Ana Xavier Silva de Magalhães

The Transcriptome of the Oncogenic HOXA9 Homeoprotein in Human Glioblastoma: Functional and Clinical Relevance
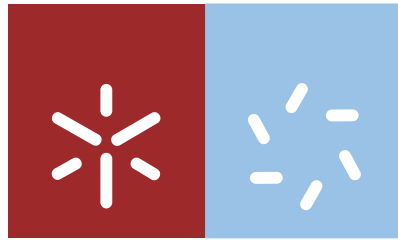
Maio de 2012

Universidade do Minho
Escola de Ciências

Ana Xavier Silva de Magalhães

The Transcriptome of the Oncogenic HOXA9 Homeoprotein in Human Glioblastoma: Functional and Clinical Relevance

Dissertação de Mestrado
Mestrado em Genética Molecular

Trabalho efetuado sob a orientação do
Doutor Bruno Marques Costa
e da
Doutora Maria João Sousa

Maio de 2012

Declaração

Nome: Ana Xavier Silva de Magalhães

Endereço Electrónico: anaxsmagalhaes@ecsaude.uminho.pt

Telefone: (+351)963153268

Número do Bilhete de Identidade: 13348361

Título da Dissertação:

The Transcriptome of the Oncogenic HOXA9 Homeoprotein in Human Glioblastoma: Functional and Clinical Relevance

O Transcriptoma da Homeoproteína Oncogénica HOXA9 no Glioblastoma Humano: Relevância Funcional e Clínica

Orientadores:

Doutor Bruno Marques Costa

Doutora Maria João Sousa

Ano de Conclusão: 2012

Designação do Mestrado: Mestrado em Genética Molecular

Universidade do Minho, ____/____/____

Assinatura: _____

Agradecimentos

Encaro esta secção como uma mera formalidade, pois não serão estas parcas palavras capazes de agradecer convenientemente a todos. Penso que os agradecimentos devem ser feitos na altura devida e, demonstrados diariamente com empenho, entusiasmo e persistência. Mesmo assim, deixo alguns agradecimentos que me parecem mais importante, sem desmerecer a contribuição e apoio de quem eventualmente não esteja aqui.

Começo por agradecer ao Doutor Bruno Costa por me ter recebido no seu grupo de investigação e pela total disponibilidade, incentivo, apoio, ensinamento e confiança transmitidos durante a execução deste trabalho.

À minha supervisora Doutora Maria João Sousa agradeço a disponibilidade e compreensão da necessidade de *timings* tão apertados.

À Marta Pojo quero agradecer a paciência, atenção, disponibilidade, ajuda e incentivo, mas acima de tudo, pelo olhar crítico e por me fazer pensar com clareza mesmo quando tudo parece estar errado. Às restantes meninas do grupo Tatiana, Carina, Joana, Céline, obrigada pela disponibilidade, paciência e apoio, mas também pelos momentos de boa disposição que se alarga a todos os elementos que passaram pelo grupo.

Márcia, as palavras não chegam para te agradecer! Obrigada por me teres indicado a este grupo, por todo o apoio, ajuda e partilha de confidências nos melhores e nos piores momentos, sem esquecer a tua energia e boa disposição contagiantes. Filipe, a discussão científica contigo é sempre frutífera, obrigada por todo o incentivo, disponibilidade e por me ouvires nos momentos de frustração.

Aos restantes elementos do laboratório, em especial à Marta, Dina, Nelma e Ricardo quero agradecer todos os momentos de boa disposição e descontracção. Assim como a todos os NeRDs pela fácil integração e relacionamento.

Sem me alongar em demasia, meus bioquímicos obrigada por tudo! Pelo apoio e amizade, mas acima de tudo pelos momentos de enorme boa disposição e descontracção, tão necessários quando tudo parece andar em sentido contrário!

Summary / Resumo

The Transcriptome of the Oncogenic HOXA9 Homeoprotein in Human Glioblastoma: Functional and Clinical Relevance

Summary

Gliomas are a heterogeneous group of neoplasias that account for the majority of primary tumors of the central nervous system, of which glioblastoma (GBM) is the most common and malignant subtype. These are dramatic diseases for which no curative therapies are yet available. The clinical responses of GBM patients are particularly poor and vary greatly among individuals, especially due to the heterogeneity of their molecular alterations. Regardless, these patients are equally treated with a standardized therapeutic approach, mainly due to the lack of well-established molecular prognostic markers. Recently, the reactivation of *HOXA9* expression in GBM has been implicated as a poor prognostic marker. As HOXA9 is a transcription factor, we hypothesized that a set of HOXA9-transcriptionally regulated genes may be its true biological effectors. In this sense, we aimed to characterize the HOXA9 transcriptome in GBM to identify novel prognostic biomarkers and putative therapeutic targets. By analyzing expression microarrays in *HOXA9*-negative and positive U87MG and hTERT/E6/E7 cells, we found a vast number of genes regulated by HOXA9 that are involved in important hallmarks of cancer as proliferation, invasion, and therapy resistance. Interestingly, we found high expression of the long non-coding RNA *HOTAIR* (*HOX transcript antisense intergenic RNA*) in *HOXA9*-positive GBM cell lines, consistent with a significant co-expression between *HOTAIR* and *HOXA9* in high-grade gliomas, particularly GBMs. Mechanistically, using chromatin immunoprecipitation and quantitative PCR analysis in GBM cell lines, we found that HOXA9 directly interacts with the promoter region of *HOTAIR* and induces its transcription. Importantly, GBM patients with high expression of *HOTAIR* had a relatively shorter overall survival, independently of other putative prognostic factors. Using *in silico* analysis we found other putative direct targets of HOXA9, and ChIP analysis proved the direct regulation of *WNT6* (*Wingless-Type MMTV Integration Site Family Member 6*) by HOXA9. Our study provides the first characterization of HOXA9 target genes in the context of GBM, and identifies new clinically-relevant prognostic biomarkers, which may be new therapeutic targets to treat this aggressive malignancy.

O Transcriptoma da Homeoproteína Oncogénica HOXA9 no Glioblastoma Humano: Relevância Funcional e Clínica

Resumo

Os gliomas são um grupo heterogéneo de neoplasias que constituem a maioria dos tumores primários do sistema nervoso central, dos quais o glioblastoma (GBM) é o simultaneamente o mais comum e o mais maligno. Actualmente não existem ainda terapias eficazes para estas neoplasias altamente agressivas. A resposta clínica dos pacientes de GBM é extremamente insatisfatória e variável entre indivíduos, sobretudo devido à heterogeneidade das alterações moleculares presente nestes tumores. Devido à falta de factores de prognóstico bem estabelecidos, todos os pacientes são tratados com a mesma abordagem terapêutica. Recentemente, a reactivação da expressão do *HOXA9* em GBM foi implicada como um marcador de pior prognóstico. Como o *HOXA9* é um factor de transcrição, colocamos a hipótese de que um grupo de genes regulados pelo HOXA9 poderão ser os seus efetores biológicos. Assim, tivemos como objectivo caracterizar o transcriptoma do HOXA9 em GBM, para identificar novos biomarcadores de prognóstico e potenciais alvos terapêuticos. Pela análise de microarrays de expressão em linhas celulares U87MG e hTERT/E6/E7, com e sem a expressão de *HOXA9*, verificámos que muitos dos genes-alvo do HOXA9 regulam características críticas no cancro, como a proliferação, a invasão e a resistência à terapia. Verificámos ainda uma expressão aumentada do RNA não-codificante *HOTAIR* (*HOX transcript antisense intergenic RNA*) em células de GBM *HOXA9*-positivas, um resultado consistente com a co-expressão entre *HOTAIR* e *HOXA9* em gliomas primários de alto-grau. Usando imunoprecipitação da cromatina (ChIP) e PCR quantitativo em linhas celulares de GBM verificámos que o HOXA9 interage directamente com a região promotora do *HOTAIR*, induzindo a sua expressão. Encontrámos também uma relevante associação entre altos níveis de expressão do *HOTAIR* e uma pior sobrevida dos pacientes de GBM. Por análises *in silico*, encontrámos outros potenciais alvos directos do HOXA9, tendo validado o WNT6 (*Wingless-Type MMTV Integration Site Family Member 6*) por ChIP. Resumindo, o nosso estudo providencia a primeira caracterização dos alvos do HOXA9 em GBM, e identifica novos biomarcadores de prognóstico, que poderão auxiliar a racionalização das decisões terapêuticas, bem como vir a ser testados como novos alvos terapêuticos.

Contents

Contents

Abbreviations List

ABC	ATP-Binding Cassette
ANGPT	Angiopoietin
ASR	Age-Standardized Incidence Rate
BAMBI	Bone Morphogenetic and Activin Membrane-bound inhibitor
BIOBASE	Biological Database
CHI3L1	Chitinase 3-Like 1
ChIP	Chromatin Immunoprecipitation
CNS	Central Nervous System
CpG	Cytosine-phospho-Guanine
CSC	Cancer Stem Cells
CXCL	Chemokine (C-X-C) motif ligand
CYBB	Cytochrome b-245-β
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	Deoxyribonucleic Acid
DOT1L	Disrupter of Telomere-Like 1
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EZH2	Enhancer of Zeste Homolog 2
FC	Fold-Change
GABRA1	Gamma-Aminobutyric Acid A Receptor
GBM	Glioblastoma
GO	Gene Ontology
H3	Histone H3
H3K27	Histone H3 Lysine 27
H3K4	Histone H3 Lysine 4
HDAC	Histone Deacetylase
HGG	High Grade Glioma
hGUS	Human β-Gluconidase
HIF	Hypoxia Inducible Factor

HOTAIR	HOX Transcript Antisense Intergenic RNA
HOX	Class I Homeobox
HSP	Heat-Shock Protein
IACR	Agency for Research on Cancer
ICAM	Intercellular Adhesin molecule
IDH	Isocitrate Dehydrogenase
IgG	Immunoglobulin G
KEGG	Kyoto Encyclopedia of Genes and Genomes
LGG	Low Grade Glioma
lncRNA	Long Non-Coding RNA
LOH	Loss of Heterozygosity
MGMT	O^6-Methylguanine Methyl Transferase
miRNA	Micro RNA
MLL	Mixed Lineage Leukemia Gene
mRNA	Messenger RNA
mTOR	Mammalian Target of Rapamycin
ncRNA	Non-Coding RNA
NDRG	N-Myc Downstream Regulated
NEFL	Neurofilament
NF	Neurofibromin
OS	Overall Survival
PCR	Polymerase Chain Reaction
PDGF	Platelet-Derived Growth Factor
PDGFR	Platelet-Derived Growth Factor Receptor
PI3K	Phosphatidylinositol 3-Kinase
PTEN	Phosphatase and Tensin Homologue
qPCR	Quantitative PCR
RAC	RAS-Related C3 Botulinum Substrate
RB	Retinoblastoma
REMBRANDT	Repository of Molecular Brain Neoplasia Data
RNA	Ribonucleic Acid

RTK	Receptor Tyrosine Kinase
RT-PCR	Reverse-Transcription PCR
SELE	E-Selectin
siRNA	Short interfering RNA
SLC12A5	Solute Carrier Family 12
sncRNA	Small Non-Coding RNA
snoRNA	Small Nucleolar RNA
SYT1	Synaptotagmin 1
TCGA	The Cancer Genome Atlas
TGF	Transforming Growth Factor
tiRNA	Transcription Initiation RNAs
Tm	Melting Temperature
TOX	Thymocyte Selection-Associated High Mobility Group Box Family
TSA	Trichostatin A
WHO	World Health Organization
WNT6	Wingless-Type MMTV Integration Site Family Member 6

1. Introduction

1. Introduction

Cancer remains a major health problem, and according to the International Agency for Research on Cancer (IARC), its global burden has more than doubled in the last 30 years[1]. Cancer is the leading cause of death in economically developed countries and the second on developing countries, as a consequence of population ageing and adoption of cancer-associated behaviors[2]. Despite the progress in treatment strategies (surgery, radiotherapy and chemotherapy), the GLOBOCAN 2008[3] estimated that 12.7 million new cancer cases and 7.6 million cancer deaths have occurred in 2008.

Cancer increasing incidence has stimulated intense research, which has created a huge body of knowledge. Cancer is thought to arise by stepwise genetic alterations, such as DNA sequence changes, copy number aberrations, chromosomal rearrangements, and epigenetic alterations. Together, these result in altered growth and survival properties of cells, and drive the development and progression of the malignancy[4,5]. During the multistep development of human tumors, cells need to attain certain biological abilities, typically termed hallmarks of cancer[6]. Hanahan and Weinberg[6] proposed ten hallmarks which have distinctive and complementary capabilities that enable tumor growth and metastatic dissemination (Figure 1.1), including active invasion and metastasis, sustained proliferative signaling, evasion from growth suppressors, resistance to cell death, replicative immortality, induction of angiogenesis, evasion from immune destruction, genome instability and mutation, deregulation of cellular energetics, and promotion of tumor inflammation[6].

The ultimate cause of cancer is the alteration on the balance of cellular networks and gene expression programs that maintain cellular homeostasis. Analysis of genetic alterations in tumors allowed the identification of several genes responsible for the maintenance of such balance. These genes were divided into oncogenes, which are activated to promote malignancy, and tumor suppressor genes that protect cells against harmful mutations and cellular regulations capable of driving malignant transformation. Several studies have been revealing mutational alterations in genes
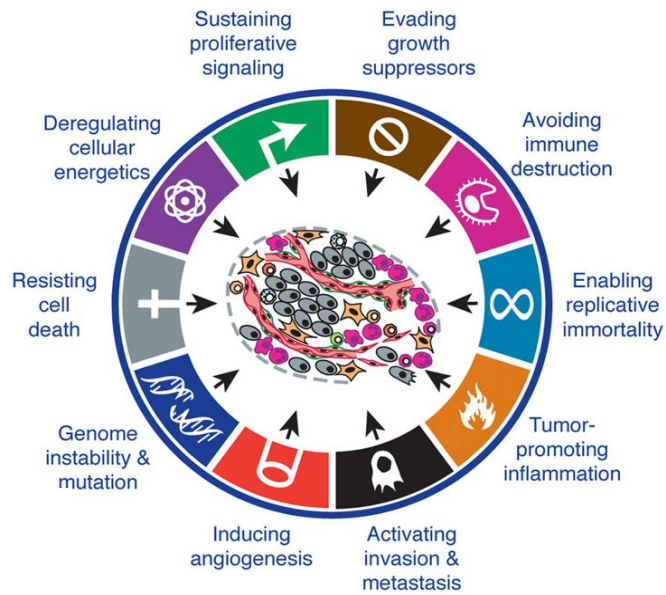
Figure 1.1| The ten hallmarks of cancer proposed by Hanahan and Weinberg. These include sustaining proliferative signaling, evading growth suppressors, resisting cell death, evading immune destruction, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, and reprogramming of energy metabolism. Underlying these hallmarks are genome instability, and inflammation. (Adapted from Hanahan and Weinberg, 2011[6])

that control critical pathways, such as cell survival, growth arrest, DNA damage response, and apoptosis. Recently, was defined a set of twelve core pathways leading to cancer progression, including: apoptosis, DNA damage control, regulation of G1-to-S phase progression, hedgehog signaling, cell adhesion, integrin signaling, c-Jun N-terminal kinase signaling, KRAS signaling, regulation of invasion, small GTPase-dependent signaling, TGF-β signaling and Wnt/Notch signaling[7-10]. Nonetheless, only the complete understanding of cancer as a multi-factorial disease, in which several pathways contribute to various stages of tumorigenesis, will allow a more rational design and use of new therapies.

1.1 General Epidemiology and Classification of Primary Brain Tumors

Tumors of the central nervous system (CNS) comprise a broad variety of entities, which range from benign to highly malignant. According to GLOBOCAN 2008[11],

the world age-standardized incidence rate (ASR; Figure 1.2) of all primary brain and CNS tumors is higher in men than in women (3.8 vs. 3.1 per 100,000 people, respectively), a tendency that is also observed in Europe (6.2 in men vs. 4.6 in women) and in Portugal (7.4 in men vs. 5.4 in women)[12]. These tumors are far more incident in industrialized countries than in developing countries (Figure 1.2), but this is probably a consequence of the discrepancy on diagnostic equipment and appropriate health care. Although primary brain and CNS tumors present low incidence rates, accounting only for approximately 2% of all primary tumors[13], they present high rates of mortality, and rank first for average of years of life lost among all tumor types[14].

Typically, the classification of these tumors is based on their localization within the CNS and the histopathological features they present, which has allowed the classification of a large number of CNS tumors by the World Health Organization (WHO)[15]. Gliomas are the most frequent CNS primary tumors in adults, encompassing a wide group of neoplasias that are classified according to the glial cell of origin or the morphological similarities between tumor and normal glial cells. These tumors consist mainly of three different tissue types[15]: ependymomas derived from ependyma or their precursors represent less than 10% of all gliomas; oligodendrogliomas derived from oligodendrocytes or their precursors account for 10-30% of all gliomas cases; astrocytomas which are derived from astrocytes and are the most common malignancies in CNS, consisting of about 70% of all diagnosed gliomas[16,17]; and oligoastrocytomas, which are a mixed lineage between oligodendrogliomas and astrocytomas, and represent 5-10% of all gliomas. The WHO grading system classifies astrocytomas in four grades of malignancy, spanning from the non-infiltrative and low-grade (WHO grade I) pilocytic astrocytoma to the malignant and highly infiltrative diffuse astrocytoma (WHO grade II), anaplastic astrocytoma (WHO grade III) and glioblastoma (WHO grade IV)[18,19].

Glioblastoma (GBM) is the most frequent and lethal of all primary tumors of the CNS in adults, accounting for more than 50% of all glial tumor types, with a global incidence rate of about 5 per 100,000 people/year[20]. Typically, the incidence peak in adults occurs between 45 and 70 years, and males are more frequently affected than females (male to female ratio of 1.7); the incidence also doubles in Caucasians as
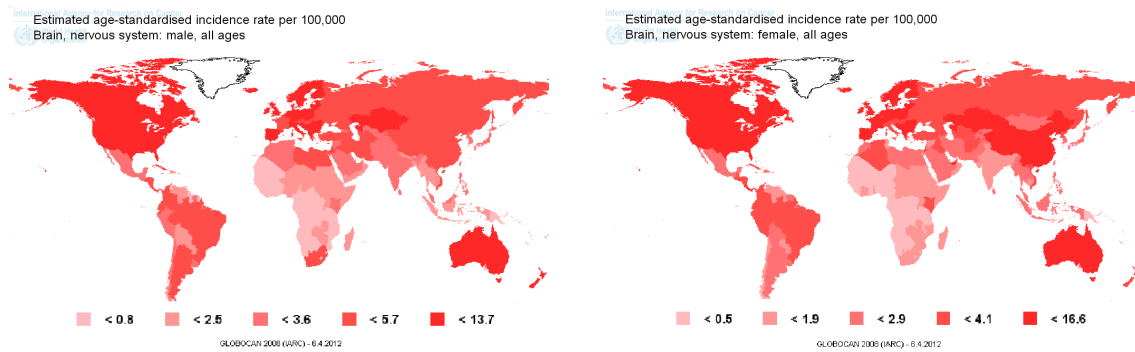
Figure 1.2| Worldwide estimated age-standardized incidence rate per 100,000 person for brain and nervous system tumors in males (left map) and in females (right map)[11].

compared to black people[21]. GBMs are characterized by rapid growth and diffuse invasiveness of the adjacent brain parenchyma, and its histopathological features include cellular polymorphism, mitotic activity, nuclear atypia, vascular thrombosis, microvascular proliferation, and necrosis[22]. From a clinical and biological point of view, GBMs may be divided into primary (or *de novo*) and secondary GBMs. Primary GBM, which is by far the most common subtype, arises as a *de novo* process, in the absence of a pre-existing lower-grade lesion, and manifest rapidly after a short clinical history. In contrast, secondary GBM develops progressively from lower-grade astrocytomas (WHO grades II or III), generally over a period of 5 to 10 years[23]. According to the report of Ohgaki and Kleihues[24], secondary GBM represent only 5% of all cases, and the patients' age distribution in both subtypes is remarkably different: the incidence peak for primary GBMs is 62 years, whereas secondary GBM develop in younger ages (peak at 45 years). Also, primary GBMs are more commonly diagnosed in males than females (male to female ratio 1.33), while secondary GBMs are more frequent in females than in males (male to female ratio 0.65)[23].

The current standard of care for GBM patients is multimodal, consisting of maximum surgical resection, combined with radiation and concomitant and adjuvant chemotherapy with the alkylating agent temozolomide[25]. However, and despite several efforts in the field of clinical neuro-oncology, the treatment remains mostly palliative, with an average survival of 15 months[24,25]. The clinical outcome of GBM patients varies greatly among individuals; nonetheless they are treated with the same standardized procedure, regardless of specific molecular alterations. So, understanding GBM as a result of several alterations, in which several pathways contribute to its

onset and aggressiveness, implies the need of holistic studies to evaluate the role of the different molecular players in gliomagenesis. Such would answer the urgent need to stratify patients in cohorts according to the molecular alterations of each tumor, and will lead to the rationalization of treatment decisions, and possibly to a patient-tailored therapy.

1.2 Molecular Pathology of Glioblastoma

Decades of molecular studies have shown that the stochastic and complex transformation of normal cells into GBMs is similar to the tumorigenesis of other human tissues, as in both cases results from the sequential accumulation of genetic aberrations and the deregulation of signaling pathways (Figure 1.3 and 1.4). The determination of the human genome sequence, complemented with the improvements in bioinformatics technologies, allowed the intense characterization of the molecular alterations promoting GBM formation[26] (Figure 1.4). Similar to other human tumors, the cancer stem cell (CSC) hypothesis is also considered as a possible mechanism for the formation of GBM. This hypothesis postulates that cancer is derived from a small set of stem cells that create a self-sustaining pool, and in fact, the adult nervous system harbors neural stem cells that are able to proliferate, self-renew, and differentiate into different mature cell types[27]. Specifically in the brain, it is believed that neural stem cells may be transformed into cancer stem cells, due to a series of selective mutations, giving rise to GBM[28], and although these cells represent only a minority of the total burden that constitute the tumor, they seem to have crucial roles in allowing tumorigenesis[28]. Interestingly, GBM stem cells are pointed as important contributors to the resistance of brain tumor to chemoradiotherapy[29]. In this sense, the complete elucidation of the differences between normal and CSC may allow the development of therapeutic approaches targeting CSC that may overcome GBM resistance to therapy, but also may spare normal brain cells from the aggressive treatments.
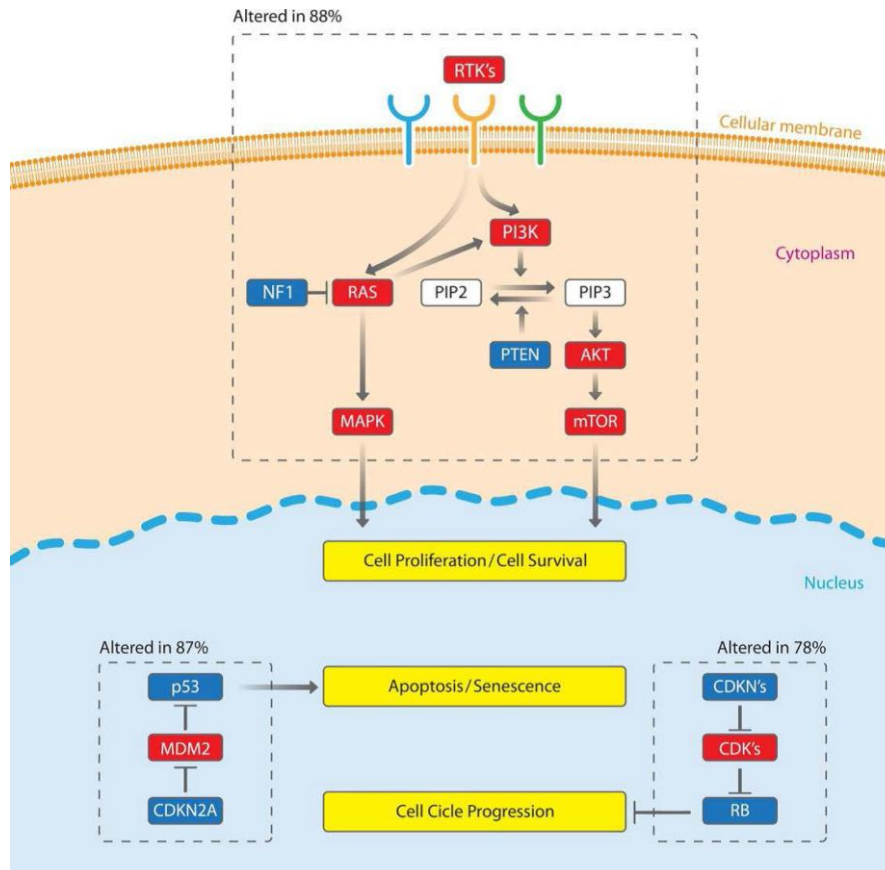
Figure 1.3| Main signaling pathways commonly altered in glioblastoma, and the alteration frequency of each pathway (dashed boxes). Red and blue colors indicate activating and inactivating protein alterations, respectively. (Adapted with permission from Pojo and Costa[30])

Among the most common and characterized molecular alterations in GBM, cytogenetic studies have identified numerous chromosomal regions with copy number alterations. In general, amplification or chromosomal gains are related with oncogenes, as they favor tumor development, while deletions are typically found in tumor suppressor genes as they inhibit tumor formation and/or progression[31]. Loss of heterozygosity (LOH) on chromosome 10 is the most frequent genetic alteration, being present in up to 80% of primary GBMs, often with loss of an entire allele (10p or 10q)[23,32,33]. The three most frequently deleted *loci* in this chromosome include the region codifying for the tumor suppressor gene phosphatase and tensin homolog (*PTEN*), suggesting the presence of this and possibly other crucial tumor suppressor genes[32,33]. Less frequent in primary GBMs, LOH on 22q, 1p and 13q have been reported as altered in 41%, and 12% for both chromosomes 1 and 13, respectively.
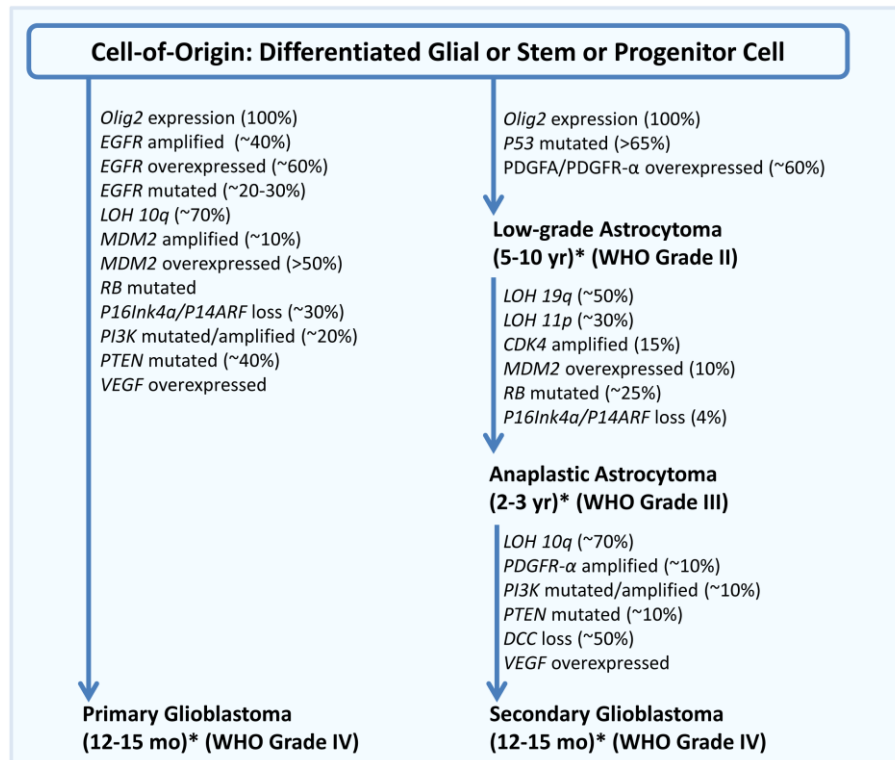
Figure 1.4| Summary of crucial molecular pathways in the development of gliomas. (Adapted from Xavier-Magalhães and Costa[34])

LOH of 13q typically includes the retinoblastoma 1 (*RB1*) locus[35]; although rare, LOH of 1p is associated with patients longer survival[36].

