



UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS GRADUAÇÃO
PROGRAMA MULTI-INSTITUCIONAL DE PÓS-GRADUAÇÃO
EM BIOTECNOLOGIA

TESE DE DOUTORADO

DIEGO MONTEIRO DE CARVALHO

ANÁLISE DA EXPRESSÃO DIFERENCIAL DE RNAs MENSAGEIROS
EM TECIDOS TIREOIDIANOS COM DIAGNÓSTICO DE BÓCIO

MANAUS

2017

DIEGO MONTEIRO DE CARVALHO

**ANÁLISE DA EXPRESSÃO DIFERENCIAL DE RNAs MENSAGEIROS
EM TECIDOS TIREOIDIANOS COM DIAGNÓSTICO DE BÓCIO**

Tese de doutorado apresentada ao Programa Multi-Institucional de Pós-Graduação em Biotecnologia área de concentração em Saúde da Universidade Federal do Amazonas, como requisito parcial para a obtenção do título de Doutor em Biotecnologia.

Orientador: Prof. Dr. João Bosco Lopes Botelho

Co-orientador: Prof. Dr. Spartaco Astolfi Filho

MANAUS

2017

Ficha Catalográfica elaborada pela bibliotecária Marilene de Sena e Silva-CRB-11-AM 561

C257_a CARVALHO, Diego Monteiro de.

Análise da expressão diferencial de RNAs mensageiros em tecidos tireoidianos com diagnóstico de bócio. Diego Monteiro de Carvalho. Manaus: UFAM, 2017.

116 p. il.

Tese (Doutorado). Universidade Federal do Amazonas (UFAM). Programa Multi-Institucional de Pós-Graduação em Biotecnologia.

Orientador: João Bosco Lopes Botelho

Co-Orientador: Spartaco Astolfi Filho

Título em Inglês: Analysis of the the differential expression of messenger RNAs in thyroids tissues with goiter.

1. Biotecnologia. 2. Bócio. 3. HOTS. I. Botelho, João Bosco Lopes (Orient.) II. Astolfi Filho, Spartaco III. Universidade Federal do Amazonas IV. Título.

CDD: 611.4

**ANÁLISE DA EXPRESSÃO DIFERENCIAL DE RNAs MENSAGEIROS EM
TECIDOS TIREOIDIANOS COM DIAGNÓSTICO DE BÓCIO**

Diego Monteiro de Carvalho

Banca de Defesa de Tese de Doutorado

Prof. Dr João Bosco Lopes Botelho
Universidade Federal do Amazonas
Orientador e Presidente da Banca

Prof^a. Dra. Antonia Queiroz Lima de Souza
Universidade Federal do Amazonas

Prof. Dr. Diego Ferreira Marques
Universidade da Amazônia – UNAMA

Prof^a. Dra. Elen Bethleen de Souza
Universidade Federal do Amazonas

Prof. Dr. Gecildo Soriano dos Anjos
Universidade Federal do Amazonas

Manaus, 24 de novembro de 2017.

DEDICATÓRIA

Esta tese é o produto de um longo esforço coletivo nos últimos trinta anos de uma família do interior da Amazônia que acreditou na força transformadora da Educação. As barreiras impostas pelo determinismo social e geográfico não foram maiores que a fé no Criador e nem tampouco maiores que a caridade em ensinar percebida nos maiores atores desta jornada: meus professores e orientadores.

Dedico esta tese ao Pai celestial, à minha esposa Lílian, aos meus pais e irmãos e a todos os meus educadores que jamais acreditaram no impossível.

AGRADECIMENTOS

A Deus e à minha família, presentes desde o início desta empreitada na pesquisa acadêmica, iniciada desde os tempos de escola: **Waldemir, Joelma, Bruno e Agnes de Carvalho;**

Agradeço à minha companheira e agora esposa, **Lilian Linhares de Carvalho**, pelas duas décadas de convívio ensinando-me diariamente que sem a paixão, o amor e a lealdade aos nossos princípios, pouca coisa avança além da esquina;

Ao Professor **Dr. João Bosco Botelho**, que ao longo destes doze anos de convívio e trabalho, orientou-me no dia-a-dia do pesquisador e colocou-me no caminho deste doutoramento, sendo antes de mais nada, um exemplo a ser seguido;

Ao Professor **Dr. Spartaco Astolfi Filho**, co-orientador deste trabalho, mão e mente amiga no laboratório e nas decisões do direcionamento deste estudo;

Aos Professores doutores **Carlos Gustavo Nunes da Silva e Edmar Vaz de Andrade**, que contribuíram com ideias e especialmente com apoio fraterno na condição de coordenadores do PPG-BIOTEC;

