Unravelling vascular tumors : combining molecular and computational biology

IJzendoorn, D.G.P. van

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

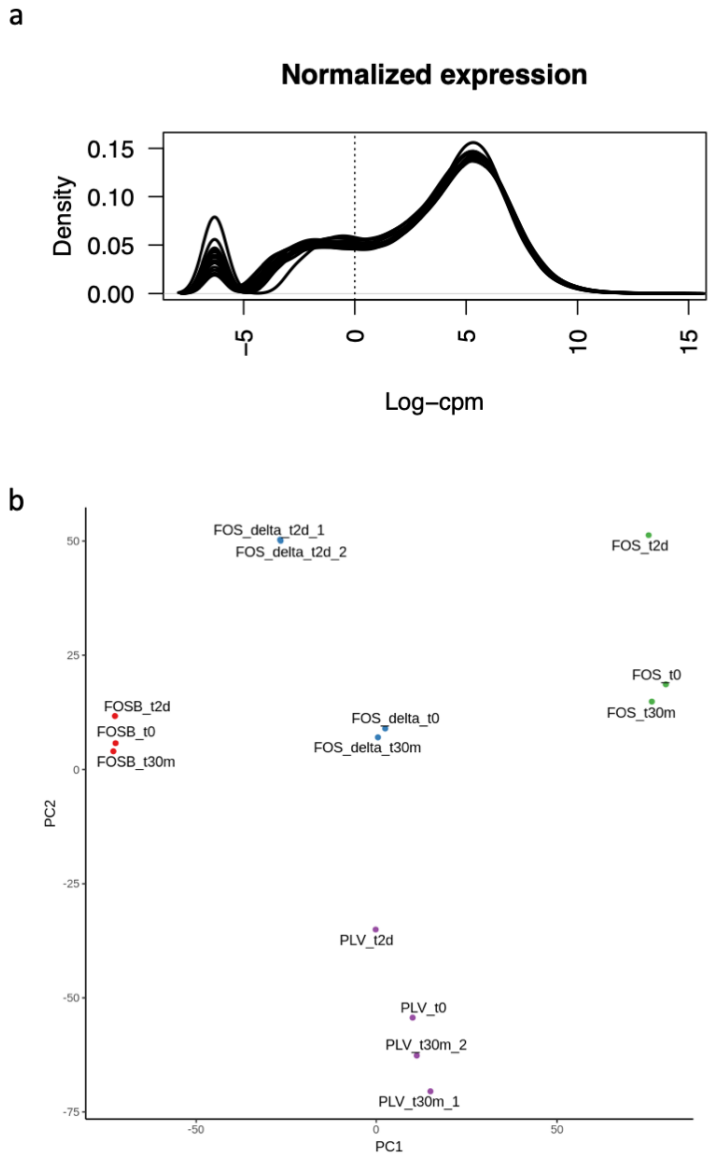

Figure 8.1: Cluster analysis of HUVEC gene expression data. (a) Transcriptome sequencing data was collected for HUVECs expressing *FOS*, *FOS*Δ, *FOSB* and an empty control plasmid. Expression data was harvested at different timepoint (0, 0.5 and 48 hours) after serum stimulation. All expression data 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).

