

UNIVERSITY OF SÃO PAULO

BIOINFORMATICS GRADUATE PROGRAM

RODRIGO GUARISCHI MATTOS AMARAL DE SOUSA

**The transcriptome of oxygen-induced retinopathy
and an angiogenesis-based prognostic gene signature
for prediction of breast cancer relapse**

São Paulo

2017

WARNING!

THIS IS A SIMPLIFIED VERSION OF THE THESIS, WHICH CONTAINS ONLY THE PREAMBLE, ABSTRACT, INTRODUCTION, CONCLUSION AND REFERENCES, IN ORDER TO PROTECT PUBLICATION RIGHTS. THE COMPLETE VERSION IS EXPECTED TO BE AVAILABLE ON OCTOBER, 2018. CONTACT THE AUTHOR IF YOU WISH TO OBTAIN IT IN ADVANCE.

RODRIGO GUARISCHI MATTOS AMARAL DE SOUSA

**The transcriptome of oxygen-induced retinopathy
and an angiogenesis-based prognostic gene signature
for prediction of breast cancer relapse**

Revised Version

Thesis presented to the Bioinformatics Graduate
Program at the University of São Paulo to obtain
the degree of Doctor of Science

Concentration Area: Bioinformatics

Advisor: Prof. João Carlos Setubal, PhD

Co-advisor: Prof. Ricardo José Giordano, PhD

São Paulo

2017

Author: Rodrigo Guarischi Mattos Amaral de Sousa

Title: The transcriptome of oxygen-induced retinopathy and an angiogenesis-based prognostic gene signature for prediction of breast cancer relapse

Thesis presented to the Bioinformatics Graduate Program at the University of São Paulo to obtain the degree of Doctor of Science.

Approved on: June 2, 2017

Examination Committee

Prof. João Carlos Setubal, PhD

Judgment: *Aprovado*

Institution: Chemistry Institute - USP

Signature: 

Prof. Paul C. Boutros, PhD

Judgment: *Approved*

Institution: Ontario Institute for Cancer Research

Signature: 

Prof. Anamaria Aranha Camargo, PhD

Judgment: *APPROVED*

Institution: Sírio-Libanês Hospital

Signature: 

Prof. Pedro A. F. Galante, PhD

Judgment: *APPROVED*

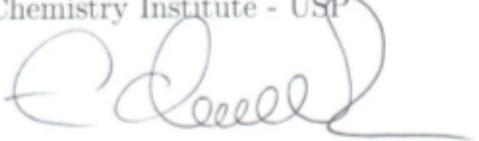
Institution: Sírio-Libanês Hospital

Signature: 

Prof. Eduardo Moraes Rego Reis, PhD

Judgment: *APROVADO*

Institution: Chemistry Institute - USP

Signature: 

Dedicatória

Dedico esta tese aos meus pais Eduardo e Kátia pelo suporte.

À minha namorada Renata Stabenow,
e seus pais, Jorge e Elizabeth, pelo carinho e companhia.

À minha irmã e amiga Renata Guarischi.

Aos meus avós Levy e Lia que são e sempre serão parte de mim.

Às minhas tias Marina e Mônica e minha priminha sapeca, Victória.

Acknowledgments

I am greatly indebted to my supervisors João Carlos Setubal and Ricardo José Giordano for offering me this opportunity, for mentoring me, for all the precious time shared with me on meetings and for the scientific counseling throughout the development of this project. I will forever be grateful for all the intellectual and professional support you were able to give me during all these years. You are two great scientists and human beings, and I will carry these lessons throughout my life.

To professor Paul C. Boutros for accepting me as an exchange student for 1 year and 3 months on his laboratory in Toronto. You provided me the most intense learning experience of my life, where I lived another culture, language and glimpsed an exciting research field on a fast-pace group. Thank you for believing and allowing me to develop myself as a professional.

To all the students and professionals that I had the pleasure of working with in Brazil and during my internship. In particular I would like to thank Thiberio, Gianluca, George, Deyvid, Jhonatas, Carlos, Javier, Ren, Connie, Veronica, Taka, Shad, Natalia for the support, laughter, scientific and non-scientific discussions.

To the São Paulo Research Foundation (FAPESP) for granting me two scholarships (2012/15197-1 and 2014/21360-8), providing financial support during almost 5 years and which were fundamental to the development of this project.

To the Bioinformatics Graduate Program and the University of São Paulo for providing equipments and technical support.

“It is hard to fail, but it is worse never to have tried to succeed.”

Theodore Roosevelt

RESUMO

Guarisch-Sousa, Rodrigo. O transcrito da retinopatia induzida por oxigênio e uma assinatura gênica prognóstica baseada em angiogênese para predição de recidiva de câncer de mama [tese]. São Paulo: Universidade de São Paulo, Programa de Pós-graduação em Bioinformática; 2017.

Angiogênese é o processo de formação de novos vasos sanguíneos a partir dos vasos existentes. É um processo vital, mas muitas doenças também dependem deste mecanismo para obter nutrientes e progredir. Estas “doenças dependentes de angiogênese” incluem cânceres, retinopias e degeneração macular. Alguns inibidores da angiogênese foram desenvolvidos na última década, com o objetivo de auxiliar no manejo dessas doenças e melhorar a qualidade de vida dos pacientes. A maioria destes compostos funciona inibindo a ligação de VEGFA/VEGFR2, que também é um elemento importante para a sobrevivência de células endoteliais quiescentes; e isso pode explicar parcialmente eventos adversos observados em alguns ensaios clínicos. Nossa hipótese é que a melhoria das terapias anti-angiogênicas depende de uma compreensão melhor e mais ampla desse processo, especialmente quando relacionada à progressão das doenças. Utilizando RNA-Seq e um modelo animal bem aceito de angiogênese, o modelo murino de Retinopatia Induzida por Oxigênio, exploramos o transcrito e identificamos 153 genes diferencialmente expressos durante a angiogênese. Uma extensiva validação de vários genes realizada por qRT-PCR e hibridização in-situ confirmou a superexpressão de Esm1 em células endoteliais de tecidos com angiogênese ativa. A análise de enriquecimento desta lista de genes confirmou a ligação da angiogênese com genes frequentemente mutados em tumores, consistente com a conhecida ligação entre câncer e angiogênese, e forneceu sugestões de fármacos já aprovados que podem ser reutilizados para controlar a angiogênese em circunstâncias patológicas. Finalmente, com base neste panorama amplo da angiogênese, fomos capazes de criar um biomarcador molecular com poder prognóstico para a predição da recidiva de câncer de mama, com aplicações clínicas promissoras. Em resumo, este trabalho revelou com sucesso genes relacionados à angiogênese e forneceu novas alternativas terapêuticas, incluindo potenciais fármacos para reposicionamento. Esse conjunto de genes diferencialmente expressos é também um recurso valioso para investigações futuras.