Aos amigos, também colegas do PPG-BIOTEC: **Dra. Lorena Nacif Marçal, Eraldo Lopes;**e em especial a **Deborah Laredo Jezini** com quem construí uma tese e uma amizade ao longos destes quatro anos;

Aos doutores, entusiastas e incentivadores: **Profa. Dra. Antonia Queiroz, Profa. Dra. Enedina Nogueira Assunção, Profa. Dra. Janice Monteiro, Profa. Dra. Helena Camarão Telles e Prof. Dr. Diego Marques;**

Aos cirurgiões de cabeça e pescoço e otorrinolaringologistas que estiveram conosco na Fundação Hospital Adriano Jorge: **Dr Lesemky Cattebeke, Dr Gecildo Soriano dos Anjos, Dr Marco Antonio Rocha, Dr Paulo Mendonça, Dra Súnia Ribeiro e Dra Viviane Saldanha;**

Aos alunos de graduação do Grupo de Pesquisa em Otorrinolaringologia – GAPO/UEA, de maneira particular aos que estiveram conosco do centro cirúrgico à bancada **Paulo Rodrigo Ribeiro Rodrigues e Andrey Alisson Mendes Maia.**

Epígrafe

“Independente de quantos cisnes brancos podemos observar, isso não justifica a conclusão de que todos os cisnes são brancos”.

Popper K, A lógica da pesquisa científica, 1972.

SUMÁRIO

1	INTRODUÇÃO.....	16
2	REVISÃO BIBLIOGRÁFICA	18
2.1	<i>A TIREOIDE E OS ESTUDOS MOLECULARES DO BÓCIO</i>	<i>23</i>
2.2	<i>O ESTUDO MOLECULAR E O RNA-seq.....</i>	<i>25</i>
2.3	<i>O ESTUDO MOLECULAR DA TIREOIDE E O SEQUENCIAMENTO DE NOVA GERAÇÃO (NGS).....</i>	<i>34</i>
3	OBJETIVOS.....	36
4	METODOLOGIA	37
4.1	<i>ASPECTOS ÉTICOS</i>	<i>37</i>
4.2	<i>ORGANIZAÇÃO E SELEÇÃO DOS TECIDOS TIREOIDEANOS.....</i>	<i>37</i>
4.3	<i>AMOSTRAS DE TECIDO TIREOIDIANO</i>	<i>38</i>
4.4	<i>EXTRAÇÃO DE RNA TOTAL</i>	<i>38</i>
4.5	<i>CONTROLE DE QUALIDADE DAS AMOSTRAS</i>	<i>39</i>
4.6	<i>SEQUENCIAMENTO</i>	<i>39</i>
4.6.1	<i>Análise das sequências</i>	<i>39</i>
4.6.2	<i>Análise biológica.....</i>	<i>41</i>
5	RESULTADOS	42
6	DISCUSSÃO.....	67
7	CONCLUSÃO.....	71
8	REFERÊNCIAS BIBLIOGRÁFICAS	72
9	ANEXOS.....	78

LISTA DE FIGURAS

Figura 1	Localização da tireoide no pescoço (ROBERTS, C. S., 1990; SMITH, T. J., 1990).	18
Figura 2	Sumário dos principais sítios de ação do hormônio tireoideano nos subtipos alfa e beta de receptores: sistema nervoso central, sistema muscular esquelético, ossos, intestino, fígado e coração (MORAN; CHATTERJEE, 2015).	19
Figura 3	A e B – Respectivamente pré-operatório e transoperatório de tireoidectomia total por bócio (Fonte: Arquivo pessoal do Prof Dr João Bosco Botelho, 2010).	20
Figura 4	Visão geral do passo a passo de uma análise transcriptômica com RNA-seq. Adaptada de WANG, 2011.	28
Figura 5	Fluxo de trabalho geral de experiências de sequenciamento de RNA de uma única célula (scRNA-seq). Um fluxo de trabalho típico de scRNA-seq inclui a maioria das seguintes etapas: 1) isolamento de células únicas, 2) lise celular, preservando o mRNA, 3) captura de mRNA, 4) transcrição reversa de RNA em DNA complementar (cDNA), 5) amplificação de cDNA, 6) preparação da biblioteca de cDNA para o sequenciamento, 7) agrupamento de bibliotecas de sequências, 8) uso de ferramentas bioinformática para avaliar qualidade e variabilidade, e 9) uso de ferramentas especializadas para analisar e apresentar os dados. Adaptado de HAQUE <i>et al</i> , 2017.	30
Figura 6	A transcrição do <i>locus</i> de <i>H19</i> resulta no <i>H19</i> lncRNA (A) e seu microRNA <i>miR-675</i> (B) codificado a partir do <i>exon</i> 1 de <i>H19</i> . Dois transcritos anti-sentido também são formados a partir do <i>locus</i> . O RNA <i>91H</i> (C) que é descrito em variados comprimentos, mas pode potencialmente ser codificado a partir da outra cadeia de DNA inteiramente abrangendo <i>H19</i> e outras porções de suas sequencias reguladoras e o RNA	32

HOTS (D) que é um transcrito a partir da maioria da sequência anti-sentido de *H19* e outras bases. *HOTS* pode ser traduzido em uma proteína nucleolar. Adaptado de POPE *et al.*, 2017.