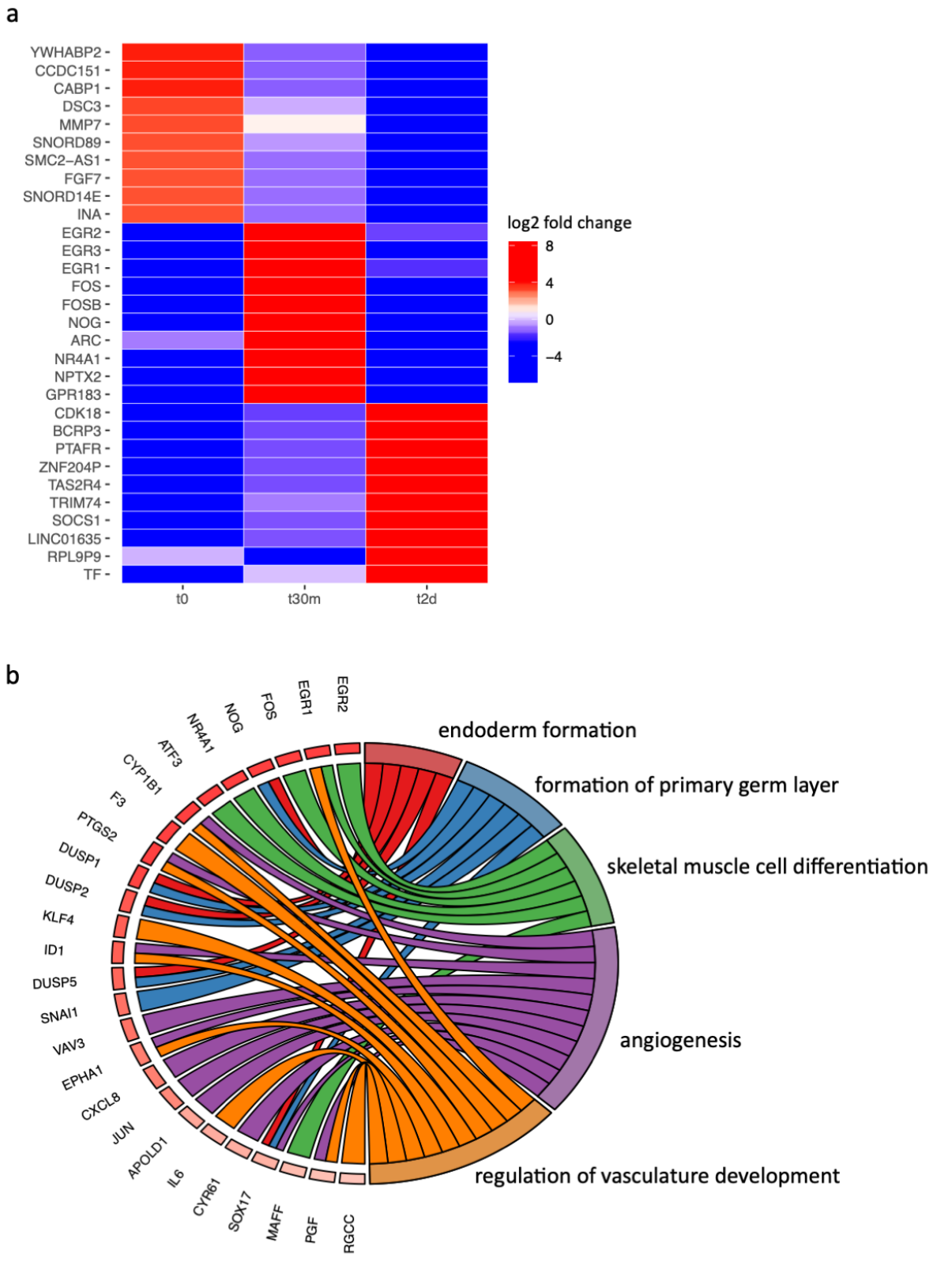

Figure 8.2: AP-1 transcription factors play an important role in the early response. (a) The heatmap shows the top ten signature genes for the control HUVECs harvested at different timepoint. The color indicates the log2 fold change compared to the other PLV samples. (b) The signature genes from the 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).