Besides the classic mutations (Figure 1.4), recent comprehensive analysis[8] allowed the discovery of unknown mutations in GBM, and the identification of a highly interconnected network of aberrations in three major pathways: receptor tyrosine kinase (RTK) signaling, retinoblastoma (RB) and p53 tumor suppressor pathways[5]. The hyperactivation of RTKs is a frequent event in human GBMs, which activates several signaling pathways involved in cellular growth and survival, but also in angiogenesis and invasion. In GBM, the overactivation of the RTK pathways may occur by several mechanisms, as gene amplification, receptors mutation resulting in its constitutive activation, or overexpression of both ligands and receptors (Figure 1.3)[37]. Also, *p53* and *RB*, which are involved in cell cycle regulation, are targets of many molecular alterations that induce their inactivation. *p53* and *RB* inactivation renders tumors high sensitivity to uncontrolled proliferation driven by mitogen activated protein kinases (MAPK; Figures 1.3 and 1.4)[38].

Recently, non-protein coding RNAs (ncRNAs) have emerged as new players on the deregulation of signaling pathways and gene expression. For a long time, RNA was delegated to just an intermediate between the DNA responsible for storing the information and the functional protein. However, the discovery that only 2% of the genome encodes protein-coding sequences[39,40], and the great evolution on whole-genome and transcriptome analysis, revealed that 90% of the genome is actively transcribed[41]. The finding that the transcriptome was more complex than first noticed once the number of transcripts do not code for a protein is four times higher than coding sequences, and that these non-protein coding transcripts can regulate gene expression, led to the knowledge of several new forms of RNA[42-45]. These ncRNAs are divided into two groups small: ncRNAs (sncRNAs) and long ncRNAs (lncRNAs)[46]. The sncRNAs are represented by a wide sort of known RNA species that include small interfering RNAs (siRNAs)/ microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), and the more recently described transcription initiation RNAs (tiRNAs), whose functions in the specific regulation of both protein-coding and non-coding genes were first elucidated. Of all sncRNAs, the functions of MiRNAs are the most studied, and are known to be key regulators of several biological processes by negatively controlling gene expression at post-transcriptional level[47]. MiRNAs alterations have been described as involved in the initiation and progression of cancer, once they may act as tumor suppressor or oncogenes depending on their target genes[48]. Interestingly, deregulation of these molecules have been detected in GBM, being involved in cell proliferation, apoptosis, cell cycle regulation, invasion, angiogenesis and glioma stem cell behavior[49]. Moreover, expression profiling of astrocytic gliomas revealed miRNA signatures that are able to differentiate histological subtypes[50]. Therefore, miRNAs represent promising diagnostic and prognostic markers, and eventually new therapeutic targets in GBM[51]. On the other hand, the characterization of the non-coding transcriptome has been revealing that the genome is replete of lncRNAs. These have been emerging as a major class of regulatory molecules in imprinting control, cell differentiation, immune responses, and tumorigenesis[43]. Even though the number of characterized lncRNAs has been increasing, most of their functions remain unknown; nonetheless, it is accepted that lncRNAs have arisen as an important component in the regulation of gene expression, and major roles on formation and progression of cancer

are anticipated[43]. Particularly in GBM, only one study assessing lncRNA expression was reported until now[52]. This report indicates that a large number of lncRNAs are differentially expressed in GBM when compared to normal brain, and implicates two lncRNAs (ASLNC22381 and ASLNC20819) in GBM pathogenesis[52]. However, this report must be carefully interpreted as it represents information of only one patient[52]. Even so, it sheds light into the putative important role of altered expression of lncRNAs in GBM, implying their urgent characterization.

1.3 Molecular Prognostic Factors of Malignant Gliomas

It is widely recognized that the stratification of GBM patients may prove crucial in rationalizing treatment decisions, for which a set of molecular markers predictive of tumor response to specific therapies and/or patient outcome are required. Recently, several biological and molecular features of GBMs, including the methylation status of O^6-methylguanine methyl transferase (*MGMT*) gene[53], isocitrate dehydrogenase (IDH) mutations[8] or class I homeobox (*HOX*) A genes expression[54], have been pointed as putative prognostic biomarkers[55]. The methylation status of *MGMT* gene is currently the most promising, although it has not yet reached clinical applicability[53,56]. *MGMT* encodes a DNA-repair protein that removes alkyl groups from the O^6 position of guanine, an important site for DNA alkylation. When DNA is left unrepaired, the lesions induced by chemotherapy trigger apoptosis and cytotoxicity[57,58]. Epigenetic silencing of *MGMT* by promoter methylation leads to the loss of *MGMT* expression and reduced DNA-repair activity, which is associated with longer overall survival in patients with GBM[59]. Patients whose *MGMT* promoter is methylated and are treated with temozolomide have an increased overall survival (median of 21.7 months), as well as a higher 2-year survival rate (46%), in comparison to patients treated with temozolomide but with unmethylated *MGMT* promoter (median survival of 12.7 months and 2-year survival of 13.8%)[58]. Thus, *MGMT* promoter methylation represents an independent and favorable predictive factor to patients' response to therapy[53].

Other important prognostic factors have been revealed by a recent genomic study, and concern the presence of mutations in *IDH1* and *2* genes (*IDH* when referring

to both)[8]. *IDH* mutations are correlated with younger age at diagnosis, and with longer survival when compared to patients with *IDH*[wt] genes, independently of age and gender[60]. As mutations in *IDH* are limited to a single amino acid, their detection for diagnostic purpose should be direct.

Strikingly, mutations in *IDH1* have been included in a GBM signature that allowed the division of GBMs into subtypes according to their recurrent genomic alterations[61]. In Verhaak and co-workers report[61], GBMs were divided in four subtypes - Classical, Mesenchymal, Proneural and Neural - each displaying different underlying genetic alterations. The identity of the Classical subtype was defined by displaying the most common genomic aberrations of GBM, with 93% of samples displaying chromosome 7 amplifications and chromosome 10 deletions, 95% showing *EGFR* amplification, and 95% with homozygous deletion on the *Ink4a/ARF* locus[61]. The Mesenchymal subtype was mainly characterized by the high expression of chitinase 3-like 1 (*CHI3L1* or *YKL-40*) and *MET*[55], and also neurofibromin 1 (*NF1*) mutation or deletion were found to be important features of this class[61]. The two major features of Proneural subtype include the amplification of *PDGFRA* and *IDH1* mutation, besides frequently presenting LOH and mutations at *TP53*. Importantly, Proneural subtype was associated with younger age and longer survival[61]. The Neural subtype was typified by the differential expression of certain genes, in this case neuron markers as gamma-aminobutyric acid A receptor (*GABRA1*), solute carrier family 12 (*SLC12A5*), neurofilament (*NEFL*) and synaptotagmin I (*SYT1*)[61].

The importance in the division of GBM into subtypes lies on the possible application of different therapeutic approaches, as treatment efficacy differs per subtype[61]. Aggressive therapy significantly reduced mortality in Classical and Mesenchymal subtypes, and a tendency to better outcome was observed for the Neural subtype, yet patients whose GBM present Proneural features do not benefit from highly aggressive therapies[61]. In this sense, some of the genetic events underlying the different GBM subtypes could become part of the clinical routine to rationalize therapeutic decisions, and ultimately lead to a personalized therapy for groups of patients with GBM. Even though *MGMT* promoter methylation status or *IDH* mutations, among other prognostic biomarkers, are well described their clinical

Table 1.1| Application of molecular markers in GBM clinical practice. (Adapted from Jansen *et al.*, 2010[62])

	Laboratory testing	Research testing for clinical trials	Clinical testing for patients
MGMT **methylation status**	No	Ready	Uncertain
EGFR pathway assessment	Uncertain	Ready	No
IDH **mutation testing**	Ready	Uncertain	Ready

application is still limited to a small number (Table 1.1)[62]. Therefore, there is the urgent need to find new and more robust prognostic biomarkers.

Recently, the aberrant expression of *HOX* genes have been implicated in several tumors; specifically in GBM, *HOXA9* has been pointed as a putative biomarker of poor patient prognosis[63]. *HOX* genes encode crucial transcription factors during embryonic development that, in humans, are gathered in four clusters (A-D) located on different chromosomes[64]. During embryonic development, *HOX* genes are sequentially expressed from 3' to 5' along the anterior-posterior (A-P) axis contributing to the temporospatial development of limbs and organs[65]. The mechanisms underlying *HOX* genes control in normal development occur according to three main principles: spatial collinearity, posterior prevalence, and temporal collinearity[66]. These were found to be altered in cancer as a consequence of three major mechanisms proposed by Abate-Shen[67]: temporospatial deregulation, gene dominance and epigenetic regulation. Different groups have been reporting the deregulation of these mechanisms in different *HOX* genes, and in different tumors. For instance, Takahashi *et al.*[68] evaluated all *HOX* genes in primary esophageal squamous cell carcinoma, reporting that normal esophagus expressed more 3' *HOX* genes than 5' *HOX* genes, a temporospatial pattern that was completely reversed in tumor samples. The dominance of *HOX* gene expression in tumor samples in comparison to normal tissue is evident in acute myeloid leukemia cells, where *HOXA9* is overexpressed when compared to normal myeloid cells. Strikingly, this differential expression is correlated with patients' poor prognosis and treatment failure[69]. Importantly, *HOX* gene expression is commonly controlled by epigenetic mechanisms, as the cytosine-phospho-guanine (CpG) islands in the promoters of silenced *HOX* genes are frequently methylated[70]. Concerning epigenetic regulation, it is known that the polycomb group proteins and the trithorax

proteins drive histone methylation, resulting in lysine 27 in histone 3 (H3K27) and lysine 4 in histone 3 (H3K4) trimethylation, respectively. The mixed lineage leukemia (*MLL*) gene is a trithorax homologue that is mutated in some leukemias, usually presenting as a fusion protein, that does not present H3K4 methylation function[71]. Instead, mutated MLL recruits the disrupter of telomere-like 1 (DOT1L) methyltransferase resulting in an altered methylation pattern that, in lymphoblastic leukemia, induces the aberrant expression of *HOXA4*, *HOXA5*, *HOXA7* and *HOXA9* genes, which correlate with worse prognosis[69].

The aberrant expression of *HOX* genes have been reported as crucial in several hallmarks of cancer, including increased proliferation, angiogenesis and invasion, and apoptosis resistance in leukemia and in several solid tumors[66,72-75]. Interestingly, in recent years, *HOX* genes aberrant expression has been implicated in gliomagenesis. Abdel-Fattah and co-workers[76] evaluated the expression of all *HOX* genes in primary astrocytomas and in non-tumor brain specimens, reporting that some *HOX* genes are abnormally expressed in malignant astrocytomas. A subsequent report by Murat *et al.*[54] identified a *HOX*-dominated gene cluster, suggestive of a signature that display self-renewal properties. These authors show that the expression of *HOXA10* gene in GBM neurospheres is consistent with a role of HOX genes in gliomas stem-like cell compartments[54]. Interestingly, the *HOX*-dominated gene signature arises along malignant progression to GBM, and is an independent predictive factor of chemo-radiotherapy resistance in patients[54]. Later, Costa and co-workers[63] showed that *HOXA* genes are predominantly activated in GBM, as compared to lower-grade gliomas and normal brain tissue, suggesting they may be a useful component of a molecular classification of gliomas. By analyzing expression microarrays data from 100 GBMs, they identified tumors with abnormal chromosomal domains of transcriptional activation, which comprise the *HOXA* cluster[63]. Pharmacological manipulation of GBM cell lines and neurospheres revealed that the expression of this *HOXA* domain was regulated by the PI3K pathway, through reversible regulation of histone modifications mediated by enhancer of zeste homologue 2 (EZH2), independently of the mammalian target of rapamycin (mTOR) activity[63]. Of all *HOXA* genes, *HOXA9* expression was predictive of worse outcome, independently of other prognostic factors[63]. The retroviral induction of *HOXA9* in GBM cell lines and immortalized human astrocytes

revealed pro-proliferative and anti-apoptotic functions, which may explain the unfavorable prognosis of GBM patients with *HOXA9* reactivation[63]. More recently, Gaspar *et al.*[77] shown a pediatric cell line that is resistant to temozolomide, although presenting *MGMT* promoter methylation, and intact mismatch repair and double-strand break repair systems[77]. Interestingly, they found that the resistant cell lines present the coordinated expression of several *HOX* genes, of which *HOXA9* and *HOXA10* were highlighted as crucial effectors in temozolomide resistance[77]. In line with Costa *et al*[63] report, Gaspar suggested that the *HOX*-enriched signature is regulated by the PI3K pathway, and interestingly, is associated with resistance to temozolomide in pediatric GBM cell lines[77]. Moreover, pediatric patients with high-grade gliomas that express *HOXA9* and *HOXA10* presented shorter survival[77].

These reports identify *HOXA9* overexpression as a marker of poor prognosis in GBM. Those allied with other studies correlating *HOXA9* expression with worse outcome in acute myelogenous leukemia patients[69], shed light into the importance of *HOXA9* reactivation in tumorigenesis. However, the mechanisms supporting the higher aggressiveness induced by HOXA9 are mostly unknown[69]. Nonetheless, it is reasonable to hypothesize that a set of genes transcriptionally regulated by HOXA9 are the true effectors of its biological functions, by affecting crucial features of cellular malignancy, as proliferation, invasion and apoptosis. In this sense, the understanding of the transcriptome of HOXA9, and the genes directly regulated by it may aid in the understanding of the mechanisms by which HOXA9 exerts its functions, and may prove crucial in identifying new molecular prognostic and therapeutic targets.

1.4 Objectives

GBM is the most malignant and common type of tumor in the adult CNS. The clinical responses of GBM patients to therapeutics are poor and vary greatly among individuals, especially due to the lack of well-established molecular prognostic markers, which could allow patient-tailored therapy. Recently, the reactivation of *HOXA9* was shown to be more frequent in high-grade rather than in low-grade gliomas, and correlated with patients' worse prognosis. Although the importance of HOXA9 in normal development, differentiation, and oncogenesis is well-recognized, little is known about the targets its transcriptome in GBM, or the genes directly regulated by it.

In this sense, we intend to:

- Characterize the full HOXA9-transcriptome in GBM;

- Identify HOXA9-target genes in the context of GBM cells and putative precursors;

- Evaluate the prognostic value of the newly identified HOXA9-target genes in the prognosis of GBM patients.

2. Materials and Methods

2. Materials and Methods

2.1 Cell lines and culture conditions

GBM cell lines A172 and U87MG, and immortalized human astrocytes (hTERT/E6/E7) were cultured in DMEM (Gibco®) supplemented with 10% Fetal Bovine Serum (Biochrom), and 1% Penicillin-Streptomycin (Invitrogen). The A172 cell line endogenously expresses *HOXA9*, whereas U87MG and hTERT/E6/E7 cells (which do not express *HOXA9* endogenously) were previously[63] genetically engineered with murine stem cell (MSCV) retroviral vectors containing the *HOXA9* coding region to obtain *HOXA9* overexpressing cells (U87MG-HOXA9 and hTERT/E6/E7-HOXA9) or with control empty vector (U87MG-MSCV and hTERT/E6/E7-MSCV, respectively). Selection of retrovirally infected cells was maintained with 500ng/µL G418 (Sigma-Aldrich®).

2.2 RNA extraction and cDNA synthesis

Total RNA was extracted from cell lines A172, U87MG-HOXA9, U87MG-MSCV, hTERT/E6/E7-HOXA9, and hTERT/E6/E7-MSCV using the TRIzoL method (Invitrogen), according to the producer protocol. In brief, cells were lysed with TRIzoL reagent and RNA was separated from DNA and proteins by chloroform-based phase separation, followed by RNA precipitation with isopropyl alcohol. The resulting pellet was washed with 75% ethanol, air-dried and dissolved in RNase-free water. RNA quantification was performed using NanoDrop™ ND-1000 spectrophotometer, using 1.5 µl of sample. cDNA was synthesized from the resulting RNA with the RT-Phusion Kit (Thermo Scientific), using random hexameres.

2.3 Microarray Validation and Interpretation

Previously, expression microarrays (Agilent, 44K, Human Whole Genome) were performed in the cell lines overexpressing *HOXA9* (U87MG-HOXA9 and hTERT/E6/E7-HOXA9) and in their *HOXA9*-negative counterparts (U87MG-MSCV and hTERT/E6/E7-MSCV), respectively.

2.3.1 Microarray Data Validation

Regarding higher log fold-changes from the microarray data, two genes among those most over and underexpressed for each cell line were selected, and reverse transcription-PCR (RT-PCR) was performed for both U87MG-HOXA9 and hTERT/E6/E7-HOXA9, and in the HOXA9-negative U87MG-MSCV and hTERT/E6/E7-MSCV. The *HOXA9* and housekeeping β-gluconidase (*hGUS)* gene expression levels were assessed, to ensure the overexpression of *HOXA9* in transfected cells, and to guarantee the equal quantity of cDNA in all reactions. PCR products were visualized on ethidium bromide-stained agarose gel. The sets of primers, specific melting temperatures (Tm), and PCR parameters for all tested genes are described in Annex I.

2.3.2 Prognostic Value of Differentially Expressed Genes

The prognostic value of the 5 genes more differentially expressed due to HOXA9 overexpression, were evaluated on the publicly available Repository of Molecular Brain Neoplasia Data (REMBRANDT)[78] platform. Microarray and clinical data were collected on October, 2011 through the online repository (https://caintegrator.nci.nih.gov/rembrandt/). The minimum cut-off for gene differential expression was established on 3 folds higher/lower than the corresponding non-tumor tissue, and the analysis were performed only in GBM patients.

2.3.3 Biological Enrichment

The Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics database allows the identification of groups of genes that share common biological functions or integrate the same pathways. DAVID was used to identify enriched biological terms on the microarray data, to cluster genes according to their annotated function, and display them in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. For these analysis, the differentially expressed genes ($p<0.05$) from the microarray data of both U87MG and hTERT/E6/E7 cell lines were used.

2.3.4 Biological Process Clustering

Biological Database (BIOBASE) software was used to perform the clustering of the differentially expressed genes of the GBM cell line U87MG-HOXA9, versus its negative counterparts, according to their gene ontology (GO) annotated biological process. The Biobase ExPlain™ software[79] was used to understand which of the differentially expressed genes of both U87MG and hTERT/E6/E7 were putative direct-targets of HOXA9. In vitro binding studies have shown that the 60-amino acids homeodomain, common to all HOX proteins, binds to AT-rich DNA sequences, particularly to a short TAAT sequence[80]. Such enabled the use of positional weight matrices for the *in silico* genome-wide search of genes that have an HOX binding site on its promoter region[80].

2.3.5 Connectivity Map Analysis

To search for drug-induced gene expression signatures similar to our microarray data, the Connectivity Map tool[81] was used. This database contains the genome-wide mRNA expression data of 453 individual instances (each instance concerns one treatment and one vehicle pair), and allows the identification of the

perturbagens (drugs, etc) responsible for a gene expression profile and its comparison against a given set of genes. The top 20 of the differentially expressed genes upregulated in U87MG-HOXA9 against its negative counterpart U87MG-MSCV, were used.

2.4 *HOTAIR* and *WNT6* Expression in Glioblastoma Cell Lines

RNA extraction and cDNA synthesis from GBM cell lines A172, U87MG-HOXA9, U87MG-MSCV and for immortalized human astrocytes hTERT/E6/E7/-HOXA9 and hTERT/E6/E7/-MSCV were performed according to the procedure stated on 2.2. *HOX* transcript antisense intergenic RNA (*HOTAIR*), wingless-type MMTV integration site family member 6 (*WNT6*) and *HOXA9* levels were evaluated by reverse transcription-PCR (RT-PCR) and by quantitative real-time PCR (qPCR) (Maxima SYBR Green, Fermentas). *HOTAIR*, *WNT6* and *HOXA9* levels were normalized to *hGUS*. The set of primers and conditions used are described in Annex I.

2.5 *HOTAIR* Expression in Oncomine and REMBRANDT

Oncomine platform is a publicly available microarray database that presents the compiled and analyzed information of gene expression profiles of human cancer samples[82]. In Oncomine[82], the categorization of *HOTAIR*-positive and *HOTAIR*-negative glioma patients was based on the Log_2 median-centered intensity values of *HOTAIR* probe. Log_2 median-centered intensity values >0 correspond to high *HOTAIR* expression, and Log_2 values ≤0 correspond to low/negative *HOTAIR* expression.

In REMBRANDT[78], the cut-off for *HOTAIR* upregulation was established on 4 folds higher than *HOTAIR* expression in non-tumor samples. For both databases, data was collected on February, 2012 at the online repository (for Oncomine https://www.oncomine.org/resource/main.html, and for REMBRANDT as stated on 2.3.2).

2.6 Genes Co-Expression Analyses in Oncomine

The human cancer microarray database Oncomine[83] was used to search for genes co-expressed with *HOTAIR* in 394 high grade gliomas (HGG; 296 primary WHO grade IV, and 98 WHO grade III) and 45 WHO grade II low grade gliomas (LGGs). This co-expression was also analyzed in the datasets of other cancer types (lung, breast, leukemia, and colorectal cancer) from TCGA available on Oncomine[83]. All datasets lodged on Oncomine were Log$_2$-transformed, median centered per array, and the standard deviation normalized to one per array[82].

2.7 Chromatin-Immunoprecipitation

ChIP experiments were done as previously described[84]. In brief, A172 cells were cross-linked with 1.42% formaldehyde for 15 minutes, followed by quenching with 125mM glycine for 5 minutes. Cells were lysed with immunoprecipitation buffer (150mM NaCl, 50mM Tris-HCl, 5mM EDTA, 0.5% NP-40, 1% Triton X-100) and chromatin was sheared by sonication (Sonics Vibra Cell VC70T, 21 cycles for 15 seconds) to obtain DNA fragments of approximately 0.5-1 kb. The volume of sheared chromatin equivalent to 2 million cells was incubated with the required antibody in an ultrasonic bath for 15 minutes followed by incubation with protein A-agarose beads (Amersham) and Chelex 100 (Bio-Rad). The following antibodies were used to immunoprecipitate chromatin: 4 µg anti-HOXA9 (Santa Cruz), 2 µg anti-Histone H3 (H3; Abcam), 3 µg anti-Immunoglobulin G (IgG; Sigma). DNA amplification was done by qPCR (Maxima SYBR Green, Fermentas). To ensure that the immunoprecipitations performed with anti-HOXA9 antibody were in fact retrieving HOXA9 direct targets, we used 3 different sets of primers to amplify 3 regions of the E-Selectin (SELE) gene promoter region that were previously described as HOXA9-binding sites[85]. Primer sets 1 and 2 were described as HOXA9-binding sites, while the third set of primers was used as negative control of HOXA9 binding[85]. Anti-histone H3 and anti-IgG were used as ChIP positive and negative controls, respectively. The input represents a control for the amount of DNA used in precipitations. The levels of *HOTAIR*, *WNT6* and *SELE*, were

calculated for each experiment using the $\Delta\Delta C_t$ method as described previously[86]. Three biological replicates were tested, and each qPCR experiment was done in triplicate. The PCR parameters and primers used were as described in Annex I.