- Figura 7 Várias condições de estresse celular poderiam causar uma instabilidade genômica com redução da expressão de p53, aumento de H19 em sub-clones estressados de células seguido de disrupção da reação celular ao estresse, culminando com proliferação desordenada. Adaptado de Raveh *et. al.* 2015. 33
- Figura 8 Fluxo de atividades em bancada para a extração de RNA. 38
- Figura 9 Eletroforese em gel de Agarose das amostras indicadas no Quadro 2 (1.0%; 0.5xTBE; 180V em 16min). M: Trans 2K Plus DNA ladder. 43
- Figura 10 Imagens histológicas dos tecidos com bócio das amostras B13 e B14 respectivamente. As imagens mostram a variabilidade de tamanho de pequenos(setas) e macro-folículos(estrela) colóides separados por densas bandas de fibrose (pentágono). 43
- Figura 11 Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.1. 45
- Figura 12 Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.1. 45
- Figura 13 Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.2. 46
- Figura 14 Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.2. 46
- Figura 15 Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.1. 47
- Figura 16 Análise do FastQC pós etapa de trimagem (remoção de 47

	adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.1.	
Figura 17	Análise FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.2.	48
Figura 18	Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.2.	48
Figura 19	Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.1.	49
Figura 20	Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.1.	49
Figura 21	Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.2.	50
Figura 22	Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.2.	50
Figura 23	Dispersão da expressão gênica entre as amostras B13 x N6.	51
Figura 24	Dispersão da expressão gênica entre as amostras B14 x N6.	51
Figura 25	Densidade da expressão gênica entre as amostras B13 x N6.	52
Figura 26	Densidade da expressão gênica entre as amostras B14 x N6.	52
Figura 27	<i>Heatmap</i> comparando os “ <i>fold values</i> ” dos 70 genes diferencialmente expressos em cada uma das amostras. Em verde representa-se os valores acima do valor de referência normalizada (Z score), em vermelho abaixo do valor referência e em tom escuro os valores próximo da referência. Cada coluna representa uma amostra e cada linha representa um gene.	54
Figura 28	Conjunto de gráficos das medidas da expressão gênica entre os genes (A) <i>up-regulated</i> e (B) <i>down-regulated</i> diferencialmente expressos no tecido com bócio vs o tecido	63

sem doença considerando $p < 0,05$ (5%).

- Figura 29 Função molecular dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos genes *down-regulated*: *Hspa6*, *C1ql4*, *Plcd4*, *Errfi1*, *Pcp4l1*, *Atrnl1*, *Mt1h*, *Abca13*, *Dnajb1*, *Ca12*, *Etv4*, *Igsf1*, *Rasd1*, *Mrpl23*, *Cpne4*, *I11r11*, *Dlec1*, *Lect1*, *Shc3*, *Sorcs1*, *Stard9*, *Tmem184a*, *Pcsk2*, *Fosb*, *Trabd2a*, *Gatm*, *Bag3*, *Fam105a*, *Hspa1b*, *Mme*, *Pde4c*, *Rtn4rl2*, *Edn3*, *Nrk*, *Egr2*, *Tnfrsf19*, *Abcc3*, *Hsd17b6*, *C15orf48*, *Gata5*, *Fras1*, *Pcdh8*, *Yjefn3*, *Sphkap*, *Tbx2*, *Kiaa1324*, *Kif5a*, *Arhgap28*, *C1orf233*, *Nrg1*, *Hspa1a*, *Thbs1*, *Plin1*, *Xkrx*. 64
- Figura 30 Classe proteica dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos genes *down-regulated*: *Hspa6*, *C1ql4*, *Plcd4*, *Errfi1*, *Pcp4l1*, *Atrnl1*, *Mt1h*, *Abca13*, *Dnajb1*, *Ca12*, *Etv4*, *Igsf1*, *Rasd1*, *Mrpl23*, *Cpne4*, *I11r11*, *Dlec1*, *Lect1*, *Shc3*, *Sorcs1*, *Stard9*, *Tmem184a*, *Pcsk2*, *Fosb*, *Trabd2a*, *Gatm*, *Bag3*, *Fam105a*, *Hspa1b*, *Mme*, *Pde4c*, *Rtn4rl2*, *Edn3*, *Nrk*, *Egr2*, *Tnfrsf19*, *Abcc3*, *Hsd17b6*, *C15orf48*, *Gata5*, *Fras1*, *Pcdh8*, *Yjefn3*, *Sphkap*, *Tbx2*, *Kiaa1324*, *Kif5a*, *Arhgap28*, *C1orf233*, *Nrg1*, *Hspa1a*, *Thbs1*, *Plin1*, *Xkrx*. 64
- Figura 31 Função molecular dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos genes *up-regulated*: *C4b*, *C2cd4c*, *Ac241377.2*, *Cpxm1*, *Naprt*, *St6gal1*, *Col14a1*. 65
- Figura 32 Classe proteica dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos 65