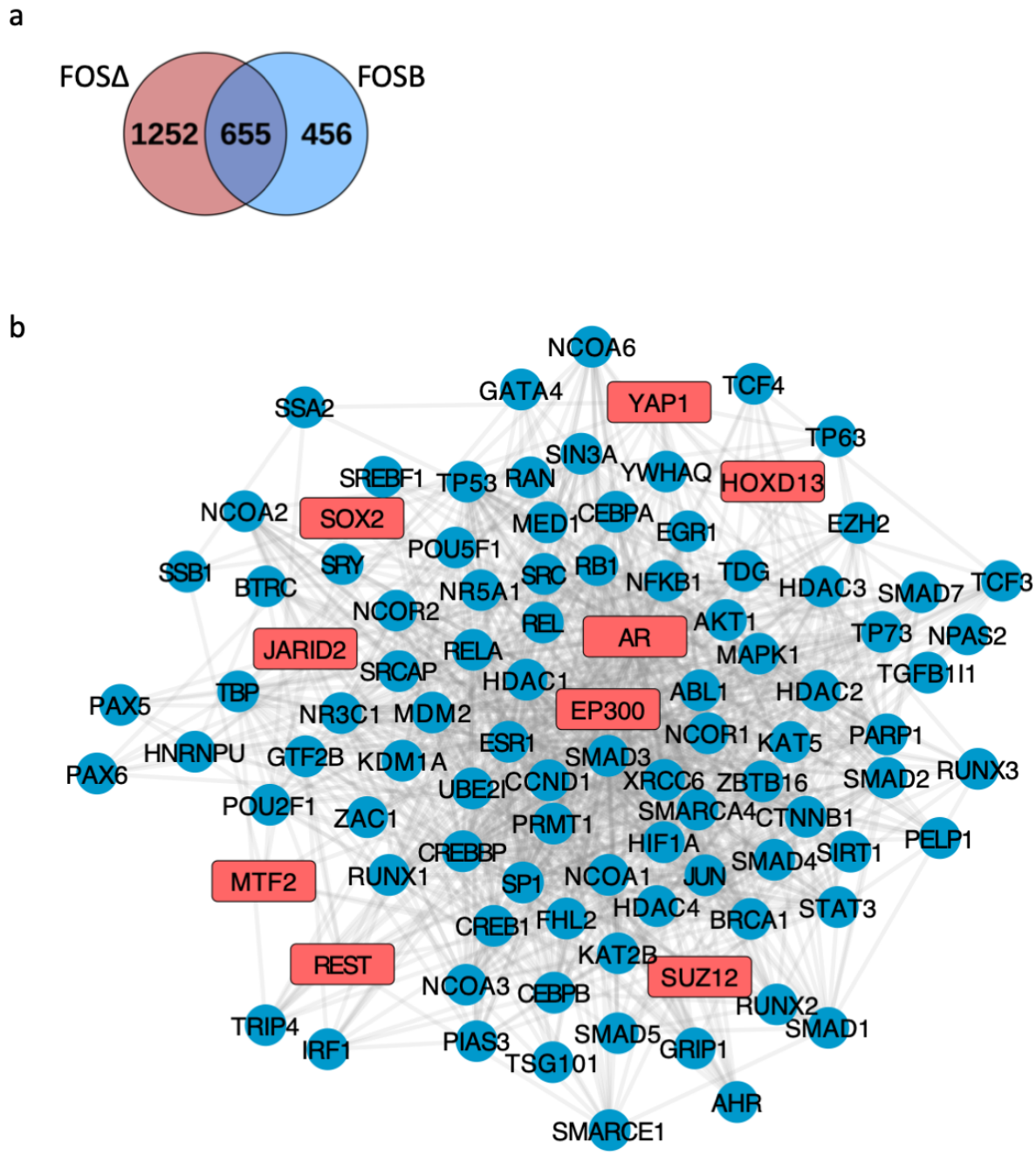

Figure 8.3: *FOSΔ* and *FOSB* regulatory network analysis. (a) The signature genes for *FOSΔ* and *FOSB* are shown in a Venn diagram. *FOSΔ* has 1907 signature genes and *FOSB* 1111. 655 signature genes overlap between *FOSΔ* and *FOSB*. (b) Using the Expression2Kinase analysis tool the gene regulatory network was reconstructed for the overlapping signature genes from *FOSΔ* 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.

