

ALTERAÇÕES NOS PADRÕES DE EXPRESSÃO ALELO-ESPECÍFICO  
SUGEREM PERDA OU GANHO DE *IMPRINTING* EM NEOPLASÍAS

**ALAN TARDIN DA SILVA**

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY  
RIBEIRO

CAMPOS DOS GOYTACAZES – RJ

FEVEREIRO DE 2018

ALTERAÇÕES NOS PADRÕES DE EXPRESSÃO ALELO-ESPECÍFICO  
SUGEREM PERDA OU GANHO DE *IMPRINTING* EM NEOPLASÍAS

**ALAN TARDIN DA SILVA**

Dissertação apresentada ao Centro de  
Biotecnologia e Biociências – CBB, da  
Universidade Estadual do Norte  
Fluminense, como parte das exigências  
para obtenção do título de Mestre em  
Biociências e Biotecnologia.

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY  
RIBEIRO

CAMPOS DOS GOYTACAZES – RJ

FEVEREIRO DE 2018

ALTERAÇÕES NOS PADRÕES DE EXPRESSÃO ALELO-ESPECÍFICO  
SUGEREM PERDA OU GANHO DE *IMPRINTING* EM NEOPLASÍAS

**ALAN TARDIN DA SILVA**

Dissertação apresentada ao Centro de Biociências e Biotecnologia – CBB, da Universidade Estadual do Norte Fluminense, como parte das exigências para obtenção do título de Mestre em Biociências e Biotecnologia.

Comissão Examinadora:

---

Dra. Regina Célia de Souza Campos Fernandes (Doutora em Doenças Infecciosas e Parasitárias) – FMC

---

Dr. Victor Martin Quintana Flores (Doutor em Engenharia Genética) -- UENF.

---

Dra. Elena Lassounskaia (Doutora em Imunobiologia de Doenças Infecciosas e Parasitárias) -- IMESP

---

Dr. Enrique Medina-Acosta (Doutor em parasitologia Médica e Molecular) – UENF.

(Orientador)

## DEDICATÓRIA

...A Deus, pela sua presença e constante zelo, por toda paz, conforto e apoio que excede todo o entendimento...

...A toda minha família por estar sempre presente em minha vida, e por toda torcida e compreensão ao longo desses anos...

...A minha querida tia Ilma, pessoa mais tenra que já conheci, que em sua breve jornada pela vida me ensinou que gentileza não faz distinção, cabe e deve ser dever de todos serem gentis com próximo, pois não sabemos a dor que pode existir ali. Tia que foi mãe, tia que foi tia, tia foi pai... Tia que foi tudo e ainda mais.

## AGRADECIMENTOS

Ao **Prof. Dr. Enrique Medina-Acosta**, que ao longo destes três anos de convivência e ensino, mais do que um orientador se tornou um amigo capaz de fazer com que eu também crescesse como pessoa. Agradeço por todo tempo dispensado ao meu crescimento científico e formação profissional, mas principalmente por ter comprado a idéia inicial deste projeto.

Á **Dra. Ana Beatriz Garcia**, por toda amizade, pelo zelo e por sempre ser tão solícita em ajudar, uma pessoa que ao longo do tempo se tornou a “mãe” de todos nós.

Á **Cristina dos Santos Ferreira**, por toda amizade, por todo incentivo, e pela colaboração na construção deste estudo, pelas discussões e ideias compartilhadas, mas principalmente pela ajuda com o Excel, a qual foi fundamental para a finalização deste trabalho.

Á **Dra. Thais Louvain de Souza**, por toda amizade, por todo incentivo, pelo apelido de Zé meningite e pela revisão deste estudo, a qual foi importante para a conclusão deste trabalho.

Ao **Dr. Filipe Brum Machado**, por toda amizade, por todo incentivo, e pelas discussões ideias de como fazer um melhor trabalho.

Aos Professores **Dra. Ana Beatriz Garcia, Dra. Thais Louvain de Souza, Dra. Elena Lassounskaia, Dr. Victor Martin Quintana Flores e Dra. Regina Célia de Souza Campos Fernandes** por aceitarem prontamente o convite para a participação como membros titulares e suplentes para a banca de defesa desta dissertação, e por todas as sugestões e correções que foram essenciais para a finalização deste trabalho.

Aos alunos de Bacharelado em Biologia: **Ronaldo da Silva Francisco Junior, Pedro Thyago Mozer Rodrigues e Douglas Terra Machado**, pela colaboração efetiva nas análises dos experimentos *in silico*, fundamentais para a finalização deste trabalho e produção do manuscrito. Em especial ao amigo **Douglas Terra Machado**, pelas discussões sobre ciência e projetos futuros, pelas ideias sobre como fazer um trabalho melhor, por toda amizade, companheirismo e espírito de equipe.

A toda família NUDIM pelo companheirismo, pelas colaborações nas discussões e análises, mas também pelo suporte na execução deste trabalho o qual foi fundamental na fase final.

Aos amigos: **Hugo Freitas Viegas Fernandes, Pamela Busch, Mirna Lucio, Nayara Seabra, Joice Moraes, Yasmim Barbosa, Yuri Rocha, Maiara Carvalho, Thiago Barbosa de Souza, Graziela Araújo de Sá Machado, Luísa Mignone Paixão, Juliana Ywasaki, Cinthya Neves, Diego Sá**, pelo convívio, companheirismo, e aprendizado mútuo ao longo desses anos. Em especial a amiga **Julia Rodine**, por toda amizade, companheirismo, pelo dom de me ouvir e me entender, por sempre ser a minha Doutora Frida quando precisei, mas principalmente por aguentar minhas pirações.

A todo o **corpo docente da Universidade Estadual do Norte Fluminense Darcy Ribeiro - UENF**, que direta ou indiretamente participaram efetivamente da minha formação acadêmica.

À **Universidade Estadual do Norte Fluminense Darcy Ribeiro– UENF**, em particular ao **Programa de pós-graduação em Biociências e Biotecnologia**, pela oportunidade concedida.

À **FAPERJ** e **CNPq** por todo auxílio técnico e financeiro.

## RESUMO

O volume de dados gerados e processados no curso da pesquisa científica moderna está crescendo exponencialmente. Grandes bancos de dados estão proliferando com vários tipos de dados biológicos, de modo que as análises computacionais vieram a alterar o pensamento atual sobre a arquitetura do tumor. Sendo assim, atualmente pouco se sabe sobre o número de genes supressores de tumor e oncogenes que atuam sobre os processos de carcinogênese, existindo aproximadamente um total de 1090 (mil e noventa) genes supressores de tumor e 800 (oitocentos) oncogenes, totalizando, assim 1890 oncogenes e genes supressores de tumor, aonde apenas os oncogenes e genes supressores de tumor imprintados vêm sendo estudados. Dessa forma, ainda não se sabe qual é a amplitude do efeito de perda ou ganho de *imprinting* em oncogenes e genes supressores de tumor em relação ao conjunto de genes imprintados. Entretanto, já se sabe que alterações aberrantes epigenéticas estão envolvidas no desenvolvimento e progressão do câncer, além de mutações, perdas cromossômicas e surgimento de algumas síndromes. Dentro desse contexto visamos contribuir para um maior entendimento dos processos que envolvem alterações nos perfis de expressão em oncogenes e genes supressores de tumor, além de determinar a amplitude do efeito da perda ou ganho de *imprinting* genômico em tecidos primários humanos com câncer. Para tal estudo realizamos uma metanálise computacional das frações de expressão alélica para polimorfismos de nucleotídeo único (SNPs) em genes sabidamente imprintados e não imprintados, segundo a literatura. As frações alélicas foram estimadas de repositórios públicos de experimentos de RNA-Seq, de tecidos normais e tumorais humanos, estabelecendo unicidade de doador. Aqui demonstramos que após as análises de RNA-Seq secundário em 1890 oncogenes e genes supressores de tumor, obtivemos 640 genes com expressão estrita/consistente com monoalélica, entre eles oncogenes e genes supressores de tumor, pelo uso do consorcio GTEx. Destes apenas 30 genes demonstraram apresentar expressão monoalélica em  $\geq 1$  tecido, de modo que 11 genes apresentaram expressão com pelo menos  $\geq 3$  SNPs monoalélicos em cada tecido, critério o qual serviu para indicar os possíveis candidatos ao *imprinting*. Para as análises de RNA-Seq primário, o número de sequências foi insuficiente ( $\leq 10$  reads) dificultando obtenção de uma análise mais confiável das razões alélicas. Para os SNPs rs4758616, rs73409814 e rs2067051 nos genes *NAP1L4*, *TSSC4* e *H19* respectivamente, foram observadas razões consistentes com expressão bialélica em tumores de fígado e pulmão, sugestivas de perda de *imprinting*. Possível sinal de ganho de *imprinting* em tumor de pulmão para os SNPs rs10840159, rs4758616 e rs2067051 nos genes *HOTS*, *NAP1L4* e *H19* respectivamente. Adicionalmente, observamos frações alélicas consistentes com a perda de *imprinting* nos SNPs rs10840159 e rs2192206 para os genes *H19* e *NDN* em câncer de mama. Os dados sugerem uma possível perda de *imprinting* para alguns genes em tecidos com câncer e possível sinal de ganho de *imprinting* para outros genes em tecido de mama com câncer. Esse estudo também trouxe novos oncogenes e genes supressores de tumor candidatos ao *imprinting*, bem como um panorama de expressão dos mesmos em diferentes tecidos.

**Palavras-chave:** *imprinting*, oncogenes, genes supressores de tumor, câncer.

## ABSTRACT

The volume of data generated and processed in the course of modern scientific research is growing exponentially. Large databases are proliferating with various types of biological data, so that computational analyzes have altered the current thinking about the architecture of the tumor. Thus, there is currently little known about the number of tumor suppressor genes and oncogenes that act on the carcinogenic processes, there being approximately a total of 1090 (one thousand and ninety) tumor suppressor genes and 800 (eight hundred) oncogenes, thus totaling 1890 oncogenes and tumor suppressor genes, where only oncogenes and imprinted tumor suppressor genes have been studied. Thus, it is not yet known what the amplitude of the loss or gain effect of imprinting on oncogenes and tumor suppressor genes in relation to the set of imprinted genes. However, aberrant epigenetic changes are already known to be involved in the development and progression of cancer, as well as mutations, chromosomal losses and the appearance of some syndromes. In this context we aim to contribute to a better understanding of the processes that involve changes in expression profiles in oncogenes and tumor suppressor genes, as well as to determine the amplitude of the effect of loss or gain of genomic imprinting in human primary tissues with cancer. For this study, we performed a computational meta-analysis of the allelic expression fractions for single nucleotide polymorphisms (SNPs) in genes known to be imprinted and not imprinted, according to the literature. Allelic fractions were estimated from public repositories of RNA-Seq experiments, from normal human and tumor tissues, establishing donor uniqueness. Here we demonstrate that after the analyzes of RNA-secondary Seq in 1890 oncogenes and tumor suppressor genes, we obtained 640 genes with strict / consistent expression with monoallelic, among them oncogenes and tumor suppressor genes, by the use of GTEx consortium. Of these, only 30 genes demonstrated to exhibit monoallelic expression in  $\geq 1$  tissue, so that 11 genes had expression with at least  $\geq 3$  monoallelic SNPs in each tissue, a criterion which served to indicate potential candidates for imprinting. For the analyzes of primary RNA-Seq, the number of sequences was insufficient ( $\leq 10$  reads) making it difficult to obtain a more reliable analysis of the allelic ratios. For the SNPs rs4758616, rs73409814 and rs2067051 in the *NAP1L4*, *TSSC4* and *H19* genes respectively, consistent ratios with biallelic expression in liver and lung tumors suggestive of loss of imprinting were observed. Possible imprinting gain signal in lung tumor for the SNPs rs10840159, rs4758616 and rs2067051 in the genes *HOTS*, *NAP1L4* and *H19* respectively. In addition, we observed allelic fractions consistent with loss of imprinting at the SNPs rs10840159 and rs2192206 for the *H19* and *NDN* genes in breast cancer. The data suggest a possible loss of imprinting for some genes in cancer tissues and possible sign of gain of imprinting for other genes in breast cancer tissue. This study also brought new oncogenes and tumor suppressor gene candidates for imprinting, as well as a panorama of expression of the same in different tissues.

**Keywords:** Genomic Imprinting, Oncogenes, RNA-seq, Tumor Suppressor

## LISTA DE FIGURAS

Figura 1. Parâmetros adotados para análise da unicidade dos tecidos. ....	30
Figura 2. Análise Ontológica dos oncogenes e genes supressores de tumor controles (28 genes).....	31
Figure 3. Análise Ontológica dos oncogenes e genes supressores de tumor alvos (1890 genes). ....	33

## TABELAS

Tabela 1. Exemplo da sequência utilizada para a interrogação de cada SNP alvo. ....	21
Tabela 2. Quantidade de amostras, antes e após a análise de cada um dos 24 tecidos primários humanos utilizados no Atlas Saudável. ....	28
Tabela 3. Quantidade de amostras antes e após a análise de cada um dos 24 tecidos primários humanos utilizados no Atlas Neoplásico. ....	29
Tabela 4. Evidência de Imprinting em vários tecidos para os 17 Oncogenes detectados como possivelmente imprintados. ....	35
Tabela 5. Evidência de Imprinting em vários tecidos para os 12 Genes Supressores de Tumor detectados como possivelmente imprintados. ....	36
Tabela 6. Análise do perfil de expressão em experimentos de RNA-Seq de fígado em oncogenes e genes supressores de tumor imprintados. ....	39
Tabela 7. Análise do perfil de expressão em experimentos de RNA-Seq de pulmão em oncogenes e genes supressores imprintados. ....	40
Tabela 8. Análise do perfil de expressão em experimentos de RNA-Seq de mama em oncogenes e genes supressores imprintados. ....	41

## TABELAS SUPLEMENTARES

Tabela 1. Lista dos Oncogenes e Genes Supressores de Tumor Imprintados.....	59
Tabela 2. Lista dos Oncogenes Alvos.....	61
Tabela 3. Lista dos Genes Supressores de Tumor Alvos.....	61
Tabela 4. Análise do perfil de expressão em experimentos de RNA-Seq em oncogenes e genes supressores imprintados.....	61

## LISTA DE ABREVIATURA

ASE - Expressão Alelo Específica (*Allelic specific expression*)

BED - Navegador de dados Extensivos (*Browser Extensible Data*)

BLAST - Ferramenta Básica de Pesquisa e Alinhamento Local (*Basic Local Alignment Search Tool*)

dbGaP - Base de Dados de Genótipos e Fenótipos (*Database of genotypes and phenotypes*)

dbSNP - Banco de Dados de Polimorfismo de Nucleotídeo Únicos (*Single Nucleotide Polymorphism database*)

DMR - Regiões Diferencialmente Metiladas (*Differentially methylated regions*)

EMR - Expressão Monoalélica Randômica (*Random monoallelic expression*)

GOI - Ganho de Imprinting (*Gain of imprinting*)

GTEX - Expressão Genotípica-Tecidual (*Genotype-Tissue Expression*)

HCC - Carcinoma Hepatocelular (*Hepatocellular carcinoma*)

ICR - Regiões Controladoras de Imprinting (*Imprinting control regions*)

IUPAC - União Internacional de Química Pura e Aplicada (*International Union of Pure and Applied Chemistry*)

LOH - Perda de Heterozigose (*Loss of heterozigose*)

LOI - Perda de Imprinting (*Loss of imprinting*)

MAF - Frequência do Menor Alelo (*Minor allele frequency*)

NCBI - Centro Nacional de Informações sobre Biotecnologia (*National Center for Biotechnology Information*)

RNA-Seq - Sequenciamento Transcriptômico de RNA (*Transcriptomic sequencing of RNA*)

SNP - Polimorfismo de Nucleotídeo Único (Single nucleotide polimorfismo)

UPD - Dissomia Uniparental (Uniparental disomy)

SRA - Arquivo de Sequência de Leitura (Sequence Read Archive)

USCS - Universidade da Califórnia de Santa Cruz (University of Califórnia Santa Cruz)

## ÍNDICE

1.	INTRODUÇÃO.....	7
2.	REVISÃO BIBLIOGRÁFICA .....	9
2.1	<i>Imprinting</i> Genômico .....	9
2.2	Perfil de Expressão Monoalélica nos mamíferos.....	11
2.3	Perda de <i>Imprinting</i> Genômico – LOI .....	12
2.4	Sequenciamento Transcriptômico: RNA-Seq .....	14
2.5	Oncogênese – Regulação e Desregulação .....	15
3	JUSTIFICATIVA.....	17
4	OBJETIVOS.....	18
4.1	Objetivo Geral.....	18
4.2	Objetivos Específicos .....	18
5	MATERIAL E MÉTODOS .....	19
5.1	Seleção de experimentos de RNA-Seq em bancos de dados públicos de RNA-Seq.....	19
5.1.1	RNA-Seq Primário.....	20
5.1.2	RNA-Seq Secundário.....	22
5.2	Seleção e Formatação de Polimorfismos em Genes alvos através de Browsers .....	24
5.3	Integração de dados com a plataforma GTEx .....	24
5.4	Ontologia dos Oncogenes e Genes Supressores de Tumor .....	26
6.	RESULTADOS .....	27
6.1	Consolidação <i>In Silico</i> do Acervo de Experimentos de RNA-Seq de Tecidos Humanos para o Estudo de <i>imprinting</i> Genômico. ....	27
6.2	Ontologia Gênica – Patologia e Funções Biológicas.....	30
6.3	Investigação de Novos Genes Candidatos ao <i>Imprinting</i> - GTEx.....	33
6.4	Análises de RNA-Seq Primário.....	37
7.	DISCUSSÃO.....	42
8.	CONCLUSÕES.....	48
9.	REFERENCIAS .....	49
10.	MATERIAL SUPLEMENTAR.....	59

## 1. INTRODUÇÃO

A mudança recente de biólogos computacionais para a atuação em bioinformática vem destacando o acompanhamento cultural e mudanças conceituais que devem ser implementadas por órgãos de financiamento e instituições acadêmicas. As abordagens computacionais começaram a revolucionar as ciências da vida uma geração atrás, quando as sequências de DNA se tornaram mais amplamente disponíveis (YANAI e CHMIELNICKI, 2017). Os métodos de análise de sequência - mais notavelmente o BLAST - foram cruciais para descobrir as bases moleculares para a homologia profunda em organismos distantes, bem como uma transferência horizontal desenfreada e anteriormente não apreciada em organismos (KOONIN, 2009). Atualmente, o volume de dados gerados e processados no curso da pesquisa científica moderna está crescendo exponencialmente (YANAI e CHMIELNICKI, 2017). Grandes bancos de dados estão proliferando com vários tipos de dados biológicos, como a plataforma *Genomic Data Commons*, que fornece aos pesquisadores acesso a dados genômicos e clínicos de pacientes com câncer (GROSSMAN *et al.*, 2016). Mais recentemente, descobriram informações sobre a biologia do câncer a partir de levantamentos de células individuais de tumores (PATEL *et al.*, 2014). Tais análises computacionais alteraram o pensamento atual sobre a arquitetura do tumor, e as ferramentas utilizadas - principalmente a análise das sequências e da expressão gênica - permearam as ciências da vida.