Palavras-chave: RNA-Seq. Transcrito. Angiogênese. Retinopatia Induzida por Oxigênio. Assinatura gênica. Biomarcador molecular. Recidiva de câncer de mama.

ABSTRACT

Guarisch-Sousa, Rodrigo. The transcriptome of oxygen-induced retinopathy and an angiogenesis-based prognostic gene signature for prediction of breast cancer relapse [thesis]. São Paulo: University of São Paulo, Bioinformatics Graduate Program; 2017.

Angiogenesis is the process of formation of new blood vessels based on existing vessels. It is a vital process but many diseases also rely on this mechanism to get nourishment and progress. These so called “angiogenesis-dependent diseases” include cancers, retinopathies and macular degeneration. Some angiogenesis inhibitors were developed in the past decade, aiming to help the management of such diseases and improve patients’ quality of life. Most of these compounds work by inhibiting VEGFA/VEGFR2 binding, which is also a key element to the survival of quiescent endothelial cells; this may partly explain unanticipated adverse events observed in some clinical trials. We hypothesize that the improvement of anti-angiogenesis therapies hinges on a better and broader understanding of the process, especially when related to diseases’ progression. Using RNA-seq and a well accepted animal model of angiogenesis, the murine model of Oxygen Induced Retinopathy, we have explored the transcriptome landscape and identified 153 genes differentially expressed in angiogenesis. An extensive validation of several genes carried out by qRT-PCR and in-situ hybridization confirmed *Esm1* overexpression in endothelial cells of tissues with active angiogenesis, providing confidence on the results obtained. Enrichment analysis of this gene list endorsed a narrow link of angiogenesis and frequently mutated genes in tumours, consistent with the known connection between cancer and angiogenesis, and provided suggestions of already approved drugs that may be repurposed to control angiogenesis under pathological circumstances. Finally, based on this comprehensive landscape of angiogenesis, we were able to create a prognostic molecular biomarker for prediction of breast cancer relapse, with promising clinical applications. In summary, this work successfully unveiled angiogenesis-related genes, providing novel therapeutic alternatives, including potential drugs for repositioning. The set of differentially expressed genes is also a valuable resource for further investigations.

Keywords: RNA-Seq. Transcriptomics. Angiogenesis. Oxygen-Induced Retinopathy. Gene signature. Molecular biomarker. Breast cancer relapse.

Chapter 1

Introduction

1.1 Hypoxia and Angiogenesis

Angiogenesis is the process of formation of new blood vessels based on pre-existing vessels. It is the main mechanism by which vertebrates react to the shortage of oxygen levels (hypoxia) during adulthood. The process begins with the transcription of Hypoxia-Inducible Factors (HIF). HIF1 complex acts as a master regulator of cellular and systemic homeostatic response to hypoxia by activating the transcription of many genes, including those involved on oxygen transportation, glucose metabolism, cell proliferation, apoptosis, angiogenesis and many others oxygen-dependent processes [1].

On the presence of oxygen, the alpha subunit of HIF1 (HIF1A) is hydroxylated and subject to ubiquitination by VHL factor, which labels it for rapid degradation on proteasome. During hypoxia, HIF1A will no longer be hydroxylated and degraded, as a result, intracellular HIF1A is translocated to the nucleus where it forms a transcriptional complex with the constitutively expressed beta subunit (HIF1B) and other accessory proteins. This complex activates the transcription of genes that will respond to hypoxia state; including growing factors and cytokines, that prompt endothelial cells (EC) to leave their basal state and migrate, forming new blood vessels. Therefore, HIF1 is a trigger and one of the main modulators for angiogenesis by inducing the production of Vascular Endothelial Growth Factor (VEGF) [2].

1.2 Cell-signaling and dynamics

VEGF is a gene family regulated by HIF1 and one of the most studied and known to be implicated in angiogenesis. This family is composed by five members: VEGFA (sometimes simply referred as VEGF), VEGFB, VEGFC, VEGFD and Placental Growth Factor (PlGF); being VEGFA the most important gene and, consequently, the most studied one [3].

VEGF produced by hypoxic cells accumulate, forming an extracellular gradient. Binding of these proteins with tyrosine kinase receptors, the VEGFRs, or neuropilin receptors located at the surface of EC induces a transphosphorylation signal that switch on an intracellular cascade of genes (Fig. 1.1).