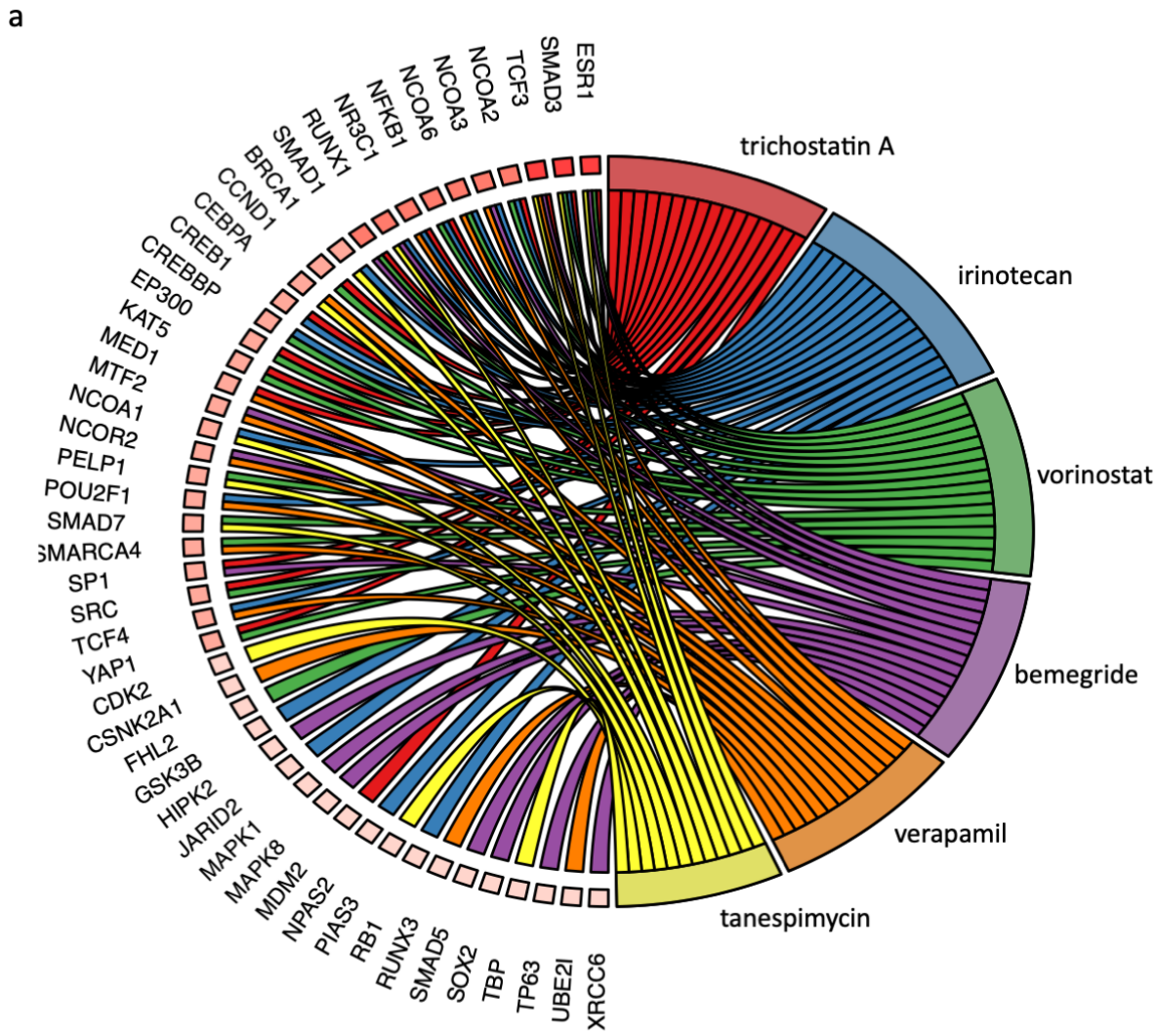

Figure 8.4: Targeted therapies identified from the *FOS*Δ and *FOSB* regulatory network. Potential targeted therapies are identified based on the regulatory network for *FOS*Δ and *FOSB*. The Chord diagram shows the potential targeted therapies on the right with the associated proteins on the left. The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