2.8 Statistical Analyses

The qPCR differences in ChIP experiments were calculated by the Student's t test using Prism GraphPad software (version 5.0a). In Oncomine[82], each gene was evaluated for differential expression using Student's t-test in the case of two-class analyses (e.g. tumor tissue versus respective normal tissue); for multiclass analyses (e.g. grade II, III, and IV gliomas) Pearson's correlation was used. Both tests were performed using the *R* statistical computing package, and *p*-values were corrected for multiple hypothesis testing using the false discovery rate method[82]. For the co-expression analysis, each dataset was filtered in order to comprise the top 50% of the most variable genes. The co-expression results were sorted by node correlation[82]. Correlation values higher than 0.3 were considered significant. To evaluate the effect of *HOTAIR* expression in the overall survival of GBM patients, Kaplan-Meier survival curves were built using SPPS 19.0 software (SPSS, Inc.). Univariate survival analyses to assess the prognostic value of *HOTAIR* and of other clinicopathological features (patient age, gender, Karnofsky performance status, and institution where the patients were treated) were performed by the Log-rank test, whenever data was available in sufficient number (at least >50% of the samples). Additionally, the independent prognostic value of *HOTAIR* was further analyzed by a multivariate Cox proportional hazard model adjusted for those potential confounding variables. All statistical tests were 2-sided, and significance was considered when *p*<0.05.

3. Results

3. Results

3.1 Microarray Data Interpretation

Previously, the GBM cell line U87MG and the putative GBM precursors human immortalized astrocytes hTERT/E6/E7 were retrovirally infected to overexpress *HOXA9*[63]. Expression microarrays have been performed (Agilent, 44K, Human Whole Genome) in basal conditions for the human GBM cell lines U87MG-HOXA9 and U87MG-MSCV (empty vector), and for immortalized human astrocytes hTERT/E6/E7-HOXA9 and hTERT/E6/E7-MSCV (empty vector). When comparing the resulting microarray data between matched *HOXA9*-overexpressing and control cell lines, a total of 3454 transcripts (top 50 listed on Supplementary Tables 2 and 3, Annex II) were differentially expressed (p-value<0.05) in U87MG cell line, consisting of 1537 transcripts upregulated and 1971 downregulated (Table 3.1); for hTERT/E6/E7, 417 differentially-expressed transcripts were found (top 50 listed on Supplementary Tables 4 and 5, Annex II), comprising 166 probes upregulated and 251 downregulated (Table 3.1). Crossing data from U87MG and hTERT/E6/E7cell lines, a total of 57 transcripts were differentially expressed in both cellular contexts, consisting of 12 transcripts upregulated and 45 downregulated (Table 3.1).

Table 3.1| Genome-wide characterization of the HOXA9 transcriptome in U87MG glioblastoma cells and in hTERT/E6/E7 human immortalized astrocytes.

	Number of Transcripts Differentially Expressed		
	Total	**Upregulated**	**Downregulated**
U87MG	3454	1537	1917
hTERT/E6/E7	417	166	251
Common to Both Cell Lines	57	12	45

3.2 Microarray Data Validation

Using the information provided on the first analysis of the microarray data, a subset of the most differentially expressed genes was selected to validate the microarray results. Genes with high log fold-changes in the array data were selected to be validated by conventional RT-PCR analyses (Table 3.2): for the U87MG cell line, the

Table 3.2| Log fold-change and *p*-values of the subset of differentially expressed genes selected to validate microarray data from both U87MG and hTERT/E6/E7 cell lines.

		Log Fold-Change	*p*-value
U87MG	**Upregulated**		
	ICAM2	3.65	5.69^{7}
	BAMBI	3.38	1.04^{6}
	Downregulated		
	ANGPT2	-6.41	1.57^{11}
	PDGFRB	-3.45	1.79^{7}
hTERT/E6/E7	**Upregulated**		
	RAC2	2.94	1.26^{6}
	CXCL1	4.5	1.43^{5}
	Downregulated		
	NDRG1	-3.59	9.06^{5}
	TOX2	-3.49	1.56^{8}

overexpressed intercellular adhesion molecule 2 (*ICAM2*) and bone morphogenetic and activin membrane-bound inhibitor (*BAMBI*), and the underexpressed angiopoietin 2 (*ANGPT2*) and platelet-derived growth factor receptor-β (*PDGFRB*) genes were selected (Table 3.2). Also, for hTERT/E6/E7 cell line, two overexpressed genes - RAS-related C3 botulinum substrate 2 (*RAC2*) and chemokine (C-X-C motif) ligand 1 (*CXCL1*) - were selected, as well as the downregulated N-myc downstream regulated 1 (*NDRG1*) and thymocyte selection-associated high mobility group box family member 2 (*TOX2*) (Table 3.2). For all genes tested and for each cell line (Figure 3.1), the RT-PCR analyses successfully validated the microarrays data, as obvious differential expression is observed in the RT-PCR results for the transcripts differentially expressed in the array data.

3.3 Prognostic Value of Differentially Expressed Genes

As *HOXA9* overexpression in GBM patients has already been associated with shorter survival[63], we performed overall survival (OS) analyses to the 5 most differentially expressed genes due to *HOXA9* overexpression, to understand if they present prognostic value. The expression and clinical information from the Repository
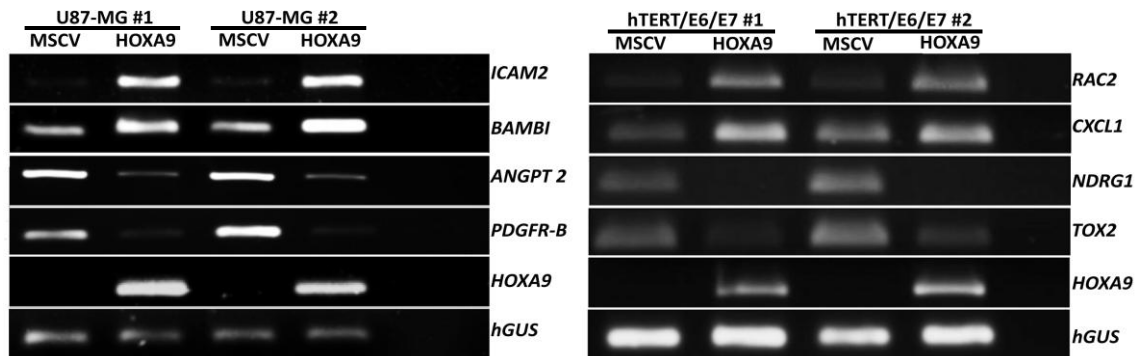
Figure 3.1| Microarray validation of U87MG-HOXA9 (left panel) and hTERT/E6/E7-HOXA9 (right panel) against its negative counterparts U87MG-MSCV and hTERT/E6/E7-MSCV, respectively. Two subsets of highly differentially expressed genes (both up and downregulated) were used to validate microarray data of both cell lines. The upregulated genes on the microarrays results show a higher expression by RT-PCR in both U87MG-HOXA9 (left panel) and hTERT/E6/E7-HOXA9 (right panel) cells as compared to their HOXA9-negative (MSCV) counterparts. The consistency between the genes expression comparing microarray data and PCR results allows the validation of microarray data.

of Molecular Brain Neoplasia Data (REMBRANDT) was used to complete these analyses. We found that GBM patients that have high expression of *NDRG1* present a statistically significant shorter survival when compared to those presenting low expression (Figure 3.2; *p*=0.009). The expression of *ICAM2*, *BAMBI*, *RAC2*, *CXCL1*, *XIST* and *TOX2* did not present statistically significant associations with survival in GBM patients (Annex III). For the remaining genes (*TOX2*, *ANGPT2*, *SDK2*, *SERPINB2*, *NPR3*, *RPL39L*, *MST4*, *C10orf35*, *CXCR7*, *C10orf10*, *FERL1L4*, *OLFML2A*, *FN1*), the number of patients presenting their expression was reduced, so no OS analyses were performed.

In fact the association of *NDRG1* expression with GBM patient survival presents interesting, as it has been associated with cellular growth, differentiation, tumorigenesis, metastasis and poor clinical outcome in several tumor types[87]. This gene warrants further analysis in GBM, as it might represent a potential prognostic biomarker.

3.4 Microarray Data Bioinformatics Analysis

In order to better understand the biological relevance and significant connections between the vast number of differentially expressed genes, several bioinformatics analysis were performed.
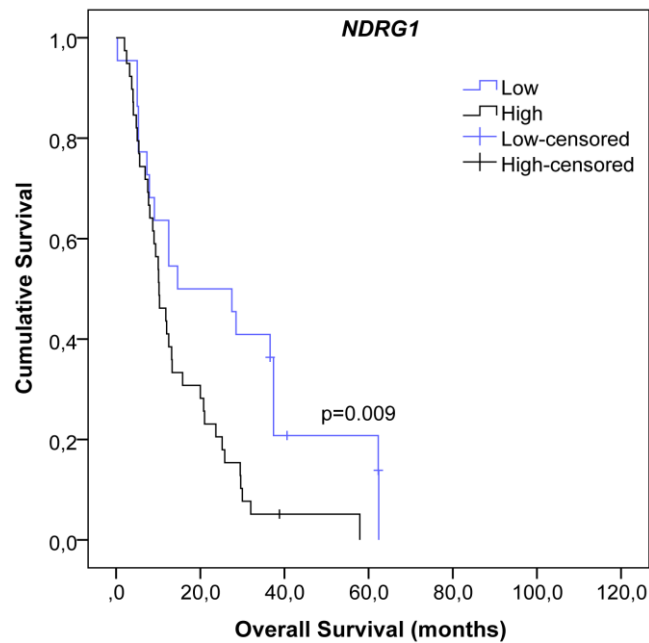
Figure 3.2| Kaplan-Meier overall survival curve of patients from REMBRANDT[78] dataset with high (n=39) and low (n=17) expression of *NDRG1*. GBM patient with higher expression of NDRG1 present a statistically significant shorter survival when compared to low NDRG1 expressing patients (*p*=0.009).

3.4.1 Biological Enrichment Analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics resource consists of an integrated biological knowledgebase and analytic tools that allow the systematical extraction of biological meaning from large gene lists[88]. The total list of the differentially expressed genes (up and downregulated) from the microarray data of both U87MG and hTERT/E6/E7 were analyzed by performing functional clustering annotation, and integrated in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways to facilitate biological interpretation in a network context (Figure 3.3).

In U87MG cells, genes upregulated by HOXA9 were enriched for several cancer-related pathways (Figure 3.3), including "pathways in cancer" and pathways related with different tumor types (e.g., small cell lung cancer and bladder cancer). Interestingly, "pathways in cancer" was found to be the most enriched pathway in U87MG-HOXA9 cells as compared to their HOXA9-negative counterparts (Figure 3.3),

Figure 3.3| DAVID functional annotation clustering analysis of differentially expressed genes due to *HOXA9* overexpression in U87MG and hTERT/E6/E7 cells, revealing cell line-specific enriched pathways.

and other important cancer hallmarks, such as the enrichment of genes involved in pathways that may drive proliferation and invasion. Moreover, genes overexpressed by HOXA9 in U87MG cells were also enriched for DNA repair pathways, which have been associated with drug resistance in some tumor types[89,90] (Figure 3.3). In agreement with an upregulated HOXA9-mediated genetic signature associated with important cancer hallmarks, genes downregulated by HOXA9 in U87MG are involved in "focal adhesion", "cell adhesion molecules" and "antigen processing and presentation" (Figure 3.3). Several of these pathways are related with crucial hallmarks of cancer[6], which may lead to increased malignancy or aggressiveness of tumor cells.

On the other hand, the *HOXA9* transcriptome in non-tumoral hTERT/E6/E7 human immortalized astrocytes cells reveals pathways that are mostly unrelated to cancer; for example, the "chemokine signaling" pathway is enriched in HOXA9-upregulated genes, and the "cytokine-cytokine receptor interaction" pathway is enriched in HOXA9-downregulated genes. These pathways are mainly involved in immune response[91], but in cancer their deregulation has been linked with increased proliferation and invasion, among others[92,93]. The fact that HOXA9-induced gene expression is so different in the U87MG GBM cell line and in the putative GBM precursors hTERT/E6/E7 immortalized human astrocytes, suggests that its transcriptome is cell-type dependent, which is supported by previous reports on the literature[94].

3.4.2 Biological Process Clustering Analysis

While the functional annotation clustering shed light into the importance of HOXA9-target genes in affecting crucial hallmarks of cancer, in order to better understand the biological significance of functionally-related genes, together as a unit, we performed biological process clustering analyses. Here, the GBM cell line U87MG HOXA9-mediated overexpressed and underexpressed genes were analyzed using gene ontology (GO) annotation terms and the Biological Database (Biobase) software.

In line with the results obtained in DAVID bioinformatics (Figure 3.3), GO analysis revealed that HOXA9-target genes are involved in some of the most important

hallmarks of cancer development and progression. For example, the HOXA9-upregulated genes were involved in "cell cycle progression", "DNA replication", "cell division", "RNA processing", and "cellular biosynthetic processes" (Table 3.3), supporting Costa and co-workers[63] report of higher proliferation in *HOXA9*-positive cells in comparison with *HOXA9*-negative cells. Additionally, deregulation of cellular energetics, a recently-established cancer hallmark[6] was also found as indicated by an enrichment of genes involved in "cellular metabolic processes" (Table 3.2). Equally interesting, genes involved in pathways related with drug-resistance were again identified (HOXA9-mediated upregulation of genes involved in "response to DNA stimulus", "DNA repair", "nucleotide excision repair" and "recombinatorial repair"; Table 3.3).

Conversely, genes downregulated by HOXA9 in U87MG cells were involved in other biological processes relevant in the context of cancer, including "antigen processing and presentation" (an important hallmark of cancer as it allows cells to avoid immune destruction, and thus represents a mechanism of resisting cell death[6]) and "extracellular matrix organization" and "cell adhesion" (relevant to allow tumor cell invasion and migration; Table 3.3). Of note, these are crucial features of GBM, as this tumor is highly invasive of the adjacent brain parenchyma[22]. While the overall results seem to support the importance of HOXA9-mediated transcriptome in favoring tumor progression and/or aggressiveness, some features that do not easily fit in that hypothesis were also found on this analysis, as is the case of the HOXA9-mediated downregulation of genes involved in "vasculature development" (Table 3.3).

Taken together, our microarray results are a strong indication that *HOXA9*-overexpression in GBM cell lines renders them a more aggressive phenotype, implying HOXA9-target genes in GBM cells as responsible for patients' worse prognosis.

3.4.3 Connectivity Map Analyses

The important occurrence of HOXA9-mediated upregulated groups of genes involved in cancer drug-resistance (Figure 3.3 and Table 3.3), and the fact that GBM is described as highly resistant to therapy, prompted us to search the Connectivity Map

Table 3.3| Biological function clustering performed using gene ontology terms for differentially expressed genes of U87MG-HOXA9 cells compared to U87MG-MSCV cells. (Redundant instances were eliminated and the most inclusive class was considered; terms are organized according to higher *p*-values, and only statistically significant classes (*p*-value <0.05) are represented.)

	GO Term	No. Hits in Group	p-Value
U87MG HOXA9-Upregulated Genes	ncRNA metabolic process	49	4.123^{-16}
	Ribonucleoprotein complex biogenesis	42	2.421^{-15}
	Cell cycle	111	3.917^{-15}
	DNA replication	46	9.807^{-15}
	DNA metabolic process	77	9.789^{-14}
	Response to DNA damage stimulus	51	6.478^{-9}
	Cellular metabolic process	501	2.892^{-7}
	Gene expression	281	2.756^{-6}
	DNA repair	37	3.207^{-6}
	Cellular response to stress	55	1.411^{-5}
	Cell division	38	1.42^{-5}
	DNA damage checkpoint	14	5.317^{-5}
	Response to ionizing radiation	11	6.544^{-5}
	Cellular biosynthetic process	293	8.309^{-5}
	Nucleotide-excision repair	6	2.485^{-4}
	RNA processing	197	4.818^{-4}
	Recombinational repair	6	8.945^{-4}
U87MG HOXA9-Downregulated Genes	Antigen processing and presentation	19	1.977^{-08}
	Extracellular matrix organization	21	$3.150E^{-07}$
	System development	201	$4.041E^{-07}$
	Cell adhesion	89	$4.412E^{-07}$
	Response to hypoxia	19	$1.509E^{-06}$
	Regulation of biological quality	127	5.242^{-5}
	Regulation of cell motion	24	9.271^{-5}
	Lipid transport	24	1.03^{-4}
	Multicellular macromolecule metabolic process	10	2.103^{-4}
	Oxidation reduction	66	2.183^{-4}
	Peptidyl-proline modification	3	3.139^{-4}
	Small GTPase mediated signal transduction	55	3.303^{-4}

Vasculature development	33	3.453^{-4}
Wound healing	24	3.998^{-4}
Nervous system development	88	5.209^{-4}
Endothelial cell proliferation	9	6.615^{-4}
Ras protein signal transduction	8	6.79^{-4}

data set [81] for drugs that induce, in cell lines, gene expression signatures similar to the one we obtained in our microarray analysis. Although we are only concerning the top 20 of the differentially expressed genes upregulated in the U87MG-HOXA9 GBM cell line as compared to its *HOXA9* negative counterpart, we found that the drug that ranked first (regarding statistically significance, *p*-value <0.00001) was LY-294002 (Table 3.4), a known phosphatidylinositol 3-kinase (PI3K) inhibitor[95]. Interestingly, the highly negative enrichment of the LY-294002 in PC3 cell line indicates that the gene expression signature we provided to the software (i.e., putative HOXA9-target genes) is highly repressed by the perturbagen (i.e., LY-294002). As inhibition of the PI3K pathway is known to inhibit *HOXA9* transcription[63], this result further supports the validity of our microarray data. The histone deacetylase (HDAC) inhibitor Trichostatin A (TSA)[96] was also found to be able to repress the gene expression signature induced by HOXA9 in MCF7 cell line. HDACs are able to deacetylate lysine residues on histones and induce transcriptional repression through chromatin condensation, and their inhibitors lead to cell cycle arrest[97] and apoptosis[97], among others[98-101]. In line with LY-294002 and TSA, Tanespimycin and its analogue Geldanamycin, the second and ninth compounds on the Connectivity map results, respectively (Table 3.4), were described to be heat-shock protein 90 (HSP90) inhibitors[102,103], which are molecular chaperones involved in the conformational maturation of proteins crucial in signaling pathways[102]. Again, on the MCF7 cell line, Tanespimycin and Geldanamycin exert a repressive effect on the query gene signature, indicating important roles for HSP90 in the mediation of HOXA9 effects in GBM cells.

Table 3.4| Connectivity Map search for drug treatments that induce gene expression signatures similar to the top 20 upregulated genes in U87MG-HOXA9 cells, as compared to U87MG-MSCV cells. (Only cancer-related drugs were considered).

Rank	Cmap Name	Cell line	n	Enrichment	*p*-Value
1	LY-294002	PC3	12	-0.634	<0.00001
2	Tanespimycin	MCF7	36	-0.434	<0.00001
3	Trichostatin A	MCF7	92	-0.399	<0.00001
9	Geldanamycin	MCF7	10	-0.56	0.002

n – number of treatments and vehicle pair; Enrichment – correlation value with the query gene signature

3.5 Expression of Non-Coding RNAs Induced by HOXA9

The fact that recent reports have implicated non-coding RNAs (ncRNAs) in cancer[104], allied with "non-coding RNA metabolic process" genes ranking first on biological function clustering (Table 3.3), drove us to understand which ncRNAs were differentially expressed in our microarray data, and which have already been associated with cancer (Table 3.5). On hTERT/E6/E7, the non-tumor cell line, only two ncRNAs were differentially expressed on *HOXA9*-positive cells in comparison to the HOXA9-negative matching part (Table 3.5); conversely, in the U87MG GBM cell line, fifteen ncRNAs were present when comparing *HOXA9*-positive and -negative cells (Table 3.5). Two ncRNAs from hTERT/E6/E7, XIST and LINC00087, were also present on U87MG (Table 3.5). Seven of the ncRNAs are yet to be characterized. Besides their uncharacterized roles in cancer, the fact that these ncRNAs are differentially expressed on *HOXA9*-positive GBM cell line makes them interesting targets to be studied in GBM. Our microarray data showed that *HOX transcript antisense intergenic RNA* (*HOTAIR*) was upregulated in *HOXA9*-overexpressing cells (Table 3.5), a long ncRNA (lncRNA) that has recently gained a central role in breast[105], colorectal[106], hepatocellular[107,108], pancreatic[109] and gastrointestinal stromal tumors[110]. Interestingly, no studies assessing *HOTAIR* status in GBM have been reported.

Table 3.5| Non-protein coding RNAs from the microarray data of both hTERT/E6/E7 and U87MG, and respective log fold-change, after *HOXA9* overexpression versus their respective MSCV counterpart.

Gene Name	FC	Tumor Type	Cancer-related Functions	Refs
hTERT/E6/E7				
LINC00087	-1.13	N/A	N/A	N/A
XIST	-6.44	Breast, ovarian, cervical, kidney, colorectal, lymphoma and testicular germ cell tumor	Tumor recurrence and therapy resistance	[111-116]
U87MG				
CECR4	-1.95	N/A	N/A	N/A
DLEU2	1.24	Chronic lymphocytic leukemia	Tumor suppressor gene	[117]
HOTAIR	1.11	Breast, colorectal, hepatocellular, pancreatic and gastrointestinal stromal tumor	Increased proliferation, invasion, drug resistance and tumor recurrence	[105-110]
LINC00085	-1.70	N/A	N/A	N/A
LINC00087	-1.22	N/A	N/A	N/A
MALAT1	-1.85	Lung adenocarcinoma, non-small cell lung, breast, pancreas, colon, prostate osteosarcoma and hepatocellular	Increased migration, invasion and drug-resistance	[118-124]
MEG3	-1.38	Hepatocellular, meningioma, acute myeloid leukemia and multiple myeloma	Tumor suppressor gene	[125-128]
MIR100HG	-2.08	N/A	N/A	N/A
NEAT1	-1.12	Ovarian	Putative apoptosis inhibitor	[129]
SNHG1	1.43	N/A	N/A	N/A
SNHG7	1.06	N/A	N/A	N/A
SNHG8	1.22	N/A	N/A	N/A
UCA1	1.25	Bladder	Increased proliferation, migration, invasion and drug resistance	[130]
XIST	-2.20	Breast, ovarian, cervical, kidney, colorectal, lymphoma and testicular germ cell tumor	Tumor recurrence and therapy resistance	[111-116]
ZNFX1-AS1	0.79	Breast	Tumor suppressor gene	[131]

FC – Log Fold- Change; N/A – Not Available

3.6 *HOTAIR* Expression in Human Gliomas

To understand if HOTAIR is expressed in primary gliomas, we analyzed its expression levels in low grade glioma (LGG) and in high grade glioma (HGG) patients using gene expression array data from the Oncomine[83] and REMBRANDT[78] datasets. We found that none of the LGG patients presented *HOTAIR* overexpression (Table 3.6), whereas 17% (76/448) of HGG patients displayed *HOTAIR* overexpression when comparing to controls (Table 3.6). This suggests that the overexpression of *HOTAIR* is increased as the glioma grade increases. Moreover, the fact that is reactivation is more frequent in HGG may also indicate that *HOTAIR* presents crucial roles in the more aggressive features of HGG compared to LGG.

3.7 *HOTAIR* Co-expression with *HOX* Genes in Human Gliomas

In an attempt to identify the mechanisms that may be regulating *HOTAIR* gene activation in gliomas, we searched for genes that are significantly co-expressed with *HOTAIR* in glioma samples using human microarray datasets from the Oncomine database[83]. We assessed four independent studies[54,55,132,133], and found that several *HOX* genes were significantly co-expressed with *HOTAIR* (correlation values >0.3; Figure 3.4 and Table 3.7) in all studies, which supports the co-expression of *HOTAIR* with these genes in GBM. Interestingly, across all studies[54,55,132,133], *HOXA* genes were the most frequently co-expressed with *HOTAIR* (Figure 3.4 and Table 3.7). In agreement with our microarray data that revealed HOXA9-induced expression of *HOTAIR*, *HOXA9* was one of the genes highly correlated with *HOTAIR* expression in primary GBMs (Figure 3.4 and Table 3.7). This co-expression was also verified in primary GBMs from the REMBRANDT dataset, where approximately 90% (43/48) of *HOTAIR*-positive GBMs (Table 3.6) were also *HOXA9*-positive, which implies a putative role for HOXA9 on *HOTAIR* regulation. Remarkably, concomitant overexpression of *HOTAIR* and *HOXA9* occurs almost exclusively in high-grade astrocytomas and oligodendrogliomas (Figure 3.4), again supporting the possible contribution of *HOTAIR* to more aggressive tumors.

Table 3.6| Clinicopathological features and *HOTAIR* expression status in patients from Oncomine[83] and REMBRANDT[78] databases.