Dentro dessa perspectiva, realizar grandes análises a respeito de processos epigenéticos, como o *imprinting* genômico, onde cerca de 1% dos genes humanos são expressos de forma origem parental dependente, evento que é conhecido como *imprinting* genômico, o qual é predominantemente encontrado em mamíferos placentários, se torna um pouco mais promissor (BARAN, SUBRAMANIAM *et al.*, 2015; MCGRATH e SOLTER, 1984; SURANI, BARTON *et al.*, 1984; (YANAI e CHMIELNICKI, 2017). Nos dias de hoje, sabe-

se que cerca de  $100 \pm 200$  genes imprintados ([/http://igc.otago.ac.nz/home.html](http://igc.otago.ac.nz/home.html)) mostram a expressão monoalélica específica dos pais, conferindo uma assimetria epigenética aos dois genomas parentais (PLASSCHAERT e BARTOLOMEI, 2014).

Sabe-se também que o padrão de expressão monoalélica é característico de genes sensíveis à dosagem. Em humanos, aproximadamente 10% dos genes autossômicos exibem este padrão de expressão (GIMELBRANT e CHESS, 2006). Em autossomos, a expressão monoalélica sob um modelo determinístico ocorre por silenciamento alélico parental-específico dentro de um mecanismo epigenético denominado *imprinting* genômico (REIK e WALTER, 2001). Neste contexto, o padrão de expressão monoalélica exerce influências importantes na determinação do fenótipo, assim como nas manifestações de doenças em um cenário de mutação devido à nulissomia funcional ou perda de heterozigose (GENDREL *et al.*, 2014; ALVES DA SILVA *et al.*, 2016).

Da mesma forma, atualmente existem poucos oncogenes e genes supressores associados com a perda de *imprinting* genômico já descritos na literatura, como: *DYRAS3*, *IGF2*, *H19*, *MEG3*, *MYC* E *PLAG1* (KELSEY e BARTOLOMEI, 2012; OKAE *et al.*, 2014; ANWAR *et al.*, 2015; BARAN, SUBRAMANIAM, BITON, TUKIAINEN, TSANG, RIVAS, PIRINEN, GUTIERREZ-ARCELUS, SMITH, KUKURBA, ZHANG, ENG, TORGERSON, URBANEK, LI, *et al.*, 2015). Sendo assim, a literatura não apresenta muitas informações a respeito dos outros genes imprintados associados com câncer, podendo ser supressores ou oncogenes, como: *IRAIN*, *NSD1*, *PAC1*, *TFPI2*, *TP53*, *TP73*, apresentando apenas dados sobre o efeito de perda de *imprinting* em genes que se tornaram um modelo clássico de estudo nessa área (HUBERTUS *et al.*, 2011; HSU *et al.*, 2016).

A teoria da perda de *imprinting* segundo a literatura sugere então uma ruptura no mecanismo que regula o *imprinting* e predispõe as células ao câncer (TIAN, TANG *et al.* 2012; ZHANG, WU *et al.* 2014). Embora a compreensão de tal mecanismo esteja longe de estar completo, os resultados dessa disfunção tem se tornado bem claro: quando os genes perdem seus grupos metil imprintados, doenças, incluindo câncer, podem resultar desse fenômeno

(FERGUSON-SMITH, 2011). Do mesmo modo, sabe-se que certos tipos de câncer parecem ser mais vulneráveis a interrupções epigenéticas (COOPER *et al.*, 2005; ABU-AMERO *et al.*, 2008).

Uma mudança no estado normal de um gene imprintado ou no seu perfil de metilação pode levar a uma perda de *imprinting*, efeito o qual foi descoberto inicialmente no tumor de Wilms, o qual demonstra um relaxamento do silenciamento normal do alelo materno de *IGF2* e, levando conseqüentemente assim a uma expressão bialélica do gene (LALANDE, 1996; GINART *et al.*, 2016);(DEJEUX *et al.*, 2009). Sabe-se que a perda da expressão de um gene imprintado tende a ocorrer em certos estados patológicos, os quais incluem distúrbios no *imprinting* humano e câncer (URIBE-LEWIS *et al.*, 2011). Outros estudos descrevem a causa da perda de *imprinting* genômico apenas em genes como *H19*, como uma metilação anormal dos alelos maternos desse gene, os quais causariam uma expressão bialélica do gene *IGF2*, sendo resultado de uma epimutação (CHEN *et al.*, 2000; BJORNSSON *et al.*, 2007; TIAN *et al.*, 2012).

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1 *Imprinting* Genômico

O *Imprinting* genômico é um fenômeno único que ocorre em mamíferos placentários, marsupiais e em um subconjunto de floração Plantas. O *imprinting* genômico desempenha um papel fundamental na manutenção da embriogênese normal e do crescimento pré-natal e pós-natal. Dentro do *imprinting* genômico, vários processos epigenéticos estão envolvidos para resultar em uma única assinatura epigenética observada em um subconjunto de *loci* no genoma (FEIL e BERGER, 2007). A metilação do DNA, por exemplo, é o principal pilar para o estabelecimento de marcas de *imprinting*, tanto nos alelos paternos quanto nos alelos maternos (FERGUSON-SMITH, 2011). Além disso, modificações de histonas e RNAs não codificantes contribuem grandemente para o processo de *imprinting* (FERGUSON-SMITH, 2011; MAUPETIT-MEHOUAS *et al.*, 2016).

O *imprinting* genômico é assim definido por um mecanismo que conduz à expressão origem parental dependente, através de uma expressão desbalanceada de cópias de um gene herdado maternalmente e paternalmente.

O *imprinting* é então um mecanismo dinâmico com diferentes graus de expressão monoalélica entre diferentes tecidos e estágios de desenvolvimento (BARAN, SUBRAMANIAM, BITON, TUKIAINEN, TSANG, RIVAS, PIRINEN, GUTIERREZ-ARCELUS, SMITH, KUKURBA, ZHANG, ENG, TORGERSON, URBANEK, CONSORTIUM, *et al.*, 2015). O reconhecimento do *imprinting* genômico ocorreu pela primeira vez em mamíferos, há mais de 30 anos atrás, quando os experimentos de transplante pronuclear mostraram que tanto o genoma materno quanto o paterno são necessários para o desenvolvimento normal de embriões de camundongos (MCGRATH e SOLTER, 1984; SURANI *et al.*, 1984).

O número de genes imprintados é pequeno, possivelmente existindo em média um pouco mais do que uma centena já conhecida (KELSEY e BARTOLOMEI, 2012; ANWAR *et al.*, 2015; BARAN, SUBRAMANIAM, BITON, TUKIAINEN, TSANG, RIVAS, PIRINEN, GUTIERREZ-ARCELUS, SMITH, KUKURBA, ZHANG, ENG, TORGERSON, URBANEK, CONSORTIUM, *et al.*, 2015). Entretanto, as pesquisas que surgiram a partir destas raras exceções levaram a avanços significativos em nossa compreensão da expressão gênica e evolução. Por exemplo, o florescente campo da epigenética que tem beneficiado muito outros trabalhos que foram realizados para descobrir como a metilação do DNA afeta a expressão de um *lócus* imprintado (FERGUSON-SMITH, 2011). Sabe-se também que muitas das hipóteses teóricas que envolvem a discussão do papel da genética na evolução do *imprinting* têm contribuído muito para a percepção de que a evolução do genoma, de um modo geral, está sujeita a uma variedade de pressões de seleção conflitantes (SPENCER e WOLF, 2014). Do mesmo modo a medicina atual também lucrou com a descoberta de erros no estado do *imprinting*, os quais podem levar a distúrbios de crescimento e câncer (LEE e BARTOLOMEI, 2013).

Atualmente já se sabe que genes sujeitos ao *imprinting* possuem uma expressão monoalélica dependente da origem parental, de modo que apenas um alelo seja expresso, paterno ou materno. Nos mamíferos, eutérios e marsupiais o mecanismo de *imprinting* tem um papel fundamental no desenvolvimento da placenta e do embrião (REIK e WALTER, 2001; ISHIDA e MOORE, 2013). Da mesma forma, também se sabe que os genes imprintados se expressam em

vários tecidos, mas particularmente na placenta e no cérebro. Embora alguns genes imprintados ocorram de forma isolada ou em pares no genoma, acredita-se que aproximadamente 60% desses genes estão localizados em clusters de até 2,3Mb, onde estes se mantêm sua estrutura conservada em ambos, humanos e camundongos (HENCKEL e ARNAUD, 2010; PETERS, 2014).

Mutações ou epimutações que afetam a função da expressão do alelo de um gene imprintado, podem resultar em distúrbios relacionados com o *imprinting*. Alguns mecanismos comuns que podem aumentar a origem de um transtorno relacionado com o *imprinting* incluem: dissomia uniparental (UPD), mutações intragênicas ou alterações que afetem diretamente a função de um gene imprintado, e mutações ou epimutações em centros de controle de *imprinting* que resultam em uma expressão alterada do gene imprintado. Nos distúrbios relacionados com o *imprinting*, tais como a síndrome de Beckwith-Wiedemann (BWS) e síndrome de Silver-Russell (SRS) a expressão alterada do gene se deve à perda ou ganho de metilação aberrante em um centro de *imprinting*, é a causa mais frequente dessas doenças (COOPER *et al.*, 2005; ABU-AMERO *et al.*, 2008).

## 2.2 Perfil de Expressão Monoalélica nos mamíferos

Atualmente sabe-se que em organismos diploides, os dois alelos de cada gene geralmente devem ser expressos em momentos e níveis semelhantes. Entretanto, uma parcela de genes pode se expressar de uma forma preferencial ou estrita a partir de um único alelo, tal processo é conhecido como expressão monoalélica. Acredita-se também que a expressão monoalélica pode surgir na ausência de polimorfismo na sequência de DNA, e isso geralmente está conectado a situações em que existe um requisito programado para regular a dosagem de genes (ECKERSLEY-MASLIN *et al.*, 2014; GENDREL *et al.*, 2014).

O fenômeno de expressão monoalélica é um fenômeno já caracterizado em mamíferos por envolver os genes ligados ao cromossomo X, através de um mecanismo que é conhecido como inativação do cromossomo X, onde um dos cromossomos, em células femininas, é marcado para ser silenciado através do silenciamento transcricional, levando assim a uma expressão alelo específica para os genes que permaneceram no X que se encontra ativo (RUSSELL, 1961;

LYON, 1962). Sabe-se também que em autossomos a expressão monoalélica pode ocorrer de forma constitutiva, para alguns genes, envolvendo a inativação e expressão, dos mesmos alelos homólogos. Sob um olhar determinístico a expressão também chamada de alelo-específico ocorre sob um efeito chamado de origem parental dependente, mecanismo que é conhecido como *imprinting* genômico (BARTOLOMEI *et al.*, 1991; REIK e WALTER, 2001; BARTOLOMEI e FERGUSON-SMITH, 2011). Entretanto, sabe-se que um modelo estocástico é mais comum em autossomos, de modo que esse fenômeno ocorre independentemente da origem parental, sendo esse conhecido como a expressão monoalélica randômica (EMR) (JEFFRIES *et al.*, 2012; ZWEMER *et al.*, 2012).

Os genes, na maioria das vezes são expressos a partir de seus dois alelos, exceto em alguns casos particulares, como a inativação aleatória de um dos dois cromossomos X em mamíferos fêmeas ou genes imprintados que são expressos apenas pelo alelo materno ou paterno. Um fenômeno menos conhecido ainda é a Expressão monoalélica randômica (EMR) de genes autossômicos, onde os genes podem ser expressos de forma estável de maneira monoalélica, a partir de qualquer um dos alelos parentais. Estudos sobre EMRs autossômicos vêm enfrentando vários desafios. O primeiro, a EMR que se baseia em mecanismos epigenéticos deve ser distinguido da expressão tendenciosa de um alelo causada por um polimorfismo da sequência de DNA em um elemento regulador. O segundo desafio, a EMR não deve ser confundida com a expressão transiente monoalélica frequentemente observada em análises de célula única e que muitas vezes corresponde a uma explosão de expressão dinâmica (GENDREL *et al.*, 2016).

### **2.3 Perda de *Imprinting* Genômico – LOI**

A mudança do estado normal de *imprinting* é considerada uma anormalidade comum em câncer, o qual envolve a perda de expressão do gene de origem específica, em um fenômeno conhecido como *Loss of Imprinting* (LOI). A LOI é então encontrada em vários tipos de neoplasias diferentes, incluindo

ovário, cólon, pulmão e bexiga, bem como leucemia mielóide crônica (TIAN *et al.*; 2012). A perda de *imprinting* descreve então os mecanismos que ou ativam um alelo normalmente silenciado ou inativam a cópia do um alelo normalmente ativo no gene. No entanto, a perda de *imprinting* nem sempre foi encontrado em tecidos normais, apenas em alguns tipos de câncer (ZHANG *et al.*, 2014).

Desse modo, sabe-se também que atualmente os estudos relacionados com perda de *imprinting* abordam apenas o uso de poucos genes sabidamente imprintados, como *CPA4*, *IGF2*, *GABRB3*, *H19*, *NNAT*, *DLK1*, *RTL1*, *MEG3*, *SLC22A3*, *SNURF*, *ZNF215*, onde muito poucos são genes supressores ou oncogenes (HUBERTUS *et al.*, 2011; HSU *et al.*, 2016) dentre o os muitos genes que já se sabem que são imprintados, os quais compreende um número de aproximadamente 140 genes (OKAE *et al.*, 2014; BARAN, SUBRAMANIAM, BITON, TUKIAINEN, TSANG, RIVAS, PIRINEN, GUTIERREZ-ARCELUS, SMITH, KUKURBA, ZHANG, ENG, TORGERSON, URBANEK, LI, *et al.*, 2015).

A perda do *imprinting* em um gene imprintado no qual sua expressão se torna dobrada, se deve ao fato do alelo reprimido se tornar ativo, alterando efetivamente a atividade da cascata de sinalização. É conhecido que genes imprintados estão frequentemente associados a neoplasias como hiperplasia, neoplasia ou displasia em tecidos como rim, tecido adrenal ou fígado, onde os sinais de transdução celular aberrante são também presentes (URIBE-LEWIS *et al.*, 2011). Sabe-se que a deleção de regiões de controle do *imprinting* (ICR) resulta em perda de *imprinting* da maioria dos genes. Pequenas eliminações dessas ICRs específicas e os erros na metilação diferencial do DNA, ou outras modificações epigenéticas também resultam em expressão aberrante de genes marcados e pode levar a uma doença, ressaltando a natureza crítica das ICRs. Nos seres humanos, a perda desse *imprinting* contribui para o desenvolvimento de um grande número de doenças, incluindo a Síndrome de Prader-Willi, Síndrome de Angelman, Síndrome de Beckwith-Wiedemann e câncer.

A melhor caracterização da relação entre câncer e um gene imprintado é a associação de perda de *imprinting* no gene *IGF2-H19* com o tumor de Wilms. O gene que codifica o fator de crescimento insulina-2 é paternalmente expresso (*IGF2*), e o gene *H19* é maternalmente expresso e não codificante. A perda de

*imprinting* congênito neste *locus* ocorre então, por mutação, epimutação (uma mutação que irá afetar apenas as marcas epigenéticas, mas não a sequência de DNA) ou herança uniparental, resultando em uma síndrome chamada de Silver-Russell ou síndrome de Beckwith-Wiedemann (ANWAR *et al.*, 2015; UREBI-LEWIS *et al.*, 2011; ZHANG *et al.*, 2014).

