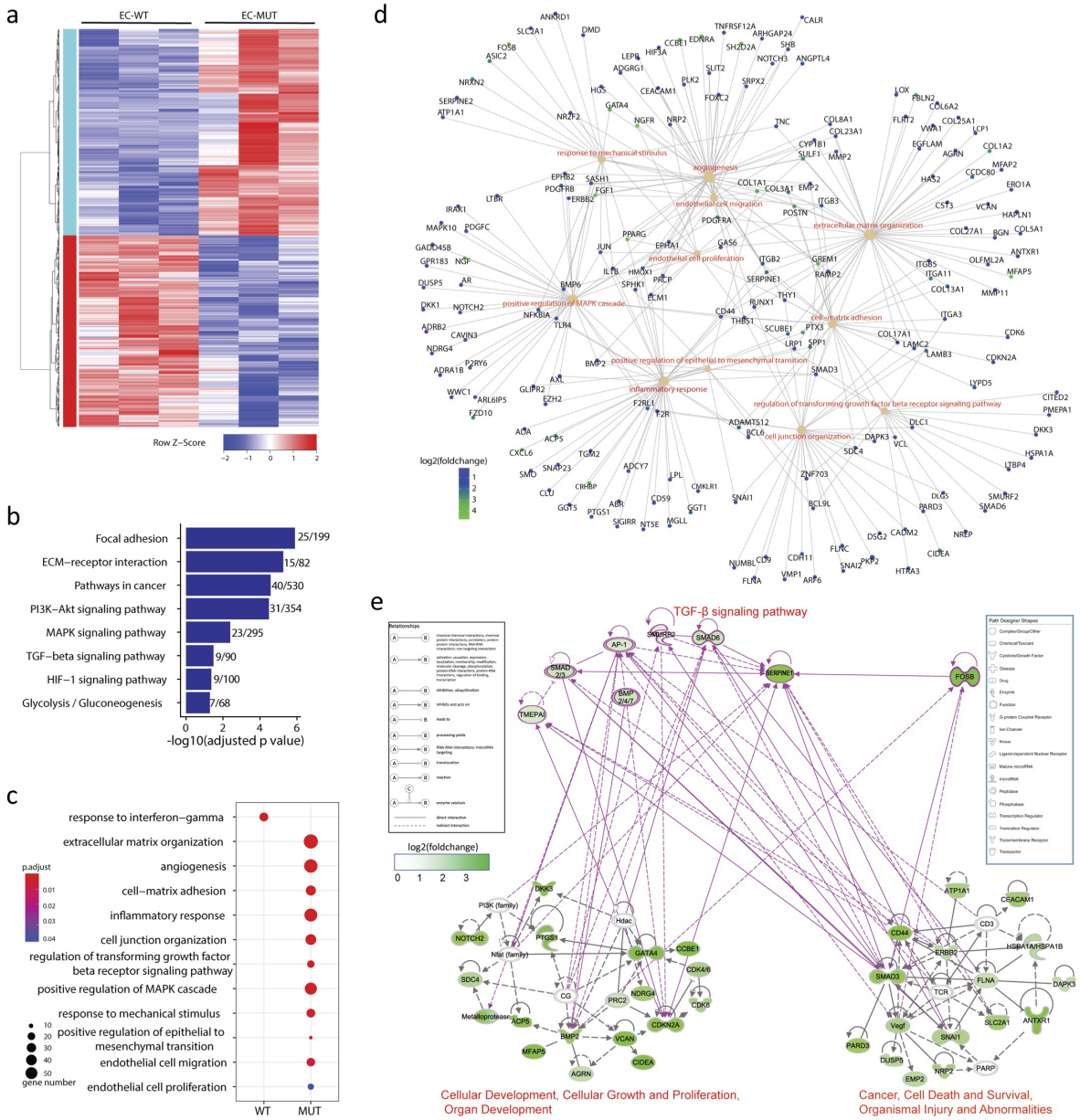


Figure 6.5: Transcriptome analysis of hiPSC-ECs with and without *SERPINE1-FOSB* translocation. (a) Hierarchical clustering analysis (HCA) of differentially expressed genes (DEGs) between hiPSC-ECs^{WT}(WT) and hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ (MUT) samples. 630 and 592 significantly upregulated and downregulated genes in MUT were identified compared to WT ECs (PFDR≤0.05). (b) Representative KEGG pathways enriched in upregulated DEGs in MUT ECs (-log10(adjusted p value)) and number of enriched genes within total genes of each pathway are shown. (c) Representative Gene ontologies (GOs) that enriched in downregulated (WT) and upregulated (MUT) DEGs. Size and color indicate gene number and adjusted p value of each GO. (d) Cnetplot of genes associated with enriched GOs (c), color indicates the log2(fold change) of gene expression in MUT compared to WT. (e) Gene interaction network constructed using all genes shown in (c) with Ingenuity pathway analysis (IPA). *SERPINE1*, *FOSB* were added manually. Interactions among *FOSB* and *TGF-β* signaling pathway and two networks were generated using IPA. Color indicates the log2(fold change) of gene expression in MUT compared to WT.

sites for the gRNAs. Any phenotypic changes observed were thus most likely associated with the *SERPINE1-FOSB* translocation.

We recently showed that overexpression of truncated FOSB in human umbilical vein ECs (HUVECs) recapitulated some features of PHE pathology (22). However, the drawback of overexpression is that they lack the regulatory elements for cell type-specific expression at levels found in tumor cells. Here we addressed the shortcomings of our original model by introducing the truncated protein under the endogenous *SERPINE1* regulatory elements via *SERPINE1-FOSB* fusion using CRISPR/Cas9-induced translocation in hiPSCs, thereby recreating the fusion with endogenous regulatory elements of *SERPINE1*. We found that the *SERPINE1-FOSB* fusion results in upregulation of FOSB expression specifically in ECs, and not in undifferentiated hiPSCs, in line with known high SERPINE1 expression in vascular cells and its function as a direct transcriptional target of the activator protein 1 (AP-1) family of proteins that includes FOSB (4, 39). Thus, we here show that self-regulation of its own promoter, and thereby of the expression of the fusion product that is considered the driver alteration in PHE, only occurs in endothelial cells and not in undifferentiated hiPSCs. These results indicate lineage restricted expression of the fusion and confirm that PHE should be considered a vascular tumor.

Vasculogenesis assays *in vivo* in mice in which hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ or hiPSC-ECs^{WT}were co-injected with bone marrow stromal cells (BMSCs) supported our *in vitro* findings and showed most strikingly the infiltrative growth pattern reminiscent of human PHE (5). Vessels from hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ were haphazardly arranged compared to hiPSC-ECs^{WT}and contained significantly higher numbers of fibrin thrombi, in two of three hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$. These results are in line with the *in vitro* barrier function analysis and suggest that the endothelium is aberrant inducing thrombi formation. However, some aspects of PHE were not recapitulated *in vivo*. The tumor cells typically do not form vessels in PHE but instead are spindle-shaped and co-express endothelial markers (CD31 and ERG) and keratin AE1AE3, features that are absent in our *in vivo* model. It might be that the 16 weeks time frame is not enough to develop these features in vivo. Invasion of hiPSC-ECs$^{SERPINE1\text{-}FOSB(D3)}$ into surrounding mouse soft tissue was observed at 16-weeks, but not at the 4-week time-point, which may suggest that the development of the phenotype takes time.

Transcriptome analysis of hiPSC-ECs with the fusion revealed differentially expressed genes associated with several pathways that are known to be related to cancer, such as TGF-beta signaling, adhesion, metabolism, inflammatory response, angiogenesis and endothelial cell migration. These are all linked to the phenotypes we observed in vitro and *in vivo* in our model, and recapitulate some aspects of PHE. Moreover, these pathways that we here identify to be regulated by the *SERPINE1-FOSB* fusion provide rationale

to develop targeted treatment strategies for inoperable multifocal PHE patients. In line with our previous report of a patient with a complete clinical remission following the multi-tyrosine kinase inhibitor telatinib, we confirm upregulated MAPK signaling and overexpression of PDGFRA and -B induced by the fusion in the current model. Moreover, we identify PI3K-Akt signaling which can be targeted using mTOR inhibitors. Indeed, anecdotal responses to mTOR inhibition in patients with PHE have been reported (40–42).

In summary, we showed that hiPSCs and hiPSC-ECs can be used to model fusion-driven tumors using CRISPR/Cas9 and a donor DNA template to introduce the translocation. The differentiated hiPSC-ECs carrying the pathognomonic translocation gave insights into the tumorigenesis of PHE, and elucidated the pathways regulated by the fusion product, that may provide rationale to develop targeted treatment strategies for inoperable multifocal PHE. Overall, this approach facilitated the elucidation of the role of specific fusion genes in the development of specific rare cancer subtypes for which cell lines are presently lacking.

Bibliography

[1] Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation; 2007. Available from: `http://www.nature.com/articles/nrc2091`.

[2] Mertens F, Antonescu CR, Mitelman F. Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. Genes Chromosomes Cancer. 2016;55(4):291–310. doi:10.1002/gcc.22335.

[3] Trombetta D, Magnusson L, von Steyern FV, Hornick JL, Fletcher CDM, Mertens F. Translocation t(7;19)(q22;q13)-a recurrent chromosome aberration in pseudomyogenic hemangioendothelioma? Cancer Genetics. 2011;204(4):211–215. doi:10.1016/j.cancergen.2011.01.002.

[4] Walther C, Tayebwa J, Lilljebjorn H, Magnusson L, Nilsson J, von Steyern FV, et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. J Pathol. 2014;232(5):534–540. doi:10.1002/path.4322.

[5] Hornick JL, Fletcher CD. Pseudomyogenic hemangioendothelioma: a distinctive, often multicentric tumor with indolent behavior. Am J Surg Pathol. 2011;35(2):190–201. doi:10.1097/PAS.0b013e3181ff0901.

[6] Hung YP, Fletcher CDM, Hornick JL. FOSB is a useful diagnostic marker for pseudomyogenic hemangioendothelioma. American Journal of Surgical Pathology. 2017;41(5):596–606. doi:10.1097/PAS.0000000000000795.

[7] Billings SD, Folpe AL, Weiss SW. Epithelioid Sarcoma-Like Hemangioendothelioma. The American journal of surgical pathology. 2003;27(1):48–57. doi:10.1097/PAS.0b013e31821caf1c.

[8] Ide YH, Tsukamoto Y, Ito T, Watanabe T, Nakagawa N, Haneda T, et al. Penile pseudomyogenic hemangioendothelioma/epithelioid sarcoma-like hemangioendothelioma with a novel pattern of SERPINE1-FOSB fusion detected by RT-PCR–report of a case. Pathol Res Pract. 2015;211(5):415–420. doi:10.1016/j.prp.2015.02.003.

[9] Sanchez-Rivera FJ, Jacks T. Applications of the CRISPR-Cas9 system in cancer biology. Nat Rev Cancer. 2015;15(7):387–395. doi:10.1038/nrc3950.

[10] Schneidawind C, Jeong J, Schneidawind D, Kim IS, Duque-Afonso J, Wong SHK, et al. MLL leukemia induction by t(9;11) chromosomal translocation in human hematopoietic stem cells using genome editing. Blood advances. 2018;2(8):832–845. doi:10.1182/bloodadvances.2017013748.

[11] Torres R, Martin MC, Garcia A, Cigudosa JC, Ramirez JC, Rodriguez-Perales S. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. Nat Commun. 2014;5:3964. doi:10.1038/ncomms4964.

[12] Torres-Ruiz R, Martinez-Lage M, Martin MC, Garcia A, Bueno C, Castaño J, et al. Efficient Recreation of t(11;22) EWSR1-FLI1+in Human Stem Cells Using CRISPR/Cas9. Stem Cell Reports. 2017;8(5):1408–1420. doi:10.1016/j.stemcr.2017.04.014.

[13] Vanoli F, Tomishima M, Feng W, Lamribet K, Babin L, Brunet E, et al. CRISPR-Cas9-guided oncogenic chromosomal translocations with conditional fusion protein expression in human mesenchymal cells. Proc Natl Acad Sci U S A. 2017;114(14):3696–3701. doi:10.1073/pnas.1700622114.

[14] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131(5):861–872. doi:10.1016/j.cell.2007.11.019.

[15] Halaidych OV, Freund C, van den Hil F, Salvatori DCF, Riminucci M, Mummery CL, et al. Inflammatory Responses and Barrier Function of Endothelial Cells Derived from Human Induced Pluripotent Stem Cells. Stem Cell Reports. 2018;10(5):1642–1656. doi:10.1016/j.stemcr.2018.03.012.

[16] Orlova VV, Drabsch Y, Freund C, Petrus-Reurer S, Van Den Hil FE, Muenthaisong S, et al. Functionality of endothelial cells and pericytes from human pluripotent stem cells demonstrated in cultured vascular plexus and zebrafish xenografts. Arteriosclerosis, Thrombosis, and Vascular Biology. 2014;34(1):177–186. doi:10.1161/ATVBAHA.113.302598.

[17] Orlova VV, Van Den Hil FE, Petrus-Reurer S, Drabsch Y, Ten Dijke P, Mummery CL. Generation, expansion and functional analysis of endothelial cells and pericytes derived from human pluripotent stem cells. Nature Protocols. 2014;9(6):1514–1531. doi:10.1038/nprot.2014.102.

[18] Vidigal JA, Ventura A. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat Commun. 2015;6:8083. doi:10.1038/ncomms9083.

[19] Kranz A, Fu J, Duerschke K, Weidlich S, Naumann R, Stewart AF, et al. An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. Genesis. 2010;48(8):512–520. doi:10.1002/dvg.20641.

[20] Rossi S, Szuhai K, Ijszenga M, Tanke HJ, Zanatta L, Sciot R, et al. EWSR1-CREB1 and EWSR1-ATF1 fusion genes in angiomatoid fibrous histiocytoma. Clinical Cancer Research. 2007;13(24):7322–7328. doi:10.1158/1078-0432.CCR-07-1744.

[21] van IJzendoorn DG, de Jong D, Romagosa C, Picci P, Benassi MS, Gambarotti M, et al. Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. Genes Chromosomes Cancer. 2015;54(9):565–574. doi:10.1002/gcc.22269.

[22] van IJzendoorn DGP, Sleijfer S, Gelderblom H, Eskens FALM, van Leenders GJLH, Szuhai K, et al. Telatinib Is an Effective Targeted Therapy for Pseudomyogenic Hemangioendothelioma. Clinical Cancer Research. 2018;24(11):2678–2687. doi:10.1158/1078-0432.CCR-17-3512.

[23] Szuhai K, Tanke HJ. COBRA: Combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping. Nature Protocols. 2006;1(1):264–275. doi:10.1038/nprot.2006.41.

[24] Van IJzendoorn DGP, Forghany Z, Liebelt F, Vertegaal AC, Jochemsen AG, Bovée JVMG, et al. Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis. Journal of Biological Chemistry. 2017;292(52):21282–21290. doi:10.1074/jbc.C117.815845.

[25] Salvatori DCF, Dorssers LCJ, Gillis AJM, Perretta G, van Agthoven T, Gomes Fernandes M, et al. The MicroRNA-371 Family as Plasma Biomarkers for Monitoring Undifferentiated and Potentially Malignant Human Pluripotent Stem Cells in Teratoma Assays. Stem Cell Reports. 2018;11(6):1493–1505. doi:10.1016/j.stemcr.2018.11.002.

[26] Sacchetti B, Funari A, Remoli C, Giannicola G, Kogler G, Liedtke S, et al. No Identical "Mesenchymal Stem Cells" at Different Times and Sites: Human Committed Progenitors of Distinct Origin and Differentiation Potential Are Incorporated as Adventitial Cells in Microvessels. Stem Cell Reports. 2016;6(6):897–913. doi:10.1016/j.stemcr.2016.05.011.

[27] Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics (Oxford, England). 2010;26(7):873–81. doi:10.1093/bioinformatics/btq057.

[28] Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005;21(9):1859–1875. doi:10.1093/bioinformatics/bti310.

[29] Hansen KD, Irizarry RA, WU Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics. 2012;13(2):204–216. doi:10.1093/biostatistics/kxr054.

[30] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England). 2010;26(1):139–40. doi:10.1093/bioinformatics/btp616.

[31] Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14(1):128. doi:10.1186/1471-2105-14-128.

[32] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research. 2016;44(W1):W90–7. doi:10.1093/nar/gkw377.

[33] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology. 2012;16(5):284–7. doi:10.1089/omi.2011.0118.

[34] Miller DW, Graulich W, Karges B, Stahl S, Ernst M, Ramaswamy A, et al. Elevated expression of endoglin, a component of the TGF-?-receptor complex, correlates with proliferation of tumor endothelial cells. International Journal of Cancer. 1999;81(4):568–572. doi:10.1002/(SICI)1097-0215(19990517)81:4<568::AID-IJC11>3.0.CO;2-X.

[35] Verbeke SL, Bertoni F, Bacchini P, Oosting J, Sciot R, Krenacs T, et al. Active TGF-beta signaling and decreased expression of PTEN separates angiosarcoma of bone from its soft tissue counterpart. Mod Pathol. 2013;26(9):1211–1221. doi:10.1038/modpathol.2013.56.

[36] Cantelmo AR, Conradi LC, Brajic A, Goveia J, Kalucka J, Pircher A, et al. Inhibition of the Glycolytic Activator PFKFB3 in Endothelium Induces Tumor Vessel Normalization, Impairs Metastasis, and Improves Chemotherapy. Cancer Cell. 2016;30(6):968–985. doi:10.1016/j.ccell.2016.10.006.

[37] Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell. 2017;168(1-2):20–36. doi:10.1016/j.cell.2016.10.044.

[38] Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. 2015;33(5):543–548. doi:10.1038/nbt.3198.

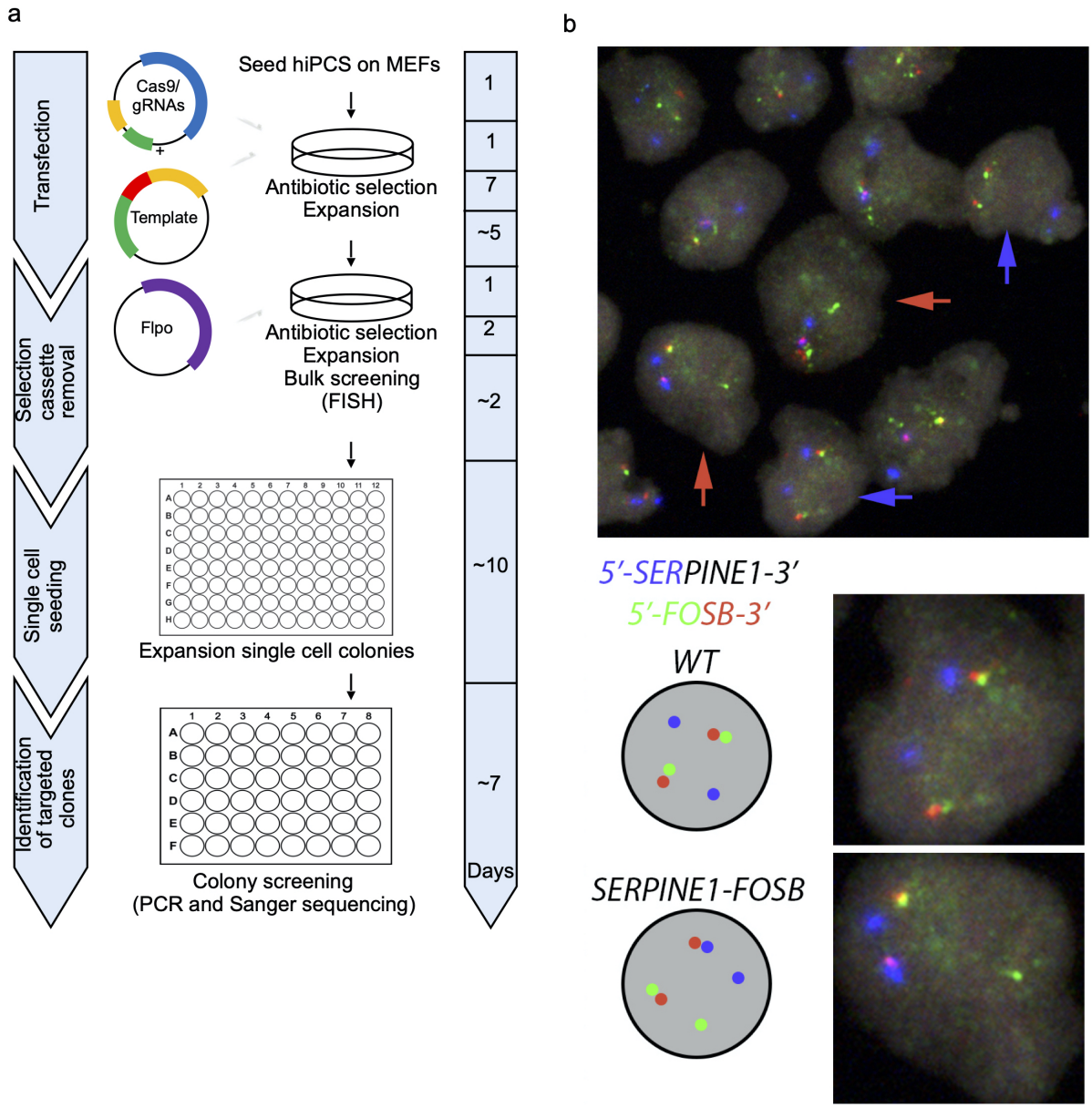
[39] Milde-Langosch K. The Fos family of transcription factors and their role in tumourigenesis. European Journal of Cancer. 2005;41(16):2449–2461. doi:10.1016/j.ejca.2005.08.008.

[40] Gabor KM, Sapi Z, Tiszlavicz LG, Fige A, Bereczki C, Bartyik K. Sirolimus therapy in the treatment of pseudomyogenic hemangioendothelioma. Pediatr Blood Cancer. 2018;65(2). doi:10.1002/pbc.26781.

