# Unravelling vascular tumors : combining molecular and computational biology

IJzendoorn, D.G.P. van

# 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<br>-Can be implemented rapidly | -Gene expression is lost over time<br>-Not equally efficient in all cell lines<br>-Generally, under control of artificial promoter |
| *Lenti-viral transduction* | -Easy to work with<br>-Efficient in most cell lines<br>-Integration into DNA | -Generally, under control of artificial promoter |
| *CRISPR/Cas9 targeting* | -Most accurate representation of actual translocations<br>-Gene regulation as found in tumors | -Designing and targeting is time consuming<br>-Off target effects<br>-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 $\alpha$ 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.