Segundo a literatura, a perda de *imprinting* do gene *TP73*, um gene supressor de tumor em estômago, pode estar associada com a deleção genética, como já descrito em alguns tipos de câncer como neuroblastoma ou adenocarcinoma, ou por uma metilação aberrante do DNA, como descrito em leucemia e linfoma (KANG *et al.*, 2000; ANWAR *et al.*, 2015). Desse modo, a hipometilação do DNA conduz a uma ativação aberrante de genes e regiões não codificantes através de uma variedade de mecanismos, podendo contribuir para o desenvolvimento de câncer e sua progressão. Desde então, vários outros genes imprintados estão envolvidos na carcinogênese de várias doenças malignas como *TCEB3C*, *TP73*, *IGF2*, *IRAIN*, *CPA4*, *H19* E *GABRB3* (HSU *et al.*, 2016).

Estudos recentes tem especificado que a perda de *imprinting* numa região diferencialmente metilada (DMR) do cluster de genes como *IGF2/H19* ou perda de heterozigose (LOH), respectivamente, bem como uma dissomia uniparental (UPD), podem ser responsáveis por uma super expressão (HUBERTUS *et al.*, 2011). Tais perfis de metilação aberrante em DMRs imprintadas têm sido observados em crianças com distúrbios de *imprinting* raros (GICQUEL *et al.*, 2005; WEKSBURG *et al.*, 2005). Desse modo, poucos estudos tem sido propostos para estabelecer uma correlação de LOH com tipos de tumores diferenciados e mais avançados , mesmo em genes humanos embrionários adultos (CHEN *et al.*, 2000).

#### **2.4 Sequenciamento Transcriptômico: RNA-Seq**

O RNA-Seq é uma abordagem recentemente desenvolvida para o processamento e armazenamento transcriptômico que utiliza tecnologias de sequenciamento de profundidade (WANG *et al.*, 2009). Recentemente, o

desenvolvimento de novos métodos de sequenciamento do DNA de alto rendimento tem proporcionado um novo método para ambos, mapeamento e quantificação de transcriptomas. Este método é denominado RNA-Seq (Sequenciamento de RNA), o qual apresenta vantagens claras sobre outras abordagens existentes. Este método já foi aplicado em *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, camundongo e células humanas oriundas de embriões com 18 ou 24 células (BARBAZUK *et al.*, 2007; LISTER *et al.*, 2008; MARIONI *et al.*, 2008; MORIN *et al.*, 2008; MORTAZAVI *et al.*, 2008; NAGALAKSHMI *et al.*, 2008; VERA *et al.*, 2008; WILHELM *et al.*, 2008).

As utilizações de RNA-Seq desenvolveram-se recentemente através da tecnologia de sequenciamento profundo, onde as leituras são tipicamente de 30-400 pb, dependendo da Sequenciamento de DNA e a tecnologia utilizada. O sequenciamento de RNA (RNA-Seq) representa, então, uma solução única e abrangente para muitos dos inconvenientes ensaios moleculares existentes.(OZSOLAK e MILOS, 2011; DAMODARAN *et al.*, 2015).

O RNA-Seq tornou-se uma forma para medir a expressão gênica em organismos-modelo e humanos (POPLAWSKI *et al.*, 2016). O RNA-Seq se transformou então em uma ferramenta de escolha massiva para análise de transcriptomas (MORTAZAVI *et al.*, 2008; HILLIER *et al.*, 2009). A técnica de RNA-Seq facilita o diagnóstico de doenças e identificação de alvos moleculares putativos de intervenção terapêutica, permitindo dessa forma uma quantificação de alterações nos níveis de expressão e detecção de *splicing* alternativo (APARICIO *et al.*, 2000; SUTHERLAND *et al.*, 2011).

## 2.5 Oncogênese – Regulação e Desregulação

O câncer é uma doença pleiotrópica e segundo Uribe (URIBE-LEWIS *et al.*, 2011), essa definição abrange bem a complexidade do termo câncer, o qual não é só causado apenas pelo crescimento anormal de células com o potencial para invadir diferentes órgãos, mas também por uma diferenciação prejudicada (GOUGH, 2013). O câncer é conhecido então por ser uma doença a qual é

marcada pela instabilidade genética e epigenética. Embora as funções de hereditariedade constitutivas e mutações genéticas adquiridas em neoplasia estejam bem descritas na literatura, pouco se sabe sobre as mudanças epigenéticas em câncer (URIBE-LEWIS *et al.*, 2011). O câncer tem então um forte componente genético e a consanguinidade, ainda muito comum em muitas partes do mundo, aumenta a homozigose do gene. A homozigose dos genes tumorais de baixa penetração e sua frequência na população podem aumentar ou diminuir o risco de câncer (SINNETT *et al.*, 2000; GAO *et al.*, 2002; STUCKER *et al.*, 2002; SOHAIL *et al.*, 2013; YAN *et al.*, 2013). Entretanto, alguns autores têm associado a causa da maioria dos cânceres a fatores epigenéticos, enquanto os restantes estão relacionados a fatores genéticos (RUSSO *et al.*, 2016). Dados mais recente tem demonstrado que existem diferenças genéticas e epigenéticas entre tumores do mesmo tipo em diferentes pacientes, bem como entre as células com câncer dentro do mesmo tumor, tendo um impacto tanto inter como intratumoral na heterogeneidade da resposta à uma terapia (JUNTTILA e DE SAUVAGE, 2013); (MICHOR e POLYAK, 2010; JUNTTILA e DE SAUVAGE, 2013).

É bem conhecido também, que um gene supressor de tumor é um gene capaz de reduzir a probabilidade de que uma célula se torne cancerosa. Entretanto, uma única mutação ou deleção de tal gene acarretará em um aumento da probabilidade de formação de um tumor (MICHOR e POLYAK, 2010). Dessa forma, os genes supressores de tumor codificam proteínas que possuem importante papel na regulação do ciclo celular e apoptose, inibindo ou podendo atuar de forma a evitar a formação de tumores (EVAN e VOUSDEN, 2001). Já os oncogenes por sua vez, são genes associados com o surgimento de tumores, tanto malignos ou benignos, assim como genes que quando deixam de funcionar normalmente, transformam uma célula normal em uma célula cancerosa. Os oncogenes por sua vez possuem uma versão funcional, os proto-oncogenes, os quais são responsáveis pelo controle da divisão celular (mitose), da diferenciação celular, da tradução proteica e no metabolismo celular (HUNTER, 1991; MARSHALL, 1991). Dessa forma, após sofrer qualquer tipo de epimutação ou mutação génica somática – translocação, amplificação ou

mutação pontual, um proto-oncogene torna-se eventualmente em um oncogene (BISHOP, 1991; CANTLEY *et al.*, 1991).

### 3 JUSTIFICATIVA

Atualmente, segundo a literatura poucos são os genes sabidamente imprintados que são utilizados para estudos de perda de *imprinting* genômico, tendo aproximadamente um total de 110 genes imprintados já descritos na literatura, da mesma forma, pouco se sabe sobre o efeito de ganho ou perda de *imprinting* genômico na literatura relacionado à oncogenes ou supressores de tumor, onde existem aproximadamente 1890 oncogenes e supressores já descritos. Dessa forma, o número conhecido de oncogenes e supressores tumorais que sofrem o efeito de *imprinting* é pequeno, tendo um total de 28 genes descritos na literatura, aonde apenas os modelos clássicos vêm sendo estudados, como *H19*, *IGF2* e *MEG3*, visto que existe um número bem maior desses genes que ainda não foram amplamente estudados, os quais desempenham funções importantes no organismo. Sendo assim, buscamos ter análises mais profundas destes processos, além de tentar estabelecer uma melhor compreensão dos processos de perda e ganho de *imprinting* genômico em oncogenes e genes supressores, sendo estes imprintados ou não.

Nesse contexto ressalta-se a importância de ter bibliotecas informativas, construídas com informações extraídas dos bancos de dados National Center for Biotechnology Information (NCBI) e University of Califórnia Santa Cruz (USCS), com a finalidade de compreender de uma forma melhor a extensão dos efeitos de *imprinting* genômico em amostras de tecidos com câncer. Além disso, o presente estudo consistiu na investigação dos padrões alélicos de expressão de genes candidatos ao *imprinting* e a sua possível correlação como fenômeno de perda e ganho de *imprinting* genômico em tecidos saudáveis e com tumor, visto que esses efeitos podem acometer outros genes que não são oncogenes e genes supressores, mas outros genes imprintados também, podendo ter um efeito expandido para outros genes.

## 4 OBJETIVOS

### 4.1 Objetivo Geral

O objetivo principal do estudo é determinar a amplitude da perda ou ganho de *imprinting* genômico em tecidos primários humanos com câncer através de uma metanálise computacional transcriptômica de oncogenes e genes supressores de tumor.

### 4.2 Objetivos Específicos

- Inferir o padrão de expressão alélica dos genes alvo e relacioná-los com possíveis efeitos de perda ou ganho de *imprinting* genômico;
- Descobrir novos genes candidatos a serem imprintados;
- Construção de um atlas informativo contendo experimentos de RNA-Seq de tecidos humanos saudáveis e neoplásicos;

## 5 MATERIAL E MÉTODOS

### 5.1 Seleção de experimentos de RNA-Seq em bancos de dados públicos de RNA-Seq

O RNA-Seq é uma abordagem desenvolvida para analisar o perfil de transcriptomas. Essa utiliza tecnologias de *deep-sequencing* que permitem avaliar quantitativa e qualitativamente um conjunto completo de transcritos em uma célula ou tecido. As análises de RNA-Seq podem ser utilizadas para comparar os possíveis efeitos nos níveis de transcrição em estágios de desenvolvimento ou em diferentes condições experimentais. Para esse projeto, os experimentos de RNA-Seq foram selecionados de bancos de dados públicos armazenados na plataforma Sequence Read Archive (SRA - <https://www.ncbi.nlm.nih.gov/sra/?term=>) do NCBI. A plataforma SRA disponibiliza dados de sequências de DNA ou RNA-Seq, de atributos das amostras e doadores. A comparação desses experimentos permite estabelecer níveis de reprodutibilidade. SRA armazena informações de sequenciamento e informações de alinhamento cruas de plataformas de sequenciamento de alto rendimento, incluindo Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics® e Pacific Biosciences SMRT®.

A procura de experimentos de RNA-Seq em diferentes tecidos foi por termos em grupos: tecidos saudáveis (controle) e com tumor (ex: RNA-Seq and Homo sapiens and "Brain"). O atlas de tecidos foi limitado a 24 sistemas (adiposo, adrenal, artéria, cérebro, mama, cólon, esôfago, tubo falopiano, coração, rim, fígado, pulmão, músculo, ovário, pâncreas, pituitária, placenta, próstata, pele, estômago, testículo, tireoide, útero e sangue). Para os experimentos de RNA-Seq foram utilizados dados de tecidos saudáveis e de 16 sistemas (adrenal, cérebro, mama, cólon, esôfago, rim, fígado, pulmão, ovário, pâncreas, pituitária, pele, estômago, testículo e útero) para experimentos de RNA-Seq de tecidos com tumor.

### 5.1.1 RNA-Seq Primário

A fonte primária de experimentos de RNA-Seq consistiu-se em 2105 experimentos de RNA-Seq, representando um atlas de 24 tecidos e também em 2424 experimentos de RNA-Seq que representam um atlas de 24 tecidos neoplásicos, selecionados nos repositórios de dados públicos de arquivo de leitura de sequência de RNA da plataforma SRA. Após a seleção desses experimentos de RNA-Seq, as informações gerais foram baixadas no formato *RunInfo* para tabular tabelas Excel de modo a organizá-los de acordo com os termos “*Experiment*”, “*Biosample*” e “*Sample*”. Logo após foi realizada uma checagem de unicidade nesses experimentos de RNA-Seq na plataforma NCBI em conjunto com o grupo SRA, onde foi selecionada a opção “*Biosample*” (<https://www.ncbi.nlm.nih.gov/biosample/>). Nessa plataforma foram inseridos então os códigos dos “*Experiments*” obtidos anteriormente pela seleção dos estudos de RNA-Seq, de modo que a checagem desses estudos refira ao respectivos BioProject. Foram excluídos da listagem os experimentos que não eram pertinentes ao estudo. Para a seleção de experimentos de RNA-Seq em tecidos primários saudáveis foi checado se as amostras presentes nesse estudo foram amostras únicas (atributo denominado “*Monoisolate*”), de modo que se tenham apenas amostras únicas ao invés de misturas (“*Multisolate*”). Essas amostras únicas eram obtidas de estudos que apresentam amostras de doadores saudáveis, ou seja, sem qualquer tipo de doença. Por fim, foram excluídos estudos de RNA-Seq em linhagens celulares.

Após o estabelecimento da unicidade para os experimentos de RNA-Seq de tecidos saudáveis, foi realizado então uma validação desse repositório (atlas de RNA-Seq) interrogando os SNPs localizados em três genes (*H19*, *SNURF*, *PLAG1*) sabidamente imprintados em vários tecidos (BISHOP, 1991). A unicidade foi assim definida para aqueles experimentos em que os perfis de expressão para os SNPs controles fossem consistentes com expressão monoalélica. Essa validação foi realizada em todos os experimentos de RNA-Seq que compõem um tecido, visando assim validar todos os tecidos ao final. A validação foi feita *on-line* na plataforma SRA, acessando a opção *SRA Run Browser* ([https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run\\_browser](https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser)). Logo após foi selecionada a aba “download – FAST/FATSQ”, onde os códigos

dos experimentos foram anexados. A interrogação de cada SNP alvo foi feita por pareamento perfeito nos transcritos utilizando três sequências de 29 nucleotídeos cada (“*query sequence string*”): A primeira sequência foi no formato da União Internacional de Química Pura e Aplicada (IUPAC), contendo a incógnita SNP e as duas outras sequências correspondentes aos alelos de referência e alternativo, respectivamente (Tabela 1).

As sequências filtradas foram tabeladas e analisadas para verificar se a soma dos alelos era igual ao número total de *reads* utilizando o código IUPAC; teor da fração alélica 1:0; classificação de verdadeiro ou falsos de acordo com o limite máximo de 5% para amostras com possível expressão bialélica; unanimidade entre os SNPs nos genes analisados. Os experimentos não eliminados foram reexaminados, um a um, quanto às informações das amostras constantes no repositório de dados BioProject do NCBI para checagem da presença de: mistura com outros tecidos; tecidos de outras espécies; algum tipo de patologia; se oriundos de linhagem celular.

**Tabela 1. Exemplo da sequência utilizada para a interrogação de cada SNP alvo.**

Alvo investigado	Sequência de 29pb utilizada
Formato IUPAC: (ex., <b>R</b> = A/G)	CAAGGCCCAGTGCG <b>R</b> TGGCTCACACCTGC
Alelo de referência	CAAGGCCCAGTGCG <b>A</b> TGGCTCACACCTGC
Alelo alternativo	CAAGGCCCAGTGCG <b>G</b> TGGCTCACACCTGC

A unicidade (isto é, a certeza de excluir misturas de amostras e de tecidos de doenças) de cada experimento SRA foi estabelecida consultando SNPs nos genes *SNURF* e *H19*, utilizando a ferramenta de busca SRA, essencialmente como relatados (ALVES DA SILVA et al., 2016). As sequências de consulta tinham 29 nt de comprimento e os alelos foram representados pelos códigos de substituição da IUPAC. Os critérios de inclusão de unicidade foram: (i) ASE  $\geq$  0,42 (consistente com a expressão monoalélica), (ii) ocorrência de alelo aleatório (ambos os alelos são observados monoalelicamente por tecido) e (iii) ausência de padrões estritamente bialélicos.

Este processo de construção, onde cada SNP teria 14 sequências de nucleotídeos flanqueando ambos os lados do SNP, foi também realizado para

todos os SNPs com MAF  $\geq 0,2$  localizados nos 28 oncogenes e genes supressores de tumor (Tabela suplementar 1). Como resultado da interrogação se obtém o número de leituras (“reads”) de RNA contendo os alelos do SNP alvo. foram admitidos experimentos com no mínimo 10 *reads* para as sequências em formato IUPAC e para os quais não houve inconsistências entre o número total de reads IUPAC e a soma dos “reads” para cada alelo ( $[\text{reads alelo1} + \text{reads alelo2}] - \text{reads IUPAC} = 0$ ). A fração alélica foi então calculada em relação ao total de “reads”. As frações alélicas são utilizadas para discriminar entre três possíveis padrões de expressão alélica: consistente com expressão bialélica (proporção 1:1 = 0.5), monoalélica (1:0 ou 0:1) e desbalanceada (desvio das proporções esperadas 1:1). Para garantir a unidade dos experimentos de RNA-Seq utilizamos por último um critério de tolerância de 5% de desvio das proporções esperadas para padrões monoalélicos de expressão (exemplo: total de 100 reads para a sequência em formato IUPAC, 95 reads para alelo 1 e 5 reads para o alelo 2).