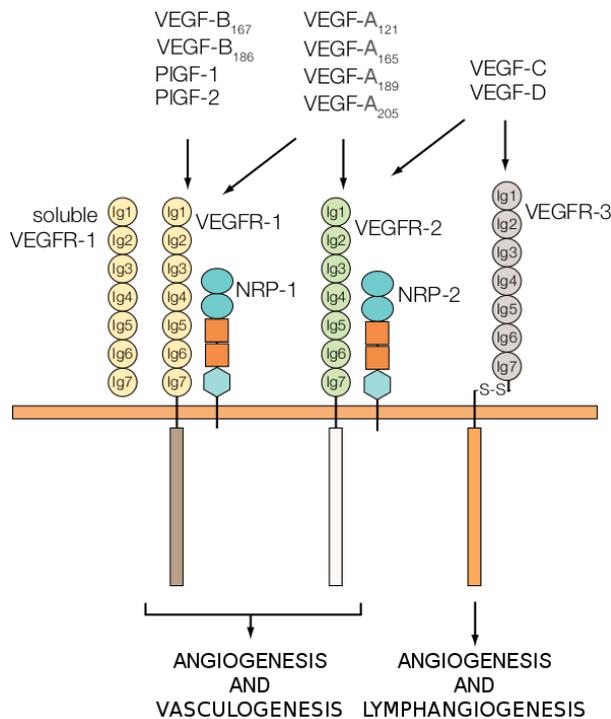


Figure 1.1: VEGF gene family and their receptors. Different VEGF transcripts are produced through alternative splicing. These isoforms differentially and selectively bind to VEGF or neuropilin receptors, triggering angiogenesis and lymphangiogenesis processes. VEGFA/VEGFR-2 binding is the main stimulator of angiogenesis and part of this cascade acts as regulators of this interaction; like soluble VEGFR-1, which works as a dummy/decoy receptor, sequestering VEGFA and inhibiting it from binding to VEGFR-2. Figure adapted from Ferreira et al. [2]

Upon activation by VEGF, EC leave their quiescent state, beginning the angiogenesis process. Interestingly, two types of EC seem to form: tip cells, which have long filopodia segments with receptors on its surface aimed for sensing the environment and coordinate migration; and stalk cells, that proliferate and create tight junctions, forming vessel's lumen and maintaining its stability [4] (Fig. 1.2).

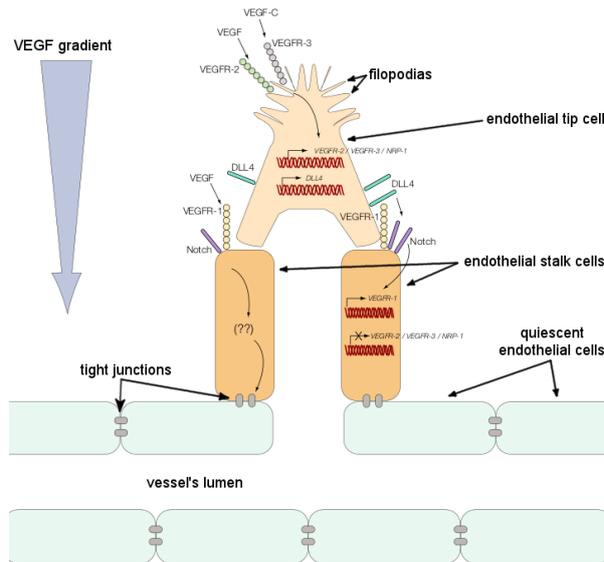


Figure 1.2: Two subtypes of EC are formed upon activation by VEGF: tip and stalk. While tip cells are aimed for sensing the gradient, stalk cells create tight junctions, forming vessel's lumen and maintaining its stability. These cell subtypes are created and maintained by a cascade that involves many growth factors, cell receptors and transcription factors. Figure adapted from Ferreira et al. [2]

1.3 Angiogenesis in adulthood and diseases

On healthy adults, angiogenesis is restricted to specific physiological processes such as: wound healing, menstruation and formation of the placenta. Where and how much angiogenesis is needed relies upon a delicate balance between stimulators and inhibitors.

Unfortunately, angiogenesis is not restricted to physiological processes and some diseases tweak the equilibrium of this signaling cascade, which supports its onset or progression. A remarkable example of this abnormal proliferation of blood vessels, known as pathological angiogenesis, is observed during the evolution of solid tumours.

1.3.1 Angiogenesis, Cancer and Metastasis

The transformation of a small bulk of cells into a tumour and further progression into metastasis relies on inducing angiogenesis and lymphangiogenesis. Like normal cells, a tumour depends on blood vessels to receive a continuous supply of oxygen and other nutrients that allows it to grow. These vessels may also be used by some cells to reach the bloodstream which, ultimately, allows the colonization of distant sites of the body during metastasis. Accordingly, angiogenesis is considered one of the hallmarks of cancer [5].

Unlike the transient physiologic processes which occurs during wound healing and female reproductive cycling, during tumour progression, an “angiogenic switch” is almost always activated and remains on, causing normally quiescent vasculature to continually sprout new vessels that help sustain expanding neoplastic growth [5, 6] (Fig. 1.3).

The blood vessels produced within tumours by chronically activated angiogenesis are