Datasets	No. cases	Age (median ± SD)	Male/female ratio	Diagnosis	WHO grade	No. HOTAIR[+]
Oncomine:						
Murat[54]	80	56.5 ± 7.8	2.81	GBM	IV	2
Phillips[55]	24	35 ± 9.9	2	AA	III	2
	76	49 ± 12.8	2.17	GBM	IV	7
Sun[133]	7	N/A	N/A	DA	II	0
	38	N/A	N/A	ODG	II	0
	19	N/A	N/A	AA	III	3
	12	N/A	N/A	ODG	III	1
	81	N/A	N/A	GBM	IV	12
Freije[132]	8	34 ± 6.8	0.13	AA	III	0
	7	32 ± 5	0.14	AOA	III	0
	18	35.5 ± 10.2	0.3	AODG	III	0
	59	47 ± 16.2	0.79	GBM	IV	1
REMBRANDT[78]	67	50 ± N/A	0.79	GBM	IV	48

AA - Anaplastic Astrocytoma; AOA - Anaplastic Oligoastrocytoma; AODG - Anaplastic Oligodendroglioma; DA - Diffuse Astrocytoma; GBM – Glioblastoma; LGG – Low Grade Glioma; ODG – Oligodendroglioma; N/A – Not Available

A

Correlation	Gene
1.000	HOTAIR
0.676	HOXC10
0.676	HOXC6
0.571	LOC400043
0.487	HOXB7
0.487	HOXA2
0.487	Loc285943
0.487	HOXB3
0.487	HOXB2
0.487	-
0.487	HOXA5
0.487	HOXA3
0.487	-
0.487	-
0.487	HOXA1
0.487	HOXA4
0.487	HOXA9
0.487	HOXA9
0.487	HOXA7
0.487	HOXA7
0.487	HOXA10

B

Correlation	Gene
1.000	HOTAIR
0.701	EN2
0.520	GALNT3
0.520	CA3
0.468	PITX2
0.372	HOXA7
0.372	HOXA10
0.372	HOXA9
0.372	SPATA17
0.372	TRIM5
0.372	APOBEC3B
0.372	DHRSX
0.372	HOXA5
0.372	HOXA1
0.372	HOXA3
0.372	-
0.372	SHOX2
0.372	SHOX2
0.372	HOXB2
0.372	HOXB3
0.302	-

C

Correlation	Gene
1.000	HOTAIR
0.309	HOXA4
0.309	HOXA3
0.309	HOXA10
0.309	HOXA10
0.309	HOXA9
0.309	HOXA9
0.285	HOXD3
0.285	-
0.285	LOC375295
0.285	HOXC10
0.285	HOXC6
0.285	-
0.285	HOXD13
0.285	HOXD11
0.285	HOXD10
0.285	-
0.285	-
0.285	TM4SF1
0.285	TM4SF1
0.285	TM4SF1

D

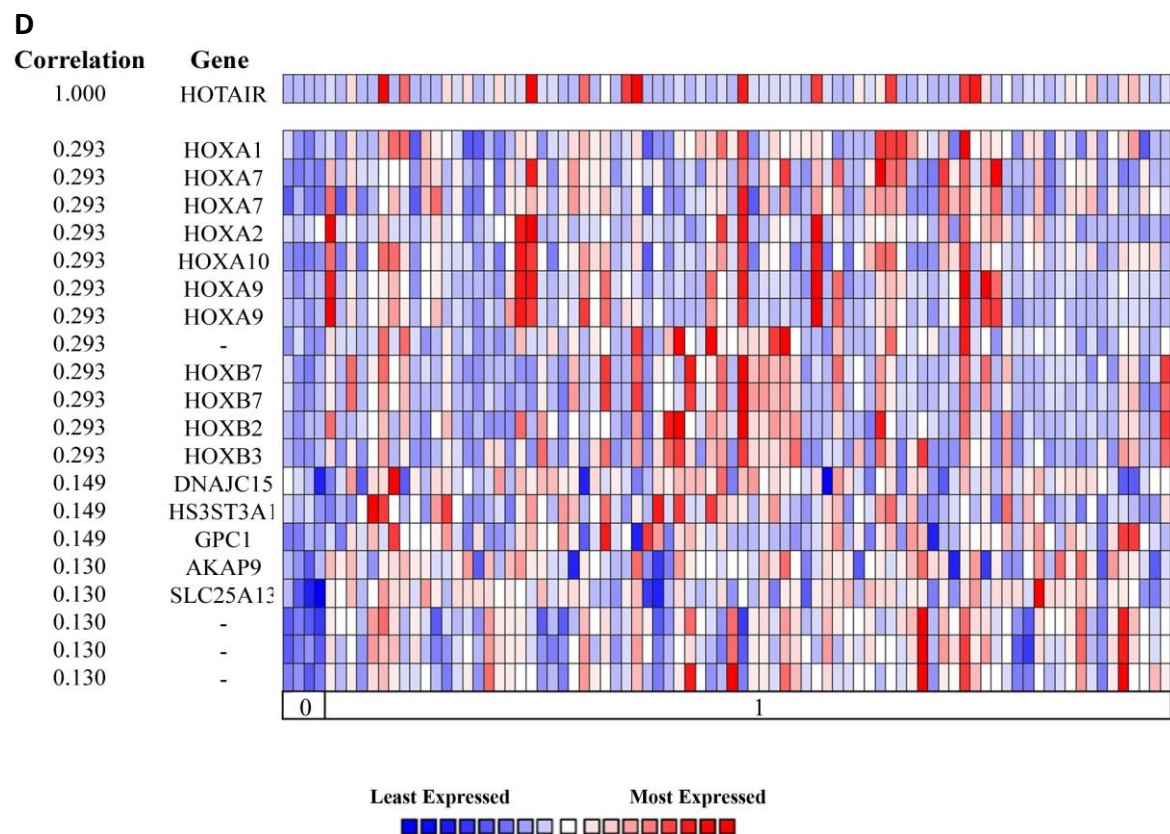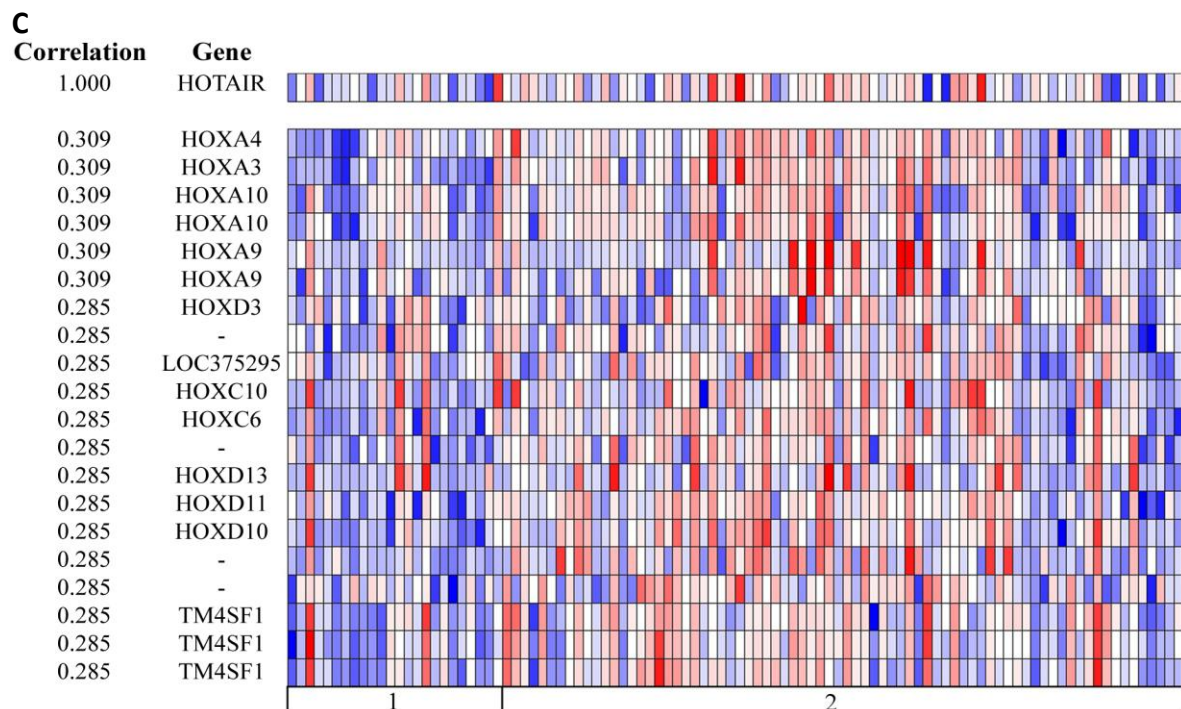| Correlation	Gene
1.000	HOTAIR
0.293	HOXA1
0.293	HOXA7
0.293	HOXA7
0.293	HOXA2
0.293	HOXA10
0.293	HOXA9
0.293	HOXA9
0.293	-
0.293	HOXB7
0.293	HOXB7
0.293	HOXB2
0.293	HOXB3
0.149	DNAJC15
0.149	HS3ST3A1
0.149	GPC1
0.130	AKAP9
0.130	SLC25A13
0.130	-
0.130	-
0.130	-

Figure 3.4| Heatmap analysis of genes co-expressed with *HOTAIR* in the Sun (A; correlation value 0.487), Freije (**B**; correlation value 0.372), Phillips (**C**; correlation value 0.309) and Murat (**D**; correlation value 0.293) studies from the Oncomine database. All sets present concomitant overexpression of *HOTAIR* and *HOXA9*, which occurs more frequently in higher WHO grades **(A - C)**. [**(A)** 0 - normal tissue (n= 23); 1 – Grade II (n=45); 2 – Grade III (n=31); 3 – Grade IV (n=81); **(B)** 0 - normal tissue (n=1); 1 - grade III (n=25); 2 - grade IV (n=59); **(C)** 1 - normal tissue (n=4); 2 – grade IV (n=80); **(D) 0** - grade III (n=24); 1 - grade IV (n=76)].

Conversely, when searching on TCGA datasets lodged on the Oncomine database[83] for co-expression between *HOTAIR* and *HOXA9* on other cancer types, no significant associations were found for lung, leukemia, colorectal, or breast cancer (Table 3.8). Taken together, our results suggest that *HOTAIR* expression is frequent in GBM, and that its expression is highly correlated with *HOXA9* expression exclusively in high-grade glioma.

Table 3.7| *HOTAIR* co-expression with several *HOX* genes, indicating a good correlation between *HOTAIR* and *HOXA9* in human gliomas (only instances with correlation values ≥0.3 are shown).

	Murat[54]	Phillips[55]	Sun[133]	Freije[132]
HOXA1	0.293	0.309	0.487	0.372
HOXA2	0.293	0.309	0.487	
HOXA3	0.293	0.309	0.487	0.372
HOXA4		0.309	0.487	
HOXA5	0.293	0.309	0.487	0.372
HOXA7	0.293	0.309	0.487	0.372
HOXA9	0.293	0.309	0.487	0.372
HOXA10	0.293	0.309	0.487	0.372
HOXA11			0.401	
HOXB2	0.293	0.309	0.487	0.372
HOXB3	0.293	0.309	0.487	0.372
HOXB6		0.309		
HOXB7	0.293	0.309	0.487	
HOXB9			0.348	
HOXC6	0.364	0.285	0.676	
HOXC9	0.699			
HOXC10	0.699	0.285	0.676	
HOXC13		0.511		
HOXD3	0.364	0.285		
HOXD4	0.364			
HOXD8	0.364			
HOXD10	0.527	0.285	0.456	
HOXD11	0.527	0.285	0.456	
HOXD13		0.285	0.367	

Table 3.8| Analysis of *HOTAIR* overexpression and co-expression with *HOXA9* in lung, leukemia, colorectal, and breast cancer datasets from TCGA, available at Oncomine[83]. Correlation values indicate lack of significant correlations between the expression of *HOTAIR* and *HOXA9* in these cancer types (Pearson correlations<0.05).

Study	No. of Patients	No. of Patients HOTAIR$^+$ (%)	Correlation HOTAIR$^+$/HOXA9$^+$
TCGA Lung[134]	167	16 (9.6%)	0.046
TCGA Leukemia[134]	197	1 (0.5%)	-0.019
TCGA Colorectal[134]	215	12 (5.6%)	0.012
TCGA Breast[134]	532	326 (61.3%)	-0.051

3.8 Mechanism Driving *HOTAIR* Expression in Glioblastoma

HOX genes encode transcription factors that are master regulators of gene expression in both normal and pathological states. Here, the interesting overexpression of *HOTAIR* in *HOXA9*-positive U87MG cell line (Table 3.5), as well as the fact that the expression of *HOTAIR* and *HOXA9* were found to be significantly correlated in primary glial tumors (Figure 3.4 and Table 3.6), led us to hypothesize that *HOTAIR* expression may be transcriptionally activated by HOXA9. To test this hypothesis, GBM cell lines A172 (endogenously expressing *HOXA9*), U87MG-HOXA9 (retrovirally infected to overexpress *HOXA9*), U87MG-MSCV (control, HOXA9-negative), and immortalized human astrocytes hTERT/E6/E7-HOXA9 (retrovirally infected to overexpress *HOXA9*) and hTERT/E6/E7-MSCV (control, HOXA9-negative) were tested for *HOTAIR* expression by RT-PCR and qPCR (Figure 3.5). All 3 tested GBM cell lines A172, U87MG-MSCV, and U87MG-HOXA9 presented *HOTAIR* expression. Importantly, in line with our hypothesis that HOXA9 activates *HOTAIR* in GBMs, U87MG-HOXA9 cells presented significantly increased expression of *HOTAIR* as compared to their HOXA9-negative counterpart (U87MG-MSCV; Figure 3.5). In contrast, non-malignant hTERT/E6/E7-MSCV and hTERT/E6/E7-HOXA9 astrocytes did not present *HOTAIR* expression (Figure 3.5), even after *HOXA9* overexpression, adding to our body of data that suggests *HOTAIR* and *HOXA9* co-expression is specific to high-grade GBM cells.
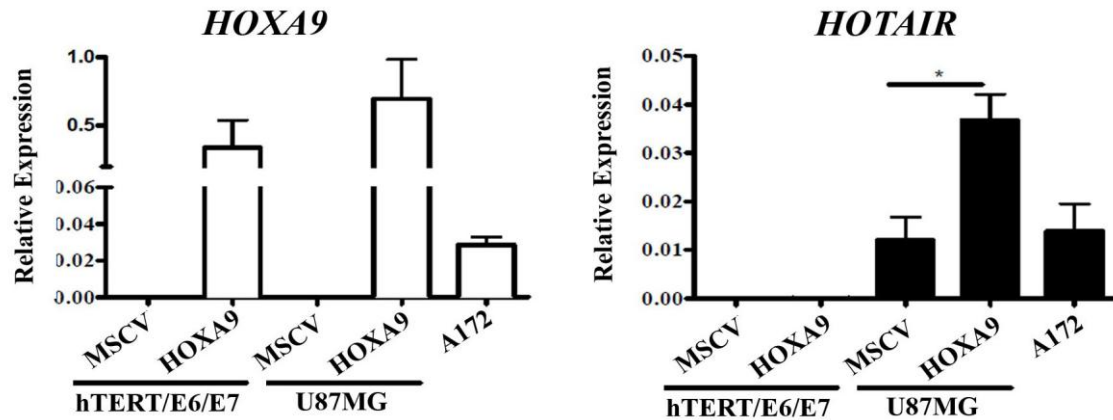
Figure 3.5| *HOTAIR* and *HOXA9* expression were evaluated by reverse-transcriptase quantitative PCR in a panel of cell lines with different levels of *HOXA9* expression. No detectable levels of *HOTAIR* are present in hTERT/E6/E7 immortalized human astrocytes (either *HOXA9*-positive or *HOXA9*-negative). The 3 GBM cell lines (U87MG-MSCV, U87MG-HOXA9, and A172) present endogenous *HOTAIR* expression, which is significantly increased after retrovirally-mediated *HOXA9* overexpression in U87MG-HOXA9 cells as compared to U87MG-MSCV (*HOXA9*-negative cell line). *HOTAIR* and *HOXA9* levels were normalized to *hGUS*. The results are representative of three independent experiments (mean and standard deviations are represented). *p=0.014.

To understand if HOXA9 directly interacts with the *HOTAIR* promoter to modulate its expression, we performed ChIP on A172, the GBM cell line endogenously expressing HOXA9, using an antibody against HOXA9. To ensure that the immunoprecipitation procedure was performed successfully before testing any putative direct-target, we used 3 different sets of primers to amplify 3 regions of the E-Selectin (SELE) gene promoter region that were previously described as HOXA9-binding sites[85]: both regions 1 and 2 of *SELE* promoter (primers sets 1 and 2, respectively) were described as HOXA9-binding sites, while region 3 (third set of primers) was described as a negative control for HOXA9 binding[85]. Accordingly, qPCR performed in HOXA9-bound chromatin from A172 cells showed that both primers sets 1 and 2 retrieved amplification of *SELE*, while the 3rd primer set presented little amplification of the fragments immunoprecipiated with anti-HOXA9 (Figure 3.6). An anti-H3 immunoprecipitated DNA was used as a positive control for all qPCR reactions (Figure 3.6). Together, these results further validate our ChIP data.
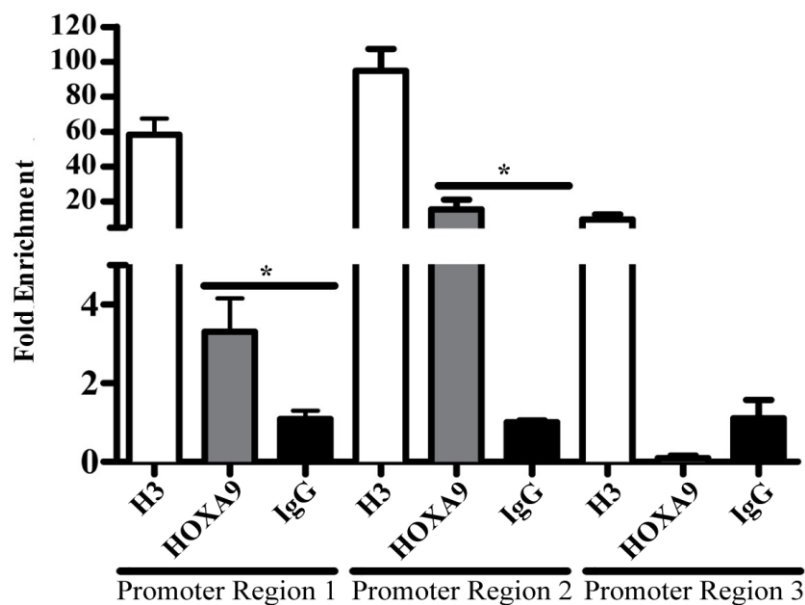
Figure 3.6| The putative binding of HOXA9 protein to the promoter region of E-SELECTIN (*SELE*) gene was assessed by chromatin immunoprecipitation (ChIP) analysis followed by qPCR in A172 cells. Chromatin immunoprecipitated with anti-HOXA9, anti-H3, and anti-IgG was amplified using 3 different sets of primers.Primer sets 1 and 2 were described as amplifying a region of *SELE* promoter bound by HOXA9, while primer set 3 amplifies a promoter region where HOXA9 does not bind[85]. As expected, sets 1 and 2 retrieved *SELE* amplification in fragments precipitated with anti-HOXA9, while set 3 presented reduced amplification. The ChIP positive control anti-histone H3 was amplified in all primers sets. IgG was used as ChIP negative control, and input DNA was not subjected to immunoprecipitation. Fold enrichment is normalized to input and to the IgG background signal from three independent experiments. * *p*=0.0333 in region 1 and *p*=0.0338 in region 2.

Then using these validated immunoprecipitates, we tested our hypothesis of *HOTAIR* direct-regulation by HOXA9, for which we used a set of primers to amplify a portion of *HOTAIR* promoter region. Quantitative PCR revealed HOXA9 occupancy of the promoter region of *HOTAIR* (Figure 3.7). Together, these results indicate that HOXA9 is a direct activator of *HOTAIR* expression in GBM.
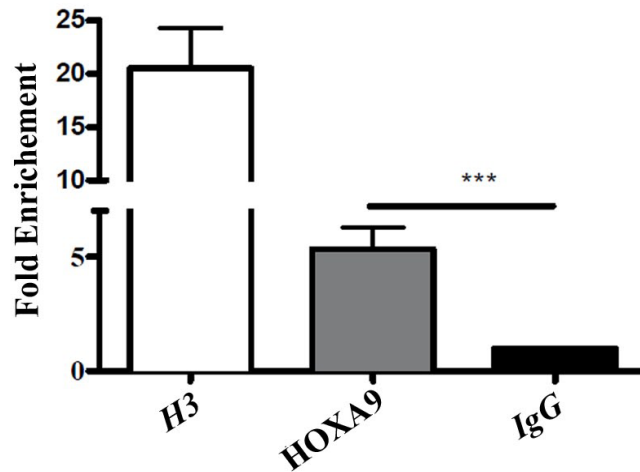
Figure 3.7| The putative binding of HOXA9 protein to the promoter region of *HOTAIR* was assessed by chromatin immunoprecipitation (ChIP) analysis followed by quantitative PCR in A172 cells. Anti-histone H3 and control IgG were used as positive and negative controls for the ChIP, respectively. Input DNA was not subjected to immunoprecipitation. Chromatin immunoprecipitated with an anti-HOXA9 antibody shows direct binding of HOXA9 to the *HOTAIR* promoter. Relative enrichment is normalized to input and to the IgG background signal from three independent experiments. ***p=0.0007.

3.9 Association of *HOTAIR* Expression with Survival of Glioblastoma Patients

HOTAIR expression was recently associated with poor prognosis of patients with breast cancer[105], hepatocellular carcinoma[108], colorectal[106], pacreatic[109], and gastrointestinal stromal tumors[110]. Additionally, *HOXA9* overexpression was shown to correlate with worse survival in GBM patients[63]. Therefore, we investigated the clinical significance of *HOTAIR* expression in GBM patients from REMBRANDT[78]. In 67 GBM patients with available survival data, a statistically significant decrease in overall survival (OS) was observed in patients with high *HOTAIR* expression (n=48), as compared to patients whose tumors present low levels of *HOTAIR* (n=19; p=0.005 Log-rank test; Figure 3.8 and Table 3.9). Univariate survival analyses of the clinicopathological data available in sufficient number (at least >50% of the samples) from this dataset, showed that patient age and institution of treatment were significantly associated with overall survival (p<0.0001 and p=0.001, respectively; Table 3.9).
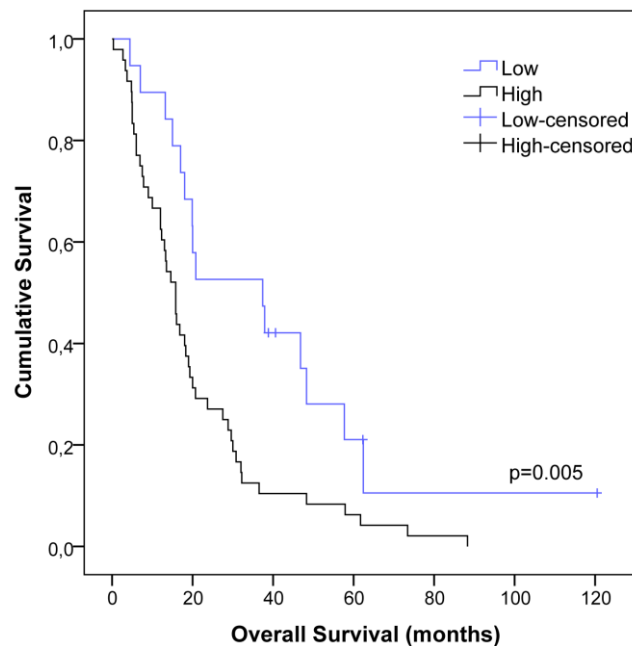
Figure 3.8| Kaplan-Meier overall survival curve of 67 patients from REMBRANDT[78] dataset indicating that patients whose tumors present high *HOTAIR* expression (n=42) show a statistically significant shorter overall survival as compared to those with *HOTAIR*-low expression tumors (n=19; *p*=0.005).

Moreover, we performed multivariate survival analyses using a Cox model (Table 3.9) to verify if *HOTAIR* expression has prognostic value independently of other putative confounding variables (as gender and patient's treatment institution). We found that high *HOTAIR* expression in GBM patients from REMBRANDT was significantly associated with decreased OS (*p*=0.034; Table 3.9), independently of the other clinicopathological features. Together, our data suggest *HOTAIR* expression as a marker of prognosis in GBM.

3.10   *WNT6* as New Direct-Target of HOXA9

As stated previously, *HOTAIR* is one of the downstream effectors of HOXA9 expression in GBM. However, the panoply of unknown HOXA9-direct targets must be enormous. Accordingly, in order to understand which of the differentially expressed genes might be directly regulated by HOXA9, we performed a genome-wide *in silico* analysis, which assessed every gene with putative binding sites for HOXA9. We found that HOXA9 may be the direct-regulator of 1907 genes. Integrating this gene list with

Table 3.9| Univariate (Log-rank) and multivariate (Cox proportional regression model) analyses of *HOTAIR* prognostic value in 67 patients from REMBRANDT[78] dataset. Patients with high *HOTAIR* expression present statistically significant shorter overall survival compared to those with *HOTAIR*-negative tumors (*p*=0.005; Log-rank test), independently of other putative prognostic factors (age, *p*=0.002, gender, *p*=0.275 and patient's treatment institution *p*=0.402).

	n	Median (95% CI)[a]	Overall survival *p*-value (Log-rank)	*p*-value (Cox)
HOTAIR expression				
High	48	15.8 (12.8-18.8)	0.005	0.034
Low	19	37.4 (11.9-62.9)		
Age				
>50	26	15.8 (9.8-21.8)	<0.0001	0.002
≤50	29	32.0 (18.5-45.5)		
Gender				
Male	16	27.5 (3.8-51.2)	0.379	0.275
Female	8	20.0 (14.5-25.5)		
Institution				
TJU	12	9 (3.9-14.1)		
NIH NOB	10	26.6 (11.0-48.2)		
PITT	2	6.9 (N/A)		
USCF	4	30.8 (N/A)		
HLMCC	5	19.0 (2.4-14.3)	0.001	0.402
UCLA	2	27.5 (N/A)		
DFCI	1	5.4 (N/A)		
HFH	27	18.0 (13.8-22.2)		
MDACC	2	3.7 (N/A)		
JHH	2	18.0 (14.3-21.7)		

n - Number of Patients; N/A – Not Available; TJU – Thomas Jefferson University; NIH NOB – National Institute of Health Neuro-Oncology Branch; PITT – University of Pittsburgh; UCSF – University of California. San Francisco; HLMCC – H. Lee Moffitt Cancer Center; UCLA – University of California. Los Angeles; DFCI – Dana-Farber Cancer Institute; HFH – Henry Ford Hospital; MDACC – M.D. Anderson Cancer Center;
[a] Median survival with 95% confidence intervals in months.

the differentially expressed genes from our microarray data, we found 173 putative-direct targets in U87MG cells, and 24 in hTERT/E6/E7 (Table 3.10), of which *CISH*, *KLHL4* and *PPP2R2B* were common to both U87MG and hTERT/E6/E7.