### 5.1.2 RNA-Seq Secundário

A fonte secundária de experimentos de RNA-Seq consistiu-se em hubs de livre acesso para o projeto Genotype-Tissue Expression (GTEx) (versão de dados V6, outubro de 2015) (The\_GTEx\_Consortium, 2015) disponível no UCSC Genome Browser e dbGaP. Os centros de trânsito GTEx, conforme projetado e fornecido pelo laboratório Lappalainen no New York Genome Center, parte do GTEx Analysis Working Group (CASTEL et al., 2015), possibilitam que os usuários acessem e utilizem todas as tabelas de suporte. O núcleo do GTEx registra os valores de medição da ASE em variantes heterozigóticas imputadas identificadas a partir de dados de transcriptomas e genótipos coletados em 51 tecidos primários de experiências de RNA-Seq (8.555 amostras de 570 doadores post-mortem). Os valores de medição ASE em GTEx referem-se à diferença média de contagens de leitura de RNA-Seq entre os dois alelos em um SNP heterozigótico imputado em doadores, com uma cobertura  $\geq 8$  leituras por site por doador. ASE é calculado como  $[0.5 - \text{Ref\_allele\_read count} / (\text{Ref\_allele\_read count} + \text{Alt\_allele\_read count})]$  (CASTEL et al., 2015). Foi utilizada a ferramenta Data Integrator para extrair informações sobre as coordenadas físicas,

sobreposição de nomes de genes, mínimos, medianos, máximos, quartis superiores (Q1 e Q3) para valores de ASE, número de doadores e cobertura de profundidade de leitura em SNPs presentes nos genes alvo. Foi realizada também uma filtragem dos SNPs utilizando um *script Short Match* em códigos R, de modo a obter-se SNPs únicos no genoma. Observamos que <5% de todos os sítios SNP (genoma-wide) são excluídos dessa maneira. Os dados de expressão foram classificados por *locus* de gene usando um script em códigos R, o qual foi desenvolvido em colaboração com o mestrando Ronaldo da Silva Francisco Junior, do programa de modelagem computacional do Laboratório Nacional de Computação Científica - LNCC e pelo Graduando Victor Ramos do Laboratório de Biologia Sistemática de Microrganismos - LaBiSisMi - da Universidade de São Paulo, e também pela Doutoranda Cristina de Souza Fernandes, do programa de Biociências e Biotecnologia na UENF. O script criado visou organizar os dados obtidos (SNPs, Genes, Tecidos), permitindo compilar os dados a partir dos parâmetros estabelecidos anteriormente. O script então organizou os dados quanto ao número de SNPs monoalélicos em relação a quantidade de tecidos para cada gene alvo, permitindo assim plotagem de gráficos que demonstrem os dados obtidos de forma a corroborar com a hipótese apresentada.

O Script utilizado também classifica a evidência de expressão em quatro categorias excluindo *allelic specific expression* (ASE). Os parâmetros ASE para cada categoria foram: Estritamente monoalélicos (valores mínimos, máximos e Q1 ASE = 0,5); consistente com monoalélico (valores ASE mínimos <0,5, Q1 e valores ASE máximos > = 0,3); estritamente bialélico (valores mínimo, máximo e Q1 ASE = 0) e consistente com bialélico (valores ASE mínimos <0,5, Q1 valores ASE <0,3 e valores ASE máximos > 0). Para a designação final do perfil de expressão (em um por gene e por base de tecido), chamamos apenas genes com pelo menos três SNPs informativos não discordantes, com profundidade  $\geq$  12 leituras por SNP.

## 5.2 Seleção e Formatação de Polimorfismos em Genes alvos através de Browsers

Para a seleção dos SNPs em genes alvos foi utilizado o Browser do NCBI - Base de dados de polimorfismos de nucleotídeo único (dbSNP - <https://www.ncbi.nlm.nih.gov/snp/?term=>). Dentro da plataforma dbSNP foram selecionadas as opções “*Organism – Homo sapiens*”, de modo que estes SNPs sejam exclusivamente de humanos, foram selecionadas também as opções “*Variation Class – SNP*”, 3’ UTR, 5’UTR, éxons, nonsense, missense e “frequência do menor alelo (MAF)”, o qual teve que apresentar valores entre 0,2 e 0,5, de modo a obter-se uma taxa mínima de heterozigose. Por fim foi selecionada a opção “*Validation Status – by 1000Genomes*”, com a finalidade de que esses experimentos fossem validados. O projeto 1000 Genomas foi criado para assegurar a usabilidade contínua dos dados gerados pelo Projeto 1000 Genomas e estender o conjunto de dados.

As informações dos SNPs foram obtidas nos formatos “*Brief*” e “*Chromosome Report*” de modo que possam ser planilhados no Excel para posterior edição e organização dos mesmos. As planilhas assim geradas contêm informações a respeito da sua posição física, número de alelos, MAF e o(s) gene(s) em que estão presentes. Para ter certeza que estes SNPs fossem únicos no genoma, foi usado um script em linguagem R criado pelo nosso grupo de pesquisa, que excluí as sequências localizadas em regiões que contém repetições. A estratégia de interrogação de cada SNP foi a mesma descrita acima para os SNPs em genes sabidamente imprintados.

## 5.3 Integração de dados com a plataforma GTEx

Inicialmente foi acessada a plataforma do UCSC para utilizar a ferramenta “*tablebrowser*” (<https://genome.ucsc.edu/cgi-bin/hgTables>), de acordo com o *table browser user's guide*, onde os genes alvos (Tabela suplementar 1) foram validados através de parâmetros presentes dentro da plataforma UCSC. Essa validação foi feita acessado o comando “*group*”, onde foi introduzido outro comando “*gene and gene predictions*” e também o comando “*track*”, onde foi introduzido o comando “*UCSC-refseq*”, de modo que podemos confirmar a presença desses genes no *browser*, além de determinar os nomes dos genes

alvos, bem como a posição física desses genes. Logo após então foi introduzido uma configuração contendo os genes alvos através da opção "*identifiers - paste list*", obtendo assim a informação necessária ao acessar o comando "*get output*".

Após a validação dos genes alvos foi criado então um "*custom track*" integrando as posições físicas dos genes alvos em formato BED na plataforma UCSC, de acordo com as instruções disponíveis (<https://genome.ucsc.edu/FAQ/FAQformat.html>). De posse desse custom track, foi construído o track que integra as posições físicas dos genes alvo, tendo como base os genes disponíveis no UCSC, onde foi selecionado a opção "*table browser*", e depois a opção "intersection". Logo após esta etapa foi selecionado então pelo comando "*GROUP*", o custom track dos genes alvo (ex: comandos: "*group*" - *custom track*/"*track*" - nome do track criado), de modo a integrá-los com os genes disponíveis na plataforma UCSC - RefSeq, obtendo assim a informação desejada após acessar a opção "*submit*". Ao final foi introduzido também o comando "*intersection*" a fim de determinar a posição física desses genes a partir do track customizado obtido anteriormente a partir dos genes alvos.

Mediante a obtenção dos dados desejados até esse ponto, a plataforma UCSC foi acessada novamente, onde a opção "*tools*" foi acessada de modo a chegar-se na opção "*data integrator*" de acordo com o *data integrator user's guide* (<https://genome.ucsc.edu/cgi-bin/hgIntegrator>). Sendo assim, foi selecionada a opção "*configure data sources*", e logo após "*add data sources*", onde foi selecionada a opção "*track group*" de forma a selecionar um *custom track*. Seguindo a próxima etapa, foi selecionada a opção "*track*", que visa selecionar o track criado anteriormente com os genes alvos e adicioná-lo com a opção "*add*" para integrá-lo ao track criado. Logo após, foi selecionado novamente a opção "*track group*", visando selecionar o comando "*GTEx Analysis Hub*" e na opção presente ao lado "*track*" foi selecionada a opção "*ASE by tissue*", e na opção seguinte "*subtract*" foram selecionadas as amostras de tecidos disponíveis na plataforma GTEx (BARAN, SUBRAMANIAM, BITON, TUKIAINEN, TSANG, RIVAS, PIRINEN, GUTIERREZ-ARCELUS, SMITH, KUKURBA, ZHANG, ENG, TORGERSON, URBANEK, CONSORTIUM, *et al.*, 2015). A integração destas amostras disponíveis juntamente com o track contendo os genes alvos foi feita de forma que sejam integrados o track com o

genes alvos e mais quatro amostras de tecidos, onde esse processo deve ser repetido até completar as 53 amostras de tecidos disponíveis pelo GTEX, gerando dados que foram planilhados com o auxílio do Excel.

Os dados obtidos trazem informações a respeito dos SNPs informativos dentro dos genes alvos, organizados de acordo com ASE entre 0 e 0,5. Uma vez organizados os dados, foram selecionados apenas os SNPs informativos que contenham ASE entre 0,42 a 0,5, para análise de perfil monoalélico dos SNPs em relação aos genes alvos e tecidos disponíveis pelo GTEX. O ASE é o valor do desequilíbrio alélico, o qual foi gerado numa base por tecidos. O desequilíbrio alélico por local (entre 0 e 0,5) é o desequilíbrio alélico mediano entre indivíduos com cobertura alélica suficiente, calculada como  $[0,5 - \text{REF\_COUNT} / (\text{REF\_COUNT} + \text{ALT\_COUNT})]$ . O ASE 0,5 informa então que o SNP de um gene tende a uma expressão monoalélica e o ASE 0 tende a uma expressão bialélica, inferindo assim para qual tecido específico aquele SNP exibe padrão de expressão consistente com monoalélico.

#### **5.4 Ontologia dos Oncogenes e Genes Supressores de Tumor**

Uma vez obtido o set de genes alvos os quais foram selecionados anteriormente, eles foram inseridos na plataforma UCSC (*tools --> table browser*) onde foi selecionada a opção “*group: gene and gene predictions / UCSC RefSeq genes*”, de modo a validar todos os genes novamente na plataforma, e por fim selecionando a opção “*past list*” para inserir os genes alvos. Logo após, foi marcada a opção “*GREAT*” e selecionado a opção “*get output*”. Por seguinte foi marcada a opção “*send query to GREAT*”, de modo que foi aberta uma nova aba, onde essa nova aba gerada visa redirecionar as informações obtidas para uma nova plataforma, onde é possível observar a quais funções biológicas do set de genes inserido estão associados, as possíveis doenças que eles estão associados, entre outras informações. Posteriormente foram plotados gráficos a partir desses resultados de modo a evidenciar a associação desses genes com suas funções/doenças associadas.

## 6. RESULTADOS

### 6.1 Consolidação *In Silico* do Acervo de Experimentos de RNA-Seq de Tecidos Humanos para o Estudo de *imprinting* Genômico.

Das 18.146 amostras iniciais, 2.105 amostras se enquadraram nos critérios e parâmetros de seleção para determinação de *imprinting* genômico (figura 1). São experimentos de RNA-Seq que compõem o atlas de experimentos saudáveis (Tabela 2). O acervo inclui experimentos de RNA-Seq referentes a 24 diferentes tecidos primários. Para os experimentos de RNA-Seq que compõem o atlas de experimentos neoplásicos obtivemos 15.802 amostras iniciais, onde ao final apenas 2.424 amostras se enquadraram nos critérios de seleção para determinação de *imprinting* genômico (Tabela 3). O acervo de experimentos neoplásicos de RNA-Seq inclui também 24 tecidos primários.

**Tabela 2. Quantidade de amostras, antes e após a análise de cada um dos 24 tecidos primários humanos utilizados no Atlas Saudável.**

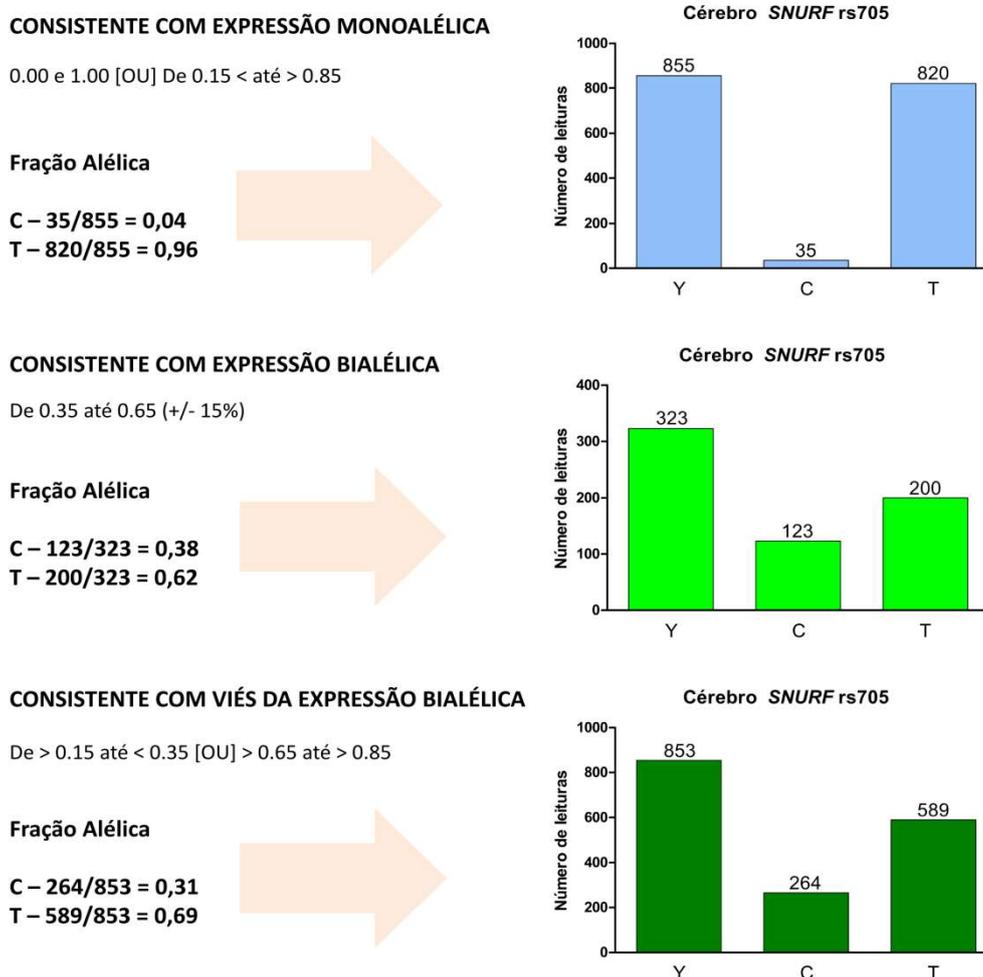
Figura adaptada de: [www.roadmapepigenomics.org](http://www.roadmapepigenomics.org)

TECIDOS HUMANOS	ANTES DAS ANÁLISES	APÓS AS ANÁLISES
ADIPOSO	454	6
ARTÉRIA	93	21
CÉREBRO	2591	720
CÓLON	477	16
CORAÇÃO	1028	132
ESÔFAGO	48	20
ESTÔMAGO	106	45
FÍGADO	706	77
GLÂNDULA ADRENAL	41	15
MAMA	2845	61
MÚSCULO	975	151
OVÁRIO	96	16
PÂNCREAS	115	11
PELE	986	132
PITUITÁRIA	59	1
PLACENTA	191	122
PRÓSTATA	922	41
PULMÃO	2180	177
RIM	1267	39
SANGUE TOTAL	2756	235
TESTÍCULO	65	30
TIREÓIDE	118	20
TROMPA DE FALÓPIO	6	4
ÚTERO	21	13
<b>TOTAL</b>	<b>18146</b>	<b>2105</b>

**Tabela 3. Quantidade de amostras antes e após a análise de cada um dos 24 tecidos primários humanos utilizados no Atlas Neoplásico.**

Figura adaptada de: [www.roadmapeigenomics.org](http://www.roadmapeigenomics.org)

<b>TECIDOS HUMANOS</b>	<b>ANTES DA ANÁLISE</b>	<b>DEPOIS DA ANÁLISE</b>
Adiposo	433	0
Adrenal	24	3
Artéria	151	0
Cérebro	2971	425
Mama	3610	517
Cólon	118	17
Esôfago	56	31
Tubo Falopiano	23	0
Coração	500	0
Rim	1384	103
Fígado	926	262
Pulmão	330	393
Músculo	1202	0
Ovário	104	14
Pâncreas	911	39
Pituitária	36	12
Placenta	10	0
Próstata	1251	442
Pele	1116	82
Estômago	110	68
Testículo	87	4
Tireoide	427	0
Útero	22	12
<b>Total</b>	<b>15802</b>	<b>2424</b>



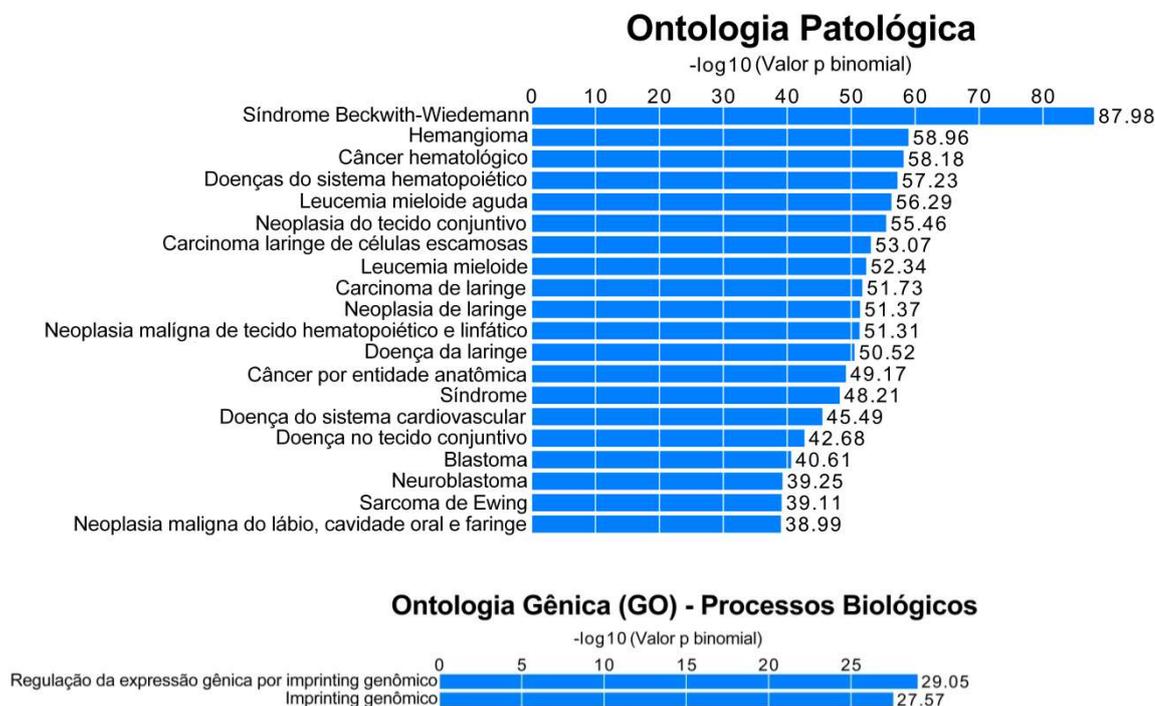
**Figura 1. Parâmetros adotados para análise da unicidade dos tecidos.**

Parâmetros para tecidos saudáveis e estudos relacionados à *imprinting* genômico após a filtragem do RNA-seq.