genes *up-regulated*: *C4b*, *C2cd4c*, *Ac241377.2*, *Cpxm1*, *Naprt*, *St6gal1*, *Col14a1*.

Figura 33 Análise de interação proteína-proteína de 63 genes funcionais diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença), utilizando o software GeneMANIA. Ao todo 423 interações gênicas (mais de 50% de co-expressão) 5 – genes: *Tnf-Rsf19*, *Ac241377.1*, *Rp11-514p8.6*, *C1ql4* E *Sphkap* não apresentavam interações identificadas com os demais. 66

LISTA DE TABELAS

Tabela 1	Listagem de <i>softwares</i> utilizados.	40
Tabela 2	Tabela de saída utilizando o protocolo Tuxedo das amostras B13 versus N6.	53

LISTA DE QUADROS

Quadro 1	Classificação Bethesda segundo risco de câncer e decisão terapêutica.	22
Quadro 2	Análise qualitativa e quantitativa de RNA total.	42

RESUMO

Introdução: O bócio continua sendo a principal indicação de tireoidectomias em regiões endêmicas. Trata-se de uma doença heterogênea e pouco compreendida molecularmente. **Objetivo:** Caracterizar a expressão diferencial gênica em tecidos tireoidianos oriundos de peças cirúrgicas diagnosticadas com bócio, comparando-os com tecido tireoidiano com ausência de doença de pacientes submetidos a tireoidectomias na cidade de Manaus-AM utilizando a tecnologia de RNA-seq. **Metodologia:** Sequenciamento do transcriptoma de fragmentos de tecido tireoideano com bócio e sem doença utilizando a plataforma de sequenciamento de nova geração (NGS) em Illumina HiSeq 2000 com protocolo de RNA-seq e análise de expressão diferencial. **Resultados:** Os resultados mostraram que há diferença entre os perfis de expressão gênica dos tecidos com bócio em comparação com um controle com ausência de doença, em que 70 sequências gênicas estiveram diferencialmente expressas. Dentre as sequências de genes *up-regulated* no bócio, a proteína nucleolar HOTS, um dos produtos da leitura do *locus H19*, esteve presente com cópias unicamente no tecido com bócio ($p < 0,05$). **Conclusão:** Estes resultados demonstraram que o perfil de expressão gênica do bócio é de uma doença pró-tumoral em que os produtos do *locus H19* merecem ser investigados quanto ao seu papel na tumorigenese.

Palavras-chave: Bócio, MRNA, H19, HOTS.

ABSTRACT

Introduction: Goiter remains a major indication for thyroidectomies in endemic regions for that disease. It is a neglected and heterogeneous molecular disease. **Objectives:** This study analyzes the differential gene expression in surgical specimens diagnosed with goiter, comparing them with thyroid tissue absent of disease from patients submitted to thyroidectomies in Manaus-AM, using a RNA-seq technology. **Methodology:** Transcriptome sequencing by New Generation Sequencing (NGS) of surgical specimens fragments with goiter and without disease, with RNA-seq protocol and differential gene expression analysis. **Results:** The results showed a difference between the gene expression profiles of the diseased tissues in comparison with a control absent of disease, in which 70 genes sequences are differentially expressed. The sequence of gene *Hots* (*locus H19*) was present with copies only in goiter tissues ($p < 0.05$). **Conclusion:** These results demonstrated the gene expression profile of goiter seems a pro-tumoral disease and that *H19* plays a central role in goiter tumorigenesis.

Key words: Goiter, MRNA, H19, HOTS

1 INTRODUÇÃO

A atual era da medicina e da biologia celular e molecular caracteriza-se pela busca da melhor definição entre o doente e o normal, em níveis cada vez mais invisíveis da matéria orgânica.

Não é sem propósito que nas últimas duas décadas após a montagem do genoma humano testemunhamos o desenvolvimento de inúmeras técnicas moleculares com o objetivo de compreender os padrões que regulam o comportamento da célula do genótipo ao fenótipo e que culminaram com o surgimento das ciências ômicas. Hoje, além das sequências de genes e suas localizações cromossômicas, importa-nos conhecer como estas sequências se relacionam entre si e com o meio celular, como e porque são reguladas diferentemente a depender do tecido e da influência do meio externo.