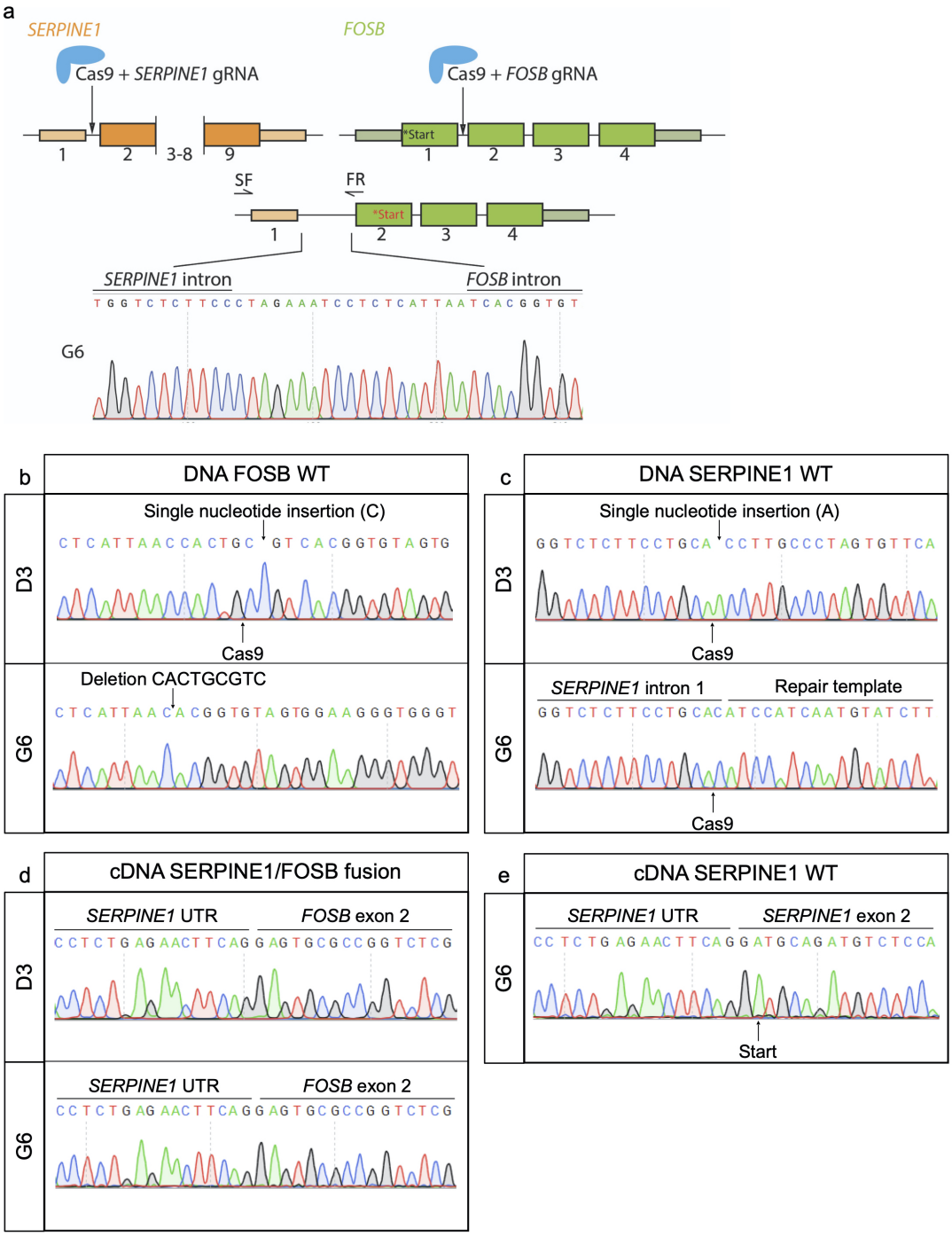
[41] Joseph J, Wang WL, Patnana M, Ramesh N, Benjamin R, Patel S, et al. Cytotoxic and targeted therapy for treatment of pseudomyogenic hemangioendothelioma. Clin Sarcoma Res. 2015;5:22. doi:10.1186/s13569-015-0037-8.

[42] Ozeki M, Nozawa A, Kanda K, Hori T, Nagano A, Shimada A, et al. Everolimus for Treatment of Pseudomyogenic Hemangioendothelioma. J Pediatr Hematol Oncol. 2017;39(6):e328–e331. doi:10.1097/MPH.0000000000000778.

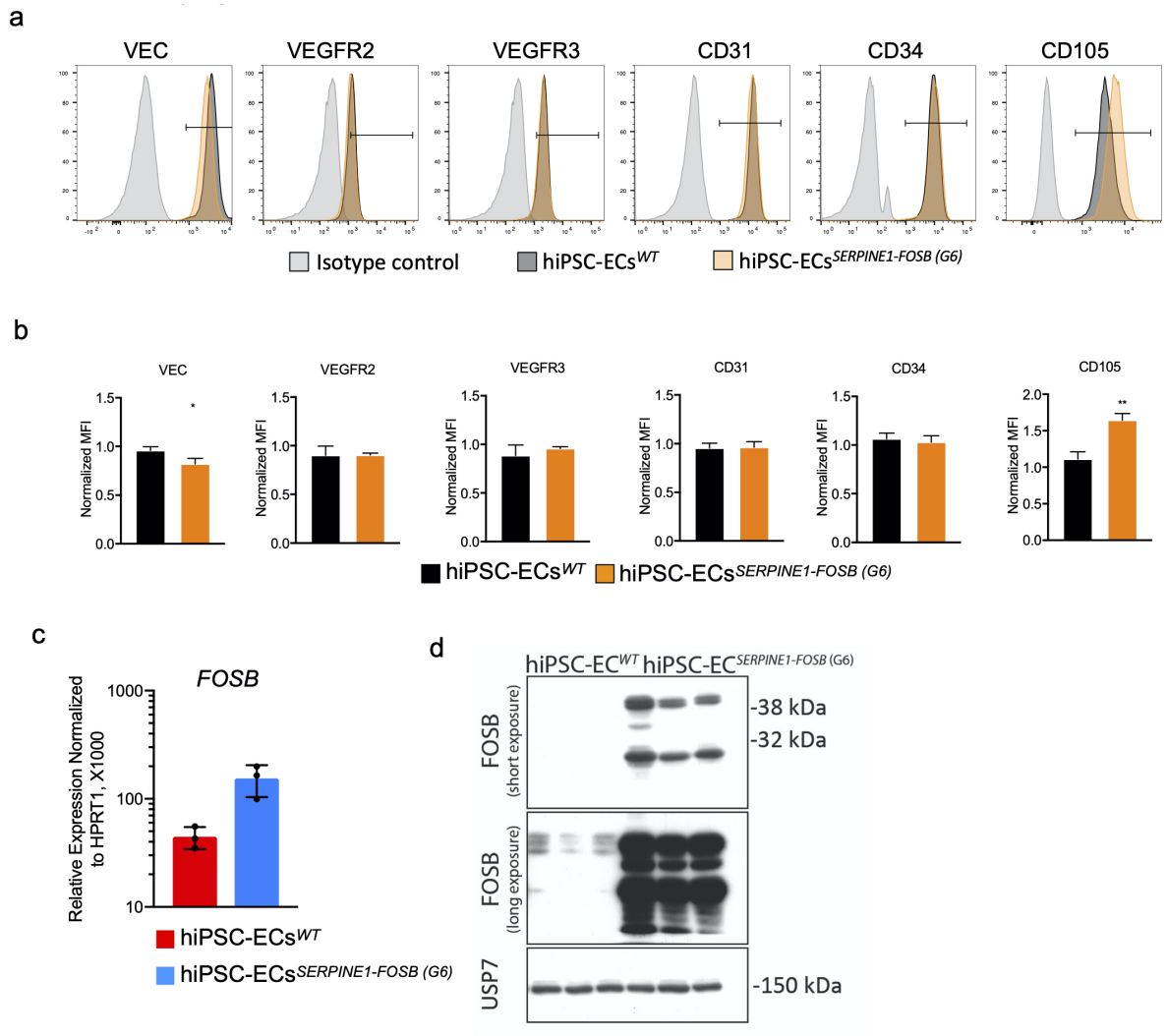
6.A Additional Figures



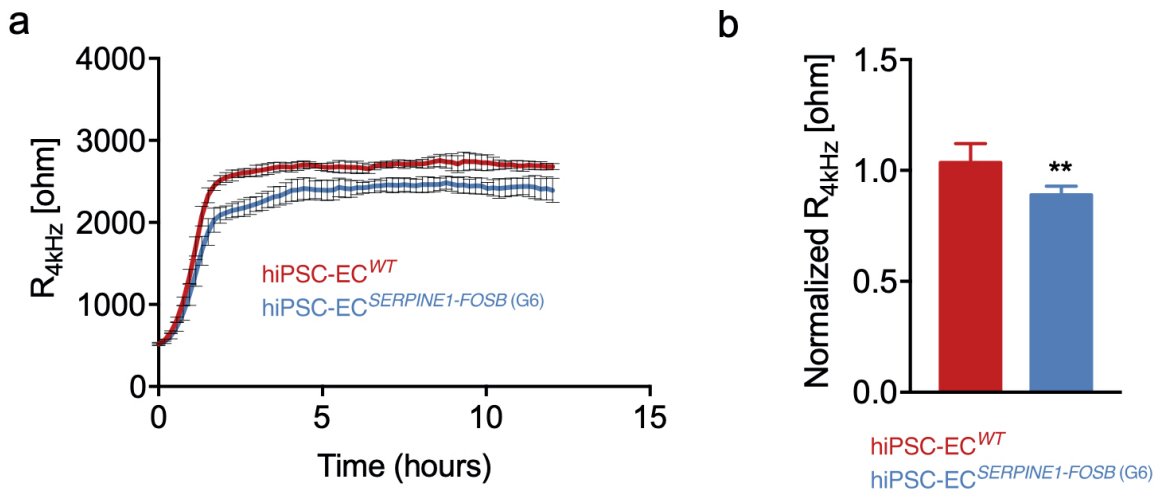
Additional Figure 6.1: Generation and characterization of hiPSCs with a *SERPINE-FOSB* fusion. (a) Schematic overview of the targeting and screening experimental workflow. (b) Three color FISH (Blue at 5' side of *SERPINE1*; green at 5' side of *FOSB* and red at 3' side of *FOSB*) for the detection of *SERPINE1-FOSB* fusion on hiPSC "bulk" culture prior to single-cell deposition. Red arrows indicate cells with the fusion, and blue arrows show wild-type cells. The right image shows a schematic and representative overview of targeted and wild-type cells, as detected with three color FISH.



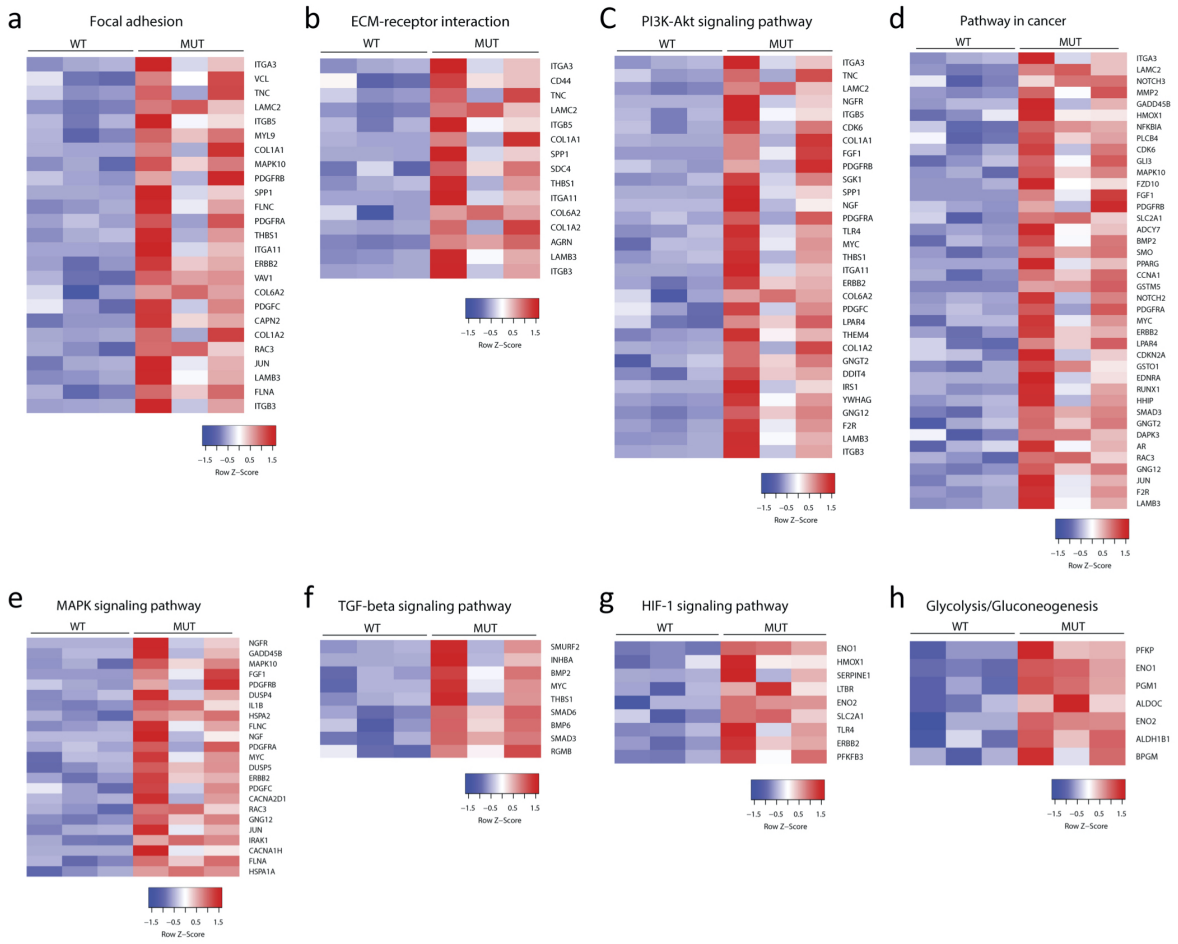
Additional Figure 6.2: Sanger sequencing of hiPSCs with a *SERPINE-FOSB* fusion. (a) Upper panel: schematic overview of NHEJ-mediated repair resulting in generation of clone G6. Bottom panel: Sanger sequencing of PCR products from G6 clone validating NHEJ recombination of *SERPINE1* and *FOSB* with a random piece of DNA inserted in the intron between the fusion. (b,c) Sanger sequencing of the non-targeted wild-type *FOSB* intron 1 and wild-type *SERPINE1* intron 1 in hiPSC clones D3 and G6. (d) Sanger sequencing of cDNA from clones D3 and G6 showing normal splicing of fusion *SERPINE1-FOSB* mRNA using forward primer on the *SERPINE1* UTR and a reverse primer on *FOSB* exon 3. (e) Sanger sequencing of the non-targeted wild-type *SERPINE1* cDNA with primers on the *SERPINE1* UTR and exon 2 showing that the insertion of the selection cassette fragment had no effect on splicing of *SERPINE1* mRNA in colony G6.



Additional Figure 6.3: Characterization of hiPSC-ECs with *SERPINE1-FOSB* fusion. (a) FACS analysis of EC markers expression on isolated hiPSC-ECs^{WT} (black filled histogram) and hiPSC-ECs$^{SERPINE\text{-}FOSB(G6)}$ (orange filled histogram) at passage 2 (P2) and relevant isotype control (gray filled histogram). (b) Quantification of relative surface expression levels (MFI) of VEC, VEGFR2, VEGFR3, CD31, CD34 and CD105. Experiments were performed in triplicate with three independent batches of hiPSC-ECs. Error bars are SD. (c) Real-time qPCR analysis of *FOSB* expression in hiPSCs^{WT}, hiPSCs^{G6}, hiPSCs, hiPSC-ECs^{WT} and hiPSC-ECs$^{SERPINE\text{-}FOSB(G6)}$. Experiments were performed with three independent batches of hiPSC-ECs. Expression is determined relative to hiPSC-ECs^{WT}, shown as log2 fold change. Error bars are SD. * $p<0.005$. (d) Western blot analysis of FOSB expression in hiPSC-ECs^{WT} and hiPSC-ECs$^{SERPINE\text{-}FOSB(G6)}$. USP7 was used as a housekeeping control.



Additional Figure 6.4: Assessment of barrier function of hiPSC-ECs with *SERPINE1-FOSB* fusion. (a) Representative absolute resistance of the hiPSC-EC monolayer in complete EC growth medium. Errors bars are shown as ±SD. (b) Normalized resistance [4 kHz] of the hiPSC-EC monolayer in complete EC growth medium. Error bars are shown as ±SD of six independent biological experiments. *** $p<0.001$. Experiments were performed in triplicate using three independent batches of hiPSC-ECs.



Additional Figure 6.5: KEGG pathways identified in hiPSC-ECsSERPINE1-FOSB (D3) upregulated DEGs. (a-h) Heatmaps of genes from eight representative KEGG pathways enriched in hiPSC-ECs$^{SERPINE\text{-}FOSB(D3)}$ upregulated DEGs.

Part III

Computational biology

Chapter 7

PyPanda: a Python package for gene regulatory network reconstruction

This chapter is based on the publication: **van IJzendoorn DGP**, Glass K, Quackenbush J, Kuijjer ML. PyPanda: a Python package for gene regulatory network reconstruction. *Bioinformatics*. 2016;32: 3363-3365.

7.1 Abstract

PANDA (Passing Attributes between Networks for Data Assimilation) is a gene regulatory network inference method that uses message-passing to integrate multiple sources of 'omics data. PANDA was originally coded in C ++. In this application note we describe PyPanda, the Python version of PANDA. PyPanda runs considerably faster than the C ++ version and includes additional features for network analysis.

7.2 Introduction

Accurately inferring gene regulatory networks is one of the most important challenges in the analysis of gene expression data. Although many methods have been proposed (1–4), computation time has been a significant limiting factor in their widespread use. PANDA (Passing Attributes between Networks for Data Assimilation) is a gene regulatory network inference method that uses message passing between multiple 'omics data types to infer the network of interactions most consistent with the underlying data (5). PANDA has been applied to understand transcriptional programs in a variety of systems (6–8). Here we introduce PyPanda, a Python implementation of the PANDA algorithm, following the approach taken in Glass et al (7). and optimized for matrix operations using NumPy (9). This approach enables the use of fast matrix multiplications using the BLAS and LAPACK functions, thereby significantly decreasing run-time for network prediction compared with the original implementation of PANDA, which was coded in C ++ and used for-loops (7). We observe further speed increase over the C ++-code because PyPanda automatically uses multiple processor-cores through the NumPy library. We have also expanded PyPanda to include common downstream analyses of PANDA networks, including the calculation of network in- and out-degrees and the estimation of single-sample networks using the recently developed LIONESS algorithm (10).

7.3 Approach

7.3.1 Comparing PANDA C ++-code to Python-code

We compared the C ++-code and Python-code versions of PANDA using several metrics. First, we assessed the two implementations by comparing the number of lines of code. Using the cloc utility we counted the number of lines of C ++-code and Python-code. The C ++-code counted 1132 lines of code. The Python-code counted 258 lines of code, significantly shorter (4.4 times) than the C ++-code. The Python-code also includes features such as the LIONESS equation and in- and out-degree calculation. Without these

features the Python-code is only 155 lines of code. Because the Python implementation is much more concise than the C ++-code it is easier to interpret and modify.

We compared the C ++-code and Python-code versions of PANDA using several metrics. First, we assessed the two implementations by comparing the number of lines of code. Using the cloc utility we counted the number of lines of C ++-code and Python-code. The C ++-code counted 1132 lines of code. The Python-code counted 258 lines of code, significantly shorter (4.4 times) than the C ++-code. The Python-code also includes features such as the LIONESS equation and in- and out-degree calculation. Without these features the Python-code is only 155 lines of code. Because the Python implementation is much more concise than the C ++-code it is easier to interpret and modify.

The speed of the network prediction was tested using simulated networks of Ne = Na dimensions, where Ne is the number of effector nodes and Na is the number of affected nodes. For each of several different network sizes (Ne = Na = 125 to Ne = Na = 2000 nodes, in steps of 125) we generated ten random 'motif data' networks according to the method described in Glass et al (7). We then ran the Python and C ++ versions of PANDA using these simulated motif data together with identity matrices for the protein-protein interaction and co-expression information. For runs on the same initial 'motif data' networks, we verified that the C ++-code and Python-code returned exactly the same output network, as expected due to the deterministic nature of PANDA.

The C ++-code only uses one CPU core. In comparing the C ++-code with the Python-code using a single core, we found a 2.07-fold speed-up relative to the C ++-code for the smallest network (Ne = Na = 125) tested. The speed increase of the Python-code over the C ++-code became larger as the network size increased. For example, the Python-code performed 12.31 times faster for the largest network (Ne = Na = 2000) (figure 7.1a). Recorded run times across the ten random networks had a standard deviation of 0.04s and 2.59s for the smallest (Ne = Na = 125) and largest (Ne = Na = 2000) networks, respectively using the C ++ code. Using the Python code these were reduced to 0.03s and 0.099s.

Given the abundance of multicore computing resources currently available, we also tested the speed increase when running the Python-code on multiple cores compared with running the Python-code on a single core. We found that for the smallest network the speed was 1.45 times faster when using 6 cores compared with using only a single core; for the largest network the speed increase was 3.7-fold (figure 7.1b).

This increase in speed enables reconstruction of networks with larger numbers of regulators and target genes. For example, using the Python-code significantly decreases the time required to infer a human gene regulatory network (Ne = 1000, Na = 20 000), from ∼18 h with the C ++-code to only about 2 h with the Python-code. This speed-up is

especially important as transcription factor motif databases are frequently updated to include more motifs. Further, the decreased running time helps to enable the estimation of network significance by making the use of bootstrapping/jackknifing methods much more feasible.

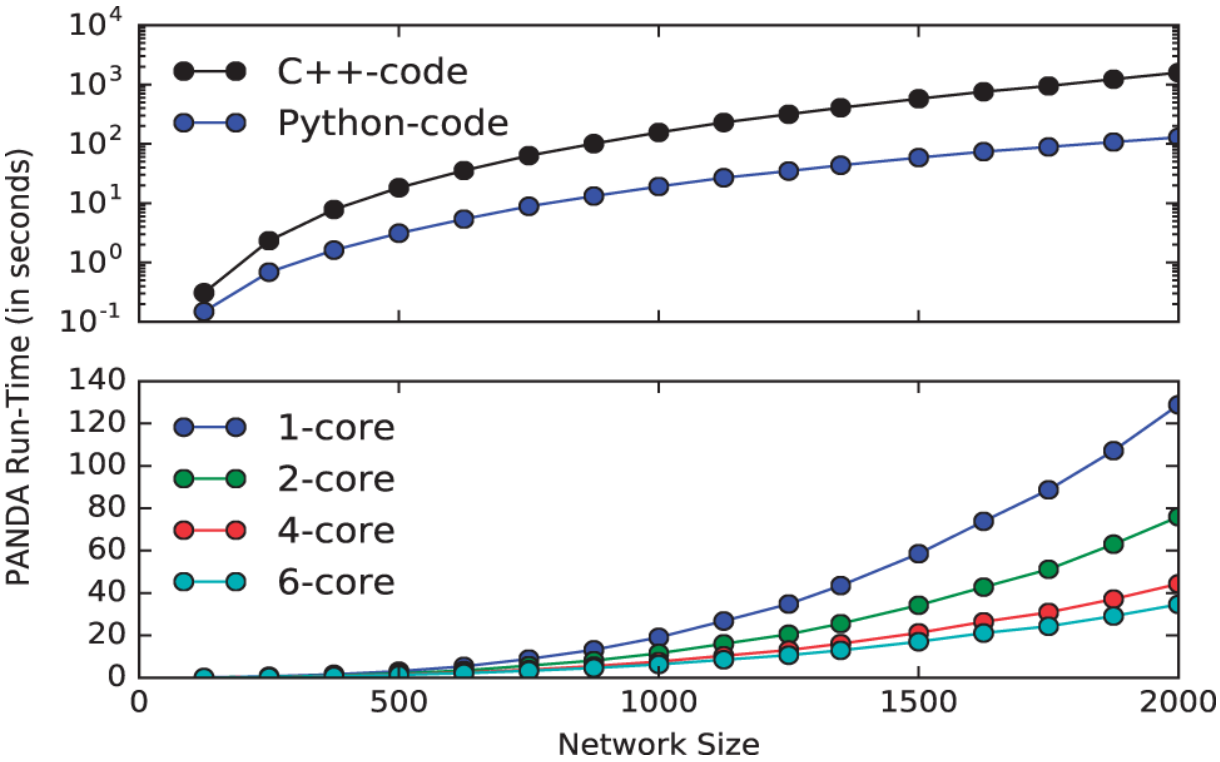


Figure 7.1: Speed comparison for network reconstruction on networks of different sizes using (a) the C ++-code and the Python-code, (b) the Python-code running on a single CPU compared with multicore (6 CPU cores).

7.3.2 Additional features

In addition to reconstructing one regulatory network based on a data set consisting of multiple samples, PyPanda can also reconstruct single-sample networks using the LIONESS algorithm (10). In PyPanda, the LIONESS method uses PANDA to infer an 'aggregate' network representing a set of N input samples, infers a network for N – 1 samples, and then applies a linear equation to estimate the network for the sample that had been removed. The process is then repeated for each sample in the original set, producing N single-sample networks. PyPanda can also use LIONESS to reconstruct single-sample networks based on Pearson correlation.

PyPanda also includes functions to calculate in-degrees (the sum of edge weights targeting a specific gene) and out-degrees (the sum of edge weights pointing out from a

regulator to its target genes). These summary metrics can be used for downstream network analysis (6).

7.4 Conclusion

PANDA is a proven method for gene regulatory network inference but, like most sophisticated network inference methods, its runtime has limited its utility. The Python implementation of PANDA uses matrix operations and incorporates the NumPy libraries, resulting in a significant simplification of the code and a dramatic increase in computing speed, even on a single processor. When applied to a test data set and run on multiple processing cores, this increase in speed was even greater, decreasing processing times by a factor of 46 relative to the original C ++-code. This creates opportunities to greatly expand the use of PANDA and to implement additional measures of network significance based on bootstrapping/jackknifing. PyPanda also includes the LIONESS method, which allows inference of single-sample networks, as well as a number of other useful network metric measures. The open source PyPanda package is freely available at `http://github.com/davidvi/pypanda`.

Bibliography

[1] Altay G, Asim M, Markowetz F, Neal DE. Differential C3NET reveals disease networks of direct physical interactions. BMC Bioinformatics. 2011;12:296. doi:10.1186/1471-2105-12-296.

[2] Ernst J, Beg QK, Kay KA, Balázsi G, Oltvai ZN, Bar-Joseph Z. A semi-supervised method for predicting transcription factor-gene interactions in Escherichia coli. PLoS Computational Biology. 2008;4(3):e1000044. doi:10.1371/journal.pcbi.1000044.

[3] Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biology. 2007;5(1):0054–0066. doi:10.1371/journal.pbio.0050008.

[4] Lemmens K, Dhollander T, De Bie T, Monsieurs P, Engelen K, Smets B, et al. Inferring transcriptional modules form ChIP-chip, motif and microarray data. Genome Biology. 2006;7(5):R37. doi:10.1186/gb-2006-7-5-r37.

[5] Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing Messages between Biological Networks to Refine Predicted Interactions. PLoS ONE. 2013;8(5):e64832. doi:10.1371/journal.pone.0064832.

[6] Glass K, Quackenbush J, Silverman EK, Celli B, Rennard SI, Yuan GC, et al. Sexually-dimorphic targeting of functionally-related genes in COPD. BMC systems biology. 2014;8:118. doi:10.1186/s12918-014-0118-y.

[7] Glass K, Quackenbush J, Spentzos D, Haibe-Kains B, Yuan GC. A network model for angiogenesis in ovarian cancer. BMC Bioinformatics. 2015;16:115. doi:10.1186/s12859-015-0551-y.

[8] Lao T, Glass K, Qiu W, Polverino F, Gupta K, Morrow J, et al. Haploinsufficiency of Hedgehog interacting protein causes increased emphysema induced by cigarette smoke through network rewiring. Genome Medicine. 2015;7(1):12. doi:10.1186/s13073-015-0137-3.

[9] van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science & Engineering. 2011;13(2):22–30. doi:10.1109/mcse.2011.37.

[10] Kuijjer ML, Tung M, Yuan G, Quackenbush J, Glass K. Estimating sample-specific regulatory networks. arXiv Molecular Networks (q-bioMN). 2015;doi:arXiv:1505.06440v2.

Chapter 8

Gene regulatory network analysis of translocation driven vascular tumors

van IJzendoorn DGP, Szuhai K, Kuijjer ML, Bovée JVMG.

8.1 Abstract

The vascular tumors epithelioid hemangioma (typical and atypical) and pseudomyogenic hemangioendothelioma are driven by translocations involving FOS family members (FOS and FOSB). FOS or FOSB, together with JUN family members form the AP-1 transcription factor complex that is involved in important pathways such as cell growth, differentiation and survival. Likely, the translocations involving FOS family members lead to prolonged and high activation of the AP-1 transcription factor.

Human Umbilical Vein Endothelial Cells with lentiviral transduced *FOS* or *FOSB* expression vectors were used as a model system. The transcriptome sequencing data were used to reconstruct the regulatory networks.

It was confirmed that AP-1 transcription factor activation is important in the early response of wild-type Human Umbilical Vein Endothelial Cells. Regulatory network reconstruction identified important regulatory proteins and a potential link to the HIPPO signaling pathway. The potential link with the HIPPO signaling pathway could explain the overlap with other vascular tumors that harbor translocations involving members of the HIPPO signaling pathway. Moreover, potential new candidate drugs for targeted therapies were identified based on the regulatory network.

8.2 Introduction

Many vascular tumors harbor recurrent translocations that potentially drive their tumorigenesis (1). This study focuses on the fusions involving members of the AP-1 transcription factor complex found in three entities: typical epithelioid hemangioma which harbors fusions involving *FOS* with different fusion partners, atypical epithelioid hemangioma which is characterized by a *ZFP36-FOSB* fusion and pseudomyogenic hemangioendotheliomas harboring either a *SERPINE1-FOSB* or a *ACTB-FOSB* translocation (2–7). As the fusions are the only known recurrent genetic alterations in these entities it is likely that their tumorigenesis is driven by these fusion genes. All three entities are very rare, and diagnosis by pathologists based on only the morphology is difficult. Correct diagnosis is however important, as the prognosis and treatment vary between the entities.

The AP-1 transcription factor consists of a FOS family member (such as FOS or FOSB) forming a heterodimer with JUN family members. AP-1 transcription factor activation is known to be involved in cell growth, differentiation and survival among other important regulatory pathways (8). For atypical epithelioid hemangioma a translocation between *ZFP36* and *FOSB* has been described (3). While for conventional epithelioid hemangioma *FOS* has been found with a number of different fusion partners which all result in an early termination, somewhere near the C-terminal, of the FOS protein (2, 9). We

showed that this early termination of the FOS protein renders the FOS protein insensitive to ubiquitin independent degradation by the proteasome, resulting in a prolonged lifespan and therefore higher activity of the FOS protein (8). Pseudomyogenic hemangioendothelioma is driven by a translocation between *SERPINE1* or *ACTB* with *FOSB* (4). This translocation leads to FOSB losing its first 48 amino acids, and attaches to the *SERPINE1* or *ACTB* promotor driving the expression. It is likely that this translocation leads to an upregulation of the FOSB protein (10). As both FOS and FOSB are members of the FOS family it is likely that AP-1 transcription factor activation plays an important role in the tumorigenesis as seen in epithelioid hemangioma, atypical epithelioid hemangioma and pseudomyogenic hemangioendothelioma.

In this study, we aimed to understand the tumorigenesis of the translocations that are found in epithelioid hemangioma, atypical epithelioid hemangioma and pseudomyogenic hemangioendothelioma. In this study it was assumed that these tumors originate somewhere in the differentiation lineage towards endothelial cells, evidence for this is that the tumor cells express CD31, ERG and FLI1, amongst other endothelial markers (1). To model the different translocations, the different genes involved in the translocations (truncated FOS and FOSB) were transduced in Human Umbilical Vein Endothelial Cells (HUVECs) using lentiviral expression vectors. Thereafter, the HUVECs transcriptome was sequenced at different timepoints after serum stimulation. We used the transcriptome data to build gene regulatory networks using the analysis tool Expression2Kinase. Expression2Kinase integrates chromatin immune-precipitation, position weight matrices, protein-protein interactions and kinase-substrate phosphorylation reactions to identify upstream regulators (11). The resulting network was matched with data from the Connectivity Map, containing data on gene expression in response to stimulation with small molecules, to identify potential new effective therapies (12).

8.3 Materials and Methods

8.3.1 Cell culture

Human umbilical vein endothelial cells (HUVECs) were cultured in EGM-2 medium (Lonza) on gelatin coated plates (0.2% gelatin in PBS for 30 minutes at 37°C). Cells were tested for mycoplasma. Serum starvation was performed by culturing the cells overnight in M199 medium (Thermo-Fisher). Thereafter EGM-2 medium was added to serum stimulate the HUVECs (for 0, 0.5 and 48 hours).

8.3.2 Gene overexpression

HUVECs were transduced with lentiviral expression vectors for *FOS*, deletion mutant *FOS* (where the last 95 amino acids are missing from the C-terminal) (8) and *FOSB*. Genes for overexpression were cloned into a PLV backbone and lentivirus was produced for transduction experiments as previously described by our group (8). As a negative control, HUVECs were transduced with empty vectors (PLV). HUVECs were serum starved and stimulated. Cells were harvested before stimulation, 30 minutes and 48 hours after serum stimulation. Transcriptome sequencing experiments were performed as single experiments.

8.3.3 RNA isolation and sequencing

RNA was isolated from the HUVECs using the Direct-zol RNA kit (Zymo Research) according to the manufacturers protocol. RIN values of the RNA were verified using the Bioanalyzer (Agilent). Next Generation Sequencing was performed on the HiSeq2500 (Illumina) at ServiceXS (Leiden, The Netherlands).

8.3.4 Gene expression normalization and analysis

Raw read counts were aligned using TopHat2 (13) to the reference genome (hg19). FeatureCounts (14) was used to determine the gene counts. First genes were removed where the median read count was zero. Next, Limma (v3.6) R package (15) was used to normalize the read counts. Normalization was done using the weight trimmed mean of M-values. Last, the log2 fold change was determined for each gene. PCA analysis was performed using the Prcomp R package with default settings.

8.3.5 Regulatory network reconstruction

The gene expression data were used to reconstruct the regulatory network using the Expression2Kinase pipeline with the default settings (11).

8.3.6 Statistical software and figures

R statistical software (v3.4.4) was used for all statistical tests (16). Chord diagrams were generated with GOplot (v1.0.2) (17). All further graphs were generated with R package ggplot2 (v2.2.1).

8.4 Results and discussion

8.4.1 Expression vectors dictate global gene expression more than time after serum stimulation

Transcriptome sequencing data was studied for the full length *FOS* gene, deletion mutant *FOS* (hereafter referred to as *FOS*Δ), *FOSB* and a negative control (2, 8, 10) at three timepoints after serum stimulation (0, 0.5 and 48 hours). All gene expression data were normalized together. Histogram analysis of the gene expression showed that the samples were comparable for further analysis (figure 8.1a). PCA cluster analysis reveals groups of HUVEC samples which associated with the different expression vectors. Interestingly, the time after serum stimulation did not seem to influence the clusters as much as the expression of the different AP-1 transcription factor family members. These results indicate that the expression vectors lead to a larger global difference in gene expression compared to the time after serum stimulation.

8.4.2 Serum stimulation leads to upregulation of AP-1 factors in normal HUVECs

To study the effect of serum stimulation on normal HUVECs we analyzed the gene expression of the control cells (with empty PLV vector) that were isolated at different timepoints (0, 0.5 and 48 hours). Gene expression data for the control cells showed that both *FOS* and *FOSB* (respectively 5.9 and 5.56 log2 fold change at 0.5 hours vs the other PLV samples, adjusted P value < 0.001) are in the top ten upregulated genes 30 minutes after serum stimulation (figure 8.2a). However, at 48 hours after serum stimulation the expression of FOS and FOSB is only slightly higher compared to the unstimulated cells (respectively unstimulated -3.5 and -3.4 log2 fold change and -2.4 and -2.2 log2 fold change at 48 hours or 0 hours vs the other PLV samples, adjusted P value < 0.001). Among the other upregulated genes are transcriptional regulators *EGR1*, *EGR2* and *EGR3*. Early Growth Response (*EGR*) genes are described to be an important modulators of AP-1, *EGR* upregulation likely stimulates AP-1 upregulation (18). This shows that AP-1 transcription factors play a central role in the first response to serum stimulation in endothelial cells. AP-1 has previously already been described as one of the main early response genes after cellular signals (19).

It is likely that the translocations which are found in epithelioid hemangioma, atypical epithelioid hemangioma and pseudomyogenic hemangioendothelioma lead to sustained high expression of *FOS*Δ and *FOSB*. This would likely result in gene expression comparable to the early response as we found in the HUVECs. Signature genes for the early

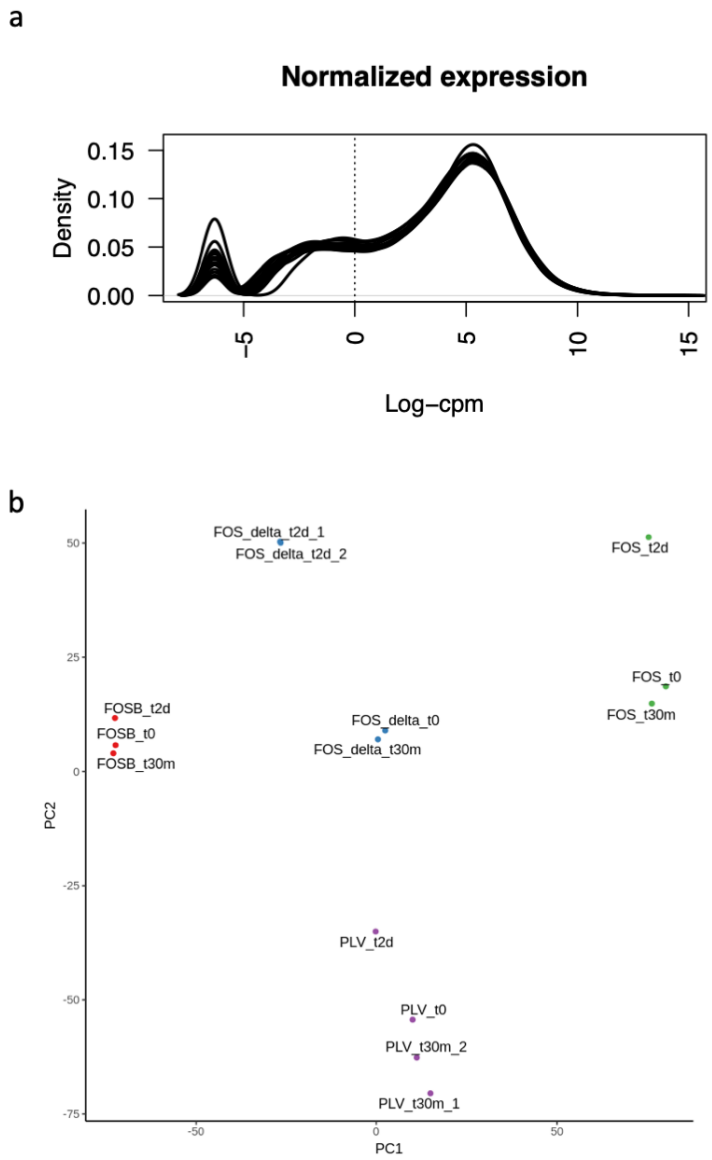


Figure 8.1: Cluster analysis of HUVEC gene expression data. (a) Transcriptome sequencing data was collected for HUVECs expressing *FOS*, *FOS*Δ, *FOSB* and an empty control plasmid. Expression data was harvested at different timepoint (0, 0.5 and 48 hours) after serum stimulation. All expression data was normalized together using Limma. The figure shows a histogram of the gene expression after normalization showing that the samples are comparable. (b) PCA analysis of the gene expression data shows different clusters. The clusters generally correlate with the expression vector that is used in the HUVECs. Time after serum stimulation did not affect the clusters.

response in the control endothelial cells were further analyzed for significant pathways in a GO term analysis (figure 8.2b). Some of the top pathways included endoderm formation (adjusted P value < 0.0001), formation of primary germ layer (adjusted P value < 0.0001), skeletal muscle cell differentiation (adjusted P value < 0.0001), angiogenesis (adjusted P value < 0.001) and regulation of vascular development (adjusted P value < 0.001).

8.4.3 Regulatory network for *FOS*Δ and *FOSB* overexpression

FOSΔ is found in epithelioid hemangioma while high *FOSB* expression is found in atypical epithelioid hemangioma and pseudomyogenic hemangioendothelioma resulting from different translocations. Signature genes were identified for *FOS*Δ and *FOSB* using a differential expression analysis, comparing either *FOS*Δ or *FOSB* to the control cells (in all timepoints) and identifying genes with a log2 fold change larger than 1. For *FOS*Δ 1907 signature genes were identified and for *FOSB* 1111 genes were found. 655 genes showed to be signature genes for both *FOS*Δ and *FOSB*. As these overlapping genes are probably regulated by both fusion genes they could be an important common factor in the tumorigenesis of epithelioid hemangioma, atypical epithelioid hemangioma and pseudomyogenic hemangioendothelioma and were therefore used for further analysis (figure 8.3a). Using the Expression2Kinase analysis tool, we constructed a regulatory network (11). This tool uses gene expression datasets as input and imputes the likely regulatory transcription factors and proteins by integrating data from chromatin immune-precipitation, position weight matrices, protein-protein interactions and kinase-substrate phosphorylation reactions. The advantage of this method is that transcription factors and proteins can be identified whose role in the tumorigenesis may not necessarily be directly detected in the transcriptome sequencing data. The regulatory network reconstruction revealed important transcription factors and other proteins that are likely master regulators based on the input signature genes. Notably YAP1, part of the HIPPO signaling pathway is regulated by the AP-1 transcription factors (figure 8.3b). As reported in the DECODE database (SABioscience) the top transcription factor binding sites in the promoter of *YAP1* are for AP-1, indicating a direct regulatory role. Of note, identified proteins also include HDAC enzymes (HDAC1, HDAC2, HDAC3 and HDAC4), indicating a role for histone deacetylase in the tumorigenesis (figure 3b). Moreover, many SMAD proteins were also identified pointing to a potential role for TGF-β signaling (20).

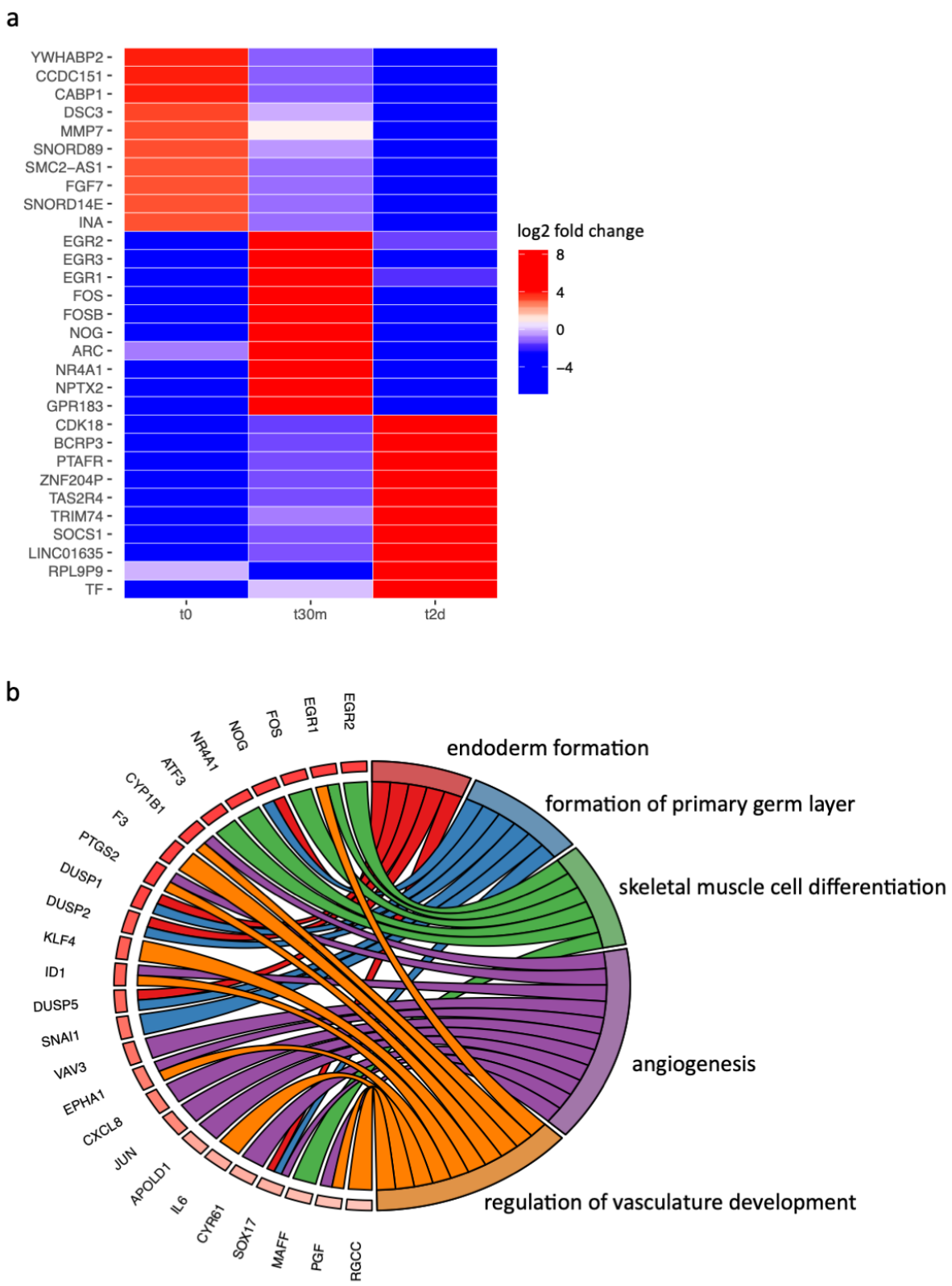


Figure 8.2: AP-1 transcription factors play an important role in the early response. (a) The heatmap shows the top ten signature genes for the control HUVECs harvested at different timepoint. The color indicates the log2 fold change compared to the other PLV samples. (b) The signature genes from the early response (30 minute) timepoint were used to identify GO terms. GO terms are shown in a Chord diagram on the right, with the associated genes on the left (sorted and colored according to log2 fold change in descending order).

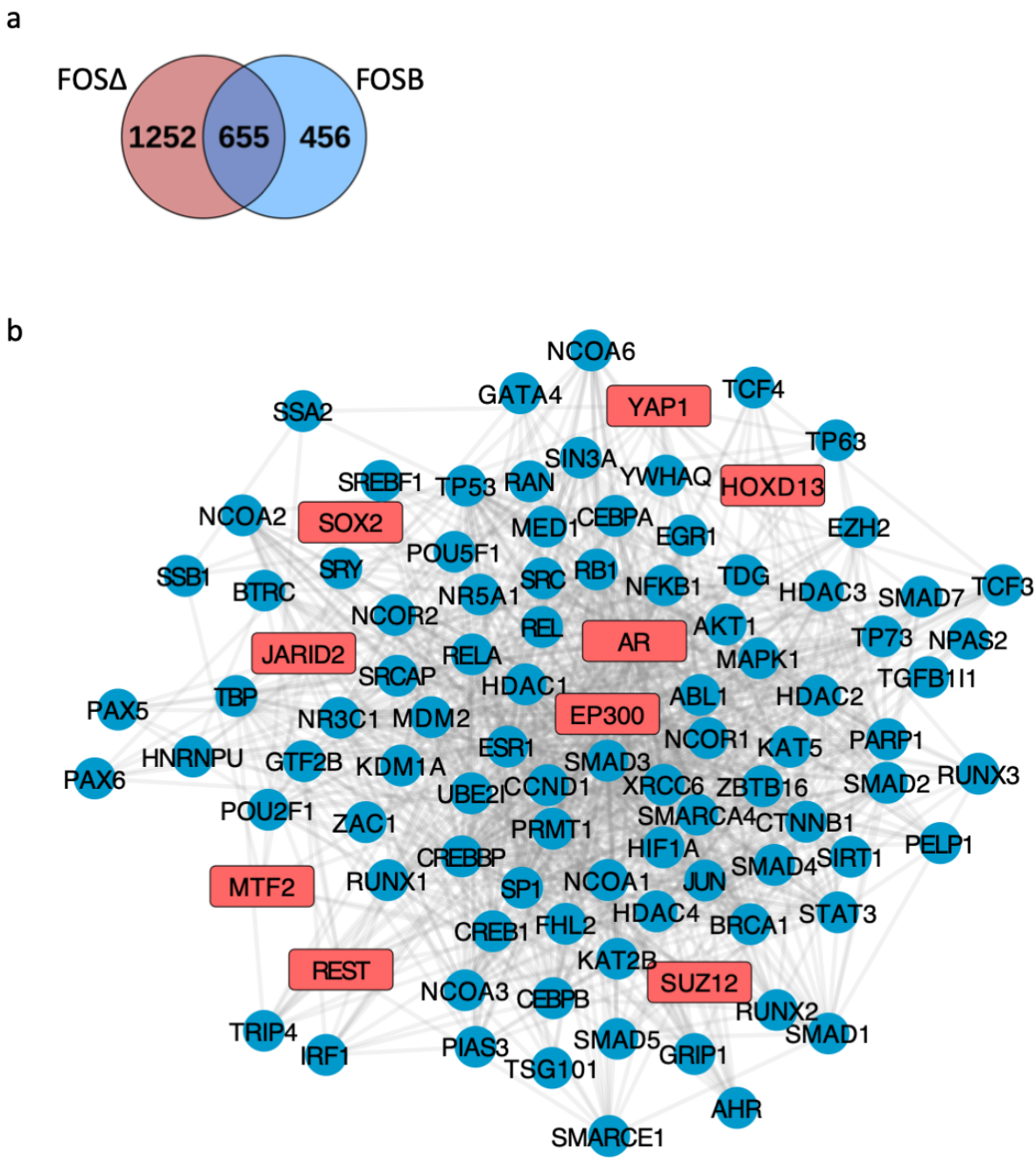


Figure 8.3: *FOS*Δ and *FOSB* regulatory network analysis. (a) The signature genes for *FOS*Δ and *FOSB* are shown in a Venn diagram. *FOS*Δ has 1907 signature genes and *FOSB* 1111. 655 signature genes overlap between *FOS*Δ and *FOSB*. (b) Using the Expression2Kinase analysis tool the gene regulatory network was reconstructed for the overlapping signature genes from *FOS*Δ and *FOSB*. The network plot shows the most important upstream regulatory proteins reconstructed from the signature genes. Important identified transcription factors are shown in red squares while regulated proteins are shown in blue circles.

8.4.4 Targeted therapies based on the *FOS*Δ and *FOSB* regulatory network

The proteins identified in the regulatory network were used to identify potential targeted therapies for AP-1 driven vascular tumors, with the HUVECs with *FOS*Δ and *FOSB* expression vectors as a model system, using the CMAP dataset (figure 8.4). Trichostatin A and vorinostat, two HDAC inhibitors, were identified as potential target drug for therapy (with a coverage of 16 and only one conflicting gene for trichostatin A). This finding is in line with the regulatory network, where we identified four HDAC family members (HDAC1, HDAC2, HDAC3 and HDAC4) to be important upstream regulatory proteins based on the signature genes. Other hits included irinotecan, a topoisomerase inhibitor used for the treatment of colon and small cell lung cancer. Tanespimycin was also identified, which is a Hsp90 inhibitor that is in a trial for solid tumors.

In summary, we confirmed AP-1 is one of the main early response transcription factors in endothelial cells. It is likely that in epithelioid hemangioma, atypical epithelioid hemangioma and pseudomyogenic hemangioendothelioma there is high activation of the AP-1 transcription factors, either because the half-life of FOS is longer or because *FOSB* becomes attached to another promoter. In our model system we showed this AP-1 transcription factor activation results in activations of pathways related to vascular development among other pathways.

Using gene regulatory network reconstruction, we identified likely important regulatory transcription factors and proteins which we consequently used to identify potential targeted therapies. Based on data from the CMAP dataset therapies such as HDAC inhibitors were identified, in line with the identified HDAC family members in the regulatory network. Other identified therapies such as irinotecan and tanespimycin, are more difficult to understand at this point as they are generally used to target tumors with high rates of cell-division (21, 22). YAP1 was found to be an important regulatory protein which was also found to harbor AP-1 binding sites in its promoter, suggesting direct regulation from AP-1. YAP1 is part of the HIPPO signaling pathway which is likely involved in other vascular tumors such as epithelioid hemangioendothelioma. Epithelioid hemangioendothelioma is driven by either *WWTR1-CAMTA* or *YAP1-TFE3* (in a small subset) fusions, both are important parts of the HIPPO signaling pathway (23–25). The direct link between AP-1 and YAP1 could be a common denominator for a group of vascular tumors with intermediate to low grade malignant behavior. The direct link with YAP1 also suggests CA3, a new YAP1 inhibitor, could be used as targeted therapy for the vascular tumors with translocations involving AP-1 transcription factor members (26), warranting further studies.

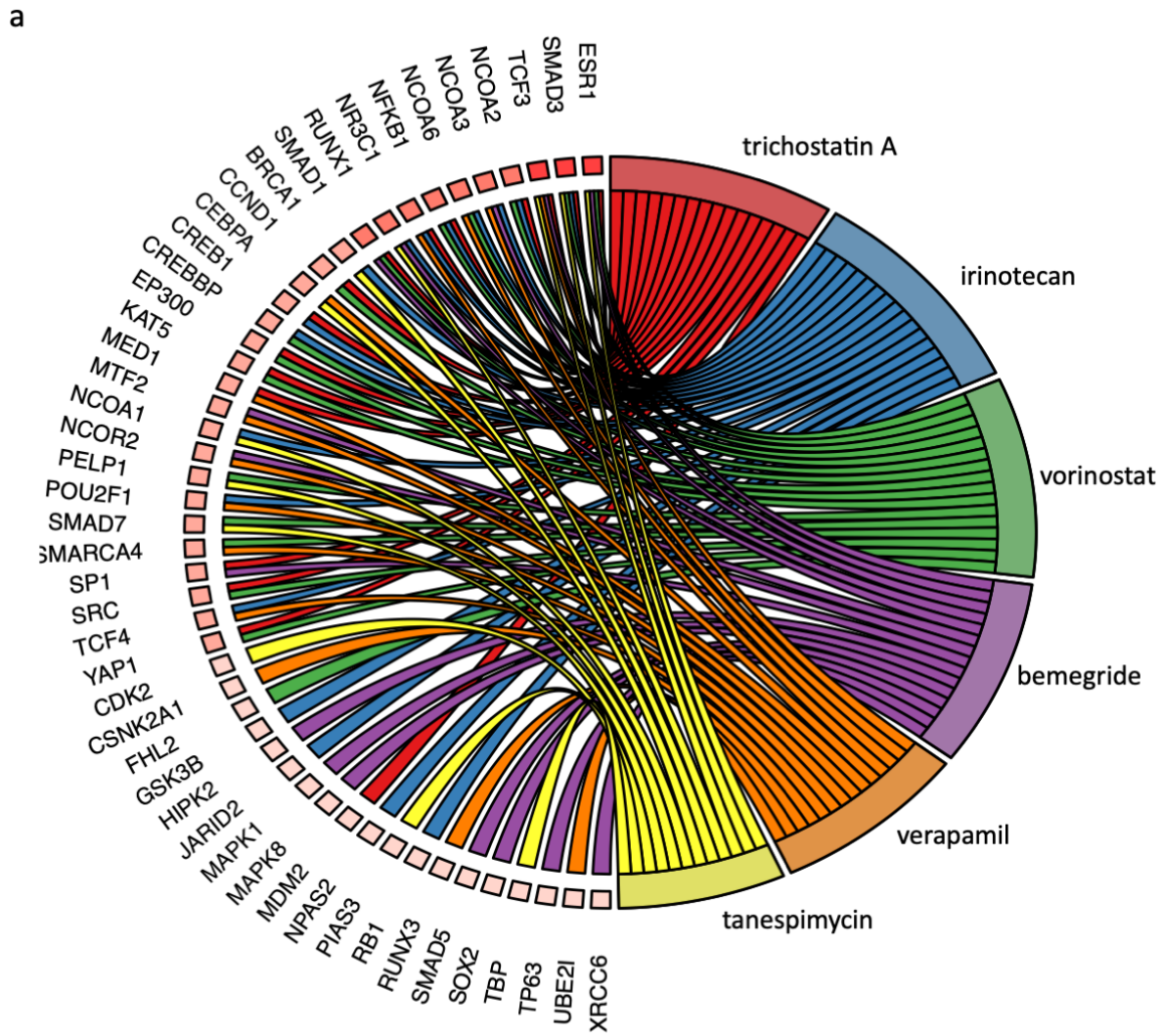


Figure 8.4: Targeted therapies identified from the *FOS*Δ and *FOSB* regulatory network. Potential targeted therapies are identified based on the regulatory network for *FOS*Δ and *FOSB*. The Chord diagram shows the potential targeted therapies on the right with the associated proteins on the left. The red gradient on the left reflects the number of therapies associated with the respective protein.

Bibliography

[1] van IJzendoorn DGP, Bovée JVMG. Vascular Tumors of Bone: The Evolvement of a Classification Based on Molecular Developments. Surgical Pathology Clinics. 2017;10(3):621–635. doi:10.1016/j.path.2017.04.003.

[2] van IJzendoorn DG, de Jong D, Romagosa C, Picci P, Benassi MS, Gambarotti M, et al. Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. Genes Chromosomes Cancer. 2015;54(9):565–574. doi:10.1002/gcc.22269.

[3] Antonescu CR, Chen HW, Zhang L, Sung YS, Panicek D, Agaram NP, et al. ZFP36-FOSB fusion defines a subset of epithelioid hemangioma with atypical features. Genes Chromosomes Cancer. 2014;53(11):951–959. doi:10.1002/gcc.22206.

[4] Walther C, Tayebwa J, Lilljebjorn H, Magnusson L, Nilsson J, von Steyern FV, et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. J Pathol. 2014;232(5):534–540. doi:10.1002/path.4322.

[5] Trombetta D, Magnusson L, von Steyern FV, Hornick JL, Fletcher CDM, Mertens F. Translocation t(7;19)(q22;q13)-a recurrent chromosome aberration in pseudomyogenic hemangioendothelioma? Cancer Genetics. 2011;204(4):211–215. doi:10.1016/j.cancergen.2011.01.002.

[6] Agaram NP, Zhang L, Cotzia P, Antonescu CR. Expanding the Spectrum of Genetic Alterations in Pseudomyogenic Hemangioendothelioma With Recurrent Novel ACTB-FOSB Gene Fusions. The American Journal of Surgical Pathology. 2018;42(12):1653–1661. doi:10.1097/PAS.0000000000001147.

[7] Zhu G, Benayed R, Ho C, Mullaney K, Sukhadia P, Rios K, et al. Diagnosis of known sarcoma fusions and novel fusion partners by targeted RNA sequencing with identification of a recurrent ACTB-FOSB fusion in pseudomyogenic hemangioendothelioma. Modern Pathology. 2019;32(5):609–620. doi:10.1038/s41379-018-0175-7.

[8] Van IJzendoorn DGP, Forghany Z, Liebelt F, Vertegaal AC, Jochemsen AG, Bovée JVMG, et al. Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis. Journal of Biological Chemistry. 2017;292(52):21282–21290. doi:10.1074/jbc.C117.815845.

[9] Huang SC, Zhang L, Sung YS, Chen CL, Krausz T, Dickson BC, et al. Frequent FOS gene rearrangements in epithelioid hemangioma: A molecular study of 58 cases with morphologic reappraisal. American Journal of Surgical Pathology. 2015;39(10):1313–1321. doi:10.1097/PAS.0000000000000469.

[10] van IJzendoorn DGP, Sleijfer S, Gelderblom H, Eskens FALM, van Leenders GJLH, Szuhai K, et al. Telatinib Is an Effective Targeted Therapy for Pseudomyogenic Hemangioendothelioma. Clinical Cancer Research. 2018;24(11):2678–2687. doi:10.1158/1078-0432.CCR-17-3512.

[11] Chen EY, Xu H, Gordonov S, Lim MP, Perkins MH, Ma'ayan A. Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. Bioinformatics. 2012;28(1):105–111. doi:10.1093/bioinformatics/btr625.

[12] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929–1935. doi:10.1126/science.1132939.

[13] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology. 2013;14(4):R36. doi:10.1186/gb-2013-14-4-r36.

[14] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–930. doi:10.1093/bioinformatics/btt656.

[15] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47. doi:10.1093/nar/gkv007.

[16] null Null. R: A Language and Environment for Statistical Computing; 2017. Available from: `http://www.r-project.org/`.

[17] Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics. 2015;31(17):2912–2914. doi:10.1093/bioinformatics/btv300.

[18] Parra E, Ferreira J, Ortega A. Overexpression of EGR-1 modulates the activity of NF-κB and AP-1 in prostate carcinoma PC-3 and LNCaP cell lines. International Journal of Oncology. 2011;39(2):345–52. doi:10.3892/ijo.2011.1047.

[19] Bahrami S, Drabløs F. Gene regulation in the immediate-early response process. Advances in Biological Regulation. 2016;62:37–49. doi:10.1016/J.JBIOR.2016.05.001.

[20] Ahmadi A, Najafi M, Farhood B, Mortezaee K. Transforming growth factor-β signaling: Tumorigenesis and targeting for cancer therapy. Journal of Cellular Physiology. 2019;234(8):12173–12187. doi:10.1002/jcp.27955.

[21] Dimopoulos MA, Mitsiades CS, Anderson KC, Richardson PG. Tanespimycin as Antitumor Therapy. Clinical Lymphoma Myeloma and Leukemia. 2011;11(1):17–22. doi:10.3816/CLML.2011.n.002.

[22] Nielsen DL, Palshof J, Brünner N, Stenvang J, Viuff BM. Implications of ABCG2 Expression on Irinotecan Treatment of Colorectal Cancer Patients: A Review. International Journal of Molecular Sciences. 2017;18(9):1926. doi:10.3390/ijms18091926.

[23] Antonescu CR, Le Loarer F, Mosquera JM, Sboner A, Zhang L, Chen CL, et al. Novel YAP1-TFE3 fusion defines a distinct subset of epithelioid hemangioendothelioma. Genes Chromosomes and Cancer. 2013;52(8):775–784. doi:10.1002/gcc.22073.

[24] Errani C, Zhang L, Sung YS, Hajdu M, Singer S, Maki RG, et al. A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. Genes Chromosomes Cancer. 2011;50(8):644–653. doi:10.1002/gcc.20886.

[25] Tanas MR, Sboner A, Oliveira AM, Erickson-Johnson MR, Hespelt J, Hanwright PJ, et al. Identification of a disease-defining gene fusion in epithelioid hemangioendothelioma. Science Translational Medicine. 2011;3(98):98ra82–98ra82. doi:10.1126/scitranslmed.3002409.

[26] Song S, Xie M, Scott AW, Jin J, Ma L, Dong X, et al. A Novel YAP1 Inhibitor Targets CSC-Enriched Radiation-Resistant Cells and Exerts Strong Antitumor Activity in Esophageal Adenocarcinoma. Molecular Cancer Therapeutics. 2018;17(2):443–454. doi:10.1158/1535-7163.MCT-17-0560.

Chapter 9

Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas

This chapter is based on the publication: **van IJzendoorn DGP**, Szuhai K, Briaire-de Bruijn IH, Kostine M, Kuijjer ML, Bovée JVMG. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLOS Comput Biol.* 2019;15: e1006826.

9.1 Abstract

Based on morphology it is often challenging to distinguish between the many different soft tissue sarcoma subtypes. Moreover, outcome of disease is highly variable even between patients with the same disease. Machine learning on transcriptome sequencing data could be a valuable new tool to understand differences between and within entities. Here we used machine learning analysis to identify novel diagnostic and prognostic markers and therapeutic targets for soft tissue sarcomas.

Gene expression data was used from the Cancer Genome Atlas, the Genotype-Tissue Expression data and the French Sarcoma Group. We identified three groups of tumors that overlap in their molecular profile as seen with unsupervised t-Distributed Stochastic Neighbor Embedding clustering and a deep neural network. The three groups corresponded to subtypes that are morphologically overlapping. Using a random forest algorithm, we identified novel diagnostic markers for soft tissue sarcoma that distinguished between synovial sarcoma and MPNST, and that we validated using qRT-PCR in an independent series. Next, we identified prognostic genes that are strong predictors of disease outcome when used in a k-nearest neighbor algorithm. The prognostic genes were further validated in expression data from the French Sarcoma Group. One of these, HMMR, was validated in an independent series of leiomyosarcomas using immunohistochemistry on tissue micro array as a prognostic gene for disease-free interval. Furthermore, reconstruction of regulatory networks combined with data from the Connectivity Map showed, amongst others, that HDAC inhibitors could be a potential effective therapy for multiple soft tissue sarcoma subtypes. A viability assay with two HDAC inhibitors confirmed that both leiomyosarcoma and synovial sarcoma are sensitive to HDAC inhibition.

In this study we identified novel diagnostic markers, prognostic markers and therapeutic leads from multiple soft tissue sarcoma gene expression datasets. Thus, machine learning algorithms are powerful new tools to improve our understanding of rare tumor entities.

9.2 Introduction

Soft tissue sarcomas are rare malignancies arising in the tissues that connect, support and surround other body structures, such as fat or muscle (1). Soft tissue sarcomas annually affect approximately one per 50 million population, and represent <1% of all malignant tumors (2). Soft tissue sarcomas can display different lines of differentiation, and as such are classified based on the tissue that they resemble most. More than 50 different subtypes have been described in the WHO classification. Even though these subtypes differ in prognosis and treatment, there is considerable morphological overlap between

the different subtypes, making differential diagnosis both difficult and important. For instance, synovial sarcoma (SS) and malignant peripheral nerve sheath tumor (MPNST) can be morphologically identical, while also their immunohistochemical profile can overlap, making molecular testing for the presence of the SS specific *SS18-SSX* fusion essential for the final diagnosis (which is laborious and time consuming). Over the last years there have been many large genetic studies generating open accessible gene expression datasets of sarcomas. One of the biggest soft tissue sarcoma sequencing projects to date is the Cancer Genome Atlas (TCGA), which recently published a detailed analysis of the driving mutations in these cancers (3). This data can be leveraged and analyzed with machine learning methodologies to better understand soft tissue sarcoma biology. Machine learning has been used previously to study gene expression patterns. Especially unsupervised algorithms, such as Principal Component Analysis (PCA) and more recently t-Distributed Stochastic Neighbor Embedding (t-SNE), have been successfully used in gene expression studies to classify cancer patients (4). Moreover, for classification of tumors, supervised algorithms such as random forest have been used previously. Gene expression signatures were shown to be effective at classifying breast cancer (5). Later, it was shown that microRNA expression patterns could be used to distinguish between a number of different tumor subtypes, ranging from brain to colorectal cancer (6). More recently, random forest analyses were used on DNA-methylation data to classify different brain tumor subtypes. The advantage of the latter is that it can be performed on paraffin embedded material (7, 8).

Previously the French Sarcoma Group used a machine learning approach on a large cohort of soft tissue sarcomas to verify a set of 67 genes (CINSARC), identified using differential expression analysis, that effectively predicted metastatic outcome in soft tissue sarcomas (9). The identified CINSARC genes were more recently found to have prognostic value for other tumor types as well, such as breast cancer (10). The CINSARC genes are mostly associated with cell proliferation and therefore lack tumor subtype specificity. Another approach to identify prognostic genes was used by the Pathology Atlas to identify tumor subtype specific prognostic genes. However, soft tissue sarcomas were not analyzed in this study (11).

In this study we used machine learning on open accessible expression data from soft tissue sarcomas to elucidate differences between and within the different entities. First, we investigated the overlap of gene expression patterns of soft tissue sarcomas with gene expression patterns of human tissues without malignancies from the GTEx project (12) using clustering with PCA and a deep neural network. Second, we identified novel diagnostic markers using a random forest approach. Third, we identified tumor subtype specific prognostic genes and showed, using a k-nearest neighbor analysis, that the iden-

tified prognostic genes are predictive of the metastasis-free interval. Last, we analyzed differential expression in the context of a regulatory network to identify novel therapies. We demonstrate that machine learning can be a powerful tool to identify novel diagnostic and prognostic biomarkers, as well as therapeutic targets, which will improve our understanding of rare soft tissue sarcomas.

9.3 Materials and methods

9.3.1 Expression data

The Cancer Genome Atlas (TCGA) RNA-seq count data was downloaded (February 2018) from the NIH GDC data portal (portal.gdc.cancer.gov/). All clinical data corresponding to the soft tissue sarcoma samples in the TCGA was recently revised by the Cancer Genome Atlas Research Network which resulted in 206 revised cases with clinical data (from the original 261 cases in the TCGA) (3). Soft tissue leiomyosarcoma (STLMS) was the most common sarcoma type with 53 samples and included cases of grade 1 (n=11), grade 2 (n=35) and grade 3 (n=7) according to the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) grading system. In addition, there were 27 uterine leiomyosarcoma (ULMS) cases. Furthermore, the TCGA included 50 dedifferentiated liposarcomas (DDLPS), 44 undifferentiated pleomorphic sarcomas (UPS), 17 myxofibrosarcomas (MFS), 10 synovial sarcomas (SS, both monophasic and biphasic) and 5 malignant peripheral nerve sheath tumors (MPNST). Second, the Genotype-Tissue Expression (GTEx) data (v7) was downloaded (`gtexportal.org`) with corresponding annotations. The data consisted of transcriptome sequencing read counts for 9662 samples. The GTEx data included expression data for 31 different tissue types (S1 table available online). Third, DDLPS (n=62) and leiomyosarcoma (LMS) (n=84) expression array data from the French Sarcoma Group was downloaded from GEO (`ncbi.nlm.nih.gov/geo`), deposited under accession number GSE21050 (public in June 2010), using GEOquery (v3.6) in R (13).

