Unravelling vascular tumors : combining molecular and computational biology

IJzendoorn, D.G.P. van

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.

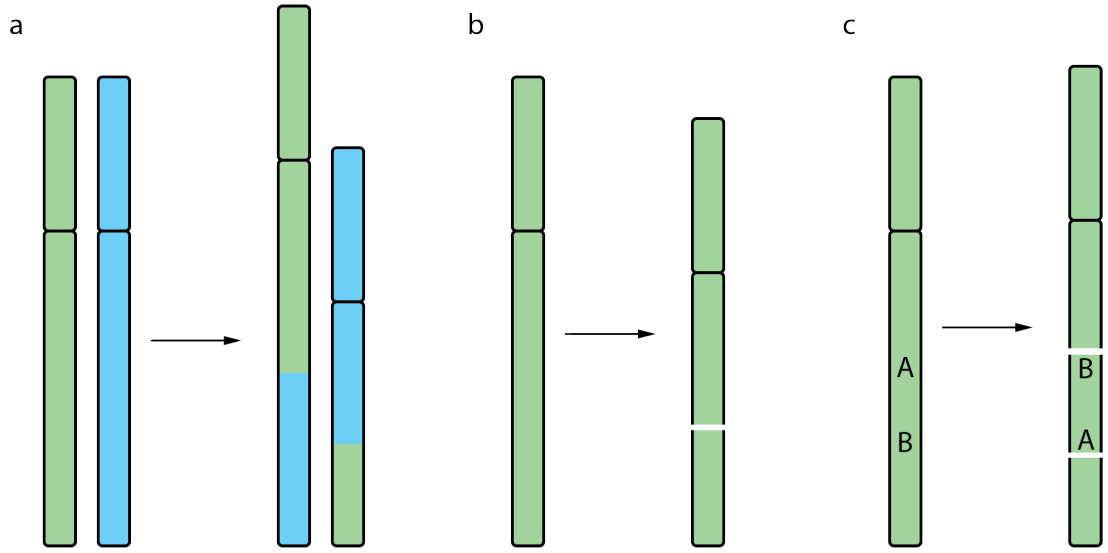

Figure 1.1: Different mechanisms leading to gene translocations. (a) Two double stranded breaks on separate chromosomes can lead to a balanced translocation. (b) Two deletions on one chromosome can lead to the loss of a piece of DNA. (c) Two deletions on one chromosome can also cause the inversion of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

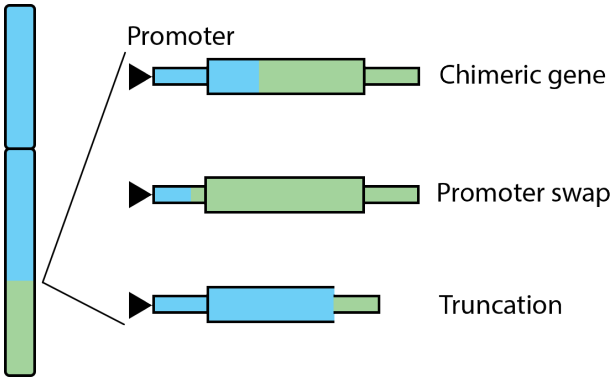

Figure 1.2: Generally, the three effects a translocation can have on the resulting fusion gene are illustrated. On the left a chromosome is depicted, while on the right resulting mRNA is shown. For the mRNA the UTR (slim bars) and coding regions (wide bars) are shown.

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-

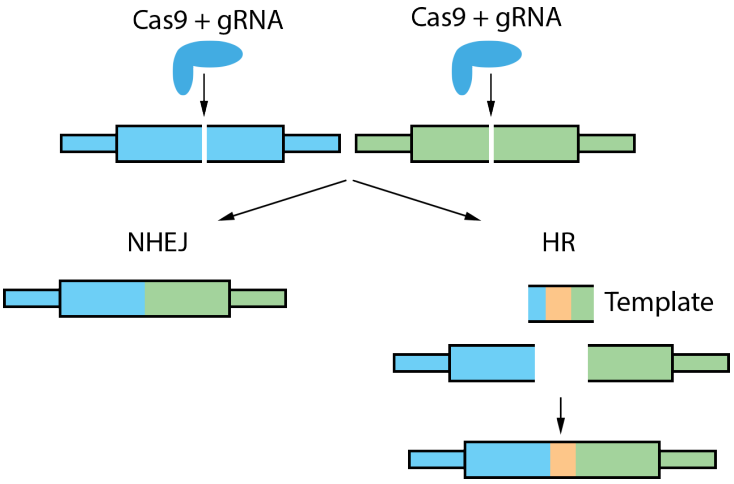

Figure 1.3: Induction of translocations using CRISPR/Cas9. Two break points are introduced and both the NHEJ and HR can lead to the formation of a translocation.

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

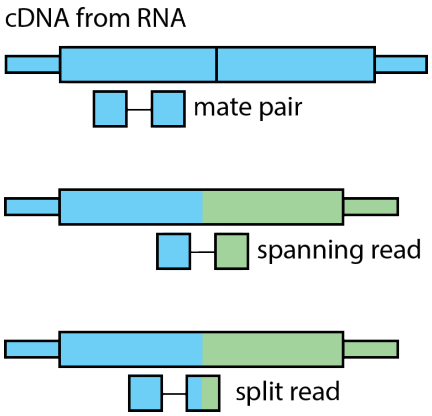

Figure 1.4: Fusion genes can be detected on mRNA (through sequencing of cDNA). Detection of spanning or split reads is generally used to discover fusion gene.

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.