## 6.2 Ontologia Gênica – Patologia e Funções Biológicas

Segundo as análises de ontologia realizadas para os 28 genes conhecidos como imprintados chamados de controles (Tabela S1), entre eles oncogenes e genes supressores de tumor, os quais já são descritos como imprintados pela literatura (*AK123878*, *HOTS*, *BLCAP*, *CDKN1C*, *DIRAS3*, *TFPI2*, *H19*, *IGF1R*, *IGF2*, *IGF2R*, *LPAR6*, *MEG3*, *NAP1L4*, *NDN*, *NSD1*, *PEG10*, *PLAGL1*, *RB1*, *SDHD*, *SLC22A18*, *SNRPN*, *STIM1*, *TP73*, *TRIM28*, *TSPAN32*, *TSSC4*, *UBE3A*, *MIR372*), observou-se que houve enriquecimento significativo ( $P = 3.64 \times 10^{-06}$ ) de seis genes nas categorias GO:0016564 (6/37) e

GO:0003714 (5/47) de ontologia para reguladores negativos da transcrição e correpressores de transcrição, respectivamente (Figura 2).

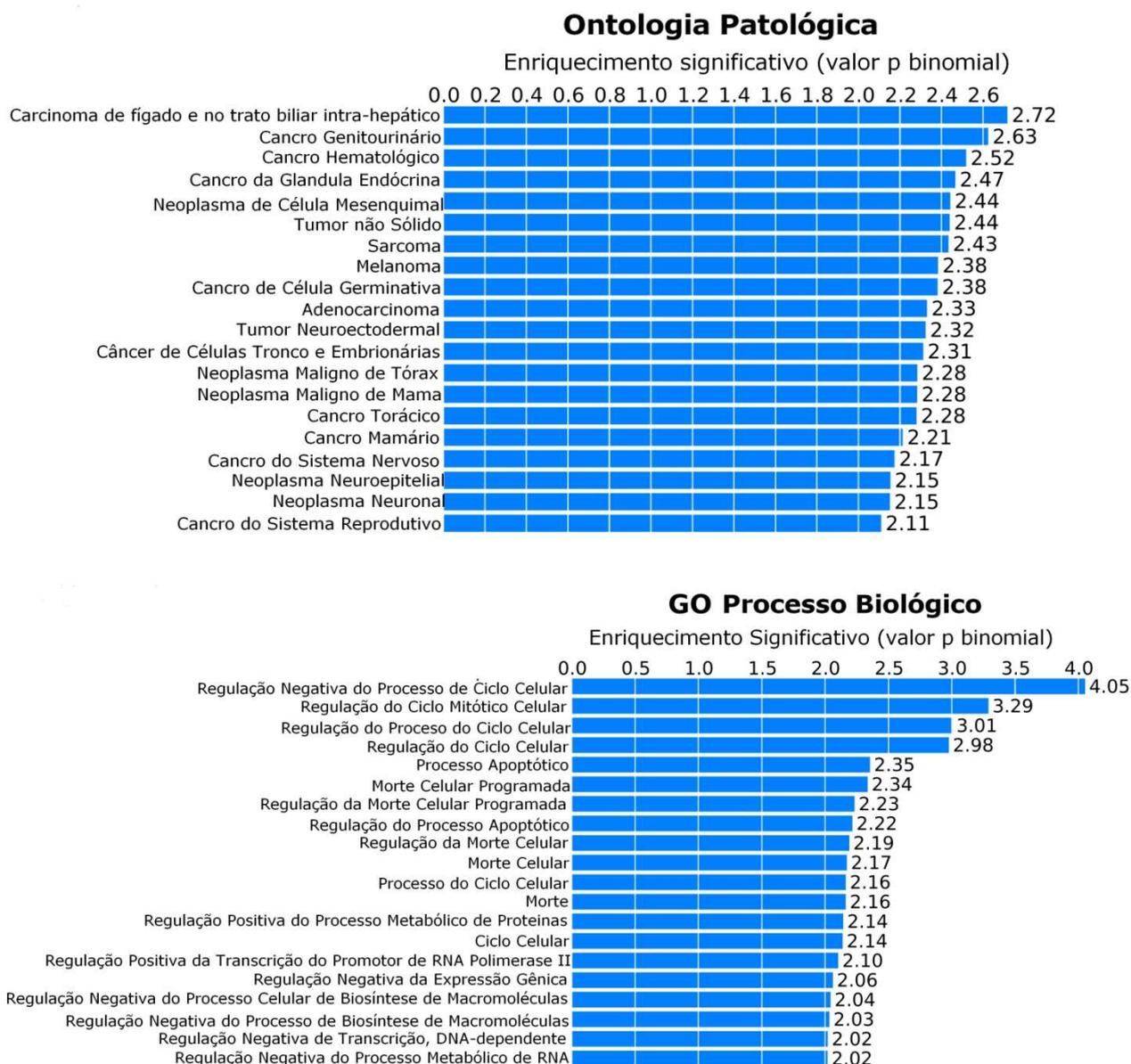


**Figura 2. Análise Ontológica dos oncogenes e genes supressores de tumor controles (28 genes).**

Observa-se um enriquecimento significativo em  $-\log_{10}$  para os genes controles quanto a funções biológicas e patologias associadas aos genes.

Para as análises de ontologia patológica realizadas foram utilizados aproximadamente 1890 genes chamados de alvos, entre eles oncogenes e genes supressores de tumor (Tabela S2 e S3), os quais ainda não foram descritos como imprintados pela literatura, observou-se que seiscentos genes estavam associados com carcinoma de fígado, outros trezentos genes para cancro hematológico e geniturinário. Os demais genes mostram-se distribuídos nas demais categorias de patologia observadas na Figura 3. A análise de ontologia patológica não revelou distinção entre oncogenes e genes supressores de tumor, ambos se apresentavam dentro das mesmas categorias avaliadas. Quanto as análises dos processos biológicos, observou-se que os genes supressores de tumor, apresentaram enriquecimento significativo ( $p = 4.18 \times 10^{-}$

<sup>06</sup>) nas categorias GO:0009987 (631/850 - processos celulares fisiológicos), GO:0050794 (643/850 - regulação dos processos celulares), GO:0042981 (210/850 – regulação dos processos apoptóticos) e GO:0043067 (213/850 – regulação da morte celular programada). Para os oncogenes foi obtido um enriquecimento significativo também ( $P = 6.23 \times 10^{-06}$ ) nas categorias GO:0009987 (978/1040 – processos celulares), GO:0008152 (694/1040 – processos metabólicos) e GO:0044260 (594/1040 – processos celulares metabólicos de macromoléculas) (Figura 3).



**Figure 3. Análise Ontológica dos oncogenes e genes supressores de tumor alvos (1890 genes).**

Observa-se um enriquecimento significativo em  $-\log_{10}$  para os genes alvos quanto a funções biológicas e patologias associadas aos genes.

### 6.3 Investigação de Novos Genes Candidatos ao Imprinting - GTE<sub>x</sub>

Dos 1890 genes utilizados para as análises, entre eles oncogenes e genes supressores de tumor (Tabela S2 e S3), muitos não foram consistentes com a evidência de *imprinting* (1250 genes) e foram classificados como bialélicos em  $\geq 1$  tecidos. Dos 640 genes com expressão estrita ou consistente com monoalélicos obtivemos 29 genes resultantes, 17 oncogenes (Tabela 4) e 12 genes supressores de tumor (Tabela 5), onde 11 genes mostraram-se

consistentes com a expressão monoalélica em  $\geq 1$  tecidos respectivamente. Sendo assim, nós identificamos 11 novos genes candidatos, prováveis ao *imprinting*. Destes 11 genes, apenas três tiveram evidências de possível *imprinting*, em Chr 14, Chr 11 e Chr 8, respectivamente *MEG3*, *INS-IGF2* e *PLAG1*.



**Tabela 5. Evidência de Imprinting em vários tecidos para os 12 Genes Supressores de Tumor detectados como possivelmente imprintados.**

A cor em laranja para os genes representa expressão predominante bialélica, em azul representa expressão predominante monoalélica. Os genes em negrito não apresentam forte evidência de expressão monoalélica. A escala de cor para os valores em cada tecido denota uma expressão bialélica em vermelho e monoalélica em azul. Análises de RNA-Seq Primário.

Gene	AdiposeSubcut	AdiposeVisceral	AdrenalGland	ArteryAorta	ArteryCoronary	ArteryTibial	Bladder	BrainAmygdala	BrainAntCinCortex	BrainCaudate	BrainCerebellHemi	BrainCerebellum	BrainCortex	BrainFrontCortex	BrainHippocampus	BrainHypothalamus	BrainNucAccumbens	BrainPutamen	BrainSpinalcord	BrainSubstanNigra	BreastMamTissue	ColonSigmoid	ColonTransverse	Ectocervix	Endocervix	EsophagusJunction	EsophagusMucosa	EsophagusMuscular	FallopianTube	HeartAtrialAppend	HeartLeftVentricle	KidneyCortex	Liver	Lung	MinorSalivGland	Musculoskeletal	NerveTibial	Ovary	Pancreas	Pituitary	Prostate	SkinExposed	SkinNotExposed	Smallintestine	Spleen	Stomach	Testis	Thyroid	Uterus	Vagina	WholeBlood	Xformedfibroblasts	Xformedlymphocytes					
<b>AHRH</b>	3	11	0	2	3	0	0	0	0	4	0	0	0	0	0	1	0	0	11	0	2	7	0	0	0	9	0	8	0	3	0	0	1	0	0	13	2	3	0	1	0	0	0	0	2	0	0	10	3	0	0	0	Monoalélico					
	3	6	3	0	7	1	11	5	2	0	0	1	7	3	7	6	4	0	0	2	13	11	8	2	1	7	3	8	0	0	0	9	0	17	3	0	13	15	0	5	4	10	8	7	16	6	24	4	5	10	0	26	26	Bialélico				
<b>BANP</b>	1	1	1	0	1	1	0	0	0	1	1	0	0	0	1	0	0	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	1	1	0	1	0	0	1	0	0	1	1	1	0	0	1	1	1	Monoalélico					
<b>CRNN/RP1-91G5.3</b>	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Bialélico			
	3	0	0	3	1	2	0	0	0	0	0	0	2	2	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	2	0	0	1	0	0	0	3	2	0	0	0	0	1	0	0	4	0	0	2	0	1	Monoalélico					
	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	2	0	2	7	5	0	0	0	0	0	2	0	0	2	0	0	7	5	0	0	0	0	1	7	0	1	0	0	Bialélico						
<b>FBXO32</b>	1	0	0	1	1	1	0	1	0	1	0	0	0	1	1	0	0	0	0	2	0	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	Monoalélico			
	20	18	14	21	21	21	9	10	8	7	0	1	14	10	11	14	8	9	19	13	21	21	21	7	4	21	21	21	7	21	21	7	0	7	17	21	21	20	10	13	21	19	20	21	10	20	6	12	21	20	5	20	2	Bialélico				
<b>H19/HOTS</b>	3	4	4	2	2	2	3	1	0	0	0	0	0	1	0	1	0	0	0	1	4	2	2	2	3	2	4	2	3	3	3	3	4	2	3	4	4	3	3	3	3	2	2	4	3	3	3	2	2	3	2	2	0	Monoalélico				
	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Bialélico	
<b>HIVEP3</b>	0	1	0	0	0	0	0	0	0	2	0	0	2	1	0	1	0	0	2	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	Monoalélico			
	1	11	1	2	0	1	0	10	12	5	22	16	16	15	8	11	5	2	13	9	14	1	0	0	0	1	12	0	0	3	1	1	0	13	12	1	1	0	2	15	6	18	20	3	2	0	0	10	1	1	1	16	23	Bialélico				
<b>ING5</b>	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1	1	0	0	1	2	0	1	1	0	0	1	2	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	Monoalélico			
	25	24	13	28	24	26	19	21	20	18	26	29	25	24	22	22	24	20	16	18	28	24	26	23	21	25	23	27	23	17	13	23	19	28	27	18	30	30	29	25	28	28	29	28	26	24	27	28	28	28	3	26	26	Bialélico				
<b>INS-IGF2/IGF2</b>	17	17	17	17	16	17	9	4	4	10	1	4	4	6	5	9	8	10	5	11	17	17	10	13	11	16	2	16	9	17	17	13	0	14	4	16	17	17	13	2	14	6	6	12	12	16	9	10	17	15	1	0	0	0	Monoalélico			
	0	0	0	0	0	0	1	1	2	0	3	3	2	0	0	0	0	0	6	0	0	0	7	0	0	1	8	1	3	0	0	0	16	2	11	1	0	0	3	13	0	10	8	3	2	0	6	6	0	0	7	17	0	Bialélico				
<b>KANK1</b>	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	Monoalélico					
	17	17	16	18	17	18	8	16	15	17	17	17	16	14	13	17	17	16	10	17	17	18	12	12	17	17	17	13	18	17	13	18	17	13	18	17	16	17	17	17	17	19	17	17	17	14	11	17	17	17	17	1	18	1	Bialélico			
<b>MEG3/RP11-123M6.2</b>	5	5	24	8	7	9	5	24	24	24	24	24	24	24	24	24	24	24	24	24	6	6	6	13	14	6	5	5	8	5	5	5	5	5	6	6	5	24	14	24	11	5	5	6	6	5	13	7	8	7	3	20	0	Monoalélico				
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Bialélico
<b>MRV1-AS1</b>	0	1	0	0	0	1	0	1	0	1	2	1	1	0	0	0	0	1	3	1	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	1	0	0	1	3	1	0	0	1	1	0	Monoalélico				
	10	9	9	10	10	9	5	9	9	9	4	8	9	9	9	9	9	9	7	9	9	9	9	7	7	9	9	8	5	8	9	7	3	10	6	7	9	9	7	8	9	9	8	9	9	9	7	9	9	7	9	8	0	Bialélico				
<b>PLAGL1</b>	3	2	3	3	3	3	3	0	0	1	1	1	1	1	1	0	1	0	0	1	3	3	3	3	2	3	3	3	2	2	2	1	0	3	2	1	2	3	3	3	3	3	3	2	0	3	2	1	3	3	0	3	0	Monoalélico				
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	Bialélico

#### 6.4 Análises de RNA-Seq Primário

Dos 24 tecidos que compunham o atlas de experimentos de RNA-Seq neoplásicos, apenas 11 tecidos foram utilizados para as análises, devido ao fato do número de experimentos de RNA-Seq para alguns tecidos terem sido insuficientes para as análises. O mesmo número e tipos de tecidos foi utilizado para o atlas de experimentos de RNA-Seq saudáveis, de modo que as análises pudessem ser comparadas entre os mesmos tecidos. As análises utilizaram 28 genes, entre eles oncogenes e genes supressores de tumor, os quais já foram descritos como imprintados segundo a literatura (CONSORTIUM, 2013; KOBAYASHI, 2014).

Os dados obtidos a partir das análises de RNA-Seq primário revelaram que apenas três dos tecidos que compunham o atlas tiveram informações significantes quanto ao efeito da possível perda ou ganho de *imprinting*: fígado, pulmão e mama respectivamente (Tabelas 6, 7 e 8). Embora alguns tecidos tenham demonstrado evidência do efeito de perda de *imprinting*, apenas os tecidos de fígado, pulmão e mama mostraram-se mais relevantes quando comparados entre si. Os demais tecidos que compunham o atlas não mostram evidências do efeito de perda de *imprinting*, entretanto trazem um novo panorama de como esses genes, os quais muitos ainda não foram estudados, estão sendo expressos em diferentes tecidos (Tabelas S4).