Besides being overexpressed in U87MG-HOXA9 cells (log fold-change of 1.61), the wingless-type MMTV integration site family member 6 (*WNT6*) was one of the putative direct targets of HOXA9 (Table 3.10). Interestingly, WNT proteins are a family

Table 3.10| *In silico* analysis using ExPlain™ software from BIOBASE to detect putative HOXA9-bindings sites at the genome-wide level. HOXA9 may directly regulate 1907 genes, of which 173 were found differentially expressed in U87MG and 24 in hTERT/E6/E7 cells. (For U87MG, examples of the genes with higher number of binding sites for HOXA9 are shown.)

U87MG		hTERT/E6/E7	
No. of Binding Sites	Gene Symbol	No. of Binding Sites	Gene Symbol
4	LRRC37A4	3	GRAMD4
3	LOC221710	2	MGLL
3	SGOL1	2	QPRT
3	TMC4	2	TRIM63
2	HLA-G	2	CPA3
2	AK1	2	KLHL4
2	SERPINB2	2	GHDC
2	UMPS	2	IRF5
2	WNT6	2	CISH
2	GRM1	2	ABHD10
2	HLA-DMA	2	CDKL2
2	CCDC132	2	CRYL1
2	ABCA8	1	LRFN2
2	CDCA3	1	PLAC8
2	IGSF10	1	ZNF585A
2	OSCAR	1	PTPRE
2	PCDHB16	1	SLC25A26
2	FASN	1	C8orf47
2	ANGPT2	1	FGF13
2	AQP9	1	LOC79999
2	BNC2	1	BAIAP2L2
2	HOXB13	1	PPAP2B
2	CCT5	1	CWF19L1
2	PLK4	1	PPP2R2B
2	RASL10B		
2	OASL		
2	KRT32		
2	LETM2		
2	LNX1		
2	MOCS1		
2	KIAA0513		
2	KLHL4		
2	SERPINB1		
2	RSC1A1		
2	KRR1		
2	PEX6		
2	SCD5		
2	CHCHD3		
2	CISH		
2	FABP4		

of glycoproteins that are able to control several signal transduction pathways and a wide range of cellular processes[135], whose aberrant expression have been implicated in several tumor types, including GBM[136-138]. However, no data assessing *WNT6* roles in GBM was reported until now. So, as WNT proteins have important roles in cancer, and *WNT6* may be directly regulated by HOXA9 in GBM cells, we assessed its expression in A172, U87MG-HOXA9, and U87MG-MSCV GBM cell lines, and in the non-tumor hTERT/E6/E7-HOXA9 and hTERT/E6/E7-MSCV immortalized human astrocytes. We found that *WNT6* is expressed in all tested cell lines, presenting no differences when comparing hTERT/E6/E7-HOXA9 and hTERT/E6/E7-MSCV. However, the levels of *WNT6* were found to be higher in U87MG-HOXA9 when comparing to the corresponding U87MG-MSCV HOXA9-negative cell line (Figure 3.9A). To verify if the putative direct regulation of *WNT6* by HOXA9 suggested by the *in silico* analysis was in fact real, we performed ChIP on A172, the endogenously HOXA9-expressing GBM cell line. Two sets of primers to amplify 2 regions of the *WNT6* promoter were used. ChIP followed by quantitative PCR revealed that HOXA9 binds to both tested regions of *WNT6* promoter (Figures 3.9B and C). Although these results are very preliminary, the integration of the microarrays, *in silico*, RT-PCR, and ChIP data suggest HOXA9 is a direct activator of *WNT6* expression in GBM.
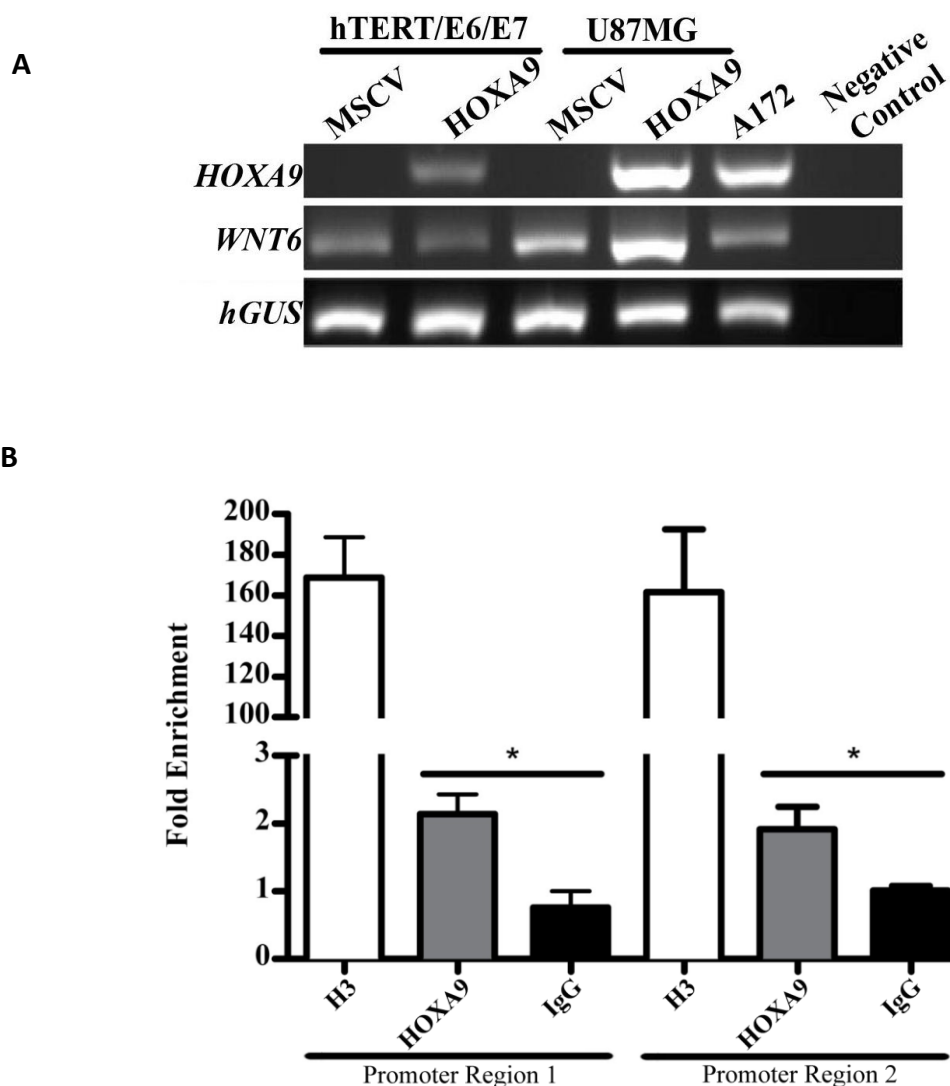
A

B

Figure 3.9| *WNT6* expression was evaluated by reverse-transcriptase PCR (**A**) in a variety of cell lines with different levels of *HOXA9* expression. All cell lines present *WNT6* expression. In hTERT/E6/E7-HOXA9 and hTERT/E6/E7 the expression level is similar; whereas in U87MG *WNT6* expression levels increased after *HOXA9* overexpression. The putative direct regulation of *WNT6* by binding of HOXA9 protein to its promoter region was evaluated by chromatin immunoprecipitation (ChIP) analysis followed by quantitative PCR (**B**) in A172 cells. Chromatin immunoprecipitated using an anti-HOXA9 antibody shows direct binding of HOXA9 to the 2 tested regions of *WNT6* promoter. Anti-histone H3 and control IgG were used as positive and negative controls for ChIP, respectively. Input DNA was not subjected to immunoprecipitation. Fold enrichment is normalized to input and to the IgG background signal from three independent experiments. * *p*=0.0174 in region 1 and *p*=0.0255 in region2.

4. Discussion

4. Discussion

Tumors of the central nervous system account for only a small percentage of all human tumor types, but comprise a wide range of distinctive neoplasias, of which glioblastoma (GBM) is the most common, malignant and aggressive subtype[15]. GBM is a particularly dramatic disease ranking first for years of life lost among all tumors and despite advances in clinical neuro-oncology, no curative therapies are available[21], remaining the prognosis of GBM patient poor, with a median survival of 15 months[25]. Patients are equally treated with the same standardized procedure (surgical resection followed by concomitant radiotherapy and chemotherapy with the alkylating agent Temozolomide[25]) regardless of specific molecular alterations, which lead to great variability in the response to treatment. In this sense, there is the urgent need to establish molecular prognostic markers that would allow the stratification of patients and the adaptation of therapeutics according to the underlying alterations. As referred previously, the methylation status of the promoter region of O^6-methylguanine methyl transferase (*MGMT*) in GBM is currently one of the most promising prognostic and therapy response predictive biomarker[58]. However, it has not reached clinical applicability and there is still the need to establish new prognostic biomarkers. Another putative biomarker concerns the deregulation of class I homeobox (*HOX*) genes in several tumor types[65,66], including primary malignant astrocytomas[76]. In GBM, it was suggested that these abnormal expression was part of a glioma stem cell-like gene expression signature, associated with worse clinical outcome of patients treated with temozolomide-based chemoradiotherapy[77]. Later, our group reported that *HOXA* genes are overexpressed more frequently in high-grade than in low-grade primary astrocytomas[63], and the reactivation of *HOXA* genes expression was found to be regulated by the PI3K pathway, through reversible regulation of EZH2-mediated histone modifications[63]. Among all *HOXA* genes, the overexpression of *HOXA9* was predictive of patient worse outcome, and the induction of *HOXA9* overexpression in GBM cell lines and neurospheres indicated pro-proliferative and anti-apoptotic roles, which may be responsible for the HOXA9-mediated worse prognosis of GBM patients[63]. More recently, a stem cell signature enriched in *HOX* genes mediated by the PI3K pathway was associated with resistance to temozolomide in pediatric GBM cell

lines[77]. Moreover, this report showed that patients with high expression of *HOXA9* and *HOXA10* genes presented shorter survival[77]. These studies, implicate the aberrant expression of the transcription factor HOXA9 in GBM as a therapeutic and prognostic marker. Such indicate the urgency in developing and testing new *HOXA9* inhibitors (both direct or indirect, as PI3K inhibitors), and demand the identification of HOXA9 direct-targets, which are likely to be the true effectors of its functions, in order to recognize new putative prognostic biomarkers and, eventually more important, to identify new putative therapeutic targets.

The transcriptome of HOXA9 in GBM is unknown; however, in hematopoietic cells, it has been shown that *HOXA9* is able to modulate the expression of a wide group of genes that have a large range of functional roles, and which were still not known to be target of any *HOX* gene[94]. In order to identify genes and pathways altered due to *HOXA9* overexpression in GBM, we performed genome-wide expression microarrays (Agilent, 44K, Human Whole Genome) for the human GBM models U87MG and hTERT/E6/E7 cells, comparing the transcriptomes of U87MG-HOXA9 vs. U87MG-MSCV, and hTERT/E6/E7-HOXA9 vs. hTERT/E6/E7-MSCV. These analyses provided a large set of differentially expressed genes that are likely to include both HOXA9-direct and indirect targets, which might be used for prognostic and therapeutic purposes. We found more than 3400 genes differentially expressed in the GBM cell line U87MG, and more than 400 in hTERT/E6/E7, indicating that HOXA9 is modulating the expression of many genes in a cell-line specific manner, acting both as an activator and a repressor of gene expression. The large number of genes might indicate that the overexpression of *HOXA9* is altering cell regulation, which may possibly lead to the higher aggressive features of *HOXA9*-positive GBM cells[63]. The fact that only 57 genes are simultaneously differentially expressed in both cell lines, implies that the effects of *HOXA9* overexpression are highly dependent on the cellular background.

Due to the lack of agreement between the techniques available to obtain microarray expression data, we validated our microarray data by reverse transcription-PCR (RT-PCR). Such validation should be global, addressing all genes differentially expressed from the microarray data[139]; however, the enormous number of genes differentially expressed obtained, makes global validation impractical. Thus, we used the strategy of choosing a subset of the most differentially expressed genes (largest

fold-changes) to perform this validation[139] (Figure 3.1). Although the chosen strategy presents pitfalls, as the agreement between microarray and RT-PCR/qPCR validation results selecting transcripts with the largest fold-changes may not be generalized to the entire set of differentially expressed genes[140], we attempted to overcome this drawback by evaluating the expression of genes with relatively mild/small fold-changes, as were the cases of *HOTAIR* and *WNT6* in U87MG cell line (Figures 3.5 and 3.9). For all genes evaluated, we obtained consistent results between the microarray data and RT-PCR, thus validating the microarray results.

As previously focused, the genes used on this analysis had their expression highly altered after *HOXA9* overexpression (Table 3.2), so they may present important roles in GBM onset and development. We found that intercellular adhesion molecule 2 (*ICAM2*) and bone morphogenetic and activin membrane-bound inhibitor homolog (*BAMBI*) were highly overexpressed due to HOXA9. It is known that the upregulation of *ICAM2* is crucial on immune response and resistance to drug-mediated apoptosis[141], while *BAMBI* has already been reported to be upregulated in several cancers, and is involved in the regulation of differentiation, migration, apoptosis, and cell proliferation[142]. The upregulation of both genes in *HOXA9*-positive U87MG cells possibly indicates that they have important roles in mediating the more aggressive phenotype induced by HOXA9 in GBM. Equally important, some of the most downregulated genes due to HOXA9 overexpression included angiopoietin 2 (*ANGPT2*) and platelet-derived growth factor receptor-β (*PDGFRB*). The expression of *ANGPT2* has been reported as increased in several tumors, and is able to promote angiogenesis and growth[143]. A very recent study showed that *ANGPT2* contributes to tumor progression and invasion[144]; however, evidence is emerging that *ANGPT2* may play different roles in the vasculature, depending on the cellular context[145]. In gastric cancer, *PDGFRB* is frequently overexpressed, and is known to be involved in the maintenance of microvessels and recruitment of pericytes[146]. So, further analyses to assess the roles of these genes in *HOXA9*-positive and negative GBM cells should be performed.

Similar to the genes validated in U87MG, those selected for the validation of GBM putative precursor hTERT/E6/E7 microarray data seem to have crucial roles in GBM. The upregulated RAS-related C3 botulinum substrate 2 (*RAC2*) is involved in

regulating the production of reactive oxygen species, and its overexpression has been reported in GBM[147], and associated with increased cell proliferation[148], migration, and neovascularization[149]. Also, chemokine (C-X-C motif) ligand 1 (*CXCL1*) was found to be upregulated in hTERT/E6/E7 cells overexpressing HOXA9; interestingly, its role in the oncogenic transformation of glioma precursors has been studied[150]. In this report, the authors implicate CXCL1 as an oncogenic factor in glioma, indicating that the deregulation of glial proliferative factors contributes to tumorigenesis[150]. The HOXA9-induced overexpression of both genes imply that these present characteristics of aggressiveness. Moreover, the increased upregulation of *CXCL1* mediated by HOXA9 may point a role for HOXA9 in the activation of genes involved in the putative oncogenic transformation of hTERT/E6/E7 in GBM, a study that would be of great interest. Concerning the downregulated N-myc downstream regulated 1 (*NDRG1*) and thymocyte selection-associated high mobility group box family member 2 (*TOX2*), their functions are still not completely understood. The function of *TOX2* is still elusive, but a very recent study reported a CpG island in the promoter region of *TOX2* that in normal cells is unmethylated, but is methylated in a significant percentage of lung (28%) and breast (23%) cancer cells. In fact, the expression of *TOX2* is correlated with the modulation of several pathways involved in tissue remodeling, inflammatory response, cell differentiation, apoptosis, cell cycle regulation, and DNA-damage response[151], so its downregulation is associated with a more aggressive phenotype. Also, the authors state that TOX2 may contribute to early malignant changes and to the modulation of the tumor microenvironment[151]. Similarly to *TOX2*, the examination of a wide variety of tumor tissues has shown that the expression level of *NDRG1* in cancer cells is usually similar or diminished when comparing to normal tissue[152-155], and has been associated with cellular growth, differentiation, tumorigenesis, metastasis[87]. Strikingly, it has been observed that the low levels of NDRG1 protein are associated with worse prognosis of glioma patients[87], which in fact is contradictory with our overall survival analysis (OS) in GBM patients from the Repository of Molecular Brain Neoplasia Data (REMBRANDT; *p*-value=0.009; Figure 3.2). However the survival analysis in Sun's report[87] does not distinguish glioma WHO grades, so in order to clear this question, the prognostic value of *NDRG1* expression in glioma patients should be assessed in other datasets, and to distinct WHO grades.

The above mentioned large list of differentially expressed genes makes the manually curated data analysis a highly time-consuming task, therefore requiring support from bioinformatics software[88]. To address this question, an increasing number of publicly available bioinformatics platforms have been arising, which are able to improve the biological analysis through the organization of massive and redundant results into comprehensible groups, and systematically ranking the most overrepresented biological terms[156]. We used two different but complementary bioinformatics software to extract biologically meaningful information from the microarray data: i) the Database for Annotation, Visualization and Integrated Discovery (DAVID)[88] bioinformatics database, and ii) the Biological Database (Biobase)[79] platform. DAVID bioinformatics was able to cluster non-redundant genes according to their function and display them in the pathways they integrate (Figure 3.3), whereas Biobase clustered differentially expressed genes according to biological processes using Gene Ontology (GO) annotation terms (Table 3.3). The overexpression of *HOXA9* in both GBM cell line U87MG and human immortalized astrocytes hTERT/E6/E7 was found to modulate the expression of genes that enrich several pathways involved in several important cellular processes. In addition to some cancer-unrelated pathways (e.g., "viral myocarditis" and "asthma", among others; Figure 3.3), an interesting result concerns the upregulation of pathways typical of cancers other than GBM, such as bladder cancer and small cell lung cancer (Figure 3.3). Also, the class comprising the highest number of overexpressed genes was found to be "pathways in cancer" (Figure 3.3), showing an association between the HOXA9-mediated upregulated gene expression signature and frequent alterations in cancer.

The cancer hallmarks represent some key alterations that cells acquire during the carcinogenic process[6], and their alteration in *HOXA9*-positive cells could be indicative of an higher aggressive phenotype. Accordingly, several hallmarks were found to be altered, as the most fundamental trait of cancer cells: the sustained chronic proliferation[6]. Contrarily to normal tissues, which carefully control production and release of growth signals, cancer cells are able to deregulate these signals in order to become self-sufficient and to proliferate uncontrollably[6]. This important feature was represented in U87MG cell line by the enrichment of upregulated genes in pathways known to lead to increased proliferation, as "cell division", "cell cycle", among others

(Figure 3.3 and Table 3.3). In agreement with our finding, the upregulation of genes involved in cell proliferation have already been reported in an HOXA9 transcriptome analysis in hematopoietic cells[157]. Furthermore, GBM is known to be highly proliferative, and the exogenous overexpression of *HOXA9* in GBM and human immortalized astrocyte cell lines was already reported as responsible for higher proliferation rates in comparison to *HOXA9*-negative cell lines[63].

Another important hallmark concerns active invasion and metastasis[6], and although GBM rarely metastasizes to other organs, it diffusely infiltrates through the adjacent brain parenchyma[21], a feature that frequently precludes its complete surgical resection[158]. Genes downregulated in GBM cell line U87MG-HOXA9 compared to its negative counterpart U87MG-MSCV enriched crucial processes during invasion, as "focal adhesion", "cell adhesion" and "extracellular matrix organization" (Figure 3.3 and Table 3.3), indicating higher invasion ability of HOXA9-positive cells. Similarly, the biological process analysis revealed that vasculature development and endothelial cell proliferation were enriched in downregulated genes from the U87MG-HOXA9 cell line (Table 3.3). Such features are unusual on GBM once it is known that this tumor present increased neovascularization and highly proliferating endothelial cells[22,159]. It is known that these characteristics might be induced by hypoxia, which is a common event in GBM[160], through the effects of hypoxia-inducible factor (HIF) 1 on its downstream targets[161]. Also, *HIF1* is known to be regulated by the oncogenic RAS protein, and when RAS is inactivated leads to the downregulation of *HIF1*[162]. Both "hypoxic response" and "RAS signal transduction pathway" were found to be enriched in downregulated genes (Figure 3.3 and Table 3.3), which is a possible reason for the negative enrichment of vasculature development and endothelial cell proliferation (Table 3.3).

For the human immortalized astrocytes hTERT/E6/E7-HOXA9 in comparison to its control hTERT/E6/E7-MSCV, the enrichment of genes upregulated in pathways as "focal adhesion" and "cell interaction with extracellular matrix" (Figure 3.3) might indicate decreased invasion capacity of these cells. This result again implicates the context-dependence of HOXA9 transcriptome.

Another important result reveals that the upregulated gene signature was enriched in genes involved in several pathways associated with metabolic processes (Figure 3.3 and Table 3.3). It is known that normal cells mainly metabolize glucose to

pyruvate for growth and survival, a process that requires oxygen, so when oxygen is restricted pyruvate is metabolized to lactate[163]. However, Otto Warburg[164] demonstrated that cancer cells are able to reprogram their glucose metabolism, and even in the presence of oxygen, cancer cells can limit their energetic metabolism mainly to glycolysis[164]. The rationale underlying the metabolic switch in cancer cells is not completely understood, however the hypothesis that increased glycolysis allows the deviation of glycolytic intermediates to several biosynthetic pathways is raising agreement[165,166]. Such pathways are related with the production of nucleosides and amino acids that facilitate the biosynthesis of macromolecules and organelles required for fast dividing cells[166]. Interestingly, we found some of these pathways enriched in our results (Figure 3.3 and Table 3.3). In this sense, our results might indicate that *HOXA9*-positive GBM cells present altered metabolism in order to manage the demands of highly proliferative cells.

A second emerging hallmark of cancer concerns the role of the immune system in resisting or terminating the formation and progression of incipient neoplasias, late-stage tumors, and micrometastases[6]. Interestingly, our results showed that "antigen processing and presentation" was one of the pathways presenting higher number of genes downregulated in U87MG-HOXA9 (Figure 3.3 and Table 3.3). The fact that cancer can still develop, even with this continuous surveillance, implies that tumors have gained the ability to avoid this detection or to limit immunological elimination[167]. Thus, our results seem to implicate that *HOXA9*-positive GBM cells are able to avoid immune system detection in a highly effective manner, precluding *HOXA9*-positive cancer cells elimination. Concerning the non-tumor human immortalized astrocytes hTERT/E6/E7-HOXA9, "cytokine-cytokine receptor interaction" and "chemokine signaling pathway" were found to be enriched in up and downregulated genes, respectively (Figure 3.3). In cancer, it is known that cytokines have a wide variety of roles, comprising cell invasion, proliferation, angiogenesis, migration, leucocyte infiltration, stimulation of neovascularization, and manipulation of the immune response; however, they may also be able to inhibit inhibit tumor growth[92]. Accordingly, the HOXA9-induced downregulation of genes involved in "cytokines interactions" might be required for the silencing those cytokines involved in growth inhibition, or even those involved in immune responses. Additionally, it is known that

chemokines act together with their cell surface receptors to direct cells to specific locations throughout the body[168]. In cancer, cell trafficking into and out of the tumor microenvironment require chemokines and their receptors[169]. Therefore, some chemokine receptors that are not present or functional in normal cells are reactivated, which is consistent with increased angiogenesis, survival, and metastatic activity[93]. In this sense, the HOXA9-induced upregulation of genes involved in "chemokine signaling pathway" might be associated with the already referred higher invasion of HOXA9-positive GBM model cells observed in our group (data not shown; manuscript in preparation). Though some of our results are not easily explained regarding our hypothesis, in general they represent crucial traits in GBM aggressiveness, as shown by the HOXA9-induced regulation of several genes that are involved in pathways associated with cancer (as increased proliferation and invasion).

It is widely recognized that one of the most important clinical features of GBM concerns to its high therapeutic resistance, and interestingly in pediatric GBM cell lines the reactivation of *HOX* genes expression was correlated with temozolomide resistance[77]. Therefore, and concerning all the features already stated that are possibly altered due to HOXA9, it seems reasonable to hypothesize that *HOXA9* overexpression might be influencing cell response to therapy. Different mechanisms are recognized to drive drug-resistance in cancer cells, and interestingly our results showed enrichment of "ATP-binding cassette (ABC) transporters", "hypoxia" and "DNA repair pathways" (Figure 3.3 and Table 3.3). Several ABC transporters are multidrug efflux pumps that play important roles in the uptake and distribution of therapeutics, so their increased expression is frequently associated with the efflux of drugs, and consequently with chemotherapy resistance[170]. It is also known that under hypoxic conditions alkylating agents may be less effective, mainly due to the increased production of nucleophilic agents that may compete with the target DNA for alkylation[171]. As HOXA9-induced downregulated genes enriched both ABC transporters and hypoxic response (Figure 3.3), seem to implicate that these are not probable mechanisms mediating GBM resistance to therapy.

Moreover, in U87MG-HOXA9 several pathways responsible for DNA damage repair were found to be enriched, as the cases of increased "DNA repair", "DNA and "nucleotide excision repair", among others (Figure 3.3 and Table 3.3). As previously

referred, the increased proliferation of cancer cells make proteins and signals involved on cell cycle triggering, important therapeutic targets. Cell proliferation may be targeted by mitotic spindle inhibitors[172], growth signaling inhibition[173], or by inducing damages on DNA (which is the most studied)[174]. These damages are able to cause cell cycle arrest and cell death either directly, or during the attempt of cells to replicate damaged DNA. Though, the effect of DNA-damaging drugs might be reduced by DNA repair pathways that eliminate DNA damages[175], consequently leading to drug resistance. Our results are highly interesting, once they may indicate that GBM cells when overexpressing *HOXA9* are more able to repair DNA damages than *HOXA9*-negative. This together with the fact that the current GBM chemotherapy is based on temozolomide[25], an alkylating agent that induces cell death by producing adducts on DNA[176], implicates that these *HOXA9*-positive GBMs may be more resistant to therapy, possibly as consequence of the upregulation of genes involved in DNA repair pathways. These hypothesis warrant validation in future studies and highlight the urgent need to investigate the alterations of the enzymes involved on these pathways, in order to understand their role in therapy resistance.

Additionally, this altered response to therapy also prompted us to understand if the differentially expressed genes have already been associated with pharmacological inhibitors. A preliminary analysis in Connectivity Map using part of the gene signature induced by HOXA9 in U87MG cell line, revealed that LY-294002, tanespimicyn and its analogue geldanamycin, and thrichostatin A (TSA) were able to reverse this gene signature (Table 3.4). In particular, LY-294002 is an inhibitor of the PI3K pathway[95], which was pointed as responsible for *HOXA9* reactivation in GBM[63]. So the use of PI3K inhibitors would indirectly lead to the silencing of *HOXA9*. As *HOXA9*-positive GBM patients present worse prognosis than those with HOXA9-negative GBM[63], it is urgent to understand if the use of PI3K inhibitors in the treatment of the HOXA9-positive GBM would ameliorate the prognosis of these patients. So, the rational design of studies involving *HOXA9*-positive GBM and PI3K inhibitors is of great importance. Tanespimicyn and its analogue geldanamycin are heat-shock protein 90 (HSP90) inhibitors[102,103]. HSP90 proteins are involved in correct folding of several proteins crucial on carcinogenic process, as p53 and several kinases[102]. Our results suggests that these drugs are able to repress part of the *HOXA9*-induced transcriptome, possibly by

inhibiting the correct maturation of one or several proteins involved in PI3K pathway, or even the correct maturation of HOXA9 downstream targets. However effects of these drugs are not specific to HOXA9 or its targets, as they may inhibit a large panoply of proteins involved in several pathways. Equally important, TSA is an inhibitor of histone deacetylase (HDAC)[96]. HDACs are able to deacetylate lysine residues on histones and induce transcriptional repression through chromatin condensation. HDAC inhibitors are responsible for cell cycle arrest in G1 and/or G2 phases[97], apoptosis[97], cell differentiation[98], transcriptional[99] and morphological[100] alterations and decreased angiogenesis[101]. So, it is possible that TSA inhibits part of the HOXA9-induced gene signature by leading to cell cycle arrest and thus to decreased proliferation. But may also point a role for HOXA9 in regulating HDACs, and consequently leading to altered acetylation levels. As TSA and other HDAC inhibitors are able to sensitize GBM cells to radiotherapy[177], and to increase the cytotoxicity of anticancer drugs targeting the DNA[178], these may be important adjuvants for the therapy currently used to treat GBM, especially of those presenting *HOXA9* expression.

A very interesting result concerns the enrichment of "non-coding RNA metabolic process" by HOXA9-induced upregulated genes from the U87MG cell line (Table 3.3). Non-protein coding RNAs (ncRNAs) have been reported as able to regulate gene expression[42-45], and are being widely associated with the carcinogenic process. We searched the differentially expressed genes of both U87MG and hTERT/E6/E7 after *HOXA9* overexpression, and found altered expression of some ncRNAs, of which several are yet to be characterized (Table 3.5). However, others have their roles on cancer well studied (Table 3.5), as the cases of *HOTAIR* and *MALAT1* whose functions on several tumors were associated with higher aggressiveness, as increased proliferation and invasion[105-108,110,118-124]. NcRNAs may be divided into small or long according to their transcript length, and specially the long non-coding RNAs (lncRNAs) are thought to be "the missing links on cancer"[179].

LncRNAs represent spliced and polyadenylated transcripts that modify the genome in a highly tissue-specific manner, which can silence gene expression through different mechanisms, such as chromatin remodeling, transcriptional and post-transcriptional regulation[43,180,181]. Most of the known lncRNAs regulate the expression of neighboring genes (*in cis*). However, the scope of lncRNAs in gene regulation had a big advance

when HOX Transcript Antisense Intergenic RNA (*HOTAIR*) was found to exhibit *trans*-regulatory functions, which allows it to reprogram chromatin organization throughout the genome[44]. *HOTAIR* is transcribed at the *HOXC* locus (between *HOXC11* and *HOXC12*) and interacts with the Polycomb Repressive Complex 2 (PRC2), a histone methylating complex, resulting in transcriptional repression of a 40 kb region of the *HOXD* cluster on chromosome 2 by epigenetic control of chromatin[180]. Though, Tsai et al.[182] reported that *HOTAIR* not only functions as a molecular scaffold to bind PRC2 complex but also to lysine specific demethylase 1/REST corepressor1/RE1-silencing transcription factor (LSD1/CoREST/REST) protein complexes. Biochemically, the 5' domain of *HOTAIR* binds to PRC2, while the 3' domain binds to the LSD1/CoREST/REST demethylating complex (Figure 4.1), mediating the enzymatic methylation promoted by PRC2 and the H3K27 methylase EZH2, and also the demethylation of histoneH3 dimethyl lysine 4 (H3K4Me2) mediated by LSD1[45,182]. The relevance of *HOTAIR* in cancer was first reported by Gupta and co-workers[105], where they report a correlation between *HOTAIR* expression and increased metastatic potential and diminished OS in breast cancer patients[105]. In patients with hepatocellular carcinoma following liver transplantation, Yang *et al.*[108] showed that *HOTAIR* levels are higher in cancer tissues than in non-cancerous adjacent regions, and that its expression was associated with tumor recurrence. Additionally, *HOTAIR* expression was associated with increased chemotherapy resistance and invasion, and decreased apoptosis[108]. Recently, *HOTAIR* expression was also evaluated in colorectal cancer[106] and gastrointestinal stromal tumors (GIST)[110], where its expression was correlated with worse patient survival and increased metastatic potential. Specifically for GIST, *HOTAIR* expression was accompanied by overexpression of *miR196a* and related with tumor cell invasiveness[110]. A very recent report associated the overexpression of *HOTAIR* in pacreatic cancer with increased proliferation and invasion, decreased apoptosis, and was associated with shorter OS[109]. Together, these studies shed light into the biological and clinical significance of *HOTAIR* expression in different tumor types. However, no studies have been performed in human gliomas until now.

Using publicly available datasets (Oncomine[83] and REMBRANDT[78]) we assess the molecular status of *HOTAIR* in glioma patients, showing that this lncRNA is
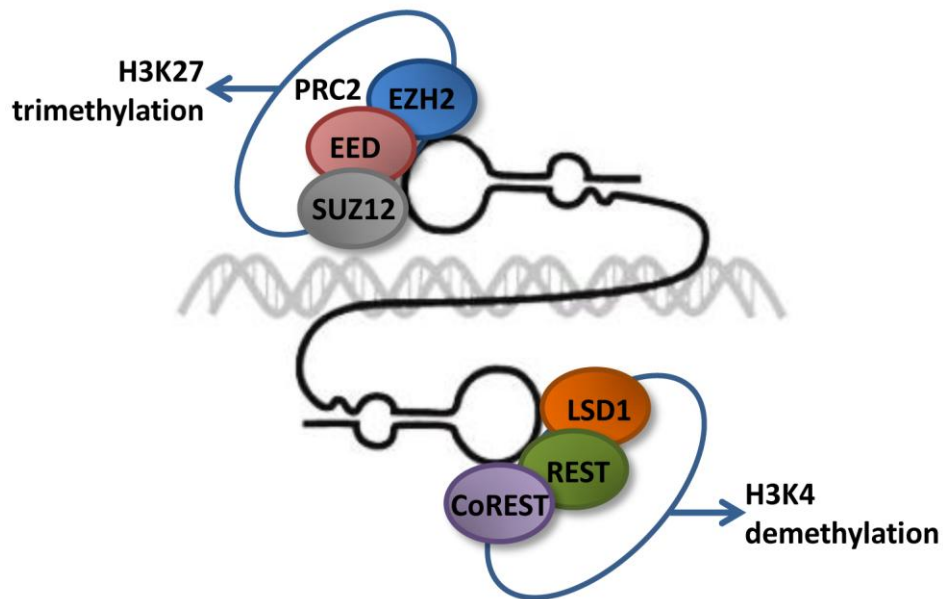
Figure 4.1| *HOTAIR* as a molecular scaffold for the binding of PRC2 and LSD1/CoREST/REST complexes. The 5' domain of *HOTAIR* binds to PRC2, while the 3' domain binds to the LSD1/CoREST/REST leading to the trimethylation of H3K27 and demethylation of H3K4, respectively. (Adapted from Prensner and Chinnaiyan, 2011[183], and Flynn and Chang, 2011[184])

overexpressed in a subset of high-grade gliomas, particularly GBMs (Table 3.5); in contrast, *HOTAIR* expression is not frequent in LGGs (Table 3.5), suggesting *HOTAIR* expression is grade-specific and associated with more malignant gliomas. Importantly, our findings have putative clinical implications as GBM patients with high *HOTAIR* expression presented a statistically significant worse prognosis when compared to low *HOTAIR*-expressing patients in 2 large independent datasets of GBM patients, independently of patients' age, sex and KPS (Figure 3.8 and Table 3.9).

Because our data suggests *HOTAIR* as a new independent prognostic biomarker in GBM, it is critical to understand the molecular mechanisms underlying *HOTAIR* activation in these tumors. To address this, we investigated the presence of copy number chromosomal aberrations and alterations in DNA methylation levels on the *HOTAIR* locus, which could lead to aberrant gene expression (data not shown; manuscript submitted). We found *HOTAIR* gene amplification is a rare event in GBM, and the methylation levels of the *HOTAIR* region (~20kb including the *HOTAIR* gene and spanning from *HOXC12* to *HOXC11*) did not significantly correlate with *HOTAIR* expression, indicating that other molecular mechanisms may be regulating *HOTAIR*

reactivation (data not shown; manuscript submitted). We next questioned which genes are frequently co-expressed with *HOTAIR* in the context of gliomas, and observed that several *HOXA*, *HOXB*, *HOXC*, and *HOXD* genes are frequently co-expressed with *HOTAIR* in these tumors, an event consistently observed in 4 independent datasets[54,55,132,133] (Figure 3.4 and Table 3.7). Not surprisingly, *HOXC* genes are among the most co-expressed genes, which is likely related to the fact that *HOTAIR* is located within the *HOXC* chromosomal cluster[180]. Similar to the microarray data, we found that the transcription factor HOXA9 is frequently co-expressed with *HOTAIR* in high-grade gliomas (Figure 3.4 and Table 3.7). Interestingly, our results show that co-expression of *HOTAIR* and *HOXA9* is associated with increased grades of malignancy, as this concomitant overexpression occurs almost exclusively in high-grade gliomas (Figure 3.4). Remarkably, we did not observe the same co-expression of *HOTAIR* and *HOXA9* in different tumor types available at TCGA (Table 3.8), indicating that this co-expression may be exclusive of gliomas. Both coding and non-coding genes share similar functional features regarding the existence of common transcription factors in their promoter regions, suggesting that they may be controlled by the same transcriptional regulatory machineries[185]. Taking this into account, together with the fact that *HOTAIR* is upregulated in *HOXA9*-overexpressing GBM cells and they are co-expressed in high-grade gliomas, we questioned if HOXA9 could directly regulate *HOTAIR* in GBM cells. Indeed, our chromatin immunoprecipitation (ChIP) results indicate that there is a direct regulation of *HOTAIR* by HOXA9 (Figure 3.6).

In contrast to the majority of *HOTAIR* studies in cancer, where the focus has been put on the downstream effects of *HOTAIR*[105,106,108,110], the present work provides novel insights on its transcriptional regulation, widening our understanding of the *HOTAIR*-associated mechanisms of aggressiveness and malignancy in gliomas. Interestingly, interpreting our findings at the light of other studies further supports the classic intrinsic tissue-specificity of lncRNAs[186]. For example, *HOTAIR* was described to be directly regulated by HOXA13 in normal fibroblasts[187]. Additionally, in contrast to Rinn and co-workers' study[180] describing *HOTAIR*-mediated silencing of *HOXD8-11* genes in fibroblasts, we found that some *HOXD* genes are frequently co-expressed with *HOTAIR* in GBM. Indeed, Rinn *et al*[180] showed *HOTAIR* represses *HOXD* genes by binding to PRC2 and recruiting it to the target loci. Considering that the PI3K pathway

is aberrantly activated in >80% of GBMs[26], which in turn inhibits the formation of PRC2 complex by phosphorylating EZH2[188], it is reasonable to speculate that despite *HOTAIR* being overexpressed in some GBMs, it will not target the same genes (e.g. *HOXD* genes) as the PRC2 partner complex is not functional. This further supports a tissue-specificity expression and function of *HOTAIR*.

In summary, these results suggest that (a) *HOTAIR* is overexpressed in a subset of high-grade gliomas, independently of gene copy number and DNA methylation levels; (b) co-expression of *HOTAIR* and *HOXA9* occurs mostly in high-grade glioma; (c) this co-expression is specific of gliomas; (d) *HOTAIR* presents prognostic value in GBM patients; and (e) *HOTAIR* and *HOXA9* may be useful biomarkers to integrate a molecularly-based stratification of GBM patients. We anticipate that *HOTAIR* may be one of the effector mechanisms by which HOXA9 creates a more aggressive and therapy-resistant form of GBM. Future studies are warranted to identify *HOTAIR* cellular functions and downstream target genes at the genome-wide level in GBM, in an attempt to better understand the mechanisms by which *HOTAIR* affects survival of patients, and ultimately investigate new therapeutic opportunities.

The interesting results obtained for the HOXA9 direct-target *HOTAIR* and the fact that HOXA9 may be mediating several potentially oncogenic pathways through its direct-targets, prompted us to find new genes directly regulated by HOXA9. Until now the number of identified HOXA9 direct-targets already identified is scarce, which has precluded the definition of the mechanism underlying HOX protein target specificity[157]. We performed a genome-wide *in silico* analysis of promoter regions, in order to understand which genes possess HOXA9 binding sites, and crossed this information with the lists of differentially expressed genes of both cell lines. Such analysis retrieved a large number of putative direct-targets of HOXA9 (Table 3.10). A few HOXA9 direct targets have already been reported and interestingly, our *in silico* analysis presented cytochrome b-245-β (*CYBB*) as a putative direct-target. In myeloid differentiation, *CYBB* has already been reported to present HOXA9 binding sites on its promoter region[189]. However, other genes that are known to have binding sites for HOXA9 on their promoter region, as *EPHB4*[190], *SOX4*[157] or *MYB*[157], were not pointed as putative direct-targets on the genome-wide analysis. Moreover, *SELE*, which was reported to be present HOXA9 binding sites[85], as well as the newly identified direct-target *HOTAIR*,

which were shown by our results to be immunoprecipitated by anti-HOXA9 antibody (Figures 3.6 and 3.7), were not identified on the *in silico* analysis. It is known that transcription factor binding sites can exhibit some variability, however they are commonly modeled as position weight matrices that have significant limitations[191]. Therefore, our analysis might be biased leading us to disregard some important putative direct-targets of HOXA9, and a future approach to circumvent these limitations would be the genome-wide prediction of HOXA9 binding sites using weight matrices calculated with different algorithms. Nonetheless, *WNT6* was considered by this *in silico* analysis as a putative direct-target of HOXA9 (Table 3.10), which was confirmed by our ChIP analysis (Figure 3.9).

Interestingly WNTs are a family of glycoproteins that control several signal transduction pathways and may promote a wide range of cellular processes, as proliferation, polarity, differentiation, adhesion and migration[135]. Also, aberrant WNT signaling have been implicated in several tumor types[136,137]; specifically in GBM, WNT pathway has been suggested as a molecular mechanism able to confer GBM radio-resistance, and is pointed as an important therapeutic target[138]. Our study concerning *WNT6* features in GBM is very preliminary, however as it is direct-target of HOXA9, its roles as a contributing effector of HOXA9 aggressiveness in GBM must be explored. Also, the potential of pharmacological inhibition of WNT6 may prove to be a feasible therapeutic approach in overcoming the worse prognosis of HOXA9-positive GBMs.

In this work, we initiated the unraveling of the complex transcriptome of HOXA9 in GBM. Our results indicate that HOXA9 is influencing several crucial features related with GBM aggressiveness (as proliferation, invasion and therapy resistance), thus implicating the importance of circumventing HOXA9 effects in GBM. For this we propose several alternatives, either via indirect inhibition of HOXA9 (as for instance PI3K inhibitors) or of its downstream targets, as *HOTAIR* and *WNT6*.

5. Concluding Remarks and Future Perspectives

5. Concluding Remarks and Future Perspectives

Glioblastoma (GBM) is the most malignant and common type of tumor in the adult CNS, for which no curative therapies are available. It is known that the lack of well-established prognostic markers contributes to GBM patient poor prognosis, as a consequence of the lack of therapeutic adaptation to each patient. So, there is the urgent need to establish new prognostic biomarkers that are able to stratify patients into cohorts regarding molecular biomarkers, which will lead to the rationalization of therapeutic decisions. In this sense, we intend to identify a set of genes differentially expressed in our microarray data, with prognostic and/or therapy response predictive value, which may help clinicians to direct patients to the most appropriate treatments. Furthermore, we anticipate that a combination of prognostic biomarkers will be more efficient in predicting patient prognosis than single genes.

Nonetheless, the reactivation of HOXA9 expression in GBM has recently emerged as a putative prognostic marker; however, the mechanisms by which HOXA9 associates with patient survival are poorly understood. The transcriptome of HOXA9 in GBM is still unknown and the understanding HOXA9´s cellular roles may explain its association with clinical outcome. Bioinformatics analysis of the microarray data of HOXA9-positive and negative GBM cell line U87MG and in human immortalized astrocytes hTERT/E6/E7 cells, allowed the identification of alterations in typical hallmarks of cancer, as increased proliferation, invasion and therapy resistance. Interestingly, our results seem to support the theory of increased therapy resistance due to the upregulation of genes involved in DNA repair pathways. As these can be pharmacologically inhibited, possibly leading to the sensitization of GBM cells to chemotherapy, their evaluation at the protein level in both HOXA9-positive and negative cell lines is important, and may help to understand the chemotherapy resistance of GBMs. Moreover, it would be of great importance to verify if the use of DNA repair enzymes inhibitors can lead to a sensitization of HOXA9-positive tumor cells to the current chemotherapy, and consequently to tumor remission. In the future, we also intend to continue exploring the microarray data of both U87MG and hTERT/E6/E7, especially by using other software, as those that enable the

understanding of relevant co-expression between genes groups in a high throughput manner (for instance, Gene Set Co-Expression Analysis).

Additionally, using the Connectivity Map platform, we provide evidence suggestive of the ability of particular drugs, including LY-294002, trichostatin A, tanespimicyn and its analogue geldamycin, to inhibit the HOXA9-induced gene expression signature in GBM. These drugs warrant further analysis, as they may prove to be crucial in the reversion of the more aggressive phenotype induced by HOXA9 in GBM cells. Furthermore, we also plan to enlarge the Connectivity Map analysis to the HOXA9-induced downregulated genes of U87MG cell lines, but also to hTERT/E6/E7.

The genome-wide *in silico* analysis revealed a vast number of putative direct targets of HOXA9, which include the *WNT6* gene. Our chromatin immunoprecipitation (ChIP) analysis proved that *WNT6* is, in fact, a direct target of HOXA9, so it might be one of the key downstream molecules mediating the higher aggressiveness of HOXA9-positive GBMs. Here we present a very preliminary analysis of *WNT6* in GBM; however, it highlights the importance of assessing its roles in mediation of HOXA9 effects, as it may represent a new therapeutic target. So, the consequences of its pharmacological inhibition on important cancer features (as proliferation, invasion and apoptosis) must be evaluated. Though, the *in silico* analysis requires further analysis as there are probably other unknown direct-targets of HOXA9 in GBM So, it is imperative to continue unraveling other genes that are directly regulated by HOXA9, as they may represent crucial prognostic and/or predictive of therapy response biomarkers, or therapeutic targets.

Our work also shows *HOTAIR*, a long non-coding RNA that has recently been implicated in the aggressiveness of several cancers, as almost exclusively overexpressed in GBMs as compared to normal and low-grade gliomas. We also show that *HOTAIR* and *HOXA9* are frequently co-expressed, especially in higher glioma grades (III and IV). Mechanistically, we proved that, HOXA9 binds to the promoter region of *HOTAIR* and induces its transcription in GBM cells. Importantly, we provide the first evidence on the molecular players regulating aberrant *HOTAIR* expression in GBM, and implicate *HOTAIR* overexpression as a novel independent prognostic marker for these patients. In this sense, *HOTAIR* and *HOXA9* may be useful biomarkers to integrate a molecularly-based stratification of GBM patients.

We intend to analyze *HOTAIR* expression in primary GBM samples, and to assess its concomitant expression with *HOXA9*. Functionally, *HOTAIR* may be controlling several pathways involved in tumor progression and aggressiveness, so we plan to modulate *HOTAIR* expression (both with silencing and overexpressing approaches) in GBM cell lines in order to understand its role in several hallmarks of cancer (including proliferation, invasion and apoptosis). From a therapeutic perspective, *HOTAIR* downstream effects may be inhibited by the use of PRC2 or LSD1 inhibitors. However, as previously referred, the PI3K pathway is overexpressed in more than 80% of GBMs, which inhibits the formation of PRC2 complex by phosphorylating EZH2. In this sense, *HOTAIR*'s effects in GBM are likely to be mainly mediated by LSD1 complex. So, the use of LSD1 inhibitors may silence *HOTAIR*, and thus limit its downstream effects and reverse its aggressive phenotype.

In summary, our study provides the first characterization of HOXA9 transcriptome in the context of GBM, suggest putative mechanisms by which HOXA9 renders GBM cells with a more malignant phenotype, and identifies new clinically-relevant prognostic biomarkers. Together, our findings may help to rationalize therapy decisions and to identify new molecular targets for GBM therapy.

6. References

6. References

1. Parkin DM, Bray F, Ferlay J, and Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74-108.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, and Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
3. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, and Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893-917.
4. Koul D. PTEN signaling pathways in glioblastoma. Cancer Biol Ther 2008;7:1321-5.
5. TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathway. Nature 2008;455:1061-68.
6. Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
7. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801-6.
8. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. Science 2008;321:1807-12.
9. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108-13.
10. Gerstung M, Eriksson N, Lin J, Vogelstein B, and Beerenwinkel N. The temporal order of genetic and pathway alterations in tumorigenesis. PLoS One 2011;6:e27136.
11. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, and Parkin DM. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2010. Available from: http://globocan.iarc.fr, accessed on 04/2012. 2010;
12. Ferlay J BF, Pisani P, Parkin DM. GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. Cancer Incidence and Mortality Worldwide (CR-rom) ed. Lyon: 2004.
13. Louis DN, Ohgaki H, Wiestler OD, and cavenee WK, Classification of tumours of the central nervous system2007, Lyon: Internation Agency for Research on Cancer.
14. Burnet NG, Jefferies SJ, Benson RJ, Hunt DP, and Treasure FP. Years of life lost (YLL) from cancer is an important measure of population burden--and should be considered when allocating research funds. Br J Cancer 2005;92:241-5.
15. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol 2007;114:97-109.
16. Kleihues P and Sobin LH. World Health Organization classification of tumors. Cancer 2000;88:2887.
17. Kyritsis AP, Yung WK, Bruner J, Gleason MJ, and Levin VA. The treatment of anaplastic oligodendrogliomas and mixed gliomas. Neurosurgery 1993;32:365-70; discussion 71.
18. Rivera AL, Pelloski CE, Sulman E, and Aldape K. Prognostic and predictive markers in glioma and other neuroepithelial tumors. Curr Probl Cancer 2008;32:97-123.
19. Zhu Y and Parada LF. The molecular and genetic basis of neurological tumours. Nat Rev Cancer 2002;2:616-26.
20. Lino MM and Merlo A. PI3Kinase signaling in glioblastoma. J Neurooncol 2010;
21. Wen PY and Kesari S. Malignant gliomas in adults. N Engl J Med 2008;359:492-507.
22. Kleihues P, Louis DN, Scheithauer BW, Rorke LB, Reifenberger G, Burger PC, et al. The WHO classification of tumors of the nervous system. J Neuropathol Exp Neurol 2002;61:215-25; discussion 26-9.

23. Ohgaki H, Dessen P, Jourde B, Horstmann S, Nishikawa T, Di Patre PL, et al. Genetic pathways to glioblastoma: a population-based study. Cancer Res 2004;64:6892-9.

24. Ohgaki H and Kleihues P. Genetic pathways to primary and secondary glioblastoma. Am J Pathol 2007;170:1445-53.

25. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 2005;352:987-96.

26. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061-8.

27. Reynolds BA and Weiss S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. Science 1992;255:1707-10.

28. Sanai N, Alvarez-Buylla A, and Berger MS. Neural stem cells and the origin of gliomas. N Engl J Med 2005;353:811-22.

29. Liu G, Yuan X, Zeng Z, Tunici P, Ng H, Abdulkadir IR, et al. Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. Mol Cancer 2006;5:67.

30. Pojo M and Costa BM, Molecular Hallmarks of Gliomas, in M. Garami (eds). Molecular Targets of CNS Tumors InTech 2011. 178-200.