9.3.2 Normalization of expression data

Genes with low expression (transcriptome sequencing read counts: cpm < 2; expression array: relative measured unit < 2) in all samples were removed. Thereafter, transcriptome sequencing read count and expression array data were normalized using Limma (v3.6) R package. For normalization, the weighted trimmed mean of M-values was used (14). Last, the data was log2 transformed and analyzed further. When indicated, data was combined and normalized. Where indicated samples were randomly subdivided into groups using

the "sample" function in R.

9.3.3 Machine learning analysis

For the deep neural network TensorFlow (v1.6) was used in combination with the Keras (v2.1.4) R package to design a neural network with one converging invisible layer. t-SNE was performed using the Rtsne (v0.13) R package. For t-SNE analysis a perplexity of 60 and a theta of 0.5 were used. Random forest analysis was performed on the normalized TCGA expression data. Data were analyzed according to Breiman's random forest algorithm, using the randomForest (v4.6) R package. Variable importance in the random forest analysis was calculated based on the Gini index, which is a measurement of variance for a given variable. For k-Nearest Neighbor analysis the Caret (v6.0) R package was used. To resample the data, the "repeatedcv" option was used and k=1-30 were tested.

9.3.4 Enrichment analysis

The EnrichR (v1.0) R package was used for Gene Ontology (GO) term enrichment analysis. GO terms were selected from the "GO biological processes 2015" database and had adjusted p values lower than 1e-4.

9.3.5 Kaplan-Meier analysis

As readout disease-free interval (DFI) was used, which was previously described as a strong measurement of outcome in soft tissue sarcomas (15). DFI is the time until relapse, including distant metastasis and loco-regional recurrence. Prognostic genes were identified using the maxstat (v0.7) R package. Maxstat determined the maximal rank statistic using a LogRank analysis, to determine the optimal gene expression cut-off. P values were calculated according to the Streitberg algorithm (16). Version 18 of the Human Protein Atlas data was downloaded to cross-check prognostic genes identified in other tumor types (`proteinatlas.org/about/download`). This dataset included genes and their association with disease outcome in common cancer types.

9.3.6 Immunohistochemistry and analysis

Immunohistochemistry (IHC) was performed on one existing tissue microarrays (TMA) and one newly constructed TMA. The TMA was constructed as previously described by our group (17). Clinicopathological details are summarized in S2 Table available online. All the specimens were coded and handled according to the ethical guidelines described in the Code for Proper Secondary Use of Human Tissue in the Netherlands of the Dutch

Federation of Medical Scientific Societies as reviewed and approved by the LUMC ethical board (B17.036). In total, seventy leiomyosarcomas could be scored for HMMR protein expression and had available clinicopathological information. The cases originated from two cohorts: the first contained 32 cases that could be scored and has been previously described by our group (17), the second consisted of 38 cases that could be scored. IHC was performed simultaneously on all cases to enable comparison between the cohorts. The 70 cases consisted of 43 females and 27 males, with a mean age of 62 years at diagnosis. Five patients had uterine LMS, the rest were soft tissue LMS. Soft tissue LMS were graded according to the FNCLCC grading system, including 10 grade 1, 23 grade 2, 31 grade 3 and for 1 grading was not available. HMMR was detected with a polyclonal rabbit antibody (Sigma-Aldrich; HPA040025). The HMMR antibody was titrated on normal testis tissue, the optimal antibody dilution was found to be 1:1000 in PBS/1%BSA/5%/non-fat dry milk. Microwave antigen retrieval was performed using citrate (pH 6.0) and immunohistochemistry was performed according to standard protocols (18). Scoring was performed using ImageJ (v1.5) in which color deconvolution was used to separate haematoxylin and 3,3'-Diaminobenzidine (DAB) staining. Haematoxylin was used to identify the core and intensity of the DAB was quantified and compared between cores. A cut-off score of 20 was used to define high and low expressing cores. The second cohort was also scored manually by a pathologist (JVMGB) blinded towards clinicopathological data and results of the automatic scoring, in which staining intensity was scored as weak (1), moderate (2) or strong (3). For the analysis, the average of the three cores per tumor were used.

9.3.7 Quantitative reverse transcriptase Polymerase Chain Reaction (qRT-PCR)

Frozen tissue from five SS and four MPNSTs was retrieved from our archive and anonymized. All selected MPNSTs were either associated with a nerve, were NF1 related or had reported loss of H3K27me3 at immunohistochemistry (19, 20). All selected synovial sarcomas were proven to be positive for the *SS18-SSX* translocation. RNA was isolated using the Direct-zol RNA isolation kit (Zymo research). cDNA was made using iScript cDNA Synthesis Kit (Bio-Rad). Real-time PCR was performed using Sybr Green (Bio-Rad) on a CFX384 real-time PCR Detection System (Bio-Rad). Real-time PCR Ct values were normalized to housekeeping gene *HPRT1* expression. The used primers, noted as 5' to 3', are listed in table 9.1.

NEURL1 _Fw	GCATCCTCGGCTCCACTATC
NEURL1 _Rv	CTGAGCAAGGGGTCAGACAG
SCD _Fw	CTTGCGATATGCTGTGGTGC
SCD _Rv	CCGGGGGCTAATGTTCTTGT
NPAS1 _Fw	CAGCTGCTACCAGTTTGTCCAC
NPAS1 _Rv	ACCCTTGTCCAGCAAGTCCAC
HPRT1 _Fw	TGACACTGGCAAAACAATGCA
HPRT1 _Rv	GGTCCTTTTCACCAGCAAGCT

Table 9.1: Used primers.

9.3.8 Cell growth and viability assay

Cells were cultured in RPMI 1640 medium (Gibco) supplemented with 10% FBS. Cells were tested for mycoplasma and Short Tandem Repeats were characterized for authentication. One SS cell-line was used (SYO-1) (21). Three LMS cell lines were included (JA192, LMS04 and LMS05). Quisinostat (Selleckchem) and trichostatin A (Selleckchem) were used for HDAC inhibition. Both compounds were dissolved in DMSO. Cells were seeded in triplicates on a 96-well plate and compounds were added after 24 hours. Cell viability was measured after 72 hours incubation with the compounds by adding Presto-Blue Cell Viability Reagent (Life Technologies) according to the manufacturers protocol. Fluorescence was measured reading the plate at 590 nm on a fluorometer (Victor3V, 1420 multi-label counter). Viability was determined in three independent experiments in triplicate.

9.3.9 Connectivity Map analysis

For Connectivity Map (CMAP) analysis the regulatory network was first determined using expression2kinase (`maayanlab.net/X2K`) based on the differentially expressed genes that were identified. Potential targeted therapies were identified based on the proteins in the regulatory network. The pipeline for identification of transcription factors and kinases is described in literature (22).

9.3.10 Statistical software and figures

R statistical software (v3.4.4) was used for all statistical tests (13). Network plots were generated with igraph (v1.2.1) R package and formatted with Cytoscape (v3.6.0) (23). Chord diagrams were generated with GOplot (v1.0.2) (24). All further graphs were generated with R package ggplot2 (v2.2.1). Cox regression was performed with the "coxph" function from the survival (v2.43) R package.

abbreviation	histology	cases
DDLPS	dedifferentiated liposarcoma	50
MFS	myxofibrosarcoma	17
MPNST	malignant peripheral nerve sheath tumor	5
SS	synovial sarcoma	10
STLMS	leiomyosarcoma - soft tissue	53
ULMS	leiomyosarcoma - gynecologic	27
UPS	undifferentiated pleomorphic sarcoma	44

Table 9.2: Soft tissue sarcoma subtypes in the TCGA.

9.4 Results

9.4.1 Results

9.4.2 Soft tissue sarcomas show different molecular profiles

Since soft tissue sarcomas are histologically classified according to their line of differentiation, we compared gene expression data from 206 soft tissue sarcoma samples in The Cancer Genome Atlas (TCGA) (table 9.2) with normal tissues from the Genotype-Tissue Expression (GTEx) project. For this we used a deep neural network approach, enabling us to find similarities between normal tissues and tumors identified through hidden layers that would not be obvious in a direct comparison (such as a PCA analysis). First the TCGA and GTEx data were combined and normalized together (supplementary figure S1a available online). Principal components were calculated for all samples, the principal components (9868 in total) for the GTEx data was used to train a neural network resulting in a prediction accuracy of 98% (supplementary figure S1b available online). The neural network was then applied to the principal components from the TCGA sarcoma data.

As might be expected, ULMS was the only sarcoma subtype showing overlap with the expression patterns of normal uterus tissue as well as normal cervical tissue (supplementary figure S1c available online). Moreover, STLMS was the only subtype showing similarity to blood vessel, which may be explained by the fact that a subset of STLMS are presumed to arise from small to medium sized veins (25). However, both ULMS and STLMS also showed overlap with skin and brain tissue which is more difficult to understand at this point. Interestingly, we found large similarities between MPNST and SS, showing expression patterns very similar to tissue derived from the nervous system (brain and nerve). In addition SS showed some overlap with salivary gland which might be explained by the fact that 2 out of 10 SS were biphasic, of which the glandular epithelial elements may have caused the found similarity with salivary gland (figure 9.1a). Sur-

prisingly, MFS, and to a lesser extent UPS, showed a large overlap with normal adipose tissue. The overlap with adipose tissue in MFS and UPS is larger than found in DDLPS, which could be due to the selective sampling of DDLPS including the dedifferentiated component. For the other soft tissue sarcoma subtypes similarities were more dispersed since no specific normal tissue showed a large overlap with the tumor gene expression (supplementary figure S1c available online).

To study the gene expression patterns of soft tissue sarcomas the TCGA expression data was normalized and differentially expressed genes (DEGs) were identified (Benjamini-Hochberg adjusted p value < 0.05 and logFC > 0) for all soft tissue sarcoma subtypes using Limma and Voom, comparing the subtypes to the other samples (supplementary figure S1d available online). The number of DEGs per subtype ranged from 331 to 7784 (in STLMS and DDLPS respectively, 3156 DEGs on average) (supplementary figure S1e available online). The DEGs were used to generate a heat map showing differences between soft tissue sarcoma subtypes. MFS and UPS showed the largest overlap in DEGs (1201 genes) followed by STLMS and ULMS (210 genes) (figure 9.1b). Using EnrichR we tested for functional enrichment of the DEGs to identify GO terms associated with each of the subtypes. The DEGs from STLMS and MPNST showed a clear relation to differentiation; GO terms for STLMS related to muscle development and for MPNST the GO terms related to neuronal development. The top GO terms associated with ULMS were not related to muscle differentiation, but with cell cycle processes. However, significant GO terms associated with muscle differentiation were identified such as "muscle system process" (adjusted p=6e-4) and "muscle contraction" (adjusted p=3e-3) matching with the GO terms found in STLMS, which suggests that proliferation was more pronounced than differentiation in the ULMS compared to the STLMS samples. We did not identify GO terms related to differentiation for DDLPS, but, as can be seen in the heat map, we found that many of the identified GO terms associated with DDLPS, UPS and MFS overlapped. These included GO terms associated with the immune system which may reflect the presence of an inflammatory infiltrate in these tumors (figure 9.1b).

9.4.3 A random forest approach can differentiate between the soft tissue sarcoma subtypes

To investigate the similarities of the molecular profiles of the different soft tissue sarcoma subtypes we performed a t-SNE analysis on the expression data (supplementary figure S2a available online). The average of the first two components for the different subtypes is shown in figure 9.2a. In the t-SNE analysis, three clusters of soft tissue sarcoma subtypes were identified. MFS, UPS and DDLPS clustered together, in line with the undifferentiated sometimes pleomorphic morphology of these tumors. ULMS and STLMS

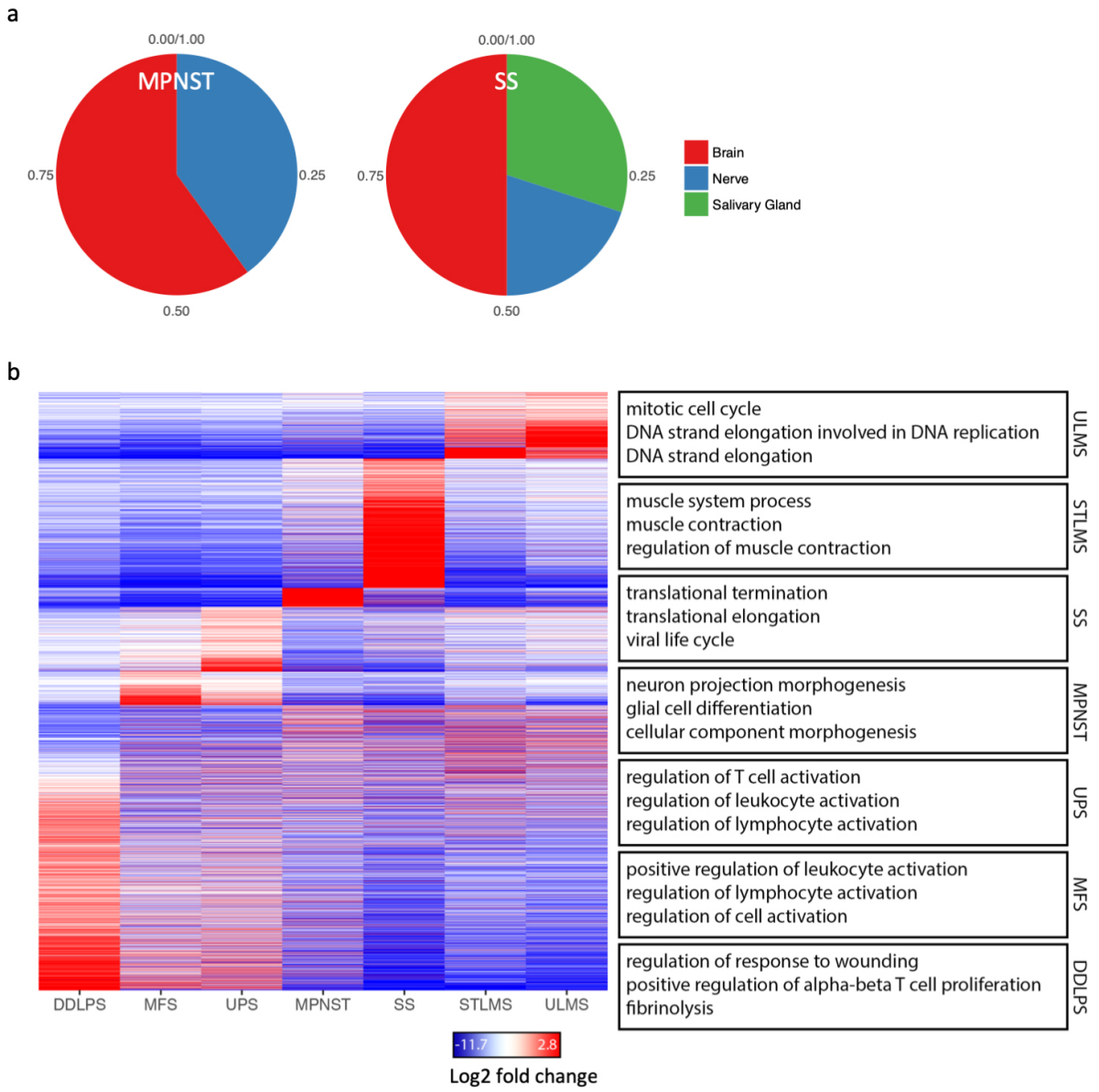


Figure 9.1: Relation to normal tissue and molecular profiles of soft tissue sarcomas. (a) A deep neural network was trained on GTEx expression data from normal tissue to investigate differentiation in the soft tissue sarcoma subtypes. MPNST and SS both showed the most specific differentiation and showed largest similarity with brain and nerve gene expression profiles. (b) Heat map plot of the identified signature genes in the different soft tissue sarcoma subtypes. The largest overlap in signature genes is seen between UPS and MFS (1201). Enriched GO terms in each of the signature genes are shown in the right panel. All GO terms have an adjusted P value lower then 1e-4.

also cluster together. The third cluster consisted of MPNST and SS, for which distinction based on morphology alone is often impossible.

As a deep neural network is not informative on the biological differences between these subtypes, we therefore used a random forest machine learning approach to identify subtype defining genes. The samples were divided into test and training groups at random. The resulting random forest reached a subtype prediction accuracy of over 95% for all groups, except in differentiating between MFS and UPS (where it reached an accuracy of 88%) (figure 9.2b).

Differentially expressed genes (adjusted $p<0.05$) were used to generate the random forest. Important genes were identified based on their variable importance index (figure 9.2c). Top genes in group 1 (STLMS and ULMS) included *HOXA11* and its anti-sense RNA (*HOXA11-AS*) were identified. *HOXA11* and *HOXA11-AS* have both been described to be important regulators of uterine development and homeostasis (26). For group 2 (MPNST and SS) genes related to neural differentiation such as *NEURL1* and *NPAS1* were identified, which were found to be upregulated in synovial sarcomas, while *SCD*, an enzyme involved in fatty acid biosynthesis, is more highly expressed in MPNST. For the third group (DDLPS, UPS and MFS), we first compared DDLPS with the UPS and MFS together. As previously described and already widely implemented in routine diagnostics, expression of *MDM2* and *CDK4* (which is part of the 12q13-15 amplification characteristic of DDLPS) were identified as diagnostic markers to identify DDLPS (27). *FRS2*, *TSPAN31* and *CTDSP2* are located near the amplified *MDM2* on chromosome 12 and therefore most likely also part of the same amplified region that characterizes DDLPS. In figure 9.2d, we visualized gene expression levels of the genes with the highest variable importance scores for each of the four comparisons. *JADE2* showed the highest variable importance score for the differentiation between UPS and MFS although expression still somewhat overlapped, confirming the large molecular and morphological similarity between the two entities (figure 9.2d).

To verify the diagnostic markers that were identified for group 2 (MPNST and SS) using the random forest algorithm we used qRT-PCR on an independent cohort of nine samples. Indeed, the expression patterns of *NEURL1*, *SCD* and *NPAS1* were similar in the independent cohort (figure 9.2e).

9.4.4 Soft tissue sarcoma subtypes have distinct prognostic genes

We identified prognostic genes for all annotated soft tissue sarcoma subtypes, except MPNST (with only five samples available). First, the optimal gene expression cutoff was calculated for all the 24168 genes that met the defined thresholds in the TCGA soft tissue sarcoma expression data. Next, disease-free interval (DFI) (time to local recurrence or

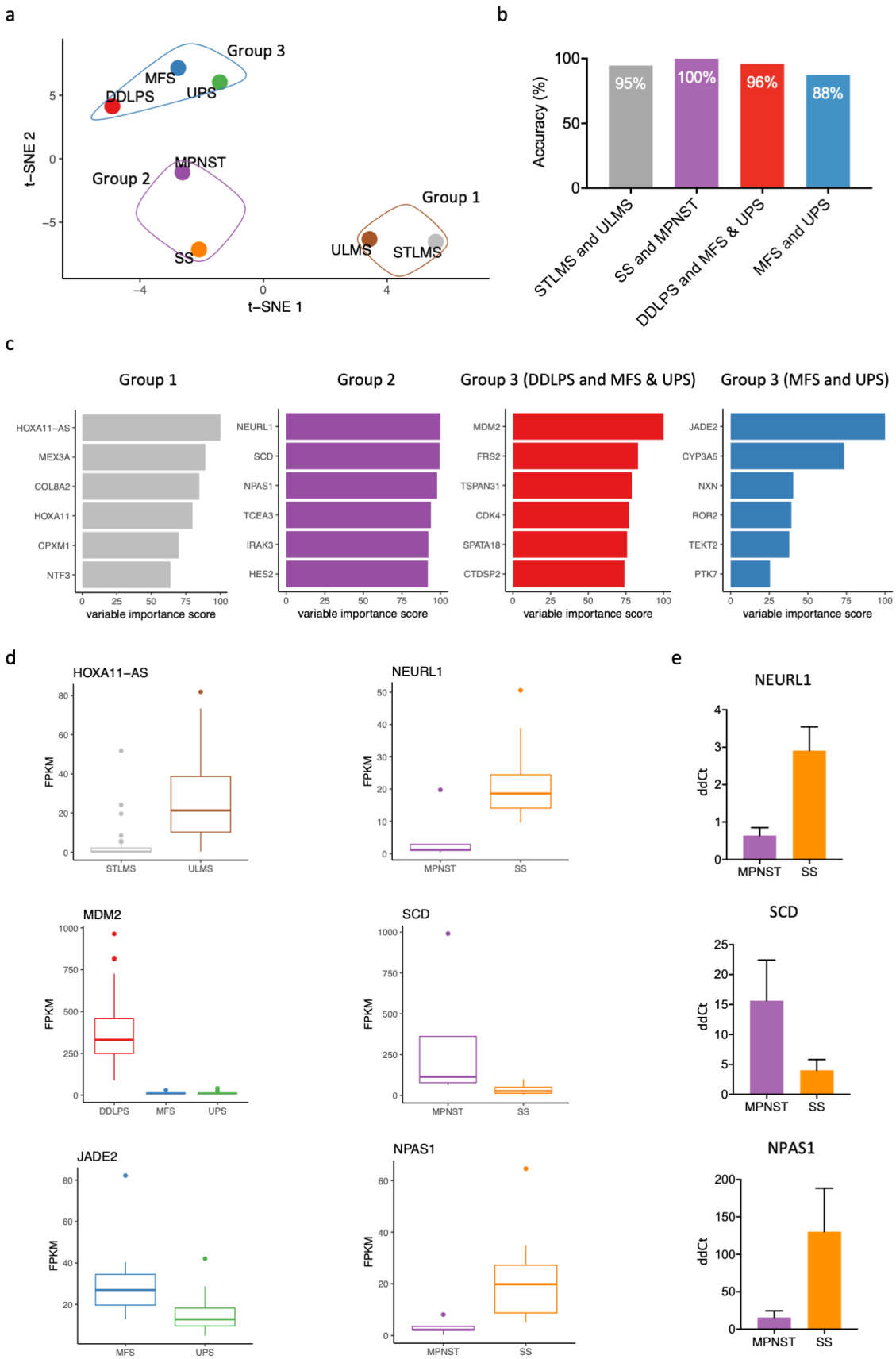


Figure 9.2: Caption on next page.

Figure 9.2: Diagnostic markers to distinguish within three subgroups. (a) T-SNE analysis of all soft tissue sarcoma subtypes in the TCGA. The first two components were used to generate the diagram. Three groups could be identified based on the molecular profile: group 1 (STLMS and ULMS); group 2 (SS and MPNST); group 3 (DDLPS, UPS and MFS). (b) A machine learning random forest analysis was trained and tested on a test dataset. Random forests were generated to differentiate between STLMS and ULMS, SS and MPNST, DDLPS and MFS with UPS and last between MFS and UPS. Within the three identified groups a prediction accuracy of over 95% was reached, except when differentiating between UPS and MFS (88%). (c) From the random forest models, the top five genes were selected based on their Gini index, score is shown relative to the best diagnostic marker. (d) Gene expression (in FPKM) for the best subtype predictor within the identified groups is shown in the boxplots on the left. On the right the top three subtype predictors are shown for group 2 (MPNST and SS), which were verified using qRT-PCR. The box shows the interquartile range from Q1 to Q3 and the mean. The whiskers show the highest and lowest values. Suspected outliers (interquartile range * 1.5) are shown as separate dots. (e) qRT-PCR validation in independent cohort: Delta-delta Ct (ddCt) values are shown for the top three diagnostic genes identified for group 2 (MPNST and SS). Expression pattern is similar to what was found in the TCGA data. Expression was normalized with a housekeeping gene (*HPRT1*).

distant metastases) was tested using the Hothorn and Lausen statistical test; DFI was used as the read-out.

In total 429 genes were found to be strong predictors (favorable or unfavorable) of DFI ($p < 0.001$). Most genes were identified for SS (166 genes) while 74 and 34 genes were identified for STLMS and ULMS respectively. Interestingly, there was very little overlap between the prognostic genes for the different subtypes. Two overlapping prognostic genes (*KLF6* and *MT1F*) were found for UPS and SS and one (*NPM2*) for ULMS and MFS. No overlapping prognostic genes were found between STLMS and ULMS (figure 9.3a). Furthermore, only one gene (*CDCA3* identified in STLMS) was found to overlap between the 67 described CINSARC genes and the soft tissue sarcoma subtype specific prognostic genes identified in the current study. From the 429 identified prognostic genes 201 were new, 228 had however been previously identified in other (non-sarcoma) tumor types in the Protein Atlas database (supplementary figure S3a available online).

To cross-check the identified prognostic genes identified for LMS, DDLPS and UPS, we used expression data from the French Sarcoma Group (9). The French Sarcoma Group array data was first normalized (supplementary figure S3b available online). The data contained information on the metastasis-free interval but not DFI as was used by us for the TCGA data. The French Sarcoma Group data was split in two groups. Genes that were significant prognostic genes for DFI in the TCGA and the metastasis-free interval in the first French Sarcoma Group cohort (both with $p < 0.05$) were considered

for further analysis (S4 table available online). From the identified genes, strong prognostic genes were used in a k-nearest neighbor analysis. For LMS *HMMR*, *MXD4* and *BRCA2* were identified, for DDLPS *KLF6* was found to be a strong prognostic gene while for UPS *PCMTD2*, *TNXA*, *TMEM65*, *SNRNP48* were identified. The k-nearest neighbor algorithm was trained on the first group and tested on the second group in the French Sarcoma samples. The k-nearest neighbor algorithm was a significant predictor for the metastasis-free interval for LMS, DDLPS and UPS in the second group (p=0.045, p=0.02 and p=0.012 respectively) (figure 9.3b), outperforming the reported CINSARC classification in the second cohort (LMS p=0.24, DDLPS p=0.14 and UPS p=0.038) (supplementary figure S3c available online).

HMMR was identified as a significant ($p<0.05$) prognostic gene for DFI and the metastasis-free interval in LMS. In an independent validation set of 70 LMS cases, we verified using immunohistochemistry with automated scoring that high protein expression of *HMMR* was associated with a shorter DFI (p=0.0061) (figure 9.3c&d). For the second cohort, manual scoring was compared with automated scoring and results were similar. Prognostic value of *HMMR* was further compared to the FNCLCC grading system. In a multivariate Cox-regression it was found that the *HMMR* staining (p=0.0039) retained significance and was a better predictor than FNCLCC histological grade (p=0.285).

9.4.5 Systems analysis of the soft tissue sarcoma subtype-specific genes identify targeted therapies

To identify novel targeted therapies gene expression data was used to infer the regulatory transcription factors and kinases in the different soft tissue sarcoma subtypes. First, the signature genes for each soft tissue sarcoma subtype were used to infer the transcription factors that were most likely to regulate those genes based on data from the ChIP-seq Enrichment Analysis (ChEA) database (22). The most important kinases regulating these transcription factors were inferred using the Kinase Enrichment Analysis (22). Based on the identified transcription factors and kinases, tumor subtype specific drugs were identified based on the Connectivity Map (CMAP) drug data (with kinases and transcription factors as input). Doxorubicin, which is commonly used as systemic treatment for STS, was identified as a potentially effective therapy for most soft tissue sarcoma subtypes, validating our analysis approach. Trichostatin A, a HDAC inhibitor, was predicted to be potentially efficient in all soft tissue sarcoma subtypes, while another HDAC inhibitor, Vorinostat, was identified for UPS and ULMS. Tanespimycin was identified for UPS, ULMS and MPNST, which is an inhibitor of Hsp90 and currently used in clinical trials for solid tumors (figure 9.4a and S5 table available online). While sensitivity to HDAC inhibition is known for translocation driven tumors like synovial sarcoma (28), for LMS

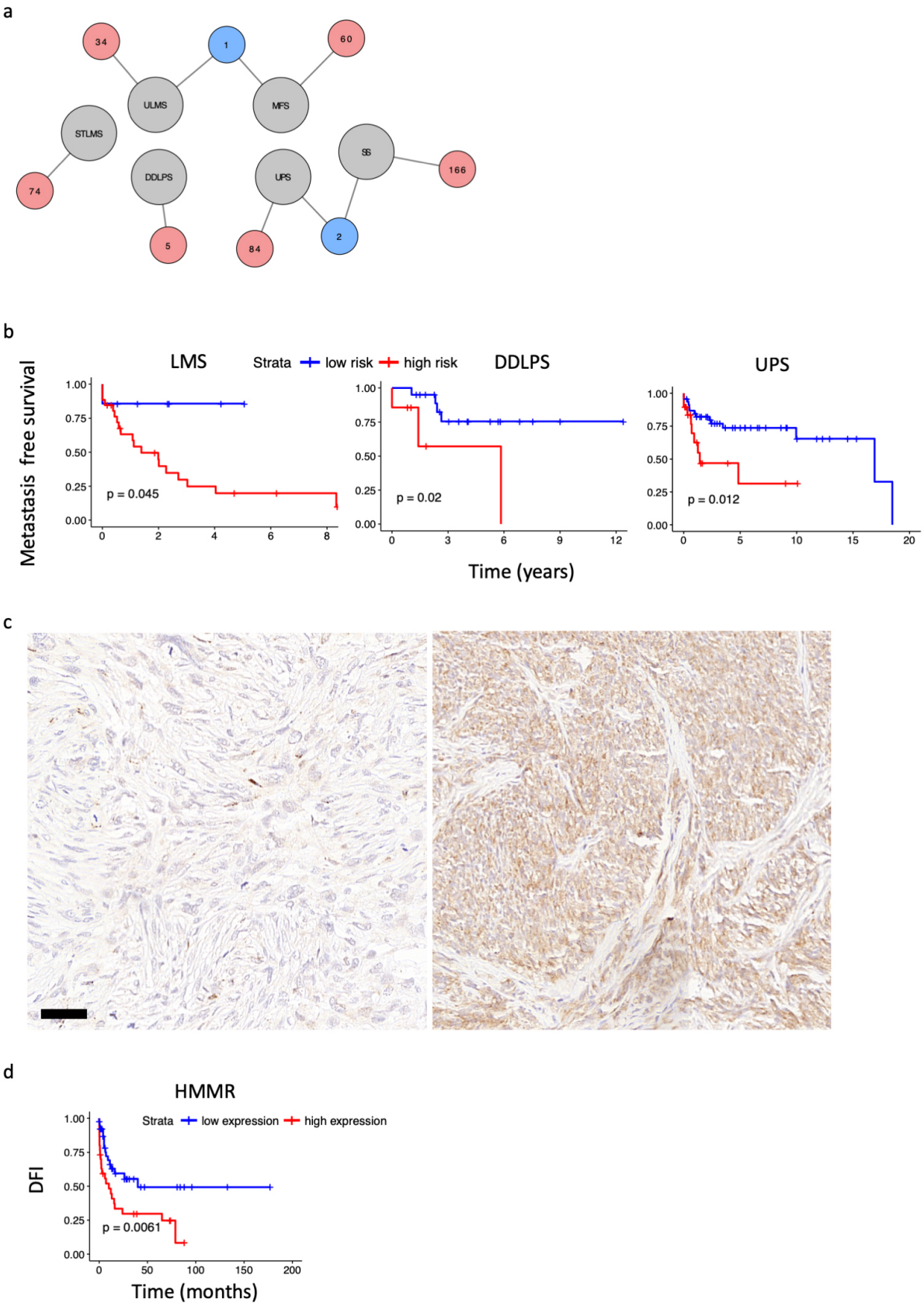


Figure 9.3: Caption on next page.

Figure 9.3: Novel prognostic biomarkers in soft tissue sarcomas. (a) All identified prognostic genes and their overlap within the different soft tissue sarcoma subtypes is shown with a network diagram. UPS and SS share two prognostic genes and ULMS and MFS share 1. Furthermore, all identified genes were specific for each sarcoma subtype. Number of prognostic genes are shown in the red circles, tumor types in the gray circles and number of overlapping prognostic genes in the blue circles. (b) The k-Nearest Neighbor algorithm was also used with expression data for the strongest prognostic genes identified in both the French Sarcoma Group and TCGA expression data. The algorithm was trained on the first and tested on the second cohort. Both were found to be significant predictors of the metastasis-free interval. (c) HMMR protein expression was tested using IHC on a LMS TMA. The left panel shows a representative sample with low expression, on the right a sample with high HMMR expression. Scale bar indicates 50 μm. (d) High HMMR protein expression as seen in an independent cohort of LMS from our archives is associated with poor outcome.

this has not been extensively studied. We thus performed cell viability assays on three LMS cell lines (JA192, LMS04 and LMS05), treated with two HDAC inhibitors (quisinostat and trichostatin A), with one SS cell line (SYO-1) as positive control (figure 9.4b). For both compounds the half maximal inhibitory concentration (IC50) was determined. For trichostatin A (TSA) an IC50 ranging from 39 to 474 nM was found (JA192: 474 nM; LMS04: 229 nM; LMS05: 178 nM; SYO-1: 39 nM). Although all cell-lines were sensitive to TSA, SYO-1 was more sensitive compared to the LMS cells. However, for quisinostat a low IC50 was found for all cell-lines; between 15 and 41 nM (JA192: 41 nM, LMS04: 34 nM; LMS05: 39 nM; SYO-1: 15 nM). These results indicate that LMS and SS cell lines are highly sensitive to HDAC inhibition by quisinostat.

9.5 Discussion

Accurate diagnosis and prediction of biological behavior is a challenge for soft tissue sarcoma pathologists. These tumors are rare and often overlap in their morphology, while subtype specific diagnostic and prognostic markers are scarce. As an increasing amount of transcriptome sequencing data becomes available, even for rare cancers such as soft tissue sarcomas, new methods need to be developed to identify novel diagnostic and prognostic biomarkers for these tumors from existing data.

Here we used machine learning algorithms to identify similarities and differences between soft tissue sarcoma subtypes and normal human tissue from the GTEx data. Using a deep neural network, we demonstrate that SS and MPNST mostly correspond to neural related tissues. MPNST often arises from or within nerves; therefore, it is likely a tumor originating from neural related tissue, while for synovial sarcoma the cell of origin and

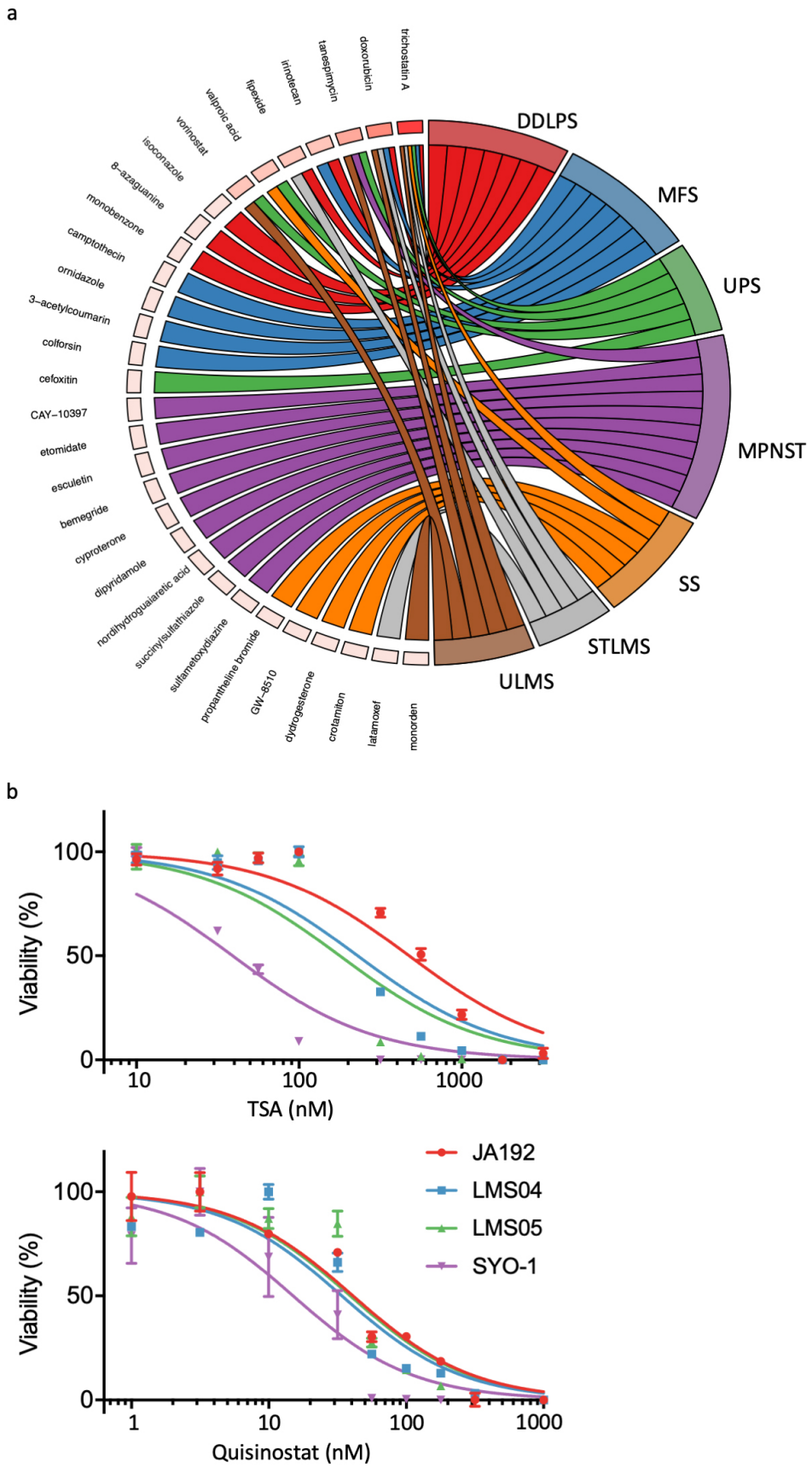


Figure 9.4: Caption on next page.

Figure 9.4: CMAP analysis to identify novel therapies. (a) CMAP analysis identifies potential drugs based on the expression profile. The chord diagram shows links between the drugs and soft tissue sarcoma subtypes. Some compounds such as trichostatin A, doxorubicin and tanespimycin show connections with multiple soft tissue sarcoma subtypes, which is illustrated by the box color for each drug (darker red indicates more connections). (b) The dose response curves are shown for both trichostatin A (TSA) and quisinostat as tested in one SS (SYO-1) and three LMS (JA192, LMS04 and LMS05) cell-lines.

line of differentiation have been unclear. Our observation of the neural related tissue as a potential tissue of origin confirms previous suggestions (29). The deep neural network also identified that cervix and uterine tissue showed the largest overlap with ULMS as is expected. Other findings however illustrate the limitations in comparing gene expression of normal tissue with tumor, such as the large overlap in gene expression between skin and adrenal gland with ULMS or the large overlap found between SS and salivary gland (that could be explained due to the biphasic SS samples displaying epithelial elements). These findings in part could be explained by the fact that the sequencing is performed on tissue containing many different cell types, including immune and stromal cells. Single cell sequencing and projects such as the Human Cell Atlas (30) could in the future shed more light on the tissue of origin for soft tissue sarcomas.

Using a random forest analysis, we identified subtype specific genes that can be used as diagnostic markers within the three groups of soft tissue sarcoma subtypes that were identified based on their molecular profile and morphology. For instance, *NEURL1* was one of the genes highly expressed in SS as compared to MPNST. *NEURL1* is an important determinant of neural tissue differentiation and functions as a tumor suppressor which is inactivated during malignant progression of astrocytic tumors (31). In line with this, the lower expression of *NEURL1* could be explained by recurring losses of chromosome 10 in 48% of MPNST (32). *SCD* was found to be highly expressed in MPNST compared to SS. SCD has been found to associate with a poor prognosis in breast and lung cancer. Moreover, *SCD* can be directly inhibited with the small molecule MF-438 which sensitized adenocarcinoma cells to cisplatin treatment (33, 34). It was previously found that when SS was treated with a HDAC inhibitor, neural differentiation was induced (28). Furthermore, treatment with BMP4 or FGF2 restored expression of neural tissue related genes in SS (29). Our study further confirms neural differentiation in SS, as shown using hidden layers in a deep neural network. Future validation studies should indicate whether the diagnostic biomarkers that we identified here can also be used immunohistochemically in the differential diagnosis.

We identified subtype specific prognostic genes using Kaplan-Meier analysis on all individual genes combined with a k-nearest neighbor algorithm to accurately predict the

disease-free interval (DFI). DFI was previously shown to be one of the strongest outcome measurements for soft tissue sarcomas (15). For all genes the cut off was determined first and the DFI for high and low expression was calculated. This Kaplan-Meier approach was previously used on 17 other cancer types, not including soft tissue sarcomas (11). Although this method results in tumor subtype specific prognostic genes that can predict outcome, a major challenge is to correct for multiple testing. Here we used an independent cohort from the French Sarcoma Group to validate strong prognostic genes for LMS, DDLPS and UPS. However, for this independent cohort only data on metastasis were available, whereas the TCGA also contained data on loco-regional recurrence. Using both data sets, overlapping prognostic genes were identified which could be considered strong prognostic genes. For the other tumor subtypes, to our knowledge, there are no available expression data sets with accurate follow up data to perform cross-validation. Interestingly we found only one gene, *CDCA3*, overlapped between the prognostic genes we identified in the TCGA soft tissue sarcoma data and the CINSARC prognosticator. We likely did not identify a larger overlap because the CINSARC study aimed to identify a general prognosticator for soft tissue sarcomas, which is not subtype specific. In addition, the outcome used was different; we used DFI as an outcome measurement while in the CINSARC study metastasis was used. Moreover, we identified subtype specific prognostic genes using a Kaplan-Meier approach which does not only take outcome but also time to events into account. Here we showed that subtype specific prognostic genes outperformed general prognostic genes.

For one of the identified genes, *HMMR*, we confirmed that high protein expression was associated with poor outcome of LMS. Further we confirmed that *HMMR* expression outperformed the FNCLCC histological grading to predict outcome. Recently it was shown that LMS displays hallmarks of "BRCAness" through identification of mutation signatures and alterations in genes related to homologous recombination (35). Here we identified strong prognostic genes for LMS, two of which were related to homologous repair (BRCA2 and HMMR). HMMR forms a complex with BRCA1 or BRCA2 together with other proteins, and high expression of HMMR was associated with poor survival in liver, pancreatic and lung cancer (11). Possibly, defects in the homologous repair pathway could result in over-expression of HMMR in an attempt to compensate for other defective proteins. The involvement of genes related to "BRCAness" and to disease outcome warrants further studies.

A regulatory network reconstruction combined with the CMAP drug data revealed not only the commonly used drug doxorubicin, but also indicated that HDAC inhibitors could be a potential treatment for many different soft tissue sarcoma subtypes. Recent studies indeed suggest that HDAC inhibitors may be effective in treating soft tissue sarcomas. In

liposarcoma it was shown that HDAC inhibitors increase apoptosis and anti-proliferation effects (36). In SS HDAC inhibitors cause disruption of the *SS18-SSX* oncoprotein resulting in apoptosis (28). Another study found HDAC inhibitors lead to apoptosis in SS cell-lines (37). In other sarcoma subtypes HDAC inhibitors have not been studied extensively. One uterine LMS cell line was tested and shown to be sensitive to the pan HDAC inhibitor ITF2357 with a synergistic effect when combined with doxorubicin (38). In this study we further investigated LMS sensitivity to HDAC inhibition using quisinostat and trichostatin A. We included three LMS cell-lines, one ULMS (LMS04) and two STLMS (LMS05 and JA192). As SS was previously found to be sensitive to HDAC inhibition we also included one SS cell-line (SYO-1) as a positive control. SS showed a greater sensitivity to TSA, however, quisinostat showed a very low IC50 (15-41 nM) in all cell lines. Thus, quisinostat might be further explored as a potential therapy for both ULMS and STLMS.

In conclusion, three groups of soft tissue sarcoma subtypes included in the TCGA study were identified based on similarities in their expression profiles, corresponding to their overlapping morphology. Using a random forest analysis, novel diagnostic markers were identified that may distinguish between soft tissue sarcoma subtypes within these three groups, including *NEURL1* that was highly expressed in SS as compared to MPNST. Next, using a Kaplan-Meier analysis, prognostic genes were identified. Of these, HMMR protein expression was confirmed to be associated with poor outcome in an independent cohort of LMS from our archives. A network reconstruction combined with CMAP data revealed that HDAC inhibitors could be effective therapy in different soft tissue sarcoma subtypes, which we confirmed in LMS and SS cell-lines. In conclusion, machine learning algorithms uncovered diagnostic biomarkers, prognostic genes and identified potential novel therapeutic targets for soft tissue sarcomas. This study thereby illustrates the power of different machine learning algorithms to improve our understanding of rare cancers using existing datasets.

Bibliography

[1] Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. Nat Rev Cancer. 2011;11(8):541–557. doi:10.1038/nrc3087.

[2] Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F. WHO Classification of Tumours of Soft Tissue and Bone. vol. WHO Classi; 2013.

[3] Abeshouse A, Adebamowo C, Adebamowo SN, Akbani R, Akeredolu T, Ally A, et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell. 2017;171(4):950–965.e28. doi:10.1016/j.cell.2017.10.014.

[4] Van Der Maaten L, Hinton G. Visualizing Data using t-SNE; 2008. Available from: `http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf`.

[5] Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences. 2001;98(19):10869–10874. doi:10.1073/pnas.191367098.

[6] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834. doi:doi:10.1038/nature03702.

[7] Röhrich M, Koelsche C, Schrimpf D, Capper D, Sahm F, Kratz A, et al. Methylation-based classification of benign and malignant peripheral nerve sheath tumors. Acta Neuropathologica. 2016;131(6):877–887. doi:10.1007/s00401-016-1540-6.

[8] Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018;555(7697):469–474. doi:10.1038/nature26000.

[9] Chibon F, Lagarde P, Salas S, Perot G, Brouste V, Tirode F, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nat Med. 2010;16(7):781–787. doi:10.1038/nm.2174.

[10] Lesluyes T, Delespaul L, Coindre JM, Chibon F. The CINSARC signature as a prognostic marker for clinical outcome in multiple neoplasms. Sci Rep. 2017;7(1):5480. doi:10.1038/s41598-017-05726-x.

[11] Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science. 2017;357(6352). doi:10.1126/science.aan2507.

[12] The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580–585. doi:10.1038/ng.2653.

[13] null Null. R: A Language and Environment for Statistical Computing; 2017. Available from: `http://www.r-project.org/`.

[14] Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology. 2010;11(3). doi:doi:10.1186/gb-2010-11-3-r25.

[15] Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018;173(2):400–416.e11. doi:10.1016/j.cell.2018.02.052.

[16] Lausen B, Hothorn T, Bretz F, Schumacher M. Assessment of optimal selected prognostic factors. Biometrical Journal. 2004;46(3):364–374. doi:10.1002/bimj.200310030.

[17] De Graaff MA, Cleton-Jansen AM, Szuhai K, Bovée JVMG. Mediator complex subunit 12 exon 2 mutation analysis in different subtypes of smooth muscle tumors confirms genetic heterogeneity. Human Pathology. 2013;44(8):1597–1604. doi:10.1016/j.humpath.2013.01.006.

[18] Baranski Z, Booij TH, Cleton-Jansen AM, Price LS, van de Water B, Bovee JV, et al. Aven-mediated checkpoint kinase control regulates proliferation and resistance to chemotherapy in conventional osteosarcoma. J Pathol. 2015;236(3):348–359. doi:10.1002/path.4528.

[19] Cleven AHG, Al Sannaa GA, Briaire-de Bruijn I, Ingram DR, Van De Rijn M, Rubin BP, et al. Loss of H3K27 tri-methylation is a diagnostic marker for malignant peripheral nerve sheath tumors and an indicator for an inferior survival. Modern Pathology. 2016;29(6):582–590. doi:10.1038/modpathol.2016.45.

[20] Prieto-Granada CN, Wiesner T, Messina JL, Jungbluth AA, Chi P, Antonescu CR. Loss of H3K27me3 Expression Is a Highly Sensitive Marker for Sporadic and Radiation-induced MPNST. Am J Surg Pathol. 2016;40(4):479–489. doi:10.1097/pas.0000000000000564.

[21] Kawai A, Naito N, Yoshida A, Morimoto Y, Ouchida M, Shimizu K, et al. Establishment and characterization of a biphasic synovial sarcoma cell line, SYO-1. Cancer Letters. 2004;204(1):105–113. doi:10.1016/j.canlet.2003.09.031.

[22] Chen EY, Xu H, Gordonov S, Lim MP, Perkins MH, Ma'ayan A. Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. Bioinformatics. 2012;28(1):105–111. doi:10.1093/bioinformatics/btr625.

[23] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–2504. doi:10.1101/gr.1239303.

[24] Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics. 2015;31(17):2912–2914. doi:10.1093/bioinformatics/btv300.

[25] Lazar A, Evans H, Shipley J. Smooth-muscle tumors. Leiomyoma of deep soft tissue. Leiomyosarcoma. In: WHO classification of tumours of soft tissue and bone. Lyon: IARC; 2013. p. 111–3.

[26] Chau YM, Pando S, Taylor HS. HOXA11 silencing and endogenous HOXA11 antisense ribonucleic acid in the uterine endometrium. J Clin Endocrinol Metab. 2002;87(6):2674–2680. doi:10.1210/jcem.87.6.8527.

[27] Binh MB, Sastre-Garau X, Guillou L, de Pinieux G, Terrier P, Lagace R, et al. MDM2 and CDK4 immunostainings are useful adjuncts in diagnosing well-differentiated and dedifferentiated liposarcoma subtypes: a comparative analysis of 559 soft tissue neoplasms with genetic data. Am J Surg Pathol. 2005;29(10):1340–1347.

[28] Laporte AN, Poulin NM, Barrott JJ, Wang XQ, Lorzadeh A, Vander Werff R, et al. Death by HDAC Inhibition in Synovial Sarcoma Cells. Mol Cancer Ther. 2017;16(12):2656–2667. doi:10.1158/1535-7163.mct-17-0397.

[29] Ishibe T, Nakayama T, Aoyama T, Nakamura T, Toguchida J. Neuronal Differentiation of Synovial Sarcoma and Its Therapeutic Application. In: Clin Orthop Relat Res. vol. 466; 2008. p. 2147–2155. Available from: `http://dx.doi.org/10.1007/s11999-008-0343-zhttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2493002/pdf/11999{_}2008{_}Article{_}343.pdf`.

[30] Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. eLife. 2017;6:e27041. doi:10.7554/eLife.27041.

[31] Nakamura H, Yoshida M, Tsuiki H, Ito K, Ueno M, Nakao M, et al. Identification of a human homolog of the Drosophila neuralized gene within the 10q25.1 malignant astrocytoma deletion region. Oncogene. 1998;16(8):1009–1019. doi:10.1038/sj.onc.1201618.

[32] Bridge RS, Bridge JA, Neff JR, Naumann S, Althof P, Bruch LA. Recurrent chromosomal imbalances and structurally abnormal breakpoints within complex karyotypes of malignant peripheral nerve sheath tumour and malignant triton tumour: a cytogenetic and molecular cytogenetic study. In: J Clin Pathol. vol. 57; 2004. p. 1172–1178. Available from: `http://dx.doi.org/10.1136/jcp.2004.019026`.

[33] Holder AM, Gonzalez-Angulo AM, Chen H, Akcakanat A, Do KA, Symmans WF, et al. High stearoyl-CoA desaturase 1 expression is associated with shorter survival in breast cancer patients. Breast Cancer Res Treat. 2013;137(1):319–327. doi:10.1007/s10549-012-2354-4.

[34] Pisanu ME, Noto A, De Vitis C, Morrone S, Scognamiglio G, Botti G, et al. Blockade of Stearoyl-CoA-desaturase 1 activity reverts resistance to cisplatin in lung cancer stem cells. Cancer Letters. 2017;406:93–104. doi:10.1016/j.canlet.2017.07.027.

[35] Chudasama P, Mughal SS, Sanders MA, Hübschmann D, Chung I, Deeg KI, et al. Integrative genomic and transcriptomic analysis of leiomyosarcoma. Nature Communications. 2018;9(1):144. doi:10.1038/s41467-017-02602-0.

[36] Ou WB, Zhu J, Eilers G, Li X, Kuang Y, Liu L, et al. HDACi inhibits liposarcoma via targeting of the MDM2-p53 signaling axis and PTEN, irrespective of p53 mutational status. Oncotarget. 2015;6(12):10510–20. doi:10.18632/oncotarget.3230.

[37] Bernhart E, Stuendl N, Kaltenegger H, Windpassinger C, Donohue N, Leithner A, et al. Histone deacetylase inhibitors vorinostat and panobinostat induce G1 cell cycle arrest and apoptosis in multidrug resistant sarcoma cell lines. Oncotarget. 2017;8(44):77254–77267. doi:10.18632/oncotarget.20460.

[38] Di Martile M, Desideri M, Tupone MG, Buglioni S, Antoniani B, Mastroiorio C, et al. Histone deacetylase inhibitor ITF2357 leads to apoptosis and enhances doxorubicin cytotoxicity in preclinical models of human sarcoma. Oncogenesis. 2018;7(2):20. doi:10.1038/s41389-018-0026-x.

Part IV

Summary and discussion

Chapter 10

Summary and discussion

Intermediate and malignant vascular tumors are extremely rare tumors that can occur in both soft-tissue and bone. A great challenge in studying these entities is the lack of adequate models, both *in vitro* and *in vivo*. To overcome this challenge, we have developed a number of models and used computational biology to find relevant and targetable pathways in vascular tumors. The methods, described in this thesis, to develop the model systems that were used to study the vascular tumors can also be used for other tumors where cell lines are not available. Moreover, some of the findings from the vascular tumors give insight into the biology of normal endothelial cells. This thesis is subdivided into three parts; diagnosis and treatment, model systems, and computational biology. **Chapter 1** introduces the vascular tumors, model systems and computational biology concepts.

10.1 Diagnosis and treatment

Tumorigenesis of vascular tumors is driven by different types of genetic alterations. Many of the vascular tumors have known balanced translocations, such as epithelioid hemangioma, epithelioid hemangioendothelioma and pseudomyogenic hemangioendothelioma. In these tumors specific chromosomal translocations have occurred leading to expression of altered genes. **Chapter 2** summarizes the current knowledge on genetic alterations, pathways, epidemiology and histopathology in vascular tumors of bone. With the increase in our understanding of the molecular changes in these entities, better classification and more accurate diagnostics has become possible. Diagnostics of vascular tumors relies on the histopathology, but also on immunohistochemical markers such as FOSB, TFE3 and CAMTA1 which can be used to detect tumor specific translocations (1–3). New techniques such as Next Generation Sequencing (NGS) could start to play an important role in the diagnostics of vascular tumors as specific translocations can be detected, even on paraffin embedded material (4, 5).

Chapter 3 focusses on epithelioid hemangioma. Using transcriptome sequencing of three index cases recurrent translocation involving the *FOS* gene were identified with different translocation partners. In all three cases the breakpoint occurred somewhere halfway the fourth exon and the predicted translocation would lead to an early stop codon and a truncation of the resulting protein. This FOS translocation was also simultaneously found and described by Huang and colleagues (6). Moreover, in multifocal tumor cases, exactly identical fusions were present in all foci. This supports the hypothesis that multiple foci originate from the same clone and can be considered multifocal regional spread. The FOS translocation was not detected in two out of seven epithelioid hemangioma cases by either PCR or FISH. A *ZFP36-FOSB* fusion was described in atypical epithelioid he-

mangiomas (7). Immunohistochemistry for FOSB has been described to be a diagnostic marker for pseudomyogenic hemangioendotheliomas which harbor a *SERPINE1-FOSB* or *ACTB-FOSB* translocation, it is likely the same immunohistochemistry could also be used to determine if the two *FOS* negative cases harbor a translocation involving *FOSB* (FOSB immunohistochemistry as a diagnostic marker was not described at the time of our study) (1). Another possibility is that there are still other genetic alterations or translocations in epithelioid hemangioma that have not been discovered, which could be discovered using Next Generation Sequencing techniques when cases are found with available fresh frozen material. We showed, in concordance with Huang and colleagues that the majority of the epithelioid hemangiomas harbor the *FOS* translocation which potentially makes it useful for diagnostics. The *FOS* translocation could be detected using break-apart FISH probes, or with a Next Generation Sequencing fusion detection panel.

The *FOS* translocations, identified in the three index cases using RNA-sequencing, involved different translocation partners that were either localized in an intron, an intergenic region or a long non-coding RNA. In all cases the fusion transcript results in a truncation of FOS through introduction of a stop codon. As fusion detection tools usually give many false positive results a common filtering strategy is to exclude fusions involving non-coding DNA. The only fusion detection tool that did not filter the fusions involving *FOS* was Defuse which does not filter for translocations involving non-coding regions (8). A downside of the Defuse approach is that it results in many false positives (in this study, in one case 65 fusions were identified). We were able to identify the correct fusion because we had previously identified the chromosomes that were involved in the translocation of our index case using Combined Binary Ratio fluorescence in situ hybridization on cultured tumor cells and one translocation breakpoint was FISH mapped to the *FOS* gene. This study therefore illustrates the benefit of running multiple fusion detection tools as each tool has its unique filtering and detection approach that may remove true positives. Liu and colleagues who compared fifteen fusion detection algorithms also concluded that multiple tools should be used to identify true gene fusions (9). In addition to using multiple fusion detection tools, chapter 3 illustrates the added value of using complementary molecular genetic techniques to identify an approximate location of the translocation. It is likely there are still many potentially important fusion genes not yet discovered in large fusion detection projects such as the TumorFusions database which relies only on a bioinformatics detection approach (10).

In **chapter 4** an unusual case of pseudomyogenic hemangioendothelioma that was inoperable but showed a remarkable response to telatinib is reported. The patient had extensive inoperable pseudomyogenic hemangioendothelioma in the head and neck which

did not respond to docetaxel. The patient was included in a multicenter phase I dose escalation study for telatinib, a small-molecule multi-tyrosine kinase inhibitor. In this study the potential mechanism of action of telatinib was elucidated using a model for pseudomyogenic hemangioendothelioma: we overexpressed truncated FOSB in endothelial cells (HUVECs). Truncated FOSB mimics the *SERPINE1-FOSB* fusion, as this fusion leads to the promoter and 5'-UTR of *SERPINE1* (resulting in sustained and high activation) attaching to the second exon of *FOSB* (leading to the loss of FOSB's first 48 amino acids). It was found that PDGFRA and FLT1 (VEGFR1) are upregulated by FOSB, both of which are known targets of telatinib. HUVECS expressing truncated FOSB that were treated with telatinib showed highly reduced growth when grown in a three-dimensional cell culture model. Telatinib was developed as a drug for advanced solid tumors where it inhibits angiogenesis by blocking VEGF signaling which is important for the tumor associated endothelial cells. Moreover, PDGFR is also blocked which thereby inhibits pericyte growth (11). This study illustrates the potential of repurposing small-molecule inhibitors, especially for rare tumors such as the vascular tumors, where developing a targeted therapy and dedicated clinical trials would likely not be feasible.

10.2 Model systems

In **chapter 5** the truncation of the FOS protein, that was identified in epithelioid hemangioma (chapter 3), was functionally explored. FOS and FOSB are members of the FOS family of proteins which either as homo- or hetero-dimer (with JUN family members) form the AP-1 transcription factor complex. The AP-1 transcription factor, through the regulation of many different genes, is known to be involved in tumorigenesis (12).It was found that the C-terminal alpha-helix that is found in the intrinsically disorganized tail of FOS leads to rapid degradation independent of ubiquitination. Loss of this alpha helix, as is the case in epithelioid hemangioma, leads to a significantly longer half-life of the FOS protein. It is therefore likely that the translocations found in epithelioid hemangioma leads to upregulation of the AP-1 transcription factor. It is tempting to speculate that inhibiting AP-1 activation in epithelioid hemangioma could be an effective targeted therapy. Of note, no mutations are reported in the *FOS* gene that would lead to an early stop-codon. This further hints at a role for the 3'-UTR in regulation of the *FOS* gene which is bypassed by the translocations we identified, which could be explored in future studies as only the resulting truncated FOS protein was studied in chapter 5.

To model the *FOSB* and *FOS* translocations in respectively chapter 4 and chapter 5 lentiviral transduction systems were used. Lentivirus overexpression systems are easy to work with and efficient at targeting difficult to transfect cells such as HUVECs (13).

Translocation modeling technique	**Pros**	**Cons**
Episomal transfection	-Easy to work with -Can be implemented rapidly	-Gene expression is lost over time -Not equally efficient in all cell lines -Generally, under control of artificial promoter
Lenti-viral transduction	-Easy to work with -Efficient in most cell lines -Integration into DNA	-Generally, under control of artificial promoter
CRISPR/Cas9 targeting	-Most accurate representation of actual translocations -Gene regulation as found in tumors	-Designing and targeting is time consuming -Off target effects -Cell of origin is often enigmatic for tumors

Table 10.1: Summary of the different techniques used to model translocation driven vascular tumors with their pros and cons.

With lentivirus gene delivery, the gene of interest is integrated into the genome of the target cell and therefore replication does not lead to loss of the integrated gene as would be the case when using a transfection or adenoviral gene system with episomal virus (as summarized in table 10.1).

The downside of using the lentivirus delivery system might be that expression can be high depending on the type of promoter selected, and often multiple copies of the gene of interest are inserted into the genome depending on the viral load used for transduction. Moreover, integration often occurs at sites with active gene transcription, potentially leading to disruption of important genes. High expression of the integrated gene can lead to oversaturation of the cellular processes and may lead to non-physiological protein-protein interaction and localization. In this study degradation was studied, and high expression could lead to oversaturation of the proteasome. However, a similar high expression of the FOS protein was observed in the patient tumor samples indicating that the proteasome was not overloaded. Another issue is that the inserted gene is generally not under the control of its own promoter. Therefore, the normal regulatory pathways no longer have an influence on the expression levels of the integrated gene. In chapter 5 the degradation of the protein was studied, however, the *FOS* gene is also regulated at mRNA level (14) and high levels of FOS protein could have a self-limiting effect if the *FOS* gene was under regulation of its own promotor, which would not have been detected using a lentivirus-based model.

Many of the shortcomings of using a lentivirus system were solved in **chapter 6** of this thesis. Using CRISPR/Cas9 two double stranded DNA breaks were introduced in chromosomes 7 and 19 of human induced Pluripotent Stem Cells (hiPSCs) at the locations where a gene fusion is found between *SERPINE1* and *FOSB* in pseudomyogenic hemangioendothelioma. By introducing a removable cassette containing a neomycin resistance gene between the fusion gene, we could select for a homogeneous population of hiPSCs with *SERPINE1-FOSB* fusion. The hiPSCs were thereafter differentiated towards endothelial cells. Starting with hiPSCs was necessary to generate this model, as HUVECs have a limited lifespan and therefore targeting with CRISPR/Cas9 would have been impossible as it requires culturing the cells for an extended period. In this model, *FOSB* expression is under control of the *SERPINE1* promoter, identical to the characteristic fusion found in pseudomyogenic hemangioendotheliomas. Using the generated cell model, it became possible to study the pathways that might play an important role in the tumorigenesis of pseudomyogenic hemangioendothelioma. In future studies this model opens up the possibility to study the effect of the fusion genes on epigenetic alterations among other possible lines of experiments. The different methods to model gene fusions are compared in table 10.1.

Another challenge in creating an accurate model to study pseudomyogenic hemangioendothelioma is that the cell of origin remains enigmatic. In chapters 5 and 6 of this thesis it was assumed that the cell of origin is somewhere in the endothelial differentiation pathway, therefore Human Umbilical Vein Endothelial cells, and hiPSCs differentiated to endothelial cells were used to model pseudomyogenic hemangioendothelioma. Our assumption that pseudomyogenic hemangioendothelioma occur somewhere in the differentiation towards endothelial cells is supported by the detected expression of vascular markers, such as CD31 and ERG. However, pseudomyogenic hemangioendothelioma also shows positivity for epithelial markers such as keratin which could point to another differentiation lineage. In this thesis evidence is provided supporting an endothelial precursor, in chapter 6 it is shown that the *SERPINE1-FOSB* fusion leads to upregulation of FOSB when the hiPSCs are differentiated to endothelial cells, and not in the precursor hiPSCs indicating that activation of the *SERPINE1* promoter and consequent expression of the fusion gene is upregulated in the endothelial differentiation lineage.

10.3 Computational biology

Next-generation sequencing is becoming faster and more affordable. Moreover, there is a trend to release sequencing data open-accessible, resulting in large public datasets. These large datasets increase the power of computational biology analysis which could lead

to novel insights into tumorigenesis and targeted therapies. Gene regulation networks could help gain insight into the pathways that are driving tumorigenesis. **Chapter 7** describes a new implementation of the PANDA algorithm in the python programming language. This implementation is much faster than the C++ and R implementation by using fast matrix operation. This increase in speed enables analysis of larger data-sets under different conditions in reasonable time. The network is reconstructed by calculating Pearson correlations between the genes. Variations in gene expression between different genes lead to generation of correlations between different genes. Unique about the PANDA algorithm is that it takes known transcription factor-gene integrations and known protein-protein interactions into account when calculating the regulatory network. A downside of this approach is that it requires large datasets, with many replicates as it relies on small variations in gene expression to reconstruct the regulatory network.

In **chapter 8** gene regulatory network reconstruction was applied to transcriptome sequencing data from Human Umbilical Vein Endothelial Cells expression truncated *FOS*, *FOSB* and wild-type controls. We confirmed that in the wild-type cells *FOS* and *FOSB* were upregulated after serum stimulation indicating that the AP-1 transcription factor complex is involved in the early serum response of normal endothelial cells. As epithelioid hemangioma and pseudomyogenic hemangioendothelioma likely have constitutively high respective *FOS* or *FOSB* expression; their gene expression pattern can be compared to the stimulated HUVECs. Interestingly, in the gene regulatory network we found regulation of *YAP1*, which is an important partner of the HIPPO signaling pathway. This could explain the similarities of epithelioid hemangioma with epithelioid hemangioendothelioma, which harbors fusions involving either *WWRT1-CAMTA* or *YAP1-TFE3*, both involving the HIPPO signaling pathway (15).

Another approach to analyze gene expression data comes from the machine learning field. Machine learning is starting to show large potential in many areas, most notable image and speech recognition. But the same algorithms, both supervised and unsupervised can also be used to analyze gene expression data. Machine learning was used in **chapter 9** on The Cancer Genome Atlas (TCGA) soft-tissue sarcoma data including 206 cases (no large public datasets for vascular tumors are available). First, we identified potential diagnostic genes using random forests to distinguish between soft-tissue sarcoma subtypes that are morphologically similar. Diagnostically relevant genes that distinguish between malignant peripheral nerve sheath tumor and synovial sarcoma were identified and verified on an independent set. Secondly, prognostic genes that can be used in a k-nearest neighbor analysis were identified that outperformed other prognosticators. HMMR immunohistochemistry was shown to be a prognosticator in leiomyosarcoma, as

high expression correlated with poor survival (in line with the *HMMR* expression found in the TCGA data). Thirdly, differentiation of soft-tissue sarcomas was studied using a neural network trained on normal tissue. Last, novel candidate therapies for soft-tissue sarcomas were identified using a network analysis approach. The use of HDAC inhibitors were identified as potential therapy for Leiomyosarcoma; a finding that was validated in cell lines. Chapter 9 illustrates the possibilities for machine learning on gene expression data as a tool to discover new biomarkers and therapies. The use of machine learning approaches, especially deep neural networks, will become even more powerful when larger sequencing datasets become publicly available (16).

10.4 Future perspectives

In this thesis we present evidence that AP-1 transcription factor activation drives the tumorigenesis of epithelioid hemangioma and pseudomyogenic hemangioendothelioma which raises a number of questions. Both tumors harbor translocations involving *FOS* and *FOSB*. Interestingly, the translocations involving the FOS gene lead to a truncation of the FOS protein resulting in a longer lifespan of the protein, as described in this thesis. Potentially, a point mutation or deletion could lead to a similar truncation of the FOS protein. Such an alteration is, to our knowledge, not yet reported in the literature or in mutation databases, which could indicate a role for the loss of the 3'-UTR of *FOS* that is difficult to understand at this point. One experiment to study this observation, using cell line modeling techniques developed in this thesis, could be to compare endothelial cells with a FOS fusion with endothelial cells harboring a mutation causing an early stop codon in the *FOS* gene. The life-span and gene expression of both cell lines could be compared to find what the function of the enigmatic translocation partner is.

As is shown in this thesis, upregulation of the AP-1 transcription factor is a potent driver of tumorigenesis, which raises the question that AP-1 transcription factor activation could also play an important role in other tumor types. Recently, fusions involving *FOS* and *FOSB* were found in osteoblastoma and osteoid osteoma which shows AP-1 transcription factor activation is likely also a potent driver of tumorigenesis in tumors which occur in other differentiation lineages (17). Although large sequencing projects such as the Cancer Genome Atlas did not reveal alterations in *FOS* or *FOSB* in other tumor types, likely excluding direct involvement of these genes, AP-1 transcription factor activation could still play an important role through indirect genetic regulation. This could be studied using gene regulation network reconstruction tools such as PANDA which could be used to find upregulation of the AP-1 transcription factor in other tumor types.

Currently drugs directly targeting the AP-1 transcription factor are lacking. T-5224

is a drug that claims to inhibit FOS/AP-1 activation (18), however, in our hands the effectivity was rather limited. In the future, new small-molecule inhibitors need to be developed to directly inhibit the AP-1 transcription factor as a targeted therapy for tumors driven by AP-1 activation. Machine learning is showing large potential for the development of small molecule inhibitors through efficient modeling of proteins and prediction of efficient small-molecule inhibitors. These technologies could lead to development of targeted therapies for AP-1 transcription factor activation driven tumors (19).

Bibliography

[1] Hung YP, Fletcher CDM, Hornick JL. FOSB is a useful diagnostic marker for pseudomyogenic hemangioendothelioma. American Journal of Surgical Pathology. 2017;41(5):596–606. doi:10.1097/PAS.0000000000000795.

[2] Shibuya R, Matsuyama A, Shiba E, Harada H, Yabuki K, Hisaoka M. CAMTA1 is a useful immunohistochemical marker for diagnosing epithelioid haemangioendothelioma. Histopathology. 2015;67(6):827–835. doi:10.1111/his.12713.

[3] Antonescu CR, Le Loarer F, Mosquera JM, Sboner A, Zhang L, Chen CL, et al. Novel YAP1-TFE3 fusion defines a distinct subset of epithelioid hemangioendothelioma. Genes Chromosomes and Cancer. 2013;52(8):775–784. doi:10.1002/gcc.22073.

[4] Lam SW, van IJzendoorn DGP, Cleton-Jansen AM, Szuhai K, Bovée JVMG. Molecular Pathology of Bone Tumors. The Journal of Molecular Diagnostics. 2019;21(2):171–182. doi:10.1016/J.JMOLDX.2018.11.002.

[5] Lam SW, Cleton-Jansen AM, Cleven AHG, Ruano D, van Wezel T, Szuhai K, et al. Molecular Analysis of Gene Fusions in Bone and Soft Tissue Tumors by Anchored Multiplex PCR–Based Targeted Next-Generation Sequencing. The Journal of Molecular Diagnostics. 2018;20(5):653–663. doi:10.1016/j.jmoldx.2018.05.007.

[6] Huang SC, Zhang L, Sung YS, Chen CL, Krausz T, Dickson BC, et al. Frequent FOS gene rearrangements in epithelioid hemangioma: A molecular study of 58 cases with morphologic reappraisal. American Journal of Surgical Pathology. 2015;39(10):1313–1321. doi:10.1097/PAS.0000000000000469.

[7] Antonescu CR, Chen HW, Zhang L, Sung YS, Panicek D, Agaram NP, et al. ZFP36-FOSB fusion defines a subset of epithelioid hemangioma with atypical features. Genes Chromosomes and Cancer. 2014;53(11):951–959. doi:10.1002/gcc.22206.

[8] McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. Defuse: An algorithm for gene fusion discovery in tumor rna-seq data. PLoS Computational Biology. 2011;7(5):e1001138. doi:10.1371/journal.pcbi.1001138.

[9] Liu S, Tsai WH, Ding Y, Chen R, Fang Z, Huo Z, et al. Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. Nucleic Acids Res. 2016;44(5):e47. doi:10.1093/nar/gkv1234.

[10] Hu X, Wang Q, Tang M, Barthel F, Amin S, Yoshihara K, et al. TumorFusions: an integrative resource for cancer-associated transcript fusions. Nucleic acids research. 2018;46(D1):D1144–D1149. doi:10.1093/nar/gkx1018.

[11] Eskens FA, Steeghs N, Verweij J, Bloem JL, Christensen O, van Doorn L, et al. Phase I dose escalation study of telatinib, a tyrosine kinase inhibitor of vascular endothelial growth factor receptor 2 and 3, platelet-derived growth factor receptor beta, and c-Kit, in patients with advanced or metastatic solid tumors. J Clin Oncol. 2009;27(25):4169–4176. doi:10.1200/JCO.2008.18.8193.

[12] Eferl R, Wagner EF. AP-1: A double-edged sword in tumorigenesis. Nature Reviews Cancer. 2003;3(11):859–868. doi:10.1038/nrc1209.

[13] Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, et al. A third-generation lentivirus vector with a conditional packaging system. Journal of virology. 1998;72(11):8463–71. doi:98440501.

[14] Rahmsdorf HJ, Schönthal A, peter Angel, Litfin M, Rüther U, Herrlich P. Posttranscriptional regulation of c-fos mRNA expression. Nucleic Acids Research. 1987;15(4):1643–1659. doi:10.1093/nar/15.4.1643.

[15] Tanas MR, Ma S, Jadaan FO, Ng CK, Weigelt B, Reis-Filho JS, et al. Mechanism of action of a WWTR1(TAZ)-CAMTA1 fusion oncoprotein. Oncogene. 2016;35(7):929–938. doi:10.1038/onc.2015.148.

[16] LeCun YA, Bengio Y, Hinton GE. Deep learning. Nature. 2015;521(7553):436–444. doi:10.1038/nature14539.

[17] Fittall MW, Mifsud W, Pillay N, Ye H, Strobl AC, Verfaillie A, et al. Recurrent rearrangements of FOS and FOSB define osteoblastoma. Nature Communications. 2018;9(1):2150. doi:10.1038/s41467-018-04530-z.

[18] Chen H, Qu J, Huang X, Kurundkar A, Zhu L, Yang N, et al. Mechanosensing by the α 6-integrin confers an invasive fibroblast phenotype and mediates lung fibrosis. Nature Communications. 2016;7:12564. doi:10.1038/ncomms12564.

[19] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discovery Today. 2018;23(6):1241–1250. doi:10.1016/J.DRUDIS.2018.01.039.

Chapter 11

Nederlandse samenvatting

Vaattumoren vormen een groep tumoren met als overeenkomst dat ze allemaal endotheliale differentiatie laten zien. Zowel intermediaire als kwaadaardige vaattumoren komen in de weke-delen of in de botten voor. Een grote uitdaging bij het bestuderen van deze tumoren is dat tot op heden goede modellen ontbraken om deze tumoren in het laboratorium te kunnen bestuderen. Dit proefschrift beschrijft een aantal geschikte tumormodellen. Van technieken uit de bioinformatica is gebruikgemaakt om nieuwe inzichten te ontwikkelen in het ontstaan van vaattumoren en andere wekedelentumoren en de behandeling daarvan. De methoden die in dit proefschrift voor het modelleren van vaattumoren zijn beschreven kunnen in de toekomst wellicht ook gebruikt worden voor andere tumoren waarvoor nog geen goede modellen beschikbaar zijn. Dit proefschrift bestaat uit drie onderdelen: diagnostiek en behandeling, tumormodellen, en bioinformatica. **Hoofdstuk 1** introduceert deze drie onderdelen.

11.1 Diagnostiek en behandeling

Verschillende genetische veranderingen kunnen ten grondslag liggen aan het ontstaan van vaattumoren. De meeste vaattumoren zijn gekenmerkt door gebalanceerde translocaties die kunnen ontstaan wanneer er twee breuken in het DNA optreden en het reparatiemechanisme van de cel de verkeerde uiteinden aan elkaar bevestigt. De samensmelting van de genen die gelokaliseerd zijn op de breukpunten kunnen dan leiden tot tumorvorming. Voorbeelden van vaattumoren die ontstaan door translocaties zijn epithelioid hemangioom, epithelioid hemangioendothelioom en pseudomyogeen hemangioendothelioom. **Hoofdstuk 2** is een samenvatting van wat bekend is over de genetische veranderingen, signaaltransductieroutes, epidemiologie en histopathologie van vaattumoren van het bot. Doordat er toenemende kennis is over de moleculaire veranderingen die ten grondslag liggen aan deze vaattumoren, wordt de diagnostiek steeds nauwkeuriger. De diagnose wordt gesteld door gebruik te maken van conventionele histologie aangevuld met immunohistochemie waarmee zelfs moleculaire veranderingen gedetecteerd kunnen worden. In de klinische praktijk wordt bijvoorbeeld gebruik gemaakt van immunohistologische markers zoals FOSB, TFE3 en CAMTA1; de overexpressie van deze eiwitten is een indicatie voor de aanwezigheid van een specifieke translocatie. In de nabije toekomst zullen waarschijnlijk nieuwere technieken zoals 'next-generation sequencing' een belangrijke rol gaan spelen in de diagnostiek van vaattumoren.

Hoofdstuk 3 beschrijft een nieuwe translocatie die wij vonden in epithelioid hemangioom door met 'next-generation sequencing' het genetisch materiaal van drie epithelioid hemangiomen te analyseren. 'Next-generation sequencing' is een techniek waarmee de sequentie van de nucleotiden van meerdere DNA-moleculen tegelijk kan worden bepaald,

waardoor uiteindelijk de gehele DNA-sequentie van een tumor achterhaald wordt. De sequentie van het RNA kan worden bepaald door RNA eerst om te zetten in cDNA. Op deze manier werd in drie tumoren een translocatie van het *FOS*-gen gevonden; in alle gevallen lag het breekpunt halverwege het vierde exon van *FOS* en dat resulteerde in een vervroegd stopcodon in het *FOS*-gen. Het epitheloid hemangioom heeft als kenmerk dat het zich in het bot vaak multifocaal presenteert. In de verschillende tumorhaarden van patiënten werden identieke translocaties gevonden waardoor duidelijk werd dat de verschillende localisaties aan elkaar gerelateerd waren en dat de tumor dus locoregionaal verspreid was en niet per toeval op meerdere plekken tegelijk was ontstaan bij dezelfde patiënt.

In **hoofdstuk 4** wordt een casus beschreven van een patiënt met inoperabel pseudomyogeen hemangioendothelioom. Bij deze patiënt had de tumor zich verspreid in het hoofdhalsgebied en de tumor reageerde niet op behandeling met docetaxel. In het kader van een fase I klinische trial werd de patiënt met telatenib, een tyrosine kinase remmer, behandeld. Dit resulteerde in een volledige remissie van de tumor. Het werkingsmechanisme van telatenib is onderzocht met behulp van een cellijnmodel voor pseudomyogeen hemangioendothelioom. Voor dit cellijnmodel werd een verkort FOSB-eiwit tot overexpressie gebracht in endotheelcellen, de meest waarschijnlijke cel van oorsprong van deze tumor. Het verkorte FOSB-eiwit werd gebruikt om de *SERPINE1-FOSB*-fusie na te bootsen die gevonden wordt in pseudomyogeen hemangioendothelioom. De hypothese was dat het eerste deel van deze fusie, *SERPINE1*, zorgt voor de hoge expressie van het eiwit maar geen eiwitverandering veroorzaakt, en dat FOSB door de fusie de eerste 48 animo zuren verliest. Met dit cellijnmodel konden we laten zien dat PDGFRA en FLT1 (VEGFR1), bekende doelwitten van telatinib, meer tot expressie kwamen door het verkorte FOSB-eiwit. Dit resulteerde in een verminderde groei bij behandeling met telatenib van het cellijnmodel. Deze studie laat het nut zien van het heroriënteren van geneesmiddelen, zeker voor zeldzame tumoren zoals de vaattumoren waarvoor het ontwikkelen van nieuwe behandelingen niet winstgevend zou zijn.

11.2 Tumormodellen

In **hoofdstuk 5** is de fusie bestudeerd zoals geïdentificeerd in het epitheloid hemangioom, waarvan het effect een verkorting van het FOS-eiwit is (zie hoofdstuk 3). Het FOS-eiwit bleek een alpha-helix in het C-terminale uiteinde te bezitten die voor snelle degradatie van het FOS-eiwit zorgt, zonder daarbij gebruik te maken van ubiquitine (de meest voorkomende manier waarop eiwitten degraderen in de cel). Eiwitdegradatie binnen de cel is belangrijk om signaaltransductie te reguleren. Het verlies van deze alpha-helix

zoals bij epithelioid hemangiomen zorgt dus voor verminderde afbraak van het FOS-eiwit. Dit leidt vervolgens tot hoge activiteit van de AP-1 transcriptie factor, hetgeen dus een belangrijke rol speelt bij het ontstaan van een epithelioid hemangioom.

In **hoofdstuk 6** werd een model voor pseudomyogeen hemangioendothelioom ontwikkeld met behulp van CRISPR/Cas9, een systeem om genetische veranderingen aan te brengen in de cel. Dit systeem bestaat uit twee componenten; het Cas9-eiwit dat door een 'guide RNA' naar een locatie op het DNA wordt gebracht waar het een breuk in het DNA veroorzaakt. Twee breuken werden in het DNA van pluripotente stamcellen aangebracht; op chromosomen 7 en 19. Deze breuken werden aangebracht op de exacte locaties waar bij patiënten met pseudomyogeen hemangioendothelioom ook de breukpunten van het fusiegen gevonden worden. Op de breukpunten zijn het *SERPINE1* en *FOSB*-gen gelokaliseerd die het fusiegen vormen. Na het creëren van het fusiegen werden de pluripotente stamcellen naar endotheelcellen gedifferentieerd. Met deze endotheelcellen kon vervolgens onderzocht worden wat het effect van het fusiegen is op endotheelcellen vergeleken met controle-endotheelcellen zonder fusiegen. We vonden dat de verbindingen tussen de cellen minder sterk werden en dat de endotheelcellen met translocatie in muizen meer invasie van deze cellen in omliggend weefsel lieten zien vergeleken met de controle-endotheelcellen. Daarnaast werd met een bioinformatica analyse aangetoond dat de genexpressie van de endotheelcellen met translocatie eigenschappen van pseudomyogeen hemangioendothelioom liet zien.

11.3 Bioinformatica

Steeds meer genetische data gegenereerd door middel van 'next-generation sequencing' worden vrij gedeeld in grote publiekelijk toegankelijke databestanden. Een manier om deze grote datasets te analyseren is door middel van gen regulatie netwerken. **Hoofdstuk 7** beschrijft een nieuwe implementatie van het PANDA-algoritme in de Python programmeertaal. Deze implementatie is sneller dan de C++ en R-implementaties door gebruik te maken van snelle matrix berekeningen. Door de hogere snelheid is het mogelijk grotere datasets met meerdere verschillende samenstellingen in korte tijd te analyseren. Door het identificeren van verschillen in genexpressie tussen de verschillende genen kunnen Pearson correlaties tussen genen, geïdentificeerd met het PANDA-algoritme, berekend worden. Om tot een beter resultaat te komen maakt het PANDA-algoritme ook gebruik van bekende transcriptiefactor - gen interacties en bekende eiwit - eiwit interacties bij het berekenen van het genregulatienetwerk.

In **hoofdstuk 8** werden genexpressiedata van endotheelcellen die verkort *FOS*, *FOSB*, en een leeg controle-plasmide tot expressie brachten, geanalyseerd door middel van netwerk-

analyse. In de controle-cellen vonden we dat FOS en FOSB hoog tot expressie kwamen na serum stimulatie, wat aangaf dat het AP-1 transcriptiefactorcomplex een belangrijke rol speelt in eerste reactie op deze stimulatie. Epithelioid hemangioom en pseudomyogeen hemangioendothelioom hebben altijd hoge expressie van *FOS* en *FOSB*, vergelijkbaar met de serum gestimuleerde endotheelcellen. De netwerkanalyse liet zien dat *YAP1* - een onderdeel van de HIPPO-signaaltransductieroute - gereguleerd werd door AP-1 activatie van de transcriptiefactor. Epithelioid hemangioendothelioom wordt gekarakteriseerd door de fusiegenen *WWTR1-CAMTA* of *YAP1-TFE3*, die beiden verbonden zijn met de HIPPO-signaaltransductieroute. De gevonden verbinding tussen AP-1 activatie en de HIPPO-signaaltransductieroute zou een verklaring kunnen zijn voor de overeenkomsten tussen epithelioid hemangioom en epithelioid hemangioendothelioom.

Een andere mogelijk manier om genexpressie te analyseren komt voort uit 'machine learning'. 'Machine learning' speelt een steeds grotere rol in diverse disciplines en op uiteenlopende terreinen, waarbij de bekendste voorbeelden het herkennen van spraak en beelden zijn. Dezelfde 'machine learning' algoritmes kunnen ook gebruikt worden om genexpressie data te analyseren. In **hoofdstuk 9** werd 'machine learning' toegepast op de genexpressie dataset van 'The Cancer Genome Atlas' (TCGA) met 206 wekedelensarcomen. Eerst identificeerden we genen die gebruikt kunnen worden om de verschillende wekedelen sarcomen te diagnosticeren welke microscopisch erg op elkaar kunnen lijken. Voorbeelden van morfologisch zeer lastig te onderscheiden tumoren zijn MPNST en synoviosarcoom waarbij de expressie van *NEURL1* en *NPAS1* gebruikt kunnen worden om onderscheid te maken. Daarnaast hebben we genen opgespoord waarvan de expressie een houvast kan bieden bij het voorspellen van de prognose. Zo konden we bijvoorbeeld bevestigen dat een immunohistochemische kleuring voor HMMR een indicatie kan geven van de prognose voor leiomyosarcoom patiënten (waarbij hoge expressie een slechte prognose betekent). Ook zochten we naar nieuwe therapieën voor wekedelensarcomen door middel van een netwerkanalyse. Met deze analyse konden we laten zien dat HDAC blokkerende medicatie een mogelijk therapie voor het leiomyosarcoom kan zijn, een bevinding die in cellijnen kon worden gevalideerd.

In dit proefschrift zijn de volgende bevindingen uniek en dragen bij aan beter begrip van vaattumoren en andere wekedelensarcomen. Voor epithelioide hemangiomen is een nieuwe translocatie beschreven die verklaart waardoor het FOS-eiwit een langere levensduur krijgt. Voor een patiënt met pseudomyogeen hemangioendothelioom die na een behandeling met telatinib volledig genezen bleek te zijn, is het mogelijke werkingsmechanisme van telatinib in pseudomyogene hemangioendotheliomen opgehelderd. Voor pseudomyogeen hemangioendothelioom hebben we een nieuw cellijn model ontwikkeld. Met CRISPR/Cas9 kon de karakteristieke translocatie in het DNA van endotheelcellen aange-

bracht worden waarmee meerdere aspecten van pseudomyogeen hemangioendothelioom gesimuleerd kunnen worden. In dit proefschrift wordt veelvuldig gebruik gemaakt van bioinformatica om de vaattumoren maar ook andere weke-delen sarcomen beter te begrijpen en te zoeken naar nieuwe behandelingen. Daarbij is gebruik gemaakt van genregulatie netwerken en machine learning, bioinformatica methoden die dit veld van biomedisch onderzoek drastisch zullen vernieuwen.

Part V

Appendices

Curriculum vitae

David Gerardus Pieter van IJzendoorn was born on December 3, 1990, in Leiden, the Netherlands. He attended bilingual pre-university education (natural sciences and technology track) at the Rijnlands Lyceum in Oegstgeest where he graduated in 2009. In 2010 he started with his medical studies at Leiden University. During the bachelor phase of his medical studies he followed the honors program and successfully finished the pre-master Biomedical sciences. A bachelor research internship at the Sir William Dunn School of Pathology at the University of Oxford supervised by Prof. dr. A.B. Hassan focused on the effect of mifamurtide on macrophages for the treatment of osteosarcoma. In a research internship during the master he participated in research into the effect of Wnt-inhibitors in osteosarcoma supervised by dr. A.M. Cleton-Jansen at the Department of Pathology of the Leiden University Medical Center. Thereafter he decided to put his masters study on hold to conduct his PhD research at the Department of Pathology of Leiden University Medical Center with prof. dr. J.V.M.G. Bovée and dr. K. Szuhai as his advisers. During his PhD research he received a travel grant from the Leiden University Fund that enabled him to spent three months at the Department of Biostatistics and Computational Biology of the Dana-Farber Cancer Institute, Harvard Medical School in Boston (MA), USA, supervised by dr. M.L. Kuijjer. The results of his PhD research are presented in this thesis. He is currently attending the clinical rotations to obtain his MD.

List of publications

- **van IJzendoorn DGP**, Salvatori DCF§, Cao X§, van den Hil F, Briaire-de Bruijn IH, de Jong D, Mei H, Mummery CL, Szuhai K, Bovée JVMG†, Orlova VV†. Pseudomyogenic hemangioendothelioma recapitulated in endothelial cells from human induced pluripotent stem cells engineered to express the *SERPINE1-FOSB* translocation. *Submitted.* §,†Shared authorship

- Lam SW, **van IJzendoorn DGP**, Cleton-Jansen AM, Szuhai K, Bovée JVMG. Molecular Pathology of Bone Tumors. *J Mol Diagnostics.* 2019;21: 171-182.

- **van IJzendoorn DGP**, Szuhai K, Briaire-de Bruijn IH, Kostine M, Kuijjer ML, Bovée JVMG. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLOS Comput Biol.* 2019;15: e1006826.

- **van IJzendoorn DGP**, Sleijfer S, Gelderblom H, Eskens FALM, van Leenders GJLH, Szuhai K, Bovée JVMG. Telatinib Is an Effective Targeted Therapy for Pseudomyogenic Hemangioendothelioma. *Clin Cancer Res.* 2018;24: 2678-2687.

- **van IJzendoorn DGP**, Bovée JVMG. Vascular Tumors of Bone: The Evolvement of a Classification Based on Molecular Developments. *Surg Pathol Clin.* 2017;10: 621-635.

- **van IJzendoorn DGP**†, Forghany Z†, Liebelt F, Vertegaal AC, Jochemsen AG, Bovée JVMG, Szuhai K, Baker DA. Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis. *J Biol Chem.* 2017;292: 21282-21290. †Shared authorship

- **van IJzendoorn DGP**, Glass K, Quackenbush J, Kuijjer ML. PyPanda: a Python package for gene regulatory network reconstruction. *Bioinformatics.* 2016;32: 3363-3365.

- **van IJzendoorn DGP**, de Jong D, Romagosa C, Picci P, Benassi MS, Gambarotti M, Daugaard S, van de Sande M, Szuhai K, Bovée JVMG. Fusion events lead to

truncation of FOS in epithelioid hemangioma of bone. *Genes Chromosom Cancer*. 2015;54: 565-574.

Acknowledgments

To my promotor prof. dr. J.V.M.G. Bovée and my co-promotor dr. K. Szuhai I owe a debt of sincere gratitude. Judith, I am grateful for the opportunities you gave me and your trust in allowing me to develop both my molecular biology and bioinformatics interests and helping me to explore my talents and passions. I am looking forward to working with you in the future on new exciting research challenges. Karoly, your endless stream of ideas for new experiments and our many interesting discussions were most inspiring.

I am grateful to my colleagues at the Department of Pathology and the Department of Cell and Chemical Biology (former Department of Molecular Cell Biology) for their support in the lab as well as their timely distractions outside of the lab. Danielle, thank you for always making time to discuss my experiments, for your indispensable help in the lab but most of all for your friendship. Inge, I thank you for your support with the technicalities of my experiments and our discussions. Renier, you made our room lively and my time as a PhD student so much more enjoyable.

I am grateful to dr. M.L. Kuijjer for the warm welcome in Boston and for her constructive feedback on my computational biology work. At the start of a career in science and at its critical junctures you sometimes need some wise advice, which prof. dr. H.J. Tanke and prof. dr. P.C.W. Hogendoorn were always willing to provide.

My loving parents, Jeanet and Rien, your advices and encouragements during our weekly dinners, many holidays and ice-skating excursions have been invaluable. Finally, Anne-Marlijn, thank you for your support and our many adventures.