O status de monoalélico com presença de *flip* foi utilizado como um indicador de expressão exclusivamente monoalélica nas amostras, o mesmo foi utilizado para o status de bialélico com presença de *flip* indicando que este teria expressão totalmente bialélica nas amostras. Uma transição desse perfil de expressão de monoalélico com *flip* para bialélico com *flip* nos forneceu evidências de uma possível perda de *imprinting* nas análises. As amostras que apresentaram o status apenas de mono/bialélico sem *flip* foram considerados como não tendo uma expressão exclusivamente mono/bialélica.

Nas amostras de experimentos de RNA-Seq de fígado os SNPs *rs4758616*, *rs73409814* e *rs2067051* nos genes *NAP1L4*, *TSSC4* e *H19* respectivamente, apresentaram razões alélicas consistentes com expressão

bialélica com *flip* em tumor de fígado. Em contrapartida, os mesmos SNPs nos mesmos genes mostram razões alélicas consistentes com expressão monoalélica com *flip* no tecido de fígado saudável, sugestivo de perda de *imprinting* (Tabela 6). Ainda no mesmo tecido, notamos que o gene *MEG3* manteve seu perfil de expressão monoalélico com *flip* em ambos os tecidos, saudável e neoplásico. Não foram observados sinais ou evidências de ganho de *imprinting* neste tecido, entretanto os genes *AK123878* e *NSD1* mostram um perfil de expressão monoalélico sem *flip* em ambos os tecidos, saudável e neoplásico. O gene *SLC22A18* e *MEG3* manteve o perfil de expressão bialélico com *flip* em ambos os tecidos, saudável e neoplásico, entretanto os demais genes não mostram perfil de expressão bialélico em ambos os tecidos, saudável e neoplásico. Para os genes restantes a maioria dos SNPs analisados, o número de sequências foi insuficiente (<10 *reads*) dificultando obter-se uma análise mais confiável das razões alélicas.

Para as amostras de experimentos de RNA-Seq de pulmão os SNPs *rs10840159*, *rs4758616* e *rs2067051* no gene *HOTS*, *NAP1L4* e *H19* respectivamente, demonstraram razões alélicas também consistentes com expressão bialélica em tumor de pulmão, indicando possível perda de *imprinting*. Em contrapartida, os mesmos SNPs nos mesmos genes mostram razões alélicas consistentes com expressão monoalélica com *flip* no tecido de pulmão saudável, sugestivo de perda de *imprinting* (Tabela 7). Os genes *STIM1* mantiveram o perfil de expressão monoalélico com *flip* também em ambos os tecidos analisados. Neste tecido foi observado um possível sinal de ganho de *imprinting*, ainda não descrito na literatura para os SNPs *rs9672965* e *rs75224805* dos genes *IGFR1* e *NSD1* respectivamente, os quais se expressavam de forma bialélica com *flip* no tecido de pulmão saudável, mas em pulmão neoplásico se expressavam de forma monoalélica sem *flip*. Para o gene *AK123878* foi exibido um perfil de expressão monoalélico sem *flip* em ambos os tecidos analisados, saudável e neoplásico. Os demais genes mantiveram o perfil de expressão bialélico com *flip* respectivamente. Para os genes restantes, a maioria dos SNPs analisados, o número de sequências foi insuficiente (<10 *reads*) dificultando obter-se uma análise mais confiável das razões alélicas. Estes não mostram perfil de

expressão bialélico em ambos os tecidos, saudável e neoplásico, o mesmo se observou para o perfil de expressão monoalélico para os mesmos genes.

**Tabela 6. Análise do perfil de expressão em experimentos de RNA-Seq de fígado em oncogenes e genes supressores de tumor imprintados.**

Para cada gene foi observado se não houve expressão bialélica (NB). Se não houve expressão monoalélica (NM). Foi observado se o status flip foi positivo (+) ou negativo (-) e se foram observados dados quanto à expressão monoalélica (Mono), bialélica (Bi), ou ausência de dados (NA).

Tecidos	Quantidade de experimentos	Status da expressão	SNPs	Fígado					
				77			262		
				Mono	Bi	Flip	Mono	Bi	Flip
Experimentos com Reads >=10		<b>AK123878</b>	<b>rs10135233</b>	7	NB	mono (-)	9	NB	mono (+)
		<b>BLCAP</b>	<b>rs6019103</b>	9	NB	mono (-)	54	NB	mono (-)
		<b>CDKN1C</b>	<b>rs191294997</b>	NA	NA	NA	NA	NA	NA
		<b>DIRAS3</b>	<b>rs982462</b>	NA	NA	NA	NA	NA	NA
		<b>H19</b>	<b>rs2067051</b>	34	1	mono (+)	105	16	mono/bi (+)
		<b>HOTS</b>	<b>rs10840159</b>	NA	NA	NA	2	NB	mono (-)
		<b>IGF1R</b>	<b>rs9672965</b>	1	NB	mono (-)	NA	NA	NA
		<b>IGF2</b>	<b>rs10770125</b>	NA	NA	NA	NA	NA	NA
		<b>IGF2R</b>	<b>rs2014620</b>	NA	NA	NA	NA	NA	NA
		<b>LPAR6</b>	<b>rs4151551</b>	NA	NA	NA	NA	NA	NA
		<b>MEG3</b>	<b>rs1053900</b>	11	10	mono/bi (+)	36	21	mono/bi (+)
		<b>MIR372</b>	<b>rs3859501</b>	NA	NA	NA	1	NB	mono (-)
		<b>NAP1L4</b>	<b>rs4758616</b>	48	NB	mono (+)	139	19	mono/bi (+)
		<b>NDN</b>	<b>rs2192206</b>	39	NB	mono (-)	175	NB	mono (-)
		<b>NSD1</b>	<b>rs75224805</b>	23	NB	mono (+)	108	NB	mono (+)
		<b>PEG10</b>	<b>rs12704734</b>	1	NB	mono (-)	21	1	mono (-)
		<b>PLAGL1</b>	<b>rs12528876</b>	1	1	mono/bi (-)	2	NB	mono (-)
		<b>RB1</b>	<b>rs4151611</b>	NA	NA	NA	NA	NA	NA
		<b>SDHD</b>	<b>rs11214077</b>	NA	NA	NA	NA	NA	NA
		<b>SLC22A18</b>	<b>rs1048047</b>	28	6	mono/bi (+)	91	47	mono/bi (+)
		<b>SNRPN</b>	<b>rs12907375</b>	NA	1	bi (-)	NA	NA	NA
		<b>STIM1</b>	<b>rs11604649</b>	NA	NA	NA	NA	NA	NA
		<b>TFPI2</b>	<b>rs1024389</b>	NA	NA	NA	NM	3	bi (-)
		<b>TP73</b>	<b>rs1885866</b>	NA	NA	NA	14	NB	mono (-)
		<b>TRIM28</b>	<b>rs77659740</b>	NA	NA	NA	NA	NA	NA
		<b>TSPAN32</b>	<b>rs2521273</b>	NA	NA	NA	NA	NA	NA
		<b>TSSC4</b>	<b>rs73409814</b>	10	NB	mono (+)	38	2	mono/bi (+)
		<b>UBE3A</b>	<b>rs12899183</b>	7	NB	mono (-)	67	NB	mono (+)

**Tabela 7. Análise do perfil de expressão em experimentos de RNA-Seq de pulmão em oncogenes e genes supressores imprintados.**

Para cada gene foi observado se não houve expressão bialélica (NB). Se não houve expressão monoalélica (NM). Foi observado se o status flip foi positivo (+) ou negativo (-) e se foram observados dados quanto à expressão monoalélica (Mono), bialélica (Bi), ou ausência de dados (NA).

Experimentos com Reads >=10	Tecidos	Quantidade de experimentos	Pulmão							
			177			393				
			Mono	Bi	Flip	Mono	Bi	Flip		
		Status da expressão								
	<b>AK123878</b>	<b>rs10135233</b>	8	NB	mono (-)	9	NB	mono (-)		
	<b>BLCAP</b>	<b>rs6019103</b>	16	NB	mono (-)	132	1	mono (+)		
	<b>CDKN1C</b>	<b>rs191294997</b>	NA	NA	NA	1	NB	mono (-)		
	<b>DIRAS3</b>	<b>rs982462</b>	1	NA	mono (-)	NA	NA	NA		
	<b>H19</b>	<b>rs2067051</b>	42	NB	mono (+)	43	10	mono/bi (+)		
	<b>HOTS</b>	<b>rs10840159</b>	1	NB	mono (-)	13	2	mono/bi (-)		
	<b>IGF1R</b>	<b>rs9672965</b>	4	1	mono (+)	7	NB	mono (+)		
	<b>IGF2</b>	<b>rs10770125</b>	NA	1	bi (-)	3	NB	mono (+)		
	<b>IGF2R</b>	<b>rs2014620</b>	NA	NA	NA	NM	1	bi (-)		
	<b>LPAR6</b>	<b>rs4151551</b>	3	NB	mono (-)	9	1	mono (+)		
	<b>MEG3</b>	<b>rs1053900</b>	NA	NA	NA	NM	1	bi (-)		
	<b>MIR372</b>	<b>rs3859501</b>	NA	NA	NA	NA	NA	NA		
	<b>NAP1L4</b>	<b>rs4758616</b>	82	NB	mono (+)	155	20	mono/bi (+)		
	<b>NDN</b>	<b>rs2192206</b>	106	2	mono/bi (-)	269	7	mono/bi (-)		
	<b>NSD1</b>	<b>rs75224805</b>	31	2	mono/bi (+)	104	NB	mono (+)		
	<b>PEG10</b>	<b>rs12704734</b>	4	NB	mono (-)	16	2	mono/bi (-)		
	<b>PLAGL1</b>	<b>rs12528876</b>	1	NB	mono (-)	3	1	mono/bi (-)		
	<b>RB1</b>	<b>rs4151611</b>	NA	NA	NA	NA	NA	NA		
	<b>SDHD</b>	<b>rs11214077</b>	1	NB	mono (-)	NA	NA	NA		
	<b>SLC22A18</b>	<b>rs1048047</b>	5	4	mono/bi (+)	51	20	mono/bi (+)		
	<b>SNRPN</b>	<b>rs12907375</b>	NA	NA	NA	1	NB	mono (-)		
	<b>STIM1</b>	<b>rs11604649</b>	NA	NA	NA	2	NB	mono (+)		
	<b>TFPI2</b>	<b>rs1024389</b>	4	NA	mono (+)	5	NB	mono (+)		
	<b>TP73</b>	<b>rs1885866</b>	NA	NA	NA	2	NB	mono (-)		
	<b>TRIM28</b>	<b>rs77659740</b>	NA	NA	NA	6	NB	mono (+)		
	<b>TSPAN32</b>	<b>rs2521273</b>	NA	NA	NA	1	NB	mono (-)		
	<b>TSSC4</b>	<b>rs73409814</b>	72	NB	mono (+)	148	1	mono (+)		
	<b>UBE3A</b>	<b>rs12899183</b>	56	NB	mono (-)	150	24	mono/bi (+)		

Nas análises no tecido de mama foram observadas frações alélicas consistentes com possível perda de *imprinting* dos SNPs *rs10840159*, e *rs2192206* para os genes *H19* e *NDN* respectivamente, foram observadas em câncer de mama. Notou-se que estes genes se expressavam de forma monoalélica em tecidos saudáveis e posteriormente de forma bialélica em tecidos neoplásicos com a presença do *flip* (Tabela 8). Observou-se também que o gene *NSD1* manteve a expressão monoalélica sem *flip*. Para os genes restantes a maioria dos SNPs analisados, o número de sequências foi insuficiente (<10 *reads*) dificultando obter-se uma análise mais confiável das razões alélicas, de modo que estes não mostram perfil de expressão bialélico

em ambos os tecidos, saudável e neoplásico, o mesmo se observou para o perfil de expressão monoalélico para os mesmos genes.

**Tabela 8. Análise do perfil de expressão em experimentos de RNA-Seq de mama em oncogenes e genes supressores imprintados.**

Para cada gene foi observado se não houve expressão bialélica (NB). Se não houve expressão monoalélica (NM). Foi observado se o status flip foi positivo (+) ou negativo (-) e se foram observados dados quanto à expressão monoalélica (Mono), bialélica (Bi), ou ausência de dados (NA).

Experimentos com Reads >=10	Tecidos	Mama					
	Quantidade de experimentos	61			517		
	Status da expressão	Mono	Bi	Flip	Mono	Bi	Flip
<i>AK123878</i>	NA	NA	NA	NA	NA	NA	NA
<i>BLCAP</i>	2	NB	mono (-)	8	NB	mono (-)	
<i>CDKN1C</i>	NA	NA	NA	NA	NA	NA	NA
<i>DIRAS3</i>	NA	NA	NA	NA	NA	NA	NA
<i>H19</i>	4	NB	mono (+)	88	6	mono/bi (+)	
<i>HOTS</i>	NA	NA	NA	1	1	mono/bi (-)	
<i>IGF1R</i>	NA	NA	NA	NA	NA	NA	NA
<i>IGF2</i>	NA	NA	NA	2	NB	mono (-)	
<i>IGF2R</i>	NA	NA	NA	NA	NA	NA	NA
<i>LPAR6</i>	NA	NA	NA	NA	NA	NA	NA
<i>MEG3</i>	NA	NA	NA	NM	1	bi (-)	
<i>MIR372</i>	NA	NA	NA	NA	NA	NA	NA
<i>NAP1L4</i>	4	NB	mono (+)	119	NB	mono/bi (+)	
<i>NDN</i>	5	NB	mono (-)	148	1	mono (+)	
<i>NSD1</i>	4	NB	mono (+)	32	NB	mono (+)	
<i>PEG10</i>	NA	NA	NA	2	1	mono/ bi (-)	
<i>PLAGL1</i>	NA	NA	NA	NA	NA	NA	NA
<i>RB1</i>	NA	NA	NA	NA	NA	NA	NA
<i>SDHD</i>	NA	NA	NA	NA	NA	NA	NA
<i>SLC22A18</i>	NM	3	bi (-)	21	8	mono/bi (+)	
<i>SNRPN</i>	NA	NA	NA	NA	NA	NA	NA
<i>STIM1</i>	NA	NA	NA	NA	NA	NA	NA
<i>TFPI2</i>	NA	NA	NA	NA	NA	NA	NA
<i>TP73</i>	NA	NA	NA	NA	NA	NA	NA
<i>TRIM28</i>	NA	NA	NA	1	1	mono/bi (-)	
<i>TSPAN32</i>	NA	NA	NA	NA	NA	NA	NA
<i>TSSC4</i>	4	NB	mono (+)	39	2	mono/bi (+)	
<i>UBE3A</i>	3	NB	mono (-)	86	1	mono (-)	

## 7. DISCUSSÃO

Nos últimos anos, o interesse no metabolismo do câncer tem sido alimentado pela identificação do envolvimento de oncogenes e genes supressores de tumor no controle de caminhos metabólicos, visto que estes podem estar influenciando nos processos metabólicos do organismo humano. Muitas dessas alterações metabólicas levam a mudanças dramáticas no balanço de bioenergia, biossíntese e equilíbrio redox dentro das células tumorais, mostrando ter uma complexa relação entre o metabolismo dessas células tumorais e o microambiente do tumor (MORISON *et al.*, 2001; AGNIHOTRI e ZADEH, 2016). Sendo assim, uma das principais forças motrizes no estudo do metabolismo do câncer é a necessidade de entender de forma mais clara como o metabolismo é reprogramado nas células cancerosas e as diferenças que existem entre o volume de células tumorais e as células raras com propriedades semelhantes a células-tronco (COLQUHOUN, 2017).

Vários outros estudos já revelaram que a ativação de certos oncogenes ou inativação de supressores de tumores podem interferir na homeostase metabólica, além de poderem estar envolvidos na promoção ou inibição da progressão do câncer (COLQUHOUN, 2017). Também é conhecido que os processos que envolvem genes supressores de tumor, bem como oncogenes exercem seu maior efeito visando reguladores particulares da progressão da fase G1 (HEICHMAN e ROBERTS, 1994; HUNTER e PINES, 1994; HAI *et al.*, 2018).

Segundo os estudos atuais sobre o uso da técnica de RNA-Seq, tanto para análises de expressão gênica em câncer quanto em outras análises de expressão gênica, esta se tornou uma abordagem desenvolvida para analisar o perfil do transcriptoma que usa tecnologias de sequenciamento profundo. Estudos usando este método já alteraram nossa visão da extensão da complexidade de análises de transcriptoma em eucariotos (EVAN e VOUSDEN, 2001). Esta é uma técnica que oferece várias vantagens em relação às tecnologias existentes, pois ao contrário das abordagens baseadas em hibridação, o RNA-Seq não se limita a detectar transcritos que correspondem à sequência genômica existente. Além disso, essa técnica apresenta outra

vantagem em relação ao uso de *microarrays* de DNA. O RNA-Seq também mostrou ser altamente preciso para quantificar os níveis de expressão, de modo que os resultados do uso de tal técnica também mostram altos níveis de reprodutibilidade (NAGALAKSHMI *et al.*, 2008; VERA *et al.*, 2008; WANG *et al.*, 2009).