Conhecer a doença pressupõe explicar a não-doença e além disso a pré-doença, certamente pouco explicáveis pela clínica e pelos exames de imagem atuais, minimamente explicáveis pela histologia e citologia, mas cada vez mais compreendidas pelos perfis de expressão genica que distinguem cada tecido e cada célula em diferentes condições.

O maior impacto desta nova era médica deve ocorrer quanto às doenças cirúrgicas não traumáticas: melhor compreensão da evolução de doenças inflamatórias agudas, tumorais benignas e malignas; identificação de fatores de risco genéticos e epigenéticos; estratificações de uma mesma doença a partir de um referencial molecular mudando as atuais diretrizes terapêuticas de quando e quem operar e, em especial, quando não operar.

Dentre as cirurgias mais praticadas na rotina médica está a remoção parcial ou total da glândula tireoide, sede de diversas doenças inflamatórias, hormonais, nodulares benignas, carcinomas e o bócio. A tireoidectomia como tratamento para doenças tireoidianas são praticadas a mais de oito séculos por cirurgiões no mundo todo, o que contribuiu para que a tireoide permanecesse como a glândula endócrina mais acessada cirurgicamente.

O bócio é uma doença uni ou multi-nodular da tireoide que ocorre normalmente sem disfunção hormonal, tem causa multifatorial além da deficiência de iodo na dieta, mas não é claro seu papel como fator de risco para os carcinomas

bem diferenciados da tireoide, embora estime-se que os carcinomas sejam mais prevalentes em áreas endêmicas de bocio. Mesmo sendo a principal indicação de tireoidectomia especialmente em áreas geográficas bociogênicas, como a Amazônia, o bocio é pouco estudado em comparação aos carcinomas tireoidianos, e por isso são escassos os trabalhos sobre o perfil molecular dos bócios e não há registros de estudos moleculares em áreas bociogênicas.

Pretendeu-se com esta tese estudar de maneira inédita em uma área geográfica endêmica para o bocio o perfil de expressão genica desta doença a partir do estudo do transcriptoma por RNA-seq de tecidos de tireoide operados na cidade de Manaus-Amazonas- Brasil.

Embora a incidência de carcinomas da tireoide seja maior em áreas onde o bocio é endêmico, esta associação ainda não é clara na literatura. Os resultados apresentados mostraram genes diferencialmente expressos no bocio que estariam envolvidos no desenvolvimento de um estado pró-tumoral, com destaque à expressão do transcrito do *gene Hots* produto do *locus* gênico imprintado *H19*.

2 REVISÃO BIBLIOGRÁFICA

A tireoide é uma glândula endócrina, de origem endodérmica, localizada no pescoço (Figura 1), diante da 5.^a a 7.^a vértebra cervical, com estroma glandular composto por dois grupos celulares distintos: células foliculares e para-foliculares (BOTELHO, 2012; GARREL; SOPHIE, 2012).

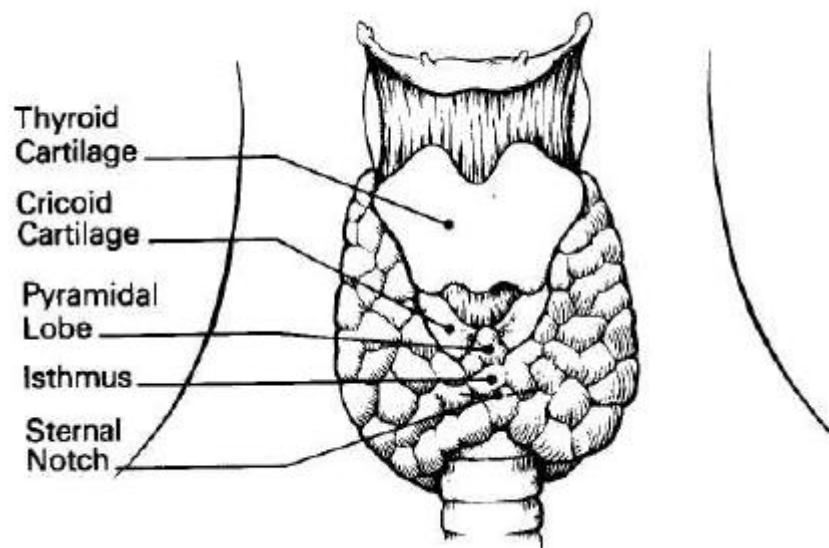


Figura 1 – Localização da tireoide no pescoço (ROBERTS, C. S., 1990; SMITH, T. J., 1990).