31. Haber D and Harlow E. Tumour-suppressor genes: evolving definitions in the genomic age. Nat Genet 1997;16:320-2.

32. Fujisawa H, Reis RM, Nakamura M, Colella S, Yonekawa Y, Kleihues P, et al. Loss of heterozygosity on chromosome 10 is more extensive in primary (de novo) than in secondary glioblastomas. Lab Invest 2000;80:65-72.

33. Ichimura K, Schmidt EE, Miyakawa A, Goike HM, and Collins VP. Distinct patterns of deletion on 10p and 10q suggest involvement of multiple tumor suppressor genes in the development of astrocytic gliomas of different malignancy grades. Genes Chromosomes Cancer 1998;22:9-15.

34. Xavier-Magalhães A and Costa BM, Molecular Determinants of Glioma Risk and Patients Prognosis2012, Saarbrücken: LAP LAMBERT Academic Publishing GmbH & Co. KG.

35. Nakamura M, Yang F, Fujisawa H, Yonekawa Y, Kleihues P, and Ohgaki H. Loss of heterozygosity on chromosome 19 in secondary glioblastomas. J Neuropathol Exp Neurol 2000;59:539-43.

36. Homma T, Fukushima T, Vaccarella S, Yonekawa Y, Di Patre PL, Franceschi S, et al. Correlation among pathology, genotype, and patient outcomes in glioblastoma. J Neuropathol Exp Neurol 2006;65:846-54.

37. Furnari FB, Fenton T, Bachoo RM, Mukasa A, Stommel JM, Stegh A, et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. Genes Dev 2007;21:2683-710.

38. Sherr CJ and McCormick F. The RB and p53 pathways in cancer. Cancer Cell 2002;2:103-12.

39. Ponting CP and Belgard TG. Transcribed dark matter: meaning or myth? Hum Mol Genet 2010;19:R162-8.

40. Stein LD. Human genome: end of the beginning. Nature 2004;431:915-6.

41. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799-816.

42. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 2007;316:1484-8.

43. Mercer TR, Dinger ME, and Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009;10:155-9.

44. Hung T and Chang HY. Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol 2010;7:582-5.

45. Gibb EA, Brown CJ, and Lam WL. The functional role of long non-coding RNA in human carcinomas. Mol Cancer 2011;10:38.

46. Brosnan CA and Voinnet O. The long and the short of noncoding RNAs. Curr Opin Cell Biol 2009;21:416-25.

47. Griffiths-Jones S. The microRNA Registry. Nucleic acids research 2004;32:D109-11.

48. Zhang B, Pan X, Cobb GP, and Anderson TA. microRNAs as oncogenes and tumor suppressors. Dev Biol 2007;302:1-12.

49. Esquela-Kerscher A and Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 2006;6:259-69.

50. Bottoni A, Piccin D, Tagliati F, Luchin A, Zatelli MC, and degli Uberti EC. miR-15a and miR-16-1 down-regulation in pituitary adenomas. J Cell Physiol 2005;204:280-5.

51. Gabriely G, Wurdinger T, Kesari S, Esau CC, Burchard J, Linsley PS, et al. MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators. Mol Cell Biol 2008;28:5369-80.

52. Han L, Zhang K, Shi Z, Zhang J, Zhu J, Zhu S, et al. LncRNA profile of glioblastoma reveals the potential role of lncRNAs in contributing to glioblastoma pathogenesis. Int J Oncol 2012;40:2004-12.

53. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 2005;352:997-1003.

54. Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou MF, et al. Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. Journal of clinical oncology 2008;26:3015-24.

55. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell 2006;9:157-73.

56. Colman H and Aldape K. Molecular predictors in glioblastoma: toward personalized therapy. Arch Neurol 2008;65:877-83.

57. Liu L, Markowitz S, and Gerson SL. Mismatch repair mutations override alkyltransferase in conferring resistance to temozolomide but not to 1,3-bis(2-chloroethyl)nitrosourea. Cancer Res 1996;56:5375-9.

58. Hegi ME, Diserens AC, Godard S, Dietrich PY, Regli L, Ostermann S, et al. Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. Clin Cancer Res 2004;10:1871-4.

59. Komine C, Watanabe T, Katayama Y, Yoshino A, Yokoyama T, and Fukushima T. Promoter hypermethylation of the DNA repair gene O6-methylguanine-DNA methyltransferase is an independent predictor of shortened progression free survival in patients with low-grade diffuse astrocytomas. Brain Pathol 2003;13:176-84.

60. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med 2009;360:765-73.

61. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010;17:98-110.

62. Jansen M, Yip S, and Louis DN. Molecular pathology in adult gliomas: diagnostic, prognostic, and predictive markers. Lancet Neurol 2010;9:717-26.

63. Costa BM, Smith JS, Chen Y, Chen J, Phillips HS, Aldape KD, et al. Reversing HOXA9 oncogene activation by PI3K inhibition: epigenetic mechanism and prognostic significance in human glioblastoma. Cancer Res 2010;70:453-62.

64. Grier DG, Thompson A, Kwasniewska A, McGonigle GJ, Halliday HL, and Lappin TR. The pathophysiology of HOX genes and their role in cancer. J Pathol 2005;205:154-71.

65. Wellik DM. Hox genes and vertebrate axial pattern. Curr Top Dev Biol 2009;88:257-78.
66. Shah N and Sukumar S. The Hox genes and their roles in oncogenesis. Nat Rev Cancer 2010;10:361-71.
67. Abate-Shen C. Deregulated homeobox gene expression in cancer: cause or consequence? Nat Rev Cancer 2002;2:777-85.
68. Takahashi O, Hamada J, Abe M, Hata S, Asano T, Takahashi Y, et al. Dysregulated expression of HOX and ParaHOX genes in human esophageal squamous cell carcinoma. Oncol Rep 2007;17:753-60.
69. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531-7.
70. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proc Natl Acad Sci U S A 2007;104:5527-32.
71. Krivtsov AV and Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer 2007;7:823-33.
72. Care A, Felicetti F, Meccia E, Bottero L, Parenza M, Stoppacciaro A, et al. HOXB7: a key factor for tumor-associated angiogenic switch. Cancer Res 2001;61:6532-9.
73. Hu YL, Fong S, Ferrell C, Largman C, and Shen WF. HOXA9 modulates its oncogenic partner Meis1 to influence normal hematopoiesis. Mol Cell Biol 2009;29:5181-92.
74. Raman V, Martensen SA, Reisman D, Evron E, Odenwald WF, Jaffee E, et al. Compromised HOXA5 function can limit p53 expression in human breast tumours. Nature 2000;405:974-8.
75. Wu X, Chen H, Parker B, Rubin E, Zhu T, Lee JS, et al. HOXB7, a homeodomain protein, is overexpressed in breast cancer and confers epithelial-mesenchymal transition. Cancer Res 2006;66:9527-34.
76. Abdel-Fattah R, Xiao A, Bomgardner D, Pease CS, Lopes MB, and Hussaini IM. Differential expression of HOX genes in neoplastic and non-neoplastic human astrocytes. J Pathol 2006;209:15-24.
77. Gaspar N, Marshall L, Perryman L, Bax DA, Little SE, Viana-Pereira M, et al. MGMT-independent temozolomide resistance in pediatric glioblastoma cells associated with a PI3-kinase-mediated HOX/stem cell gene signature. Cancer Res 2010;70:9243-52.
78. National Cancer Institute REMBRANDT home page: http://rembrandt.nci.nih.gov. 2005.
79. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 2006;34:D108-10.
80. Mann RS, Lelli KM, and Joshi R. Hox specificity unique roles for cofactors and collaborators. Curr Top Dev Biol 2009;88:63-101.
81. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 2006;313:1929-35.
82. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 2007;9:166-80.
83. Oncomine™ (Compendia Bioscience, Ann Arbor, MI) was used for analysis and visualization.
84. Nelson JD, Denisenko O, and Bomsztyk K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. Nat Protoc 2006;1:179-85.
85. Bandyopadhyay S, Ashraf MZ, Daher P, Howe PH, and DiCorleto PE. HOXA9 participates in the transcriptional activation of E-selectin in endothelial cells. Mol Cell Biol 2007;27:4207-16.

86. Livak KJ and Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;25:402-8.

87. Sun B, Chu D, Li W, Chu X, Li Y, Wei D, et al. Decreased expression of NDRG1 in glioma is related to tumor progression and survival of patients. *J Neurooncol* 2009;94:213-9.

88. Huang da W, Sherman BT, and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44-57.

89. Selvakumaran M, Pisarcik DA, Bao R, Yeung AT, and Hamilton TC. Enhanced cisplatin cytotoxicity by disturbing the nucleotide excision repair pathway in ovarian cancer cell lines. *Cancer Res* 2003;63:1311-6.

90. Wu X, Zhao H, Wei Q, Amos CI, Zhang K, Guo Z, et al. XPA polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. *Carcinogenesis* 2003;24:505-9.

91. Wilson J and Balkwill F. The role of cytokines in the epithelial cancer microenvironment. *Semin Cancer Biol* 2002;12:113-20.

92. Zhu VF, Yang J, Lebrun DG, and Li M. Understanding the role of cytokines in Glioblastoma Multiforme pathogenesis. *Cancer Lett* 2012;316:139-50.

93. Zlotnik A, Burkhardt AM, and Homey B. Homeostatic chemokine receptors and organ-specific metastasis. *Nat Rev Immunol* 2011;11:597-606.

94. Dorsam ST, Ferrell CM, Dorsam GP, Derynck MK, Vijapurkar U, Khodabakhsh D, et al. The transcriptome of the leukemogenic homeoprotein HOXA9 in human hematopoietic cells. *Blood* 2004;103:1676-84.

95. Matter WF, Brown RF, and Vlahos CJ. The inhibition of phosphatidylinositol 3-kinase by quercetin and analogs. *Biochem Biophys Res Commun* 1992;186:624-31.

96. Yoshida M, Kijima M, Akita M, and Beppu T. Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J Biol Chem* 1990;265:17174-9.

97. Yamashita Y, Shimada M, Harimoto N, Rikimaru T, Shirabe K, Tanaka S, et al. Histone deacetylase inhibitor trichostatin A induces cell-cycle arrest/apoptosis and hepatocyte differentiation in human hepatoma cells. *Int J Cancer* 2003;103:572-6.

98. Munster PN, Troso-Sandoval T, Rosen N, Rifkind R, Marks PA, and Richon VM. The histone deacetylase inhibitor suberoylanilide hydroxamic acid induces differentiation of human breast cancer cells. *Cancer Res* 2001;61:8492-7.

99. Grunstein M. Histone acetylation in chromatin structure and transcription. *Nature* 1997;389:349-52.

100. Hoshikawa Y, Kwon HJ, Yoshida M, Horinouchi S, and Beppu T. Trichostatin A induces morphological changes and gelsolin expression by inhibiting histone deacetylase in human carcinoma cell lines. *Exp Cell Res* 1994;214:189-97.

101. Kwon HJ, Kim MS, Kim MJ, Nakajima H, and Kim KW. Histone deacetylase inhibitor FK228 inhibits tumor angiogenesis. *Int J Cancer* 2002;97:290-6.

102. Sauvageot CM, Weatherbee JL, Kesari S, Winters SE, Barnes J, Dellagatta J, et al. Efficacy of the HSP90 inhibitor 17-AAG in human glioma cell lines and tumorigenic glioma stem cells. *Neuro Oncol* 2009;11:109-21.

103. Fukuyo Y, Hunt CR, and Horikoshi N. Geldanamycin and its anti-cancer activities. *Cancer Lett* 2010;290:24-35.

104. Ruan K, Fang X, and Ouyang G. MicroRNAs: novel regulators in the hallmarks of human cancer. *Cancer Lett* 2009;285:116-26.

105. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010;464:1071-6.

106. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* 2011;71:6320-6.

107. Geng YJ, Xie SL, Li Q, Ma J, and Wang GY. Large Intervening Non-coding RNA HOTAIR is Associated with Hepatocellular Carcinoma Progression. J Int Med Res 2011;39:2119-28.

108. Yang Z, Zhou L, Wu LM, Lai MC, Xie HY, Zhang F, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol 2011;18:1243-50.

109. Kim K, Jutooru I, Chadalapaka G, Johnson G, Frank J, Burghardt R, et al. HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. Oncogene 2012;

110. Niinuma T, Suzuki H, Nojima M, Nosho K, Yamamoto H, Takamaru H, et al. Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors. Cancer Res 2012;

111. Huang KC, Rao PH, Lau CC, Heard E, Ng SK, Brown C, et al. Relationship of XIST expression and responses of ovarian cancer to chemotherapy. Mol Cancer Ther 2002;1:769-76.

112. Kawakami T, Zhang C, Taniguchi T, Kim CJ, Okada Y, Sugihara H, et al. Characterization of loss-of-inactive X in Klinefelter syndrome and female-derived cancer cells. Oncogene 2004;23:6163-9.

113. Ganesan S, Silver DP, Greenberg RA, Avni D, Drapkin R, Miron A, et al. BRCA1 supports XIST RNA concentration on the inactive X chromosome. Cell 2002;111:393-405.

114. Looijenga LH, Gillis AJ, van Gurp RJ, Verkerk AJ, and Oosterhuis JW. X inactivation in human testicular tumors. XIST expression and androgen receptor methylation status. Am J Pathol 1997;151:581-90.

115. Wu ZS, Lee JH, Kwon JA, Kim SH, Han SH, An JS, et al. Genetic alterations and chemosensitivity profile in newly established human renal collecting duct carcinoma cell lines. BJU Int 2009;103:1721-8.

116. McDonald HL, Gascoyne RD, Horsman D, and Brown CJ. Involvement of the X chromosome in non-Hodgkin lymphoma. Genes Chromosomes Cancer 2000;28:246-57.

117. Lerner M, Harada M, Loven J, Castro J, Davis Z, Oscier D, et al. DLEU2, frequently deleted in malignancy, functions as a critical host gene of the cell cycle inhibitory microRNAs miR-15a and miR-16-1. Exp Cell Res 2009;315:2941-52.

118. Perez DS, Hoage TR, Pritchett JR, Ducharme-Smith AL, Halling ML, Ganapathiraju SC, et al. Long, abundantly expressed non-coding transcripts are altered in cancer. Hum Mol Genet 2008;17:642-55.

119. Li L, Feng T, Lian Y, Zhang G, Garen A, and Song X. Role of human noncoding RNAs in the control of tumorigenesis. Proc Natl Acad Sci U S A 2009;106:12956-61.

120. Schmidt LH, Spieker T, Koschmieder S, Humberg J, Jungen D, Bulk E, et al. The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. J Thorac Oncol 2011;6:1984-92.

121. Xu C, Yang M, Tian J, Wang X, and Li Z. MALAT-1: a long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. Int J Oncol 2011;39:169-75.

122. Tano K, Mizuno R, Okada T, Rakwal R, Shibato J, Masuo Y, et al. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. FEBS Lett 2010;584:4575-80.

123. Fellenberg J, Bernd L, Delling G, Witte D, and Zahlten-Hinguranage A. Prognostic significance of drug-regulated genes in high-grade osteosarcoma. Mod Pathol 2007;20:1085-94.

124. Lai MC, Yang Z, Zhou L, Zhu QQ, Xie HY, Zhang F, et al. Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. Med Oncol 2011;

125. Benetatos L, Dasoula A, Hatzimichael E, Georgiou I, Syrrou M, and Bourantas KL. Promoter hypermethylation of the MEG3 (DLK1/MEG3) imprinted gene in multiple myeloma. Clin Lymphoma Myeloma 2008;8:171-5.

126. Benetatos L, Hatzimichael E, Dasoula A, Dranitsaris G, Tsiara S, Syrrou M, et al. CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes. Leuk Res 2010;34:148-53.

127. Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, et al. microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. Oncogene 2011;30:4750-6.

128. Mawrin C and Perry A. Pathological classification and molecular genetics of meningiomas. J Neurooncol 2010;99:379-91.

129. Kim YS, Hwan JD, Bae S, Bae DH, and Shick WA. Identification of differentially expressed genes using an annealing control primer system in stage III serous ovarian carcinoma. BMC Cancer 2010;10:576.

130. Wang F, Li X, Xie X, Zhao L, and Chen W. UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. FEBS Lett 2008;582:1919-27.

131. Askarian-Amiri ME, Crawford J, French JD, Smart CE, Smith MA, Clark MB, et al. SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. Rna 2011;17:878-91.

132. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, et al. Gene expression profiling of gliomas strongly predicts survival. Cancer Res 2004;64:6503-10.

133. Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. Cancer Cell 2006;9:287-300.

134. The Cancer Genome Atlas (TCGA). [cited 2012; Available from: http://cancergenome.nih.gov/].

135. Fearon ER and Cadigan KM. Cell biology. Wnt signaling glows with RNAi. Science 2005;308:801-3.

136. Karim R, Tse G, Putti T, Scolyer R, and Lee S. The significance of the Wnt pathway in the pathology of human cancers. Pathology 2004;36:120-8.

137. Mazieres J, He B, You L, Xu Z, and Jablons DM. Wnt signaling in lung cancer. Cancer Lett 2005;222:1-10.

138. Kim Y, Kim KH, Lee J, Lee YA, Kim M, Lee SJ, et al. Wnt activation is implicated in glioblastoma radioresistance. Lab Invest 2012;92:466-73.

139. Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, et al. Post-analysis follow-up and validation of microarray experiments. Nat Genet 2002;32 Suppl:509-14.

140. Miron M, Woody OZ, Marcil A, Murie C, Sladek R, and Nadon R. A methodology for global validation of microarray experiments. BMC Bioinformatics 2006;7:333.

141. Ishigami T, Uzawa K, Fushimi K, Saito K, Kato Y, Nakashima D, et al. Inhibition of ICAM2 induces radiosensitization in oral squamous cell carcinoma cells. Br J Cancer 2008;98:1357-65.

142. Pils D, Wittinger M, Petz M, Gugerell A, Gregor W, Alfanz A, et al. BAMBI is overexpressed in ovarian cancer and co-translocates with Smads into the nucleus upon TGF-beta treatment. Gynecol Oncol 2010;117:189-97.

143. Fagiani E, Lorentz P, Kopfstein L, and Christofori G. Angiopoietin-1 and -2 exert antagonistic functions in tumor angiogenesis, yet both induce lymphangiogenesis. Cancer Res 2011;71:5717-27.

144. Holopainen T, Saharinen P, D'Amico G, Lampinen A, Eklund L, Sormunen R, et al. Effects of angiopoietin-2-blocking antibody on endothelial cell-cell junctions and lung metastasis. J Natl Cancer Inst 2012;104:461-75.

145. Yuan HT, Khankin EV, Karumanchi SA, and Parikh SM. Angiopoietin 2 is a partial agonist/antagonist of Tie2 signaling in the endothelium. Mol Cell Biol 2009;29:2011-22.

146. Carmeliet P. Angiogenesis in health and disease. Nat Med 2003;9:653-60.

147. Hwang SL, Lieu AS, Chang JH, Cheng TS, Cheng CY, Lee KS, et al. Rac2 expression and mutation in human brain tumors. Acta Neurochir (Wien) 2005;147:551-4; discussion 54.

148. Guo F, Cancelas JA, Hildeman D, Williams DA, and Zheng Y. Rac GTPase isoforms Rac1 and Rac2 play a redundant and crucial role in T-cell development. Blood 2008;112:1767-75.

149. Bhatti P, Stewart PA, Hutchinson A, Rothman N, Linet MS, Inskip PD, et al. Lead exposure, polymorphisms in genes related to oxidative stress, and risk of adult brain tumors. Cancer Epidemiol Biomarkers Prev 2009;18:1841-8.

150. Zhou Y, Zhang J, Liu Q, Bell R, Muruve DA, Forsyth P, et al. The chemokine GRO-alpha (CXCL1) confers increased tumorigenicity to glioma cells. Carcinogenesis 2005;26:2058-68.

151. Tessema M, Yingling CM, Grimes MJ, Thomas CL, Liu Y, Leng S, et al. Differential Epigenetic Regulation of TOX Subfamily High Mobility Group Box Genes in Lung and Breast Cancers. PLoS One 2012;7:e34850.

152. Ando T, Ishiguro H, Kimura M, Mitsui A, Kurehara H, Sugito N, et al. Decreased expression of NDRG1 is correlated with tumor progression and poor prognosis in patients with esophageal squamous cell carcinoma. Dis Esophagus 2006;19:454-8.

153. Bandyopadhyay S, Pai SK, Gross SC, Hirota S, Hosobe S, Miura K, et al. The Drg-1 gene suppresses tumor metastasis in prostate cancer. Cancer Res 2003;63:1731-6.

154. Bandyopadhyay S, Pai SK, Hirota S, Hosobe S, Takano Y, Saito K, et al. Role of the putative tumor metastasis suppressor gene Drg-1 in breast cancer progression. Oncogene 2004;23:5675-81.

155. Strzelczyk B, Szulc A, Rzepko R, Kitowska A, Skokowski J, Szutowicz A, et al. Identification of high-risk stage II colorectal tumors by combined analysis of the NDRG1 gene expression and the depth of tumor invasion. Ann Surg Oncol 2009;16:1287-94.

156. Khatri P and Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics 2005;21:3587-95.

157. Huang Y, Sitwala K, Bronstein J, Sanders D, Dandekar M, Collins C, et al. Identification and characterization of Hoxa9 binding sites in hematopoietic cells. Blood 2012;119:388-98.

158. Brandes AA, Tosoni A, Franceschi E, Reni M, Gatta G, and Vecht C. Glioblastoma in adults. Crit Rev Oncol Hematol 2008;67:139-52.

159. Lopes MB. Angiogenesis in brain tumors. Microsc Res Tech 2003;60:225-30.

160. Kaur B, Khwaja FW, Severson EA, Matheny SL, Brat DJ, and Van Meir EG. Hypoxia and the hypoxia-inducible-factor pathway in glioma growth and angiogenesis. Neuro Oncol 2005;7:134-53.

161. Carmeliet P, Dor Y, Herbert JM, Fukumura D, Brusselmans K, Dewerchin M, et al. Role of HIF-1alpha in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. Nature 1998;394:485-90.

162. Blum R, Jacob-Hirsch J, Amariglio N, Rechavi G, and Kloog Y. Ras inhibition in glioblastoma down-regulates hypoxia-inducible factor-1alpha, causing glycolysis shutdown and cell death. Cancer Res 2005;65:999-1006.

163. Marie SK and Shinjo SM. Metabolism and brain cancer. Clinics (Sao Paulo) 2011;66 Suppl 1:33-43.

164. Warburg O, Wind F, and Negelein E, The Metabolism of Tumours: Investigations from the Kaiser Wilhelm Institute for Biology., (eds). Metabolism of Tumours in the Body, London: Constable & Co. 1930.

165. Potter VR. The biochemical approach to the cancer problem. Fed Proc 1958;17:691-7.

166. Vander Heiden MG, Cantley LC, and Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science 2009;324:1029-33.

167. Ochsenbein AF, Klenerman P, Karrer U, Ludewig B, Pericin M, Hengartner H, et al. Immune surveillance against a solid tumor fails because of immunological ignorance. Proc Natl Acad Sci U S A 1999;96:2233-8.

168. Raman D, Sobolik-Delmaire T, and Richmond A. Chemokines in health and disease. Exp Cell Res 2011;317:575-89.

169. Balkwill F. Cancer and the chemokine network. Nat Rev Cancer 2004;4:540-50.

170. Gottesman MM, Fojo T, and Bates SE. Multidrug resistance in cancer: role of ATP-dependent transporters. Nat Rev Cancer 2002;2:48-58.

171. Vaupel P, Thews O, and Hoeckel M. Treatment resistance of solid tumors: role of hypoxia and anemia. Med Oncol 2001;18:243-59.

172. Bhalla KN. Microtubule-targeted anticancer agents and apoptosis. Oncogene 2003;22:9075-86.

173. Sathornsumetee S and Reardon DA. Targeting multiple kinases in glioblastoma multiforme. Expert Opin Investig Drugs 2009;18:277-92.

174. Helleday T, Petermann E, Lundin C, Hodgson B, and Sharma RA. DNA repair pathways as targets for cancer therapy. Nat Rev Cancer 2008;8:193-204.

175. Markovits J, Pommier Y, Kerrigan D, Covey JM, Tilchen EJ, and Kohn KW. Topoisomerase II-mediated DNA breaks and cytotoxicity in relation to cell proliferation and the cell cycle in NIH 3T3 fibroblasts and L1210 leukemia cells. Cancer Res 1987;47:2050-5.

176. Liu L and Gerson SL. Targeted modulation of MGMT: clinical implications. Clin Cancer Res 2006;12:328-31.

177. Kim JH, Shin JH, and Kim IH. Susceptibility and radiosensitization of human glioblastoma cells to trichostatin A, a histone deacetylase inhibitor. Int J Radiat Oncol Biol Phys 2004;59:1174-80.

178. Kim MS, Blake M, Baek JH, Kohlhagen G, Pommier Y, and Carrier F. Inhibition of histone deacetylase increases cytotoxicity to anticancer drugs targeting DNA. Cancer Res 2003;63:7291-300.

179. Huarte M and Rinn JL. Large non-coding RNAs: missing links in cancer? Hum Mol Genet 2010;19:R152-61.

180. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007;129:1311-23.

181. He S, Liu S, and Zhu H. The sequence, structure and evolutionary features of HOTAIR in mammals. BMC Evol Biol 2011;11:102.

182. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science 2010;329:689-93.

183. Prensner JR and Chinnaiyan AM. The emergence of lncRNAs in cancer biology. Cancer Discov 2011;1:391-407.

184. Flynn RA and Chang HY. Active chromatin and noncoding RNAs: an intimate relationship. Curr Opin Genet Dev 2011;

185. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 2004;116:499-509.

186. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, and Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A 2008;105:716-21.

187. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 2009;106:11667-72.

188. Cha TL, Zhou BP, Xia W, Wu Y, Yang CC, Chen CT, et al. Akt-mediated phosphorylation of EZH2 suppresses methylation of lysine 27 in histone H3. Science 2005;310:306-10.

189. Bei L, Lu Y, and Eklund EA. HOXA9 activates transcription of the gene encoding gp91Phox during myeloid differentiation. J Biol Chem 2005;280:12359-70.

190. Bruhl T, Urbich C, Aicher D, Acker-Palmer A, Zeiher AM, and Dimmeler S. Homeobox A9 transcriptionally regulates the EphB4 receptor to modulate endothelial cell migration and tube formation. Circ Res 2004;94:743-51.

191. Siddharthan R. Dinucleotide weight matrices for predicting transcription factor binding sites: generalizing the position weight matrix. PLoS One 2010;5:e9722.

Annex I

Supplementary Table 1| Sequence of primers used for PCR analyses.

Gene	Primer Sense	Primer Antisense	Primer Tm (ºC)
ICAM2	GGATCCCAGAGCTACCCTTC	CGTGTCATGGGAGATGTTTG	59
BAMBI	CTTGCAAGCACGACAGACAT	GAAGTCAGCTCCTGCACCTT	58
ANGPT2	ATAAGCAGCATCAGCCAACC	CCTTGAGCGAATAGCCTGAG	61
PDGFRB	ATAAGCAGCATCAGCCAACC	CCTTGAGCGAATAGCCTGAG	57
RAC2	CAGCACACCCATCATCCTG	CCTCTCTGGGTGAGAGCTGA	61
CXCL1	AGGGAATTCACCCCAAGAAC	TGTTCAGCATCTTTTCGATGA	60
NDRG1	CTCGCTGAGGCCTTCAAGTA	AGAGAAGTGACGCTGGAACC	60
TOX2	CTTCCCGCACATCTCTGAGT	TGAGGTAGAGCGATTTGTCC	58
HOXA9	GCCCGTGCAGCTTCCAGTCC	GAGCGCGCATGAAGCCAGTTG	61
hGUS	CCTGTGACCTTTGTGAGCAA	GTGCCCGTAGTCGTGATACC	57
HOTAIR	CAGTGGGGAACTCTGACTCG	GTGCCTGGTGCTCTCTTACC	60
WNT6	CGAAATGGAGGCAGCTTCT	GACGAGAAGTCGAGGCTCTTT	60
HOTAIR (ChIP)	ATGGACGCTCTCGTTTGTTC	CGGGTGCAAGATAAACCACT	60
WNT6 1	CAGGGGCATCAAAGACATTT	TCAAGAGATCGAGGGGTCAG	60
WNT6 2	CAGGCCAACTTCCTCTCTTG	GAAGGGCTGGGAAGAAGAGT	60
SELE 1	GCATCGTGGATATTCCCGGGAAAG	CAGCTGAACACTACTTCGGCTGAGG	68
SELE 2	CTACCACAACTACATGAGAGACACTAC	CTTTCCCGGGAATATCCACGATGC	58
SELE 3	ATCTACCTTGTGAGTCATTC	TAGTTGTGGTAGTAATTAGAAT	46

For all genes tested, the PCR parameters were as follows: 4 minutes at 94ºC, 35 cycles of denaturation for 30 seconds at 94ºC, annealing for 30 seconds (at specific primer Tm temperature), extension at 72ºC for 30 seconds, and final extension at 72ºC for 8 minutes. For qPCR, all parameters were identical to conventional PCR, except the number of cycles that was extended to 45, and final extension was performed by increasing the temperature in 1ºC each 5 seconds from 65ºC to 95ºC.

Annex II

Supplementary Table 2| Top 50 of the differentially expressed genes upregulated on the microarray data of U87MG-HOXA9 against its negative counterpart U87MG-MSCV.

Gene Name	Gene Symbol	Log Fold-Change	*p*-value
NK2 homeobox 1	NKX2-1	5,72	3,10E-05
intercellular adhesion molecule 2	ICAM2	3,65	5,69E-07
sidekick homolog 2 (chicken)	SDK2	3,48	3,09E-05
BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	BAMBI	3,38	1,04E-06
serpin peptidase inhibitor, clade B (ovalbumin), member 2	SERPINB2	3,29	7,19E-04
serine peptidase inhibitor, Kazal type 1	SPINK1	3,17	1,23E-05
hypothetical protein LOC283454	LOC283454	3,13	3,03E-06
chromosome 1 open reading frame 133	C1orf133	3,06	5,00E-06
serpin peptidase inhibitor, clade B (ovalbumin), member 2	SERPINB2	3,03	4,83E-05
FYVE, RhoGEF and PH domain containing 3	FGD3	3,00	1,78E-05
transmembrane protein 47	TMEM47	2,94	1,23E-06
protein phosphatase 1, regulatory (inhibitor) subunit 14C	PPP1R14C	2,90	1,97E-04
harakiri, BCL2 interacting protein (contains only BH3 domain)	HRK	2,86	2,58E-08
SH2 domain containing 5	SH2D5	2,85	8,44E-06
ring finger protein 43	RNF43	2,74	1,10E-06
C1q and tumor necrosis factor related protein 2	C1QTNF2	2,70	7,49E-06
selenocysteine lyase	SCLY	2,69	7,85E-06
uridine phosphorylase 1	UPP1	2,68	1,97E-07
heparan sulfate (glucosamine) 3-O-sulfotransferase 2	HS3ST2	2,65	6,68E-06
frizzled homolog 8 (Drosophila)	FZD8	2,62	3,79E-05
lymphocyte cytosolic protein 1 (L-plastin)	LCP1	2,62	2,06E-07
prostate transmembrane protein, androgen induced 1	PMEPA1	2,57	1,00E-05
claudin 14	CLDN14	2,56	6,57E-04
protein phosphatase, Mg2+/Mn2+ dependent, 1F	PPM1F	2,51	1,43E-07
C1q and tumor necrosis factor related protein 2	C1QTNF2	2,50	8,70E-05
RRS1 ribosome biogenesis regulator homolog (S. cerevisiae)	RRS1	2,48	2,09E-07

chemokine (C-C motif) ligand 3	CCL3	2,45	2,00E-05
matrix metallopeptidase 1 (interstitial collagenase)	MMP1	2,40	9,63E-08
interleukin 22 receptor, alpha 1	IL22RA1	2,38	6,99E-05
sphingosine kinase 1	SPHK1	2,37	1,59E-04
pleckstrin homology domain containing, family G (with RhoGef domain) member 3	PLEKHG3	2,36	3,17E-07
G0/G1switch 2	G0S2	2,35	1,47E-05
histone cluster 1, H1a	HIST1H1A	2,34	5,46E-04
hypothetical LOC284344	LOC284344	2,34	3,76E-05
ribosomal RNA processing 9, small subunit (SSU) processome component, homolog (yeast)	RRP9	2,32	7,94E-04
SMAD family member 7	SMAD7	2,28	1,41E-05
BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	BAMBI	2,27	2,98E-07
solute carrier family 7, (cationic amino acid transporter, y+ system) member 11	SLC7A11	2,24	9,82E-06
pregnancy specific beta-1-glycoprotein 4	PSG4	2,24	2,20E-07
vitamin D (1,25- dihydroxyvitamin D3) receptor	VDR	2,18	1,09E-04
protein phosphatase, Mg2+/Mn2+ dependent, 1F	PPM1F	2,17	7,48E-06
spindle and kinetochore associated complex subunit 1	SKA1	2,15	5,16E-06
tribbles homolog 3 (Drosophila)	TRIB3	2,15	4,24E-04
chromosome 20 open reading frame 82	C20orf82	2,14	7,06E-06
glycine-N-acyltransferase-like 1	GLYATL1	2,12	6,99E-07
lymphocyte cytosolic protein 1 (L-plastin)	LCP1	2,08	1,84E-03
Rho family GTPase 3	RND3	2,08	3,2E-03
chromosome 16 open reading frame 57	C16orf57	2,08	3,50E-07
chromosome 14 open reading frame 34	C14orf34	2,05	4,49E-05
PSMC3 interacting protein	PSMC3IP	2,04	1,32E-04

Supplementary Table 3| Top 50 of the differentially expressed genes downregulated on the microarray data of U87MG-HOXA9 against its negative counterpart U87MG-MSCV.

Gene Name	Gene Symbol	Log Fold-Change	*p*-value
angiopoietin 2	ANGPT2	-6,41	1,57E-11
fer-1-like 4 (C. elegans)	FER1L4	-5,31	2,84E-07
chromosome 10 open reading frame 10	C10orf10	-4,66	1,86E-06
chromosome 10 open reading frame 10	C10orf10	-4,65	1,67E-05
olfactomedin-like 2A	OLFML2A	-4,61	1,41E-06
fibronectin 1	FN1	-4,22	1,28E-05
fibronectin 1	FN1	-4,20	1,49E-07
family with sequence similarity 133, member A	FAM133A	-4,06	2,09E-07
selenium binding protein 1	SELENBP1	-3,84	5,70E-08
yippee-like 4 (Drosophila)	YPEL4	-3,82	4,01E-08
family with sequence similarity 133, member A	FAM133A	-3,81	1,73E-07
apolipoprotein H (beta-2-glycoprotein I)	APOH	-3,72	1,26E-06
sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5B	SEMA5B	-3,66	1,79E-07
proline-rich transmembrane protein 2	PRRT2	-3,65	3,31E-04
apolipoprotein E	APOE	-3,60	7,63E-07
lymphocyte-specific protein 1	LSP1	-3,50	3,58E-06
nuclear factor (erythroid-derived 2), 45kDa	NFE2	-3,49	4,82E-06
C-type lectin domain family 3, member B	CLEC3B	-3,48	3,04E-07
Janus kinase 3	JAK3	-3,46	2,91E-05
platelet-derived growth factor receptor, beta polypeptide	PDGFRB	-3,45	1,79E-07
death associated protein-like 1	DAPL1	-3,32	4,01E-08
major histocompatibility complex, class II, DR alpha	HLA-DRA	-3,27	3,94E-06
acetyl-CoA acetyltransferase 2	ACAT2	-3,26	7,62E-09
cadherin 7, type 2	CDH7	-3,24	1,59E-07
zona pellucida glycoprotein 1 (sperm receptor)	ZP1	-3,23	4,20E-05
MACRO domain containing 2	MACROD2	-3,23	2,61E-06

WNT1 inducible signaling pathway protein 2	WISP2	-3,23	9,50E-05
hypothetical LOC197187	MGC23284	-3,22	1,01E-07
MRS2 magnesium homeostasis factor homolog (S. cerevisiae)	MRS2	-3,15	2,91E-05
odz, odd Oz/ten-m homolog 2 (Drosophila)	ODZ2	-3,15	2,28E-05
SH3 and cysteine rich domain 2	STAC2	-3,12	1,13E-06
aldolase C, fructose-bisphosphate	ALDOC	-3,10	4,44E-05
kelch-like 4 (Drosophila)	KLHL4	-3,07	1,80E-08
DIRAS family, GTP-binding RAS-like 1	DIRAS1	-3,07	3,30E-08
family with sequence similarity 46, member A	FAM46A	-3,04	7,83E-07
carbonic anhydrase IX	CA9	-3,02	4,95E-04
fibronectin 1	FN1	-2,98	3,33E-05
calcyphosine	CAPS	-2,95	2,57E-05
WAP four-disulfide core domain 10B	WFDC10B	-2,95	1,96E-07
SLIT and NTRK-like family, member 6	SLITRK6	-2,94	3,08E-07
collagen, type I, alpha 1	COL1A1	-2,92	3,49E-05
chitinase 3-like 1 (cartilage glycoprotein-39)	CHI3L1	-2,90	1,08E-05
CAP-GLY domain containing linker protein 3	CLIP3	-2,90	5,07E-07
eEF1A2 binding protein	DKFZp434B1231	-2,89	2,28E-03
neurexin 2	NRXN2	-2,87	4,52E-06
amiloride-sensitive cation channel 2, neuronal	ACCN2	-2,85	1,73E-04
glucosidase, alpha; acid	GAA	-2,84	9,12E-04
apolipoprotein C-I	APOC1	-2,81	5,35E-05
sperm associated antigen 4	SPAG4	-2,79	7,31E-06
interleukin 1 receptor accessory protein-like 1	IL1RAPL1	-2,79	1,02E-04

Supplementary Table 4| Top 50 of the differentially expressed genes upregulated on the microarray data of hTERT/E6/E7-HOXA9 against its negative counterpart hTERT/E6/E7-MSCV.

Gene Name	Gene Symbol	Log Fold-Change	*p*-value
chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	CXCL1	4,51	1,43E-05
natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	NPR3	4,50	1,95E-04
ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	RAC2	4,12	2,31E-07
ribosomal protein L39-like	RPL39L	3,97	3,40E-09
serine/threonine protein kinase MST4	MST4	3,75	2,08E-06
elastin microfibril interfacer 2	EMILIN2	3,68	1,15E-05
placenta-specific 8	PLAC8	3,60	4,24E-06
placenta-specific 8	PLAC8	3,54	1,32E-05
collagen, type III, alpha 1	COL3A1	2,99	1,98E-07
natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	NPR3	2,95	5,06E-05
ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	RAC2	2,94	1,26E-06
hypothetical LOC100506305	LOC100506305	2,85	1,91E-06
cytochrome b5 reductase 2	CYB5R2	2,76	1,76E-05
O-6-methylguanine-DNA methyltransferase	MGMT	2,72	1,53E-07
DENN/MADD domain containing 2D	DENND2D	2,67	6,47E-07
beta-site APP-cleaving enzyme 2	BACE2	2,64	1,73E-05
platelet-derived growth factor receptor-like	PDGFRL	2,62	4,68E-08
beta-site APP-cleaving enzyme 2	BACE2	2,60	2,75E-05
collagen, type I, alpha 2	COL1A2	2,49	2,02E-05
syndecan 2	SDC2	2,24	1,04E-05
protocadherin 17	PCDH17	2,19	1,52E-05
chromosome X open reading frame 57	CXorf57	2,11	8,22E-06
desmocollin 3	DSC3	2,05	3,91E-04
desmoplakin	DSP	1,97	1,87E-04
zinc finger protein 558	ZNF558	1,92	6,25E-07
melanoma antigen family D, 4B	MAGED4B	1,89	2,81E-04
ankyrin 1, erythrocytic	ANK1	1,85	1,94E-04

cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	CDKN2B	1,84	3,2E-04
zinc finger protein 469	ZNF469	1,84	1,31E-04
chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)	CXCL6	1,82	1,40E-05
hypothetical LOC100287221	LOC100287221	1,79	7,77E-06
surfactant associated 1 (pseudogene)	SFTA1P	1,75	4,61E-05
phosphatidic acid phosphatase type 2B	PPAP2B	1,72	1,67E-05
cadherin 6, type 2, K-cadherin (fetal kidney)	CDH6	1,71	5,31E-05
ABI family, member 3 (NESH) binding protein	ABI3BP	1,70	2,14E-04
cadherin 6, type 2, K-cadherin (fetal kidney)	CDH6	1,60	6,01E-09
sarcoglycan, alpha (50kDa dystrophin-associated glycoprotein)	SGCA	1,53	1,24E-05
hairy and enhancer of split 1, (Drosophila)	HES1	1,52	6,69E-05
arylsulfatase family, member J	ARSJ	1,50	1,90E-05
zinc finger, C4H2 domain containing	ZC4H2	1,50	5,24E-07
hypothetical protein LOC731479	LOC731479	1,48	2,58E-04
collagen, type XI, alpha 1	COL11A1	1,39	1,13E-06
inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	ID3	1,36	5,49E-06
AE binding protein 1	AEBP1	1,35	3,28E-04
coiled-coil domain containing 80	CCDC80	1,32	4,23E-04
hypothetical LOC100505894	LOC100505894	1,31	1,01E-04
chemokine (C-X-C motif) ligand 5	CXCL5	1,28	2,46E-05
inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	ID4	1,28	2,38E-04
desmocollin 3	DSC3	1,26	1,30E-07
chromodomain helicase DNA binding protein 7	CHD7	1,25	4,20E-05

Supplementary Table 5| Top 50 of the differentially expressed genes downregulated on the microarray data of hTERT/E6/E7-HOXA9 against its negative counterpart hTERT/E6/E7-MSCV.

Gene Name	Gene Symbol	Log Fold-Change	*p*-value
X (inactive)-specific transcript (non-protein coding)	XIST	-6,44	1,71E-10
chromosome 10 open reading frame 35	C10orf35	-4,52	2,74E-07
chemokine (C-X-C motif) receptor 7	CXCR7	-4,41	3,51E-06
N-myc downstream regulated 1	NDRG1	-3,59	9,06E-05
TOX high mobility group box family member 2	TOX2	-3,49	1,56E-08
HtrA serine peptidase 3	HTRA3	-3,46	3,47E-08
mannosidase, alpha, class 1C, member 1	MAN1C1	-3,30	2,95E-06
transmembrane protein 158 (gene/pseudogene)	TMEM158	-3,25	1,68E-05
insulin-like growth factor 2 (somatomedin A)	IGF2	-3,15	3,50E-07
CUB domain containing protein 1	CDCP1	-3,14	4,88E-06
extracellular leucine-rich repeat and fibronectin type III domain containing 2	ELFN2	-3,12	1,78E-07
G0/G1switch 2	G0S2	-3,03	3,34E-06
cellular repressor of E1A-stimulated genes 1	CREG1	-3,03	2,29E-06
G protein-coupled receptor 68	GPR68	-2,99	2,03E-07
docking protein 7	DOK7	-2,96	9,07E-05
paraneoplastic antigen MA3	PNMA3	-2,94	5,06E-05
bradykinin receptor B1	BDKRB1	-2,87	6,03E-05
IMP (inosine 5'-monophosphate) dehydrogenase 1	IMPDH1	-2,80	1,80E-05
bradykinin receptor B1	BDKRB1	-2,78	0,000169435
paralemmin	PALM	-2,76	4,12E-06
tumor protein p53 inducible protein 11	TP53I11	-2,69	0,00026756
eEF1A2 binding protein	DKFZp434B1231	-2,62	4,77E-05
coagulation factor III (thromboplastin, tissue factor)	F3	-2,62	3,71E-06
leucine zipper, putative tumor suppressor 1	LZTS1	-2,56	3,17E-06
glucose-fructose oxidoreductase domain containing 1	GFOD1	-2,53	7,98E-06
protein tyrosine phosphatase, receptor type, E	PTPRE	-2,49	0,000383438
collagen, type XIII, alpha 1	COL13A1	-2,49	0,00018848

|			
collagen, type XIII, alpha 1	COL13A1	-2,46	0,000209884
inhibin, beta B	INHBB	-2,41	1,32E-06
cathepsin H	CTSH	-2,37	5,38E-07
sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1	SPOCK1	-2,35	2,98E-05
midline 2	MID2	-2,33	4,10E-05
cell growth regulator with EF-hand domain 1	CGREF1	-2,33	1,38E-05
family with sequence similarity 134, member B	FAM134B	-2,25	6,57E-05
RALBP1 associated Eps domain containing 2	REPS2	-2,22	1,42E-08
hypothetical gene supported by BC013438	LOC375295	-2,20	6,99E-05
glutathione S-transferase mu 3 (brain)	GSTM3	-2,16	9,20E-06
glutathione S-transferase mu 3 (brain)	GSTM3	-2,16	4,81E-05
bradykinin receptor B2	BDKRB2	-2,16	2,61E-06
carboxypeptidase A3 (mast cell)	CPA3	-2,13	2,39E-06
chromosome 14 open reading frame 132	C14orf132	-2,12	1,05E-06
vitamin D (1,25- dihydroxyvitamin D3) receptor	VDR	-2,06	0,000339465
ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 5	ST6GALNAC5	-2,04	8,42E-08
interferon regulatory factor 5	IRF5	-2,01	2,86E-08
solute carrier family 30 (zinc transporter), member 3	SLC30A3	-1,95	7,95E-05
lectin, galactoside-binding, soluble, 3	LGALS3	-1,93	0,000265902
Kv channel interacting protein 3, calsenilin	KCNIP3	-1,92	6,06E-05
Down syndrome critical region gene 1-like 2	DSCR1L2	-1,91	7,04E-06
colony stimulating factor 2 (granulocyte-macrophage)	CSF2	-1,90	2,97E-06
solute carrier organic anion transporter family, member 4A1	SLCO4A1	-1,90	2,62E-05

Annex III

Annex III

A
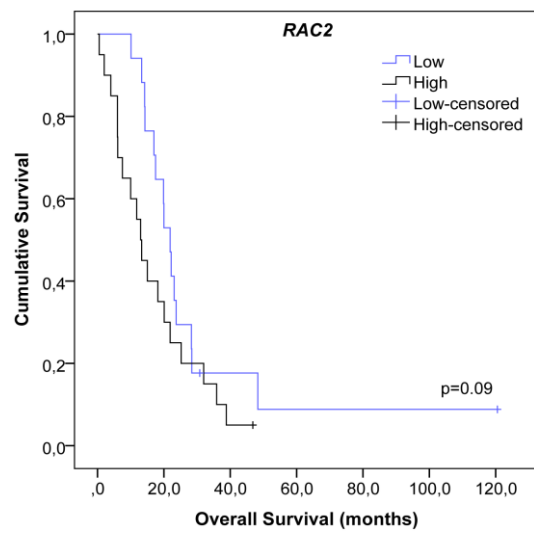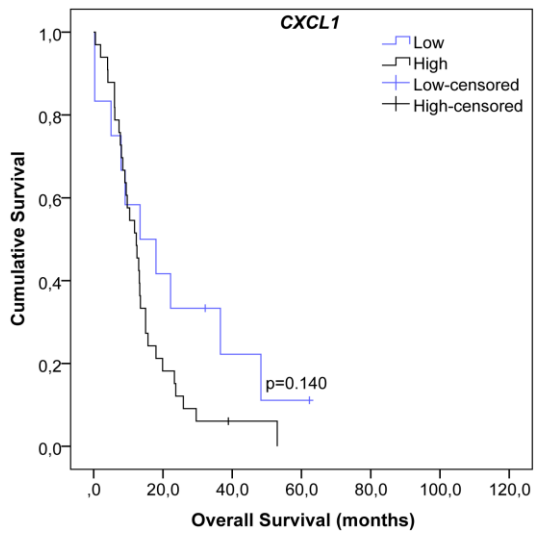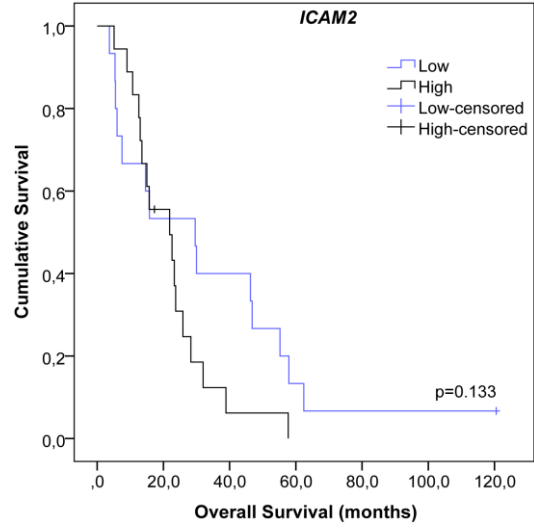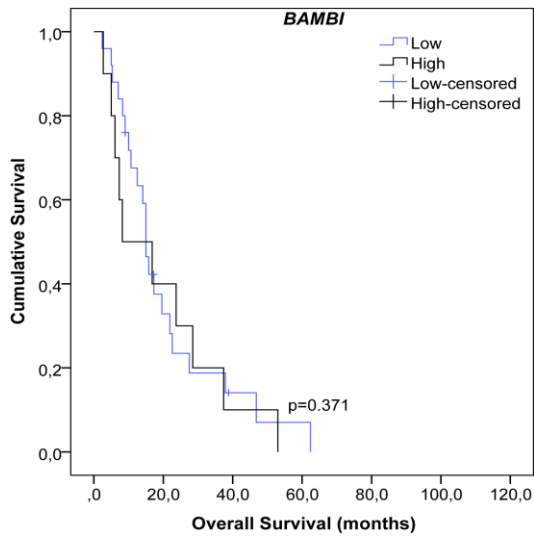
B
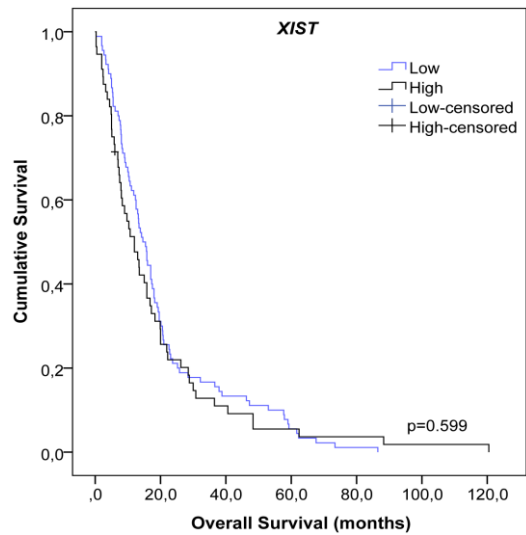
Supplementary Figure 1 (previous page)| Kaplan-Meier overall survival curves of patients from REMBRANDT[78] dataset that presented expression of the genes with higher log fold-change in the microarray data of U87MG **(A)** and hTERT/E6/E7 **(B)** cell lines. The overexpression of *ICAM2*, *BAMBI*, *CXCL1, RAC2* and *XIST* in GBM patients is not associated with overall survival (*ICAM2*, *p*=0.133; *BAMBI*, *p*=0.371; *CXCL1*, *p*=0.140; *RAC2, p*=0.09; *XIST*, *p*=0.599).