A construção de uma biblioteca de experimentos de RNA-Seq por sua vez então se mostra um método robusto para realizar análises de transcriptoma, além de permitir análises quanto à expressão gênica. Embora existam alguns percalços durante a criação da mesma, como desafios quanto a bioinformática, devido ao grande volume de dados gerados para serem analisados. Dessa forma, assim como outras tecnologias de sequenciamento de alto rendimento, o RNA-Seq enfrenta vários desafios informáticos, incluindo o desenvolvimento de métodos eficientes para armazenar, recuperar e processar grandes quantidades de dados, de modo a evitar obter-se baixos níveis de *reads* (CLOONAN *et al.*, 2008; MORTAZAVI *et al.*, 2008; NAGALAKSHMI *et al.*, 2008; WANG *et al.*, 2009).

Levando em consideração que muitos estudos tem como objetivo caracterizar o *imprinting* no desenvolvimento de seu papel funcional, seja na função placentária ou no crescimento do feto, os dados obtidos aqui geram evidências dos padrões de expressão monoalélica e também dos padrões de *imprinting* em adultos, os quais também podem ter consequências funcionais (WU *et al.*, 2008; UBEDA e GARDNER, 2011). Dessa forma os padrões de expressão monoalélicos bem como os padrões de *imprinting* obtidos neste estudo fornecem também dados para avaliar teorias e hipóteses para questões evolutivas sobre o *imprinting* (BARTOLOMEI e FERGUSON-SMITH, 2011; DENT e ISLES, 2014), pois em princípio, é possível estimar os perfis de transcrição alélica em sítios hipoteticamente polimórficos determinando as frações das sequências de leitura (*reads*) para cada possível alelo. Desde que Luedi e colaboradores divulgaram seu estudo sobre predições de novos genes que poderiam ser ou estarem imprintados, poucas pesquisas tem se esforçado na validação experimental do padrão de expressão destes genes (PATTEN *et al.*, 2014).

É válido também lembrar que existem limitações para análises de perfil de expressão monoalélica ou a respeito do *imprinting* em seres humanos, pois assim como já foi dito por Baran e colaboradores (2015), os resultados de *imprinting* das análises baseadas em populações precisam ser validados em dados familiares em larga escala para provar formalmente que os efeitos detectados são verdadeiramente relacionados aos parentes e não conduzidos por fatores de confusão técnicos ou biológicos. E devem permitir também a detecção de *imprinting*, ainda que sutil e que não leve a expressão monoalélica, e por fim permitir análises formais de sensibilidade *versus* especificidade (LUEDI *et al.*, 2007). A abordagem utilizada neste estudo é conservadora, pois o foco principal foi obter-se baixos números de falsos positivos, entretanto ainda é possível que haja perda de alguns efeitos verdadeiros, o que pode incluir a heterogeneidade de genes imprintados.

Em geral, a expressão e o *imprinting* de novos *loci* são restritos a menor quantidade de tecidos o que pode explicar porque eles não foram previamente detectados. Dos 105 genes previamente identificados como imprintados, com confiança em seres humanos, de acordo com o banco de dados de Otago (BARAN, SUBRAMANIAM, BITON, TUKIAINEN, TSANG, RIVAS, PIRINEN, GUTIERREZ-ARCELUS, SMITH, KUKURBA, ZHANG, ENG, TORGERSON, URBANEK, LI, *et al.*, 2015) poucos genes não possuíam dados em nossa análise. Esses 11 genes candidatos ao *imprinting* mostraram ter três ou mais SNPs por tecidos, critério o qual foi utilizado para determinar se estes seriam candidatos ao *imprinting* genômico de acordo com Baran e colaboradores (2015). A falta de suporte para o *status* de *imprinting* em nossos dados pode ser devido a vários fatores: os catálogos que são derivados de fontes heterogêneas e podem incluir alguns falsos positivos, o tecido imprintado pode não estar presente ou bem coberto por nossos dados, e as fases iniciais de desenvolvimento não são capturados em nossas amostras adultas. Um exemplo desses padrões é o *locus* GNAS complexo que tem sido relatado como imprintado de forma específica de tipo celular (MORISON *et al.*, 2001) e, portanto, a mistura de tipos de células nos tecidos GTEx, a falta de tecidos chave e a incapacidade de distinguir os transcritos sobrepostos pode contribuir para o

sinal de expressão bialélica. No geral, os resultados obtidos sugerem que os catálogos anteriores de *imprinting* humano não demonstram de forma perfeita o *imprinting* em adultos.

De acordo com a literatura e alguns autores como Onyango (1997), Jay (2008) e Kobayashi (2014), sabe-se que os genes *HOTS*, *H19*, *NAP1L4*, *NSD1* e *NDN* são genes imprintados, onde o gene *HOTS* é o que eles chamam de supressor de tumor oposto ao gene *H19*. Já o gene *H19*, este é bem conhecido por ser um supressor tumor modelo para estudos de Tumor de Wilms e imprintado em vários tecidos como cérebro, fígado e outros, enquanto o gene *NDN* é conhecido por ter um perfil de expressão tecido específico em cérebro, fibroblastos e vários outros tecidos, sendo um gene supressor de tumor. O gene *NAP1L4* por sua vez codifica um membro da família de proteínas de montagem de nucleossomas (NAP). Este gene é um dos vários localizados perto do domínio de *imprinting* 11p15.5, uma importante região do gene supressor de tumor. Já o gene *NSD1* é um oncogene, o qual codifica uma proteína contendo um domínio SET. Esta proteína pode atuar como um fator transcricional básico e também como um regulador transcriptional bifuncional. (JAY *et al.*, 1997; RODRIGUEZ *et al.*, 2004; BASTEPE, 2007; ONYANGO e FEINBERG, 2011; KOBAYASHI, 2014; BARAN, SUBRAMANIAM, BITON, TUKIAINEN, TSANG, RIVAS, PIRINEN, GUTIERREZ-ARCELUS, SMITH, KUKURBA, ZHANG, ENG, TORGERSON, URBANEK, LI, *et al.*, 2015).

A partir dos dados obtidos neste estudo foi possível observar uma possível perda de *imprinting* genômico para os genes *H19*, *NAP1L4* e *HOTS* em tecidos como fígado, pulmão e mama com câncer, os quais segundo alguns autores já se sabe que sofrem este fenômeno de perda de *imprinting*. De acordo com Schwienbacher e colaboradores (2000) já foi descrito a provável ocorrência de perda de *imprinting* para o gene *H19* em hepatocarcinomas, mas não só nesse tipo de câncer, pois segundo Hibi (*et al.* 1996) esse mesmo fenômeno já foi observado para câncer de esôfago, de modo que este gene já foi frequentemente super expressado em ambos os tipos de câncer em comparação com os tecidos normais correspondentes, e o fenômeno de perda de *imprinting* (LOI) de *H19* sempre esteve associada à sua super expressão, corroborando assim os dados observados quanto a possível perda de *imprinting* deste gene. Da mesma forma,

assim como o gene *HOTS* pode funcionar como um gene supressor de tumor, ele também está sujeito à perda epigenética (LOI) e genética durante a tumorigênese, mas ainda não se tem descrito se tal gene tem sofrido o efeito de LOI em algum tipo de tecido acometido por câncer, de modo que os dados aqui demonstrados são os primeiros dados obtidos que trazem tal evidência quanto a este gene. Além disso, nos casos de LOI de *IGF2*, o gene *H19* é silenciado e a expressão do gene *HOTS* é perdida, sugerindo que *HOTS* poderia ser o gene supressor de tumor expressado maternalmente pensado para residir na região 11p15 (HIBI *et al.*, 1996; SCHWIENBACHER *et al.*, 2000; HAN *et al.*, 2017).

Quanto ao gene *NDN*, sabe-se que este codifica uma proteína putativa homóloga à proteína NECDIN específica do cérebro de camundongos. Aqui neste estudo demonstramos que este poderia ter uma possível LOI em tecido de mama com câncer, pois sendo este um gene imprintado exclusivamente expressando o alelo paterno em tecidos como cérebro, fígado, coração, pulmão, mas não em mama, podemos inferir que esteja ocorrendo uma possível perda do *imprinting* genômico neste gene (JAY *et al.*, 1997; ONYANGO e FEINBERG, 2011).

Um dos primeiros sinais da descrição do fenômeno de ganho de *imprinting* genômico na literatura foi constatado por Schwienbacher e colaboradores (2000), ao demonstrar o ganho de *imprinting* (GOI) somático em genes da região 11p15 expressados pela mãe, de modo que os genes normalmente ativos são reprimidos. Posteriormente o mesmo fenômeno foi descrito por Gallagher e colaboradores (2006), onde este descrevia tal fenômeno como uma anormalidade mais frequente nos carcinomas hepatocelulares, o que mostrou resultar em uma redução concomitante da expressão de outros genes, de modo que o GOI afeta tanto as transcrições que ocorrem em *sense* e *antisense* neste local (SCHWIENBACHER *et al.*, 2000; GALLAGHER *et al.*, 2006).

Segundo Schwienbacher e colaboradores (2000), em tecidos somáticos, a expressão de um gene imprintado é detectada como uma diferença no nível de expressão entre os dois alelos, como observado nos dados obtidos para ambos os genes *IGFR1* e *NSD1*. Para provar essa diferença no nível de expressão, Schwienbacher realizou por sua vez análises em genes como

*SLC22A1L* e *CDKN1C* em fígado saudável e revelou um grau variável de expressão (SCHWIENBACHER *et al.*, 2000). Aqui neste estudo demonstramos então sinais de possível GOI para os genes *IGFR1* e *NSD1* em pulmão, os quais ainda não foram descritos na literatura com associação a tal efeito. Entretanto estudos mais recentes tem demonstrado que alterações epigenéticas no tecido neoplásico que venham a causar perda completa de expressão é melhor descrita como um "ganho de *imprinting*" somático neste *locus*, uma vez que a expressão para ambos os genes *IGFR1* e *NSD1* se mostrou de forma monoalélica sem *flip* neste estudo (com presença de apenas um alelo), mostrando então sinais de perda de expressão nos dados obtidos (CHRISTOFORI *et al.*, 1995; DAO *et al.*, 1999; SMILINICH *et al.*, 1999).

O fenômeno de GOI já foi detectado em uma grande fração de hepatocarcinomas e pode representar um mecanismo patogênico, possivelmente ocorrendo em outras neoplasias ou durante o desenvolvimento. Em princípio, a perda de expressão de alelos maternos em tumores pediátricos tem a mesma consequência que o GOI em hepatocarcinomas, de inativar os genes (HENRY *et al.*, 1993; TYCKO, 2000), além disso, este mecanismo lembra os casos de Síndrome de Beckwith-Wiedemann com dissomia uniparental paterna (SLATTER *et al.*, 1994; SCHWIENBACHER *et al.*, 2000). Dessa forma as descobertas sobre GOI segundo a literatura, levantam a possibilidade de que tal fenômeno possa ser um evento inicial na tumorigênese (HENRY *et al.*, 1993).

## 8. CONCLUSÕES

O acervo de experimentos de RNA-Seq, os quais compõem os atlas saudável e neoplásico, irão proporcionar consultar a expressão alélica de genes candidatos ao *imprinting* genômico, de modo a obter-se as razões alélicas dos mesmos.

Os dados obtidos neste estudo sugerem possível perda de *imprinting* para alguns genes como *H19*, *NSD1*, *NAP1L4*, *TSSC4*, *HOTS* e *NDN* em tecidos neoplásicos de fígado, pulmão e mama. Os dados também sugerem um possível sinal de ganho de *imprinting* para os genes *IGFR1* e *NSD1* em tecido de pulmão neoplásico.

Para muitos dos genes estudados o número de sequências obtidas para cada SNP foi insuficiente (<10 *reads*), dificultando obter-se assim uma análise mais confiável das razões alélicas.

Para alguns dos tecidos analisados não foi possível observar se os genes apresentavam expressão monoalélica ou bialélica, talvez devido ao fato de o número de experimentos de RNA-Seq que compõem esses tecidos terem sido insuficientes.

Os Dados obtidos a partir das análises de RNA-Seq secundário neste estudo na plataforma GTEx, nos fornecem evidências de possível *imprinting* para os genes *CRNN*, *DCD* e *TBC1D7-LOC100130357*, os quais demonstraram expressão consistente com monoalélica em tecidos como tireoide, estômago e cérebro respectivamente.

## 9. REFERENCIAS

ABU-AMERO, S. et al. The genetic aetiology of Silver-Russell syndrome. **J Med Genet**, v. 45, n. 4, p. 193-9, Apr 2008.

AGNIHOTRI, S.; ZADEH, G. Metabolic reprogramming in glioblastoma: the influence of cancer metabolism on epigenetics and unanswered questions. **Neuro Oncol**, v. 18, n. 2, p. 160-72, Feb 2016.

ALVES DA SILVA, A. F. et al. Trisomy 21 Alters DNA Methylation in Parent-of-Origin-Dependent and -Independent Manners. **PLoS One**, v. 11, n. 4, p. e0154108, 2016.

ANWAR, S. L. et al. Loss of DNA methylation at imprinted loci is a frequent event in hepatocellular carcinoma and identifies patients with shortened survival. **Clin Epigenetics**, v. 7, p. 110, 2015.

APARICIO, S. A.; CALDAS, C.; PONDER, B. Does massively parallel transcriptome analysis signify the end of cancer histopathology as we know it? **Genome Biol**, v. 1, n. 3, p. REVIEWS1021, 2000.

BARAN, Y. et al. The landscape of genomic imprinting across diverse adult human tissues. **Genome Res**, v. 25, n. 7, p. 927-36, Jul 2015. ISSN 1549-5469 (Electronic)

1088-9051 (Linking).

BARAN, Y. et al. The landscape of genomic imprinting across diverse adult human tissues. **Genome Res**, v. 25, n. 7, p. 927-36, Jul 2015.

BARBAZUK, W. B. et al. SNP discovery via 454 transcriptome sequencing. **Plant J**, v. 51, n. 5, p. 910-8, Sep 2007.

BARTOLOMEI, M. S.; FERGUSON-SMITH, A. C. Mammalian genomic imprinting. **Cold Spring Harb Perspect Biol**, v. 3, n. 7, Jul 1 2011.

BARTOLOMEI, M. S.; ZEMEL, S.; TILGHMAN, S. M. Parental imprinting of the mouse H19 gene. **Nature**, v. 351, n. 6322, p. 153-5, May 9 1991.

BASTEPE, M. The GNAS Locus: Quintessential Complex Gene Encoding Galpha, XLalphas, and other Imprinted Transcripts. **Curr Genomics**, v. 8, n. 6, p. 398-414, Sep 2007.

BISHOP, J. M. Molecular themes in oncogenesis. **Cell**, v. 64, n. 2, p. 235-48, Jan 25 1991.

BJORNSSON, H. T. et al. Epigenetic specificity of loss of imprinting of the IGF2 gene in Wilms tumors. **J Natl Cancer Inst**, v. 99, n. 16, p. 1270-3, Aug 15 2007.

CANTLEY, L. C. et al. Oncogenes and signal transduction. **Cell**, v. 64, n. 2, p. 281-302, Jan 25 1991.

CHEN, C. L. et al. Loss of imprinting of the IGF-II and H19 genes in epithelial ovarian cancer. **Clin Cancer Res**, v. 6, n. 2, p. 474-9, Feb 2000.

CHRISTOFORI, G.; NAIK, P.; HANAHAHAN, D. Deregulation of both imprinted and expressed alleles of the insulin-like growth factor 2 gene during beta-cell tumorigenesis. **Nat Genet**, v. 10, n. 2, p. 196-201, Jun 1995.

CLOONAN, N. et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. **Nat Methods**, v. 5, n. 7, p. 613-9, Jul 2008.

COLQUHOUN, A. Cell biology-metabolic crosstalk in glioma. **Int J Biochem Cell Biol**, v. 89, p. 171-181, Aug 2017.

CONSORTIUM, G. T. The Genotype-Tissue Expression (GTEx) project. **Nat Genet**, v. 45, n. 6, p. 580-5, Jun 2013. ISSN 1546-1718 (Electronic) 1061-4036 (Linking).

COOPER, W. N. et al. Molecular subtypes and phenotypic expression of Beckwith-Wiedemann syndrome. **Eur J Hum Genet**, v. 13, n. 9, p. 1025-32, Sep 2005.

DAMODARAN, S.; BERGER, M. F.; ROYCHOWDHURY, S. Clinical tumor sequencing: opportunities and challenges for precision cancer medicine. **Am Soc Clin Oncol Educ Book**, p. e175-82, 2015.

DAO, D. et al. Multipoint analysis of human chromosome 11p15/mouse distal chromosome 7: inclusion of H19/IGF2 in the minimal WT2 region, gene specificity of H19 silencing in Wilms' tumorigenesis and methylation hyper-dependence of H19 imprinting. **Hum Mol Genet**, v. 8, n. 7, p. 1337-52, Jul 1999.

DEJEUX, E. et al. Hypermethylation of the IGF2 differentially methylated region 2 is a specific event in insulinomas leading to loss-of-imprinting and overexpression. **Endocr Relat Cancer**, v. 16, n. 3, p. 939-52, Sep 2009.

DENT, C. L.; ISLES, A. R. Brain-expressed imprinted genes and adult behaviour: the example of Nesp and Grb10. **Mamm Genome**, v. 25, n. 1-2, p. 87-93, Feb 2014.

ECKERSLEY-MASLIN, M. A. et al. Random monoallelic gene expression increases upon embryonic stem cell differentiation. **Dev Cell**, v. 28, n. 4, p. 351-65, Feb 24 2014.

EVAN, G. I.; VOUSDEN, K. H. Proliferation, cell cycle and apoptosis in cancer. **Nature**, v. 411, n. 6835, p. 342-8, May 17 2001.

FEIL, R.; BERGER, F. Convergent evolution of genomic imprinting in plants and mammals. **Trends Genet**, v. 23, n. 4, p. 192-9, Apr 2007.

FERGUSON-SMITH, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. **Nat Rev Genet**, v. 12, n. 8, p. 565-75, Aug 2011.

GALLAGHER, E. et al. Gain of imprinting of SLC22A18 sense and antisense transcripts in human breast cancer. **Genomics**, v. 88, n. 1, p. 12-7, Jul 2006.

GAO, C. M. et al. Glutathione-S-transferases M1 (GSTM1) and GSTT1 genotype, smoking, consumption of alcohol and tea and risk of esophageal and stomach cancers: a case-control study of a high-incidence area in Jiangsu Province, China. **Cancer Lett**, v. 188, n. 1-2, p. 95-102, Dec 15 2002.

GENDREL, A. V. et al. Developmental dynamics and disease potential of random monoallelic gene expression. **Dev Cell**, v. 28, n. 4, p. 366-80, Feb 24 2014.

GENDREL, A. V. et al. Random monoallelic expression of genes on autosomes: Parallels with X-chromosome inactivation. **Semin Cell Dev Biol**, v. 56, p. 100-110, Aug 2016.

GICQUEL, C. et al. Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver-Russell syndrome. **Nat Genet**, v. 37, n. 9, p. 1003-7, Sep 2005.

GIMELBRANT, A. A.; CHESS, A. An epigenetic state associated with areas of gene duplication. **Genome Res**, v. 16, n. 6, p. 723-9, Jun 2006.

GINART, P. et al. Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. **Genes Dev**, v. 30, n. 5, p. 567-78, Mar 1 2016.

GOUGH, N. R. Focus issue: From genomic mutations to oncogenic pathways. **Sci Signal**, v. 6, n. 268, p. eg3, Mar 26 2013.

GROSSMAN, R. L. et al. Toward a Shared Vision for Cancer Genomic Data. **N Engl J Med**, v. 375, n. 12, p. 1109-12, Sep 22 2016.

HAI, M. F. et al. Expression and potential mechanism of metabolism-related genes and CRLS1 in non-small cell lung cancer. **ONCOLOGY LETTERS**, 2018.

HAN, J. Y. et al. Identification of a novel de novo nonsense mutation of the NSD1 gene in monozygotic twins discordant for Sotos syndrome. **Clin Chim Acta**, v. 470, p. 31-35, Jul 2017.

HEICHMAN, K. A.; ROBERTS, J. M. Rules to replicate by. **Cell**, v. 79, n. 4, p. 557-62, Nov 18 1994.

HENCKEL, A.; ARNAUD, P. Genome-wide identification of new imprinted genes. **Brief Funct Genomics**, v. 9, n. 4, p. 304-14, Jul 2010.

HENRY, I. et al. Somatic mosaicism for partial paternal isodisomy in Wiedemann-Beckwith syndrome: a post-fertilization event. **Eur J Hum Genet**, v. 1, n. 1, p. 19-29, 1993.

HIBI, K. et al. Loss of H19 imprinting in esophageal cancer. **Cancer Res**, v. 56, n. 3, p. 480-2, Feb 1 1996.

HILLIER, L. W. et al. Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. **Genome Res**, v. 19, n. 4, p. 657-66, Apr 2009.

HSU, C. M. et al. Altered Expression of Imprinted Genes in Squamous Cell Carcinoma of the Head and Neck. **Anticancer Res**, v. 36, n. 5, p. 2251-8, May 2016.

HUBERTUS, J. et al. Altered expression of imprinted genes in Wilms tumors. **Oncol Rep**, v. 25, n. 3, p. 817-23, Mar 2011.

HUNTER, T. Cooperation between oncogenes. **Cell**, 1991.

HUNTER, T.; PINES, J. Cyclins and cancer. II: Cyclin D and CDK inhibitors come of age. **Cell**, v. 79, n. 4, p. 573-82, Nov 18 1994.

ISHIDA, M.; MOORE, G. E. The role of imprinted genes in humans. **Mol Aspects Med**, v. 34, n. 4, p. 826-40, Jul-Aug 2013.

JAY, P. et al. The human necdin gene, NDN, is maternally imprinted and located in the Prader-Willi Syndrome chromosomal region. **Nature**, 1997.

JEFFRIES, A. R. et al. Stochastic choice of allelic expression in human neural stem cells. **Stem Cells**, v. 30, n. 9, p. 1938-47, Sep 2012.

JUNTTILA, M. R.; DE SAUVAGE, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. **Nature**, v. 501, n. 7467, p. 346-54, Sep 19 2013.

KANG, M. J. et al. Loss of imprinting and elevated expression of wild-type p73 in human gastric adenocarcinoma. **Clin Cancer Res**, v. 6, n. 5, p. 1767-71, May 2000.

KELSEY, G.; BARTOLOMEI, M. S. Imprinted genes ... and the number is? **PLoS Genet**, v. 8, n. 3, p. e1002601, 2012.

KOBAYASHI, H. Imprinting genes associated with endometriosis. **Excli J**, v. 13, p. 252-64, 2014.

KOONIN, E. V. Darwinian evolution in the light of genomics. **Nucleic Acids Res**, v. 37, n. 4, p. 1011-34, Mar 2009.

LALANDE, M. Parental imprinting and human disease. **Annu Rev Genet**, v. 30, p. 173-95, 1996.

LEE, J. T.; BARTOLOMEI, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. **Cell**, v. 152, n. 6, p. 1308-23, Mar 14 2013.

LISTER, R. et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. **Cell**, v. 133, n. 3, p. 523-36, May 2 2008.

LUEDI, P. P. et al. Computational and experimental identification of novel human imprinted genes. **Genome Res**, v. 17, n. 12, p. 1723-30, Dec 2007.

LYON, M. F. Sex chromatin and gene action in the mammalian X-chromosome. **Am J Hum Genet**, v. 14, p. 135-48, Jun 1962.

MARIONI, J. C. et al. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. **Genome Res**, v. 18, n. 9, p. 1509-17, Sep 2008.

MARSHALL, C. J. Tumor suppressor genes. **Cell**, v. 64, n. 2, p. 313-26, Jan 25 1991.

MAUPETIT-MEHOUAS, S. et al. Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. **Nucleic Acids Res**, v. 44, n. 2, p. 621-35, Jan 29 2016. ISSN 1362-4962 (Electronic) 0305-1048 (Linking).

MCGRATH, J.; SOLTER, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. **Cell**, v. 37, n. 1, p. 179-83, May 1984.

MICHOR, F.; POLYAK, K. The origins and implications of intratumor heterogeneity. **Cancer Prev Res (Phila)**, v. 3, n. 11, p. 1361-4, Nov 2010.

MORIN, R. et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. **Biotechniques**, v. 45, n. 1, p. 81-94, Jul 2008.

MORISON, I. M.; PATON, C. J.; CLEVERLEY, S. D. The imprinted gene and parent-of-origin effect database. **Nucleic Acids Res**, v. 29, n. 1, p. 275-6, Jan 1 2001.

MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Methods**, v. 5, n. 7, p. 621-8, Jul 2008.

NAGALAKSHMI, U. et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. **Science**, v. 320, n. 5881, p. 1344-9, Jun 6 2008.

OKAE, H. et al. Genome-wide analysis of DNA methylation dynamics during early human development. **PLoS Genet**, v. 10, n. 12, p. e1004868, Dec 2014.

ONYANGO, P.; FEINBERG, A. P. A nucleolar protein, H19 opposite tumor suppressor (HOTS), is a tumor growth inhibitor encoded by a human imprinted H19 antisense transcript. **Proc Natl Acad Sci U S A**, v. 108, n. 40, p. 16759-64, Oct 4 2011.

OZSOLAK, F.; MILOS, P. M. RNA sequencing: advances, challenges and opportunities. **Nat Rev Genet**, v. 12, n. 2, p. 87-98, Feb 2011.

PATEL, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. **Science**, v. 344, n. 6190, p. 1396-401, Jun 20 2014.

PATTEN, M. M. et al. The evolution of genomic imprinting: theories, predictions and empirical tests. **Heredity (Edinb)**, v. 113, n. 2, p. 119-28, Aug 2014.

PETERS, J. The role of genomic imprinting in biology and disease. **Nat. Rev. Genet**, v. 15, p. 517-530, 2014.

PLASSCHAERT, R. N.; BARTOLOMEI, M. S. Genomic imprinting in development, growth, behavior and stem cells. **Development**, v. 141, n. 9, p. 1805-13, May 2014.

POPLAWSKI, A. et al. Systematically evaluating interfaces for RNA-seq analysis from a life scientist perspective. **Brief Bioinform**, v. 17, n. 2, p. 213-23, Mar 2016.

REIK, W.; WALTER, J. Genomic imprinting: parental influence on the genome. **Nat Rev Genet**, v. 2, n. 1, p. 21-32, Jan 2001.

RODRIGUEZ, P. et al. NAP-2 is part of multi-protein complexes in HeLa cells. **J Cell Biochem**, v. 93, n. 2, p. 398-408, Oct 1 2004.

RUSSELL, L. B. Genetics of mammalian sex chromosomes. **Science**, v. 133, n. 3467, p. 1795-803, Jun 9 1961.

RUSSO, M. et al. Understanding genistein in cancer: The "good" and the "bad" effects: A review. **Food Chem**, v. 196, p. 589-600, Apr 1 2016.

SCHWIENBACHER, C. et al. Gain of imprinting at chromosome 11p15: A pathogenetic mechanism identified in human hepatocarcinomas. **Proc Natl Acad Sci U S A**, v. 97, n. 10, p. 5445-9, May 9 2000.

SINNETT, D.; KRAJINOVIC, M.; LABUDA, D. Genetic susceptibility to childhood acute lymphoblastic leukemia. **Leuk Lymphoma**, v. 38, n. 5-6, p. 447-62, Aug 2000.

SLATTER, R. E. et al. Mosaic uniparental disomy in Beckwith-Wiedemann syndrome. **J Med Genet**, v. 31, n. 10, p. 749-53, Oct 1994.

SMILINICH, N. J. et al. A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. **Proc Natl Acad Sci U S A**, v. 96, n. 14, p. 8064-9, Jul 6 1999.

SOHAIL, A. et al. Effects of glutathione-S-transferase polymorphisms on the risk of breast cancer: a population-based case-control study in Pakistan. **Environ Toxicol Pharmacol**, v. 35, n. 2, p. 143-53, Mar 2013.

SPENCER, H. G.; WOLF, J. B. Genomic imprinting: theories and data. **Heredity (Edinb)**, v. 113, n. 2, p. 93-5, Aug 2014.

STUCKER, I. et al. Genetic polymorphisms of glutathione S-transferases as modulators of lung cancer susceptibility. **Carcinogenesis**, v. 23, n. 9, p. 1475-81, Sep 2002.

SURANI, M. A.; BARTON, S. C.; NORRIS, M. L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. **Nature**, v. 308, n. 5959, p. 548-50, Apr 5-11 1984.

SUTHERLAND, G. T.; JANITZ, M.; KRIL, J. J. Understanding the pathogenesis of Alzheimer's disease: will RNA-Seq realize the promise of transcriptomics? **J Neurochem**, v. 116, n. 6, p. 937-46, Mar 2011.

TIAN, F. et al. Loss of imprinting of IGF2 correlates with hypomethylation of the H19 differentially methylated region in the tumor tissue of colorectal cancer patients. **Mol Med Rep**, v. 5, n. 6, p. 1536-40, Jun

\_\_\_\_\_. Loss of imprinting of IGF2 correlates with hypomethylation of the H19 differentially methylated region in the tumor tissue of colorectal cancer patients. **Mol Med Rep**, v. 5, n. 6, p. 1536-40, Jun 2012.

TYCKO, B. Epigenetic gene silencing in cancer. **J Clin Invest**, v. 105, n. 4, p. 401-7, Feb 2000.

UBEDA, F.; GARDNER, A. A model for genomic imprinting in the social brain: adults. **Evolution**, v. 65, n. 2, p. 462-75, Feb 2011.

URIBE-LEWIS, S. et al. Molecular mechanisms of genomic imprinting and clinical implications for cancer. **Expert Rev Mol Med**, v. 13, p. e2, 2011.

VERA, J. C. et al. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. **Mol Ecol**, v. 17, n. 7, p. 1636-47, Apr 2008.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet**, v. 10, n. 1, p. 57-63, Jan 2009.

WEKSBERG, R.; SHUMAN, C.; SMITH, A. C. Beckwith-Wiedemann syndrome. **Am J Med Genet C Semin Med Genet**, v. 137C, n. 1, p. 12-23, Aug 15 2005.

WILHELM, B. T. et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. **Nature**, v. 453, n. 7199, p. 1239-43, Jun 26 2008.

WU, J. Q. et al. Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. **Genome Biol**, v. 9, n. 1, p. R3, Jan 3 2008.

YAN, X. S. et al. Adiposity, inflammation, genetic variants and risk of post-menopausal breast cancer findings from a prospective-specimen-collection, retrospective-blinded-evaluation (PRoBE) design approach. **Springerplus**, v. 2, p. 638, 2013.

YANAI, I.; CHMIELNICKI, E. Computational biologists: moving to the driver's seat. **Genome Biol**, v. 18, n. 1, p. 223, Nov 23 2017.

ZHANG, M. et al. Loss of imprinting of insulin-like growth factor 2 is associated with increased risk of primary lung cancer in the central China region. **Asian Pac J Cancer Prev**, v. 15, n. 18, p. 7799-803, 2014.

ZWEMER, L. M. et al. Autosomal monoallelic expression in the mouse. **Genome Biol**, v. 13, n. 2, p. R10, 2012.



## 10. MATERIAL SUPLEMENTAR

**Tabela suplementar 1. Lista dos Oncogenes e Genes Supressores de Tumor Imprintados.** Locus dos Oncogenes e Genes Supressores de Tumor Imprintados, sua descrição e posição no genoma.

<i>Locus</i>	Descrição	Cromossomo	Posição de início (hg19)	Posição final (hg19)
<i>DIRAS3</i>	DIRAS family GTPase 3	1	68045962	68051762
<i>TP73</i>	tumor protein p73	1	3652565	3736201
<i>NSD1</i>	nuclear receptor binding SET domain protein 1	5	177133079	177300213
<i>IGF2R</i>	insulin like growth factor 2 receptor	6	159969099	160111504
<i>PLAGL1</i>	PLAG1 like zinc finger 1	6	143940300	144064599
<i>PEG10</i>	paternally expressed 10	7	94656325	94669695
<i>TFPI2</i>	tissue factor pathway inhibitor 2	7	93885397	93890991
<i>CDKN1C</i>	cyclin dependent kinase inhibitor 1C	11	2883218	2885804
<i>H19</i>	H19, imprinted maternally expressed transcript (non-protein coding)	11	1995176	2001466
<i>IGF2</i>	insulin like growth factor 2	11	2129112	2149603
<i>NAP1L4</i>	nucleosome assembly protein 1 like 4	11	2944430	2992377

<i>PHLDA2</i>	pleckstrin homology like domain family A member 2	11	2928273	2929420
<i>SDHD</i>	succinate dehydrogenase complex subunit D	11	112086824	112095801
<i>SLC22A18</i>	solute carrier family 22 member 18	11	2899721	2925246
<i>STIM1</i>	stromal interaction molecule 1	11	3855703	4093210
<i>TSPAN32</i>	tetraspanin 32	11	2299480	2318201
<i>TSSC4</i>	tumor suppressing subtransferable candidate 4	11	2400409	2403878
<i>RB1</i>	RB transcriptional corepressor 1	13	48303747	48481890
<i>MEG3</i>	maternally expressed 3 (non-protein coding)	14	100826108	100861026
<i>IGF1R</i>	insulin like growth factor 1 receptor	15	98648539	98964530
<i>NDN</i>	necdin, MAGE family member	15	23685407	23687303
<i>UBE3A</i>	ubiquitin protein ligase E3A	15	25337234	25439086
<i>MIR372</i>	microRNA 372	19	53787890	53787956
<i>TRIM28</i>	tripartite motif containing 28	19	58544469	58550715
<i>BLCAP</i>	bladder cancer associated protein	20	37517417	37527931
<i>HOTS</i>	H19 opposite tumor suppressor	11	1992408	1996294
<i>LPAR6</i>	lysophosphatidic acid receptor 6	13	48985182	48987653
<i>AK123878</i>	Homo sapiens cDNA FLJ41884 fis, clone OCBBF2023162	14	101320998	101323960

**Tabela suplementar 2. Lista dos Oncogenes Alvos.** Nome, citobanda, e função dos 800 Oncogenes alvos.

**Tabela suplementar 3. Lista dos Genes Supressores de Tumor Alvos.** Nome, citobanda e função dos 1090 Genes Supressores de Tumor.

**Tabela S4. Análise do perfil de expressão em experimentos de RNA-Seq em oncogenes e genes supressores imprintados.** Para cada gene foi observado se não houve expressão bialélica (NB). Se não houve expressão monoalélica (NM). Foi observado se o status flip foi positivo (+) ou negativo (-) e se foram observados dados quanto a expressão monoalélica (Mono), bialélica (Bi), ou ausência de dados (NA).

<https://sites.google.com/view/dissertao-alan-tardin-da-silva/p%C3%A1gina-inicial>