Nos folículos tireoidianos ocorre a iodação dos resíduos de tirosina, processo essencial para a síntese dos hormônios da tireoide (T3 ou tri-iodotironina e T4 ou tiroxina). Das células foliculares é que se originam os bóciós, os tumores benignos, os tumores bem diferenciados (papilífero e folicular) e os indiferenciados da tireoide (COOPER; ROSS, 2013; GARREL; SOPHIE, 2012).

A outra população celular: células para-foliculares, ou células C, produzem e secretam a calcitonina, um peptídeo que tem atividade antagônica ao paratormônio, inibindo a reabsorção óssea pelos osteoclastos e reduzindo a concentração de

cálcio no sangue, os tumores medulares da glândula tireoide originam-se das células C (GARREL; SOPHIE, 2012).

A secreção hormonal da tireoide é controlada por fatores tróficos secretados tanto pelo hipotálamo quanto pela adeno-hipófise, e são hormônios que atuam em diversas cascatas celulares, incluindo a estimulação da transcrição gênica, e, por conseguinte, do metabolismo celular e orgânico do sistema nervoso central, do sistema musculoesquelético, também do metabolismo ósseo, hepático e cardíaco inclusive no período fetal (Figura 2) (MORAN; CHATTERJEE, 2015).

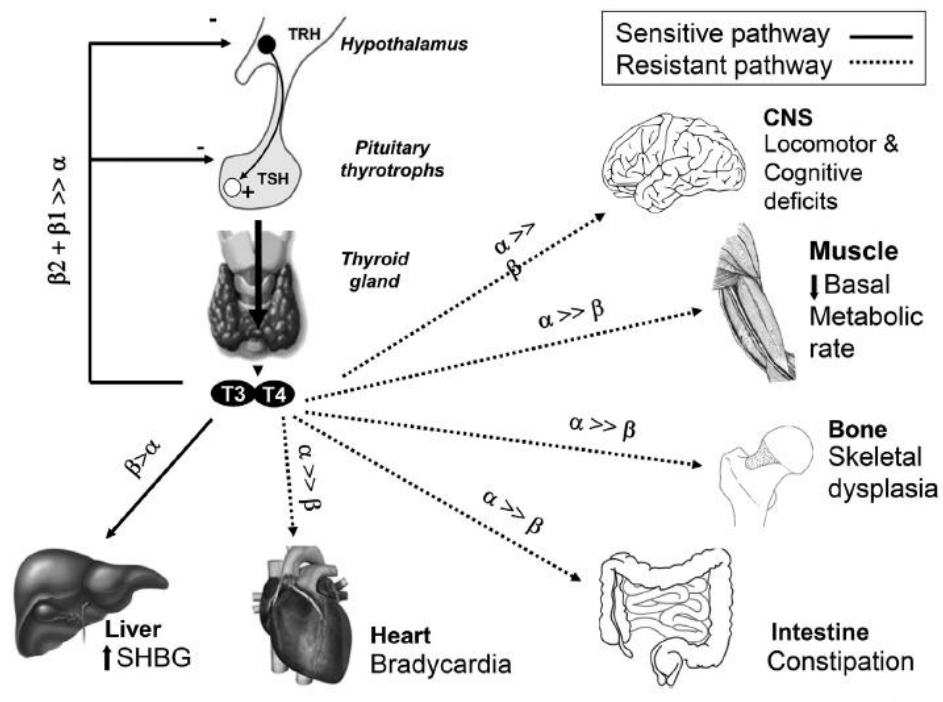


Figura 2 – Sumário dos principais sítios de ação do hormônio tireoideano nos subtipos alfa e beta de receptores: sistema nervoso central, sistema muscular esquelético, ossos, intestino, fígado e coração (MORAN; CHATTERJEE, 2015).

A tireoide possui expressiva quantidade de hormônios, nucleoproteínas, armazenados em seus folículos em forma de coloide, e, além disso, é fortemente reativa a fatores de crescimento e de inibição, que atuam por meio de seus receptores de membrana e transdutores citoplasmáticos (COLIN et al., 2013).

O rompimento do equilíbrio desses sinais leva a um distúrbio de crescimento e proliferação que culmina com o aparecimento das neoplasias e hiperplasias da glândula (BRENNER; HOISTAD; HAIN, 2004; CORDAS et al., 2012).

As mais representativas doenças do parênquima tireoidiano incluem os bócios (uni ou multinodulares), os adenomas foliculares, os carcinomas medulares, os carcinomas bem diferenciados (papilífero e folicular) e as neoplasias anaplásicas, estas últimas mais agressivas e muitas vezes fatais (BOTELHO, 2012; GARREL; SOPHIE, 2012).

Sem dúvida, os bócios são os principais diagnósticos das patologias nodulares tireoideanas em áreas endêmicas (GRAF, 2014; ROSÁRIO et al., 2013).