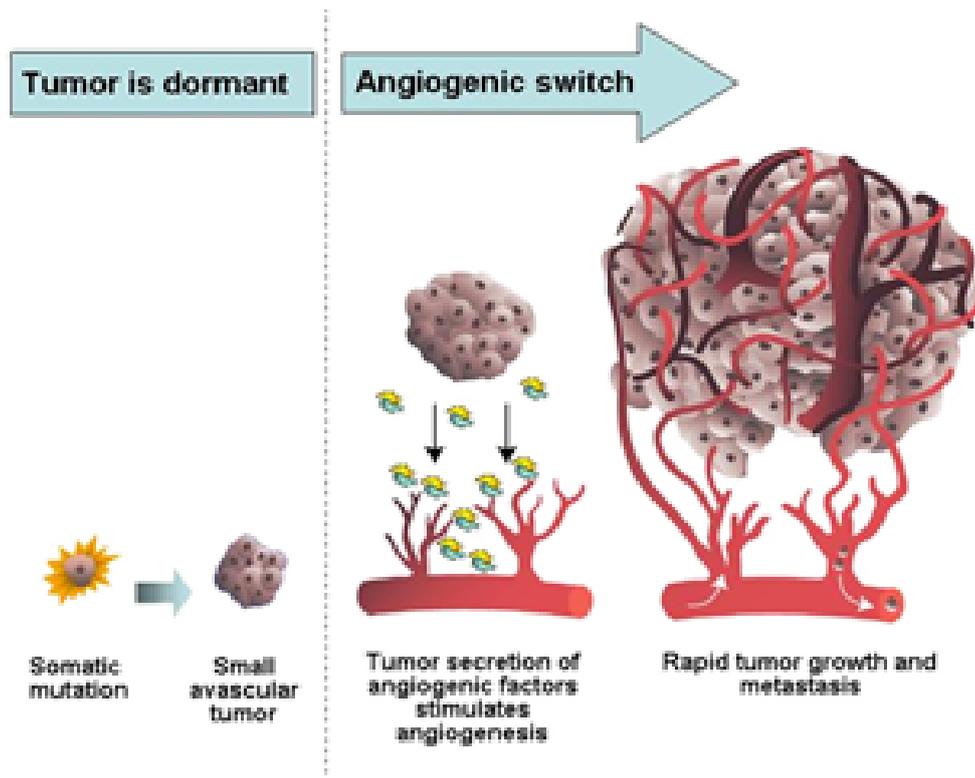


Figure 1.3: Angiogenesis and tumour growth. Like normal tissues, tumours need nutrients, oxygen and ways to evacuate carbon dioxide and metabolic wastes. Accordingly, cells of tumours need to induce angiogenesis in order to obtain nutrients and support their rapid expansion. Some cells may also use these vessels to reach the bloodstream, in order to colonize distant sites during metastasis. Figure from <http://www.clinicaloptions.com/>

typically aberrant: tumour neovasculature is marked by precocious capillary sprouting, convoluted and excessive vessel branching, distorted and enlarged vessels, erratic blood flow, microhemorrhaging, leakiness, abnormal levels of EC proliferation and apoptosis [5].

It has been observed that avascular tumours are severely restricted in their growth potential when on lack of a blood supply. Therefore, its inhibition is considered a valuable approach for cancer therapies [7, 3].

Although anti-angiogenic therapies have been evaluated more extensively in cancer settings, there is a growing body of evidence that suggests that many human diseases share angiogenesis as a common denominator. Either by excessive or insufficient growth, these “Angiogenesis-dependent diseases” also include macular degeneration, retinopathies, chronic wounds and are estimated to affect millions of people worldwide [8, 3]

1.3.2 Neovascular Age-Related Macular Degeneration

Age-Related Macular Degeneration (AMD) is a medical condition which may result in blurred or no vision in the center of the visual field. The disease onset might be asymptomatic but, over time, some people experience a gradual worsening of vision that may affect one or both eyes. Loss of central vision can make activities of daily life such as recognizing faces, driving and reading much harder.

It most commonly occurs in people over the age of fifty. In United States, it affects about 0.4% of adults between 50 and 60 years old. Incidence raises with aging, reaching nearly 12% of people over 80 years old [9]. Globally is thought to affect over 23.5 million people, being 13.4 million on its moderate or severe forms [10].

Late stages of AMD are classified on two different types: geographic atrophy and neovascular. Neovascular AMD is caused by abnormal blood vessels that grow underneath the retina. These vessels can leak fluid and blood, which may lead to swelling and a rapid and severe damage of the macula [9].

Treatment of advanced neovascular AMD typically involves monthly intra-ocular injections of anti-VEGFA drugs, which is normally overly expressed and stimulate the growth of abnormal blood vessels. Anti-VEGFA injection therapy blocks this growth and stabilizes disease progression [9, 10] (Sec. 1.4).

1.3.3 Human Retinopathies

Retinopathy of Prematurity (ROP) is a disease that affects preterm babies while receiving neonatal care and may have consequences throughout the whole life. Due to premature development of lungs, oxygen-therapy is frequently used. This oxygen-rich environment delays vascular development in the eye and partially regress existing vessels. If not controlled properly, the abrupt transition to room air with the lack of a fully formed blood network creates hypoxia, inducing VEGFA expression and abnormal blood vessels formation [11].

These abnormal blood vessels spread throughout the retina but are fragile and can leak, producing fibrous scar extending from the retina to the vitreous gel and lens. When these scars shrink, they pull retina out of position, sometimes causing retinal detachment, which is the main cause of visual impairment and blindness in ROP. Although the understanding of risk factors is much better nowadays, the number of infants with ROP has increased,

most likely due to the increased survival of very low birth weight babies [11].

Diabetic Retinopathy (DR), a secondary microvascular complication of diabetes mellitus, is the primary cause of visual loss in working-age persons in developed countries. Globally, it is estimated that diabetes mellitus affects 4% of world's population and almost half of these patients have some degree of DR at any given time [12].

Chronically increased levels of blood glucose are thought to have a structural and physiological effect on retinal capillaries, causing them to be both functionally and anatomically incompetent. High blood glucose is thought to induce hypoxia in the retina, thus leading to VEGFA production and abnormal blood vessels proliferation [13]. Visual loss primarily occurs by either proliferation of new retinal vessels (proliferative diabetic retinopathy) or increased permeability of capillaries, that leads to swelling in the macula and causes Diabetic Macular Edema (DME) [12].

As a shared denominator to many human diseases, anti-angiogenic therapies have great clinical potential. Accordingly, some anti-angiogenic therapies have been developed and tested on the past decade, especially for cancer and ophthalmic diseases. Either as primary or adjuvant to conventional treatments, clinical trials revealed varying levels of efficacy, which stimulated the development of many different drugs [3].

1.4 Anti-angiogenic therapies

Under normal physiological conditions, it is estimated that more than 99% of EC remain on quiescent state. This stability led to the belief that the signaling cascade of angiogenesis would be restrict to blood vessels under development, specifically those implicated on pathological angiogenesis. This would allow the development of therapies with mild side-effects, halting disease progression and possibly starving cells of solid tumours [2]

This hypothesis boosted the development of many pharmacological compounds targeting proteins associated with the proliferation and survival of EC. The first VEGFA inhibitor, bevacizumab (branded as Avastin), was approved by the US Food and Drug Administration (FDA) in 2004 for the first-line treatment of metastatic colorectal cancer. Pegaptanib (branded as Macugen) and Ranibizumab (branded as Lucentis) also received approval from the same regulatory agency for ophthalmic applications shortly thereafter [3].

Alternative approaches to inhibit VEGFA/VEGFR2 signaling have also been exploited, and small molecule inhibitors targeting the kinase domain of VEGFR2 were developed. In 2005, Sorafenib (branded as Nexavar) gained FDA approval for metastatic renal cell carcinoma. Today, this drug is also used for hepatocellular carcinoma and on thyroid carcinoma [3].

Sunitinib, Ranibizumab, Aflibercept and at least ten other compounds are available today targeting angiogenesis at different moments of the signaling cascade. Nevertheless, although anti-angiogenesis therapies have resulted in significant improvements in vision and quality of life for patients with AMD, DME, DR and cancer, drug resistance and other unanticipated adverse events have been reported during clinical trials.

In phase I clinical trials, bevacizumab monotherapy was generally well tolerated, with no severe adverse events, typically hypertension and mild proteinuria. Some studies also reported clinical improvements in a number of tumour types when used adjuvant to most conventional chemotherapy regimens. Importantly, bevacizumab did not markedly increase toxicity when used in combination with a range of chemotherapeutic agents. However, subsequent studies revealed that a fraction of patients suffered from gastrointestinal perforations, nephrotic syndrome or arterial thromboembolic complications such as myocardial infarction and stroke [3].

Patients with neovascular AMD showed suboptimal treatment response in approximately 40% of cases, and higher doses are not likely to be helpful. Data from Phase III studies indicate that current approved doses are at or near the top of the dose response curves for AMD and DME [3].

It is known today, that VEGFA and other parts of this signaling pathway are also survival factors to quiescent EC, with roles during hematopoiesis, myelopoiesis, homeostasis and maintenance of vascular survival [2]. Possibly a significant fraction of observed side-effects are related to inhibition of key elements to survival of quiescent EC and physiological angiogenesis. Therefore, the improvement of anti-angiogenesis therapies hinges on a better and broader understanding of pathological angiogenesis, and animal models that mimic angiogenesis *in vivo* are great tools of study. A model such as the Oxygen-Induced Retinopathy (OIR) enables the investigation of molecular and cellular mechanisms that resembles more closely a disease environment.

1.5 Oxygen-Induced Retinopathy model

OIR model is one of the animal models that best mimic angiogenesis in vivo. Briefly, newborn mice are kept under hyperoxia (75% oxygen) between days 7 and 12 postnatal, a critical period for vascular maturation. The high levels of oxygen reduce VEGF expression, halting neovascularization and even inducing a mild capillary depletion in the central area of the retina. When removed from oxygen chamber at day 12, the then existing network of vessels is no longer sufficient to supply oxygen to the demanding photoreceptor cells. The tissue becomes highly hypoxic leading to abnormally high levels of VEGFA and an exacerbated neovascularization, which persists until day 17, when blood vessels slowly begin to regress (Fig. 1.4) [14, 15].

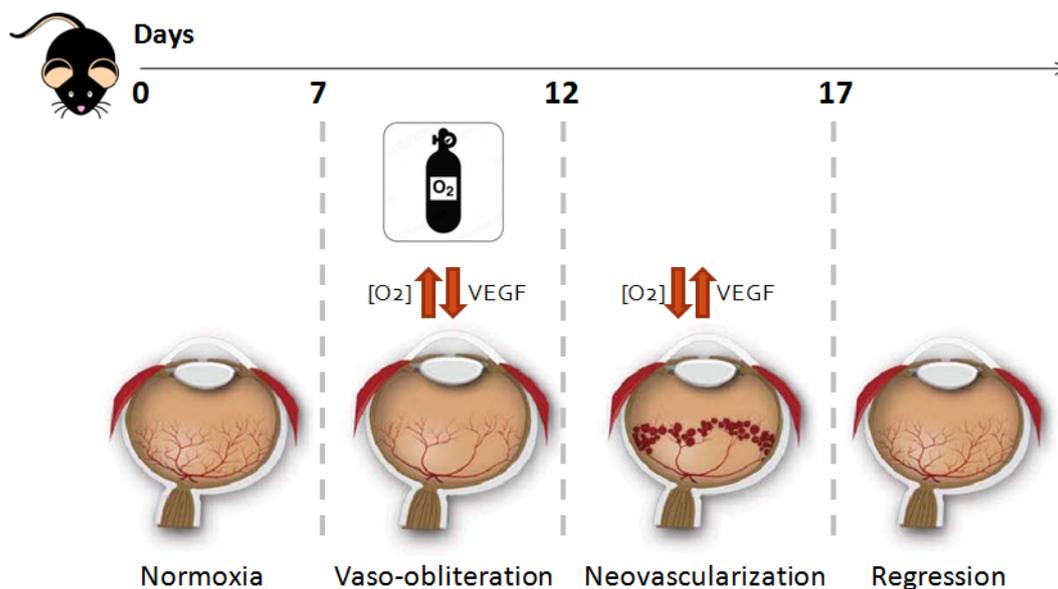


Figure 1.4: OIR model. Newborn mice are kept under hyperoxia (75% O_2) between days 7 and 12, this impairs physiological vasculature development and leads to capillary depletion. When returned to normoxia ($\sim 20,8\%$ O_2) there is an exacerbated neovascularization in the retina of mouse pups which persists until day 17. Figure adapted from Connor et al. [15].

OIR has some advantages over other angiogenesis models. Firstly, blood vessels grow inside a living tissue (retina) and not an artificial matrix such as Matrigel, used in other angiogenesis assays. Therefore, signaling between different cells is preserved as it would in a living organism. In addition, the induction of angiogenesis is modulated by oxygen and not by chemicals or other external compounds added artificially, which would otherwise exacerbate individual signaling cascades. And lastly, the experimental setting of this model resembles closely the ROP disease, that affects millions of preterm babies (Sec. 1.3.3).

Thus, the OIR has become the most widely used animal model to study angiogenesis, and has played a pivotal role in our understanding of retinal angiogenesis and ocular immunology. OIR model is also a hallmark in the development of groundbreaking therapeutics such as anti-VEGFA injections for neovascular AMD [16] (Sec. 1.3.2).

Angiogenic retinas of OIR model may be used to visualize blood vessels under development by immunostaining, evaluate outcome after treatment with drugs in vivo or even quantify expression of genes of interest, among other assays.

1.6 Gene expression and RNA-Seq

Measurement of gene expression is possible by different methods. Some technologies may be used to profile expression of a single gene, while high-throughput methods are able to assess the expression of multiple or even the whole set of genes, the transcriptome, at once [17].

Some high-throughput technologies are based on hybridization, such as microarrays, where there is a large numbers of microscopic DNA spots (probes) attached to a solid surface, each reflecting a gene of interest. Relative expression levels are measured simultaneously based on the intensity of fluorescence released after hybridization of cDNA from samples and probes [17].

Alternatively, the exponential growth in sequencing throughput, driven by the introduction of Massive Parallel Sequencing platforms, allowed a significant decrease on the cost per sequenced base, enabling applications such as RNA-Seq. Briefly, the RNA from samples of interest is converted to cDNA and sequenced, the read content reflects the nucleotide sequence of a transcript while the number of reads, its expression level. As DNA sequencing does not rely on information a priori, this method allows the measurement of transcripts with known or still unknown sequence, not being limited to a reference annotation, unlike microarrays which are limited to the set of probes used. As such, RNA-Seq provides a wealth of information that can be used for many different types of analysis [17].

A typical simplified data generation protocol of RNA-Seq involves steps of:

1. *RNA extraction*: sample is collected and material is isolated.
2. *DNAse treatment*: reduces the amount of contaminant genomic DNA.

3. *Integrity assessment*: gel and capillary electrophoresis are used to verify degradation of input material.
4. *RNA selection or depletion*: rRNA dominates transcriptome landscape. RNA of interest can be enriched either by 3' polyadenylated tail (poly-A) probes, given that rRNA lack a poly-A tail; or oligomers complementary to rRNA sequences may be used to deplete it. This step seeks to enrich the sample with non-ribosomal RNAs.
5. *Library preparation*: RNA is reversely transcribed to cDNA, fragmented and size selection is performed to purify sequences that are of appropriate length for the sequencing machine. Finally, auxiliary sequences flanking the sequence of interest, known as adapters, are ligated; these sequences are used during sequencing.
6. *Sequencing*: biological information is transformed into digital data, typically a FASTQ file.
7. *Analysis*: digital information is transformed into knowledge, answering biological questions like: Which genes are the most expressed? Which genes are differentially expressed genes between treated and control conditions? Are there non-annotated genes or isoforms? Are there gene fusions?

The core component of this thesis is sustained on the employment of RNA-Seq on the OIR model to compare the transcriptome of angiogenesis-induced and physiological retinas, in order to identify genes involved on angiogenesis control.

Details of objectives, materials and methods will be presented on chapters 2 and 3. The differentially expressed genes (DEG) in angiogenesis, experimental validation of results, as well as the computational pipeline used will be presented on chapter 4. Chapter 5 will describe a follow-up study that resulted on the creation of an angiogenesis-based gene signature with prognostic power to predict patients more likely to have a breast cancer relapse after surgery. Chapters 6 and 7 will explore the contributions of this work to the scientific community and describe other achievements during collaborations.

Chapter 6

Conclusion

Our general goal was to obtain a comprehensive landscape of gene expression in angiogenesis. To that, we have used the OIR model and carried out a RNA-Seq on retinas of mouse pups at four stages of intense neovascularization (days 12, 12.5, 15 and 17), identifying 153 genes at least twice more expressed on angiogenesis-induced compared to control conditions. This list includes some genes well-known as being involved on angiogenesis, such as growth factors (Vegfa and Pdgfb), chemokines (Ccl2), integrins (Itga1); but also many not well characterized ones, including some predicted genes (Gm15983, Gm12802, Gm43620 and Gm35040), which still lack a better description on annotation databases.

An in-depth hybridization-based validation was conducted and an excellent overall correlation was observed, confirming the robustness of results obtained and validity of the pipeline proposed for processing RNA-Seq data.

In-situ hybridization of an identified DEG (Esm1) verified its overexpression on blood vessels under development at tissues of active angiogenesis, representing a promising therapeutic target for controlling angiogenesis on circumstances when it is exacerbated.

Enrichment analysis of the list of DEG enabled the creation of a pathway map with many angiogenesis-associated biological terms, confirming a big overlap with results on the literature. Post-enrichment analysis suggested alternative drugs that may be repurposed to control angiogenesis under pathological circumstances and endorsed a narrow link with commonly mutated cancer-genes, which stimulated us to assess the performance of an angiogenesis-based gene signature to predict tumour relapse.

A literature search revealed nine distinct angiogenesis signatures, created from eight

distinct cancer types, and only a mild degree of commonalities between them, partly explained by distinct methodologies employed. When tested on a cancer cohort with almost 2.000 samples, no prognostic power was observed to predict breast cancer relapse. Based on the same cancer dataset, the list of DEG during angiogenesis and machine learning techniques, we were able to create a signature which showed great prognostic power on assigning patients in risk groups.

Although our classifier performed remarkably well on basal and luminal breast cancer subtypes, more studies are necessary to check the performance reproducibility on different cohorts and to evaluate how the proposed signature perform compared to FDA-cleared breast cancer biomarkers, such as MammaPrint [79]. Another interesting investigation would be the assessment of the performance of the same gene signature on angiogenesis-rich cancers such as glioblastomas and renal carcinomas; or even, the creation of other gene signatures based on the same set of angiogenesis-associated genes.

In summary, our work presents three significant scientific contributions. First, a list of DEG on OIR model, an extensively used animal model for the study of angiogenesis. Secondly, a drug set that may be used to target these genes, bringing a strong potential for drug repositioning, which represents a shortcut over traditional drug development process. Finally, a gene signature with prognostic power to predict breast cancer relapse, which may assist treatment guidance. Worth noting that this list of genes associated with angiogenesis may also benefit other research groups studying other “angiogenesis-dependent diseases” (e.g.: retinopathies and macular degeneration), not only cancers, being a valuable resource for further investigations.

Bibliography

- [1] E B Rankin and A J Giaccia. The role of hypoxia-inducible factors in tumorigenesis. *Cell Death and Differentiation*, 15(4):678–685, 2008.
- [2] Carlos Gil Ferreira and José Cláudio Rocha. *Oncologia Molecular*. Atheneu, 2 edition, 2010.
- [3] Napoleone Ferrara and Anthony P Adamis. Ten years of anti-vascular endothelial growth factor therapy. *Nature Reviews Drug Discovery*, 15(6):385–403, 2016.
- [4] L.-K. Phng and Holger Gerhardt. Angiogenesis: A Team Effort Coordinated by Notch. *Developmental Cell*, 16(2):196–208, 2009.
- [5] Douglas Hanahan and Robert A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674, 2011.
- [6] Douglas Hanahan and Judah Folkman. Patterns and Emerging Mechanisms of the Angiogenic Switch during Tumorigenesis. *Cell*, 86(3):353–364, 1996.
- [7] Judah Folkman. Role of angiogenesis in tumor growth and metastasis. *Seminars in Oncology*, 29(6Q):asonc02906q0015, 2002.
- [8] J Folkman. Angiogenesis-dependent diseases. *Seminars in Oncology*, 28(6):536–542, 2001.
- [9] Sonia Mehta. Age-Related Macular Degeneration. *Primary Care: Clinics in Office Practice*, 42(3):377–391, 2015.
- [10] Raul Velez-Montoya, Scott C. N. Oliver, Jeffrey L. Olson, et al. Current knowledge and trends in age-related macular degeneration. *Retina*, 34(3):423–441, 2014.
- [11] Jing Chen and Lois E. H. Smith. Retinopathy of prematurity. *Angiogenesis*, 10(2):133–140, 2007.
- [12] Lloyd Paul Aiello. Angiogenic Pathways in Diabetic Retinopathy. *New England Journal of Medicine*, 353(8):839–841, 2005.
- [13] Talia Crawford, D. Alfaro III, John Kerrison, and Eric Jablon. Diabetic Retinopathy and Angiogenesis. *Current Diabetes Reviews*, 5(1):8–13, 2009.
- [14] A Scott and M Fruttiger. Oxygen-induced retinopathy: a model for vascular pathology in the retina. *Eye*, 24(3):416–421, 2009.
- [15] Kip M Connor, Nathan M Krah, Roberta J Dennison, et al. Quantification of oxygen-induced retinopathy in the mouse: a model of vessel loss, vessel regrowth and pathological angiogenesis. *Nat. Protoc.*, 4(11):1565–1573, 2009.
- [16] Clifford Kim, Patricia D’Amore, and Kip Connor. Revisiting the mouse model of oxygen-induced retinopathy. *Eye and Brain*, page 67, 2016.

- [17] Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63, 2009.
- [18] Alexander Dobin, Carrie A Davis, Felix Schlesinger, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.
- [19] Andrew Yates, Wasiu Akanni, M Ridwan Amode, et al. Ensembl 2016. *Nucleic Acids Res.*, 44(D1):710–6, 2016.
- [20] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15(12):550, 2014.
- [21] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, 2015.
- [22] S. Anders, A. Reyes, and W. Huber. Detecting differential usage of exons from RNA-seq data. *Genome Research*, 22(10):2008–2017, 2012.
- [23] Steffen Durinck, Paul T Spellman, Ewan Birney, and Wolfgang Huber. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.*, 4(8):1184–1191, 2009.
- [24] Steffen Durinck, Yves Moreau, Arek Kasprzyk, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, 21(16):3439–3440, 2005.
- [25] Mihaela Pertea, Geo M Pertea, Corina M Antonescu, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3):290–295, 2015.
- [26] A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.
- [27] Jo Vandesompele, Katleen De Preter, Filip Pattyn, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, 3(7):RESEARCH0034, 2002.
- [28] N Schaeren-Wiemers and A Gerfin-Moser. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: in situ hybridization using digoxigenin-labelled cRNA probes. *Histochemistry*, 100(6):431–440, 1993.
- [29] Jüri Reimand, Reimand Jüri, Arak Tambet, et al. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.*, 44(W1):W83–W89, 2016.
- [30] Paul Shannon, Andrew Markiel, Owen Ozier, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11):2498–2504, 2003.
- [31] Daniele Merico, Ruth Isserlin, Oliver Stueker, Andrew Emili, and Gary D Bader. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*, 5(11):e13984, 2010.
- [32] David S Wishart, Craig Knox, An Chi Guo, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, 34(Database issue):668–72, 2006.
- [33] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, 6(269):11, 2013.

- [34] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, 2(5):401–404, 2012.
- [35] Christina Curtis, Sohrab P Shah, Suet-Feung Chin, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403):346–352, 2012.
- [36] David R Cox. Regression Models and Life-Tables. In *Springer Series in Statistics*, pages 527–541. 1992.
- [37] Vinayak Bhandari and Paul C Boutros. Comparing continuous and discrete analyses of breast cancer survival information. *Genomics*, 108(2):78–83, 2016.
- [38] Hemant Ishwaran and Udaya B Kogalur. Random survival forests for R. *R News*, 7:25–31, 2007.
- [39] P. C. Boutros, S. K. Lau, M. Pintilie, et al. Prognostic gene signatures for non-small-cell lung cancer. *Proceedings of the National Academy of Sciences*, 106(8):2824–2828, 2009.
- [40] M. H. Starmans, M. Pintilie, M. Chan-Seng-Yue, et al. Integrating RAS Status into Prognostic Signatures for Adenocarcinomas of the Lung. *Clinical Cancer Research*, 21(6):1477–1486, 2015.
- [41] Cole Trapnell, Adam Roberts, Loyal Goff, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3):562–578, 2012.
- [42] Ann E. Loraine, Ivory Clabaugh Blakley, Sridharan Jagadeesan, et al. Analysis and Visualization of RNA-Seq Expression Data Using RStudio, Bioconductor, and Integrated Genome Browser. pages 481–501. 2015.
- [43] Ernesto Picardi, Anna Maria D’Erchia, Antonio Montalvo, and Graziano Pesole. Using REDItools to Detect RNA Editing Events in NGS Datasets. In *Current Protocols in Bioinformatics*, pages 1–12. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2015.
- [44] Daehwan Kim, Geo Pertea, Cole Trapnell, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36, 2013.
- [45] Ana Conesa, Pedro Madrigal, Sonia Tarazona, et al. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1):13, 2016.
- [46] Pär G Engström, Tamara Steijger, Botond Sipos, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, 10(12):1185–1191, 2013.
- [47] Giacomo Baruzzo, Katharina E Hayer, Eun Ji Kim, et al. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature Methods*, 14(2):135–139, 2016.
- [48] Tamara Steijger, Josep F Abril, Pär G Engström, et al. Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods*, 10(12):1177–1184, 2013.
- [49] F. Seyednasrollah, A. Laiho, and L. L. Elo. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics*, 16(1):59–70, 2015.
- [50] Nicholas J. Church, Pietá Schofield, Marek Gierliński, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, 22(6):839–851, 2016.

- [51] Katharina E. Hayer, Angel Pizarro, Nicholas F. Lahens, John B. Hogenesch, and Gregory R. Grant. Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data. *Bioinformatics*, page btv488, 2015.
- [52] W J Kent, C W Sugnet, T S Furey, et al. The human genome browser at UCSC. *Genome Res.*, 12(6):996–1006, 2002.
- [53] Susana F Rocha, Maria Schiller, Ding Jing, et al. Esm1 modulates endothelial tip cell behavior and vascular permeability by enhancing VEGF bioavailability. *Circ. Res.*, 115(6):581–590, 2014.
- [54] Arpita S. Bharadwaj, Binoy Appukuttan, Phillip A. Wilmarth, et al. Role of the retinal vascular endothelial cell in ocular disease. *Progress in Retinal and Eye Research*, 32:102–180, 2013.
- [55] Jun-Hao Li, Shun Liu, Hui Zhou, Liang-Hu Qu, and Jian-Hua Yang. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, 42(D1):D92–D97, 2014.
- [56] J.-H. Yang, J.-H. Li, P. Shao, et al. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Research*, 39(Database):D202–D209, 2011.
- [57] A Kozomara and S Griffiths-Jones. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, 39(Database issue):D152–157, 2011.
- [58] Ana Kozomara and Sam Griffiths-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42(D1):D68–D73, 2014.
- [59] Matthew K Iyer, Yashar S Niknafs, Rohit Malik, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics*, 47(3):199–208, 2015.
- [60] M H W Starmans, N G Liewwes, P N Span, et al. Independent and functional validation of a multi-tumour-type proliferation signature. *British Journal of Cancer*, 107(3):508–515, 2012.
- [61] Maud H.W. Starmans, Kenneth C. Chu, Syed Haider, et al. The prognostic value of temporal in vitro and in vivo derived hypoxia gene-expression signatures in breast cancer. *Radiotherapy and Oncology*, 102(3):436–443, 2012.
- [62] Jiangting Hu, Fabrizio Bianchi, Mary Ferguson, et al. Gene expression signature for angiogenic and nonangiogenic non-small-cell lung cancer. *Oncogene*, 24(7):1212–1219, 2005.
- [63] Marta Mendiola, Jorge Barriuso, Andrés Redondo, et al. Angiogenesis-related gene expression profile with independent prognostic value in advanced ovarian carcinoma. *PLoS One*, 3(12):e4051, 2008.
- [64] Stefan Bentink, Benjamin Haibe-Kains, Thomas Risch, et al. Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. *PLoS One*, 7(2):e30269, 2012.
- [65] Massimo Masiero, Filipa Costa Simões, Hee Dong Han, et al. A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis. *Cancer Cell*, 24(2):229–241, 2013.
- [66] Tak L Khong, Ngayu Thairu, Helene Larsen, et al. Identification of the angiogenic gene signature induced by EGF and hypoxia in colorectal cancer. *BMC Cancer*, 13:518, 2013.

- [67] D J Pinato, T M Tan, S T K Toussi, et al. An expression signature of the angiogenic response in gastrointestinal neuroendocrine tumours: correlation with tumour phenotype and survival outcomes. *Br. J. Cancer*, 110(1):115–122, 2014.
- [68] Elena Sanmartín, Rafael Sirera, Marta Usó, et al. A Gene Signature Combining the Tissue Expression of Three Angiogenic Factors is a Prognostic Marker in Early-stage Non-small Cell Lung Cancer. *Ann. Surg. Oncol.*, 21(2):612–620, 2013.
- [69] Benoit Langlois, Falk Saupe, Tristan Rupp, et al. AngioMatrix, a signature of the tumor angiogenic switch-specific matrisome, correlates with poor prognosis for glioma and colorectal cancer patients. *Oncotarget*, 5(21):10529–10545, 2014.
- [70] Ingunn M Stefansson, Maria Raeder, Elisabeth Wik, et al. Increased angiogenesis is associated with a 32-gene expression signature and 6p21 amplification in aggressive endometrial cancer. *Oncotarget*, 6(12):10634–10645, 2015.
- [71] Jen-Tsan Chi, Zhen Wang, Dimitry S A Nuyten, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med.*, 3(3):e47, 2006.
- [72] Gareth P Elvidge, Louisa Glenny, Rebecca J Appelhoff, et al. Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways. *J. Biol. Chem.*, 281(22):15215–15226, 2006.
- [73] Stuart C Winter, Francesca M Buffa, Priyamal Silva, et al. Relation of a hypoxia meta-gene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res.*, 67(7):3441–3449, 2007.
- [74] Renaud Seigneuric, Maud H W Starman, Glenn Fung, et al. Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother. Oncol.*, 83(3):374–382, 2007.
- [75] Zhiyuan Hu, Cheng Fan, Chad Livasy, et al. A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med.*, 7:9, 2009.
- [76] F M Buffa, A L Harris, C M West, and C J Miller. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br. J. Cancer*, 102(2):428–435, 2010.
- [77] Brita Singers Sørensen, Kasper Toustrup, Michael R Horsman, Jens Overgaard, and Jan Alsner. Identifying pH independent hypoxia induced genes in human squamous cell carcinomas in vitro. *Acta Oncol.*, 49(7):895–905, 2010.
- [78] Joel S Parker, Michael Mullins, Maggie C U Cheang, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.*, 27(8):1160–1167, 2009.
- [79] Laura J. van 't Veer, Hongyue Dai, Marc J. van de Vijver, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871):530–536, 2002.