Os nódulos tireoidianos são observados em 4% a 7% da população adulta, sendo 5-10% destes malignos, e a maioria é não-tumoral, são os bócios (Figura 3) (ROSÁRIO et al., 2013).

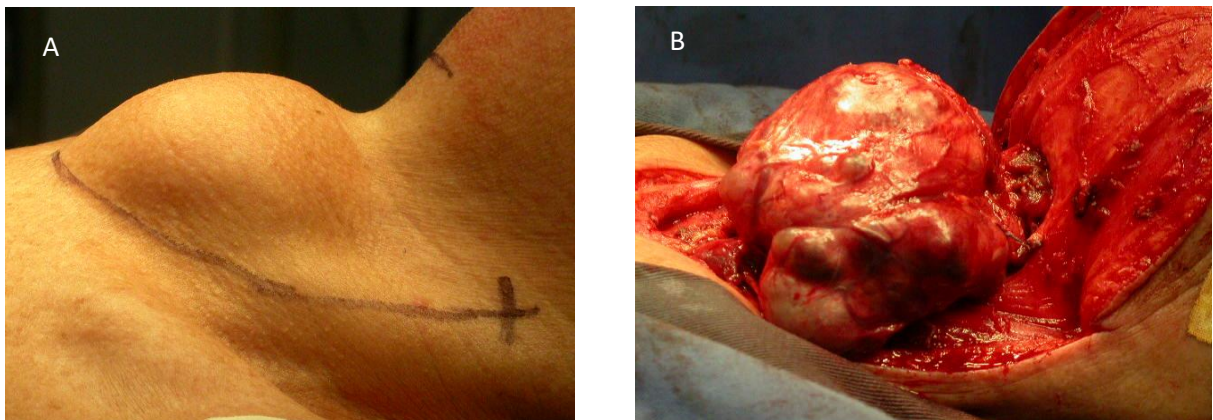


Figura 3 A e B – Respectivamente pré-operatório e transoperatório de tireoidectomia total por bócio (Fonte: Arquivo pessoal do Prof Dr João Bosco Botelho, 2010).

A predominância do bócio é até nove vezes maior nas mulheres em relação aos homens, na faixa etária entre a menarca e a menopausa, e a presença de nódulos solitários em homens tem maior potencial de malignidade (NAGAR et al., 2014; ROSÁRIO et al., 2013).

As primeiras cirurgias para os bócios oficialmente descritas, foram realizadas por Rogério, em 1180, na segunda edição do livro *A Prática Cirúrgica*, em que retrata a abordagem das fístulas do pescoço e da garganta, do bócio e das doenças da úvula no capítulo XIII e descreve a tireoidectomia. Ao trabalho de Rogério

seguiram-se os trabalhos de Guy de Chauliac (1300-1368) em 1363, no livro *La Grand Chirurgie*; dois séculos mais tarde com Juan de Vigo (1460-1517?), em 1501, e João Rodrigues de Castelo Branco, conhecido como Amatus Lusitanus (1511-1568) em 1550, que publicaram experiências exitosas com menores taxas de mortalidade (BOTELHO, 2012).

Nos séculos 16 e 17, Félix Platter (1536-1614), Wolfgang Hoefer (1614-1681) e Malachias Geiger (1606-1671), também fizeram narrativas a respeito do bócio. Ainda no século 16, Vesalius (1514-1564), propôs as primeiras teorias sobre a fisiologia da tireoide, além de acrescentar maiores informações descritivas da anatomia glandular. Em 1656, a glandula foi denominada tireoide (do grego *thyreo* = escudo) por Thomas Wharton (1614-1673) (BOTELHO, 2012).

A decisão em propor cirurgias para a remoção parcial ou completa da tireoide (as tireoidectomias), desde de suas primeiras descrições, até a década de 90, era pautada exclusivamente em critérios clínicos que incluíam o volume dos nódulos e da glândula, a presença de alterações hormonais, a extensão do bócio para compartimentos fora do pescoço, como o mediastino, e a compressividade nas estruturas vizinhas, em especial os vasos cervicais e a traqueia (DEGROOT; STANBURY, 1975; SMITH, 1990).

O critério clínico e macroscópico utilizado na indicação de tireoidectomias, perdurou como único do século XII até o final do século XX. Somente após os trabalhos envolvendo a análise citológica dos produtos obtidos de nódulos tireoideanos por punção aspirativa com agulha fina (PAAF), seguido pelo PAAF guiado por ultrassonografia, que o critério citológico passou a compor critérios nas indicações de cirurgia (KIHARA *et al.*, 2013; YOKOZAWA *et al.*, 1995).

Afim de padronizar a interpretação citológica, criou-se o sistema Bethesda para laudos de citopatologia da tireoide, que resultou de uma conferência realizada no National Institutes of Health em 2007. Este sistema levou à padronização da PAAF com base em seis categorias diagnósticas: I = amostra não diagnóstica, II = benigna, III = atipia / lesão folicular de significância indeterminada, IV = neoplasia folicular/ suspeita de uma neoplasia folicular, V = suspeita de malignidade, e VI = maligno (GARREL; SOPHIE, 2012; KIHARA *et al.*, 2013).

O critério Bethesda tenta prever as chances de uma lesão nodular ser maligna ou não, sendo assim, altamente sensível e pouco específico para o bócio (Quadro 1) (ZHANG; LIN, 2016).

Quadro 1- Classificação Bethesda segundo risco de câncer e decisão terapêutica.

Classe Bethesda	Risco de Câncer	Decisão clínica
I	?	Repetir a PAAF com base no tipo de lesão ao ultrassom
II	0- 3%	Controle com ultrassom com intervalos de 6-18 meses
III	5 - 15%	Nova PAAF com intervalo de 3 a 6 meses
IV	15 - 30%	Controle cirúrgico
V	60 - 75%	Controle cirúrgico
VI	97 – 99%	Controle cirúrgico

Fonte: Adaptado de Garrel & Sophie, 2012.

Atualmente, a todos os pacientes com nódulos superiores à 1cm deve ser oferecida a PAAF como a investigação de escolha para identificar o risco de malignidade (COOPER; ROSS, 2013; ROSÁRIO et al., 2013).

Assim, todos os doentes com Bethesda IV à VI são candidatos a tireoidectomias. Os de classe I à III são reavaliados periodicamente antes da decisão cirúrgica, que leva em consideração também: o volume dos nódulos, compressividade local, histórico de malignidade na família ou exposição do pescoço à radioterapia (GARREL; SOPHIE, 2012).

Essa lacuna deixada pelo sistema de Bethesda nas lesões classificadas entre I e IV (mais frequentes diagnósticos), permite decisões subjetivas na condução

clínica dos pacientes com lesões nodulares expondo-os a cirurgias nem sempre necessárias ou ao estresse emocional das repetidas PAAF's e ultrassonografias (CHOI et al., 2013).

Mesmo que se compreenda a gênese endócrino-metabólica dos bóciós, incluindo os fatores favorecedores sociais, como a privação de iodo, ainda permanece obscura a ausência de doença em indivíduos expostos aos mesmos estímulos, e, mais ainda, àqueles com manifestações clinicamente diferentes da mesma doença, isto é, bóciós de pequenos e grandes volumes (BOTELHO, 2013).

Diferentes grupos em todo o mundo desenvolvem linhas de pesquisa nesta área médico-cirúrgica. É cada vez mais frequente a utilização das ferramentas de biotecnologia, especialmente sobre as lesões malignas bem diferenciadas (folicular variantes papilífero e folicular) (DOMINGUEZ et al., 2011; MAZEH et al., 2013; STOLF et al., 2005).

Permanecem inexpressivos os grupos de trabalho com bóciós, doenças também cirúrgicas, com variável taxa de morbidade e endêmicas em regiões como a Amazônia (CORRÊA FILHO et al., 2002).

A literatura é clara quanto à associação de alguns fatores alimentares – como a privação do iodo na dieta – e a maior prevalência dos bóciós, nas áreas geográficas bociogênicas (COOPER; ROSS, 2013; HEGEDÜS; BRIX; PASCHKE, 2009, 2009).

O desafio atual concentra-se no estabelecimento de métodos de diagnóstico e prognóstico eficientes para esses pacientes com bócio, com mínimo grau de previsibilidade de cura ou de comportamento da doença, apostando nas atividades de rastreio de biomarcadores em potencial (DE MONTGOLFIER; AUDET; CYR, 2011; KANT; RAVI, 2013; UGURLUOGLU et al., 2015).

2.1 A TIREOIDE E OS ESTUDOS MOLECULARES DO BÓCIO

São inúmeros os potenciais biomarcadores propostos para as doenças tireoidianas, que incluem citocinas, quimiocinas, fatores de crescimento, fatores angiogênicos e proteínas de membrana: galactina 3, HBME-1, CK19, NPNT, MLEC, PRX6, ANXA5, SELENBP, IFN- α , IL 10, IL12, IL13, IL 15, IL17, IL4, IL 5, IL 6, MCP-1, MIG, RANTES, DR5, EGF, bFGF, HGF, G-CFS e VEGF (BAN *et al.*, 2012; CORDAS *et al.*, 2012; LINKOV *et al.*, 2008; SOFIADIS *et al.*, 2012).

Em sua maioria, são trabalhos desenvolvidos em áreas onde o bócio não é endêmico, e concentram-se na aplicação destes achados aos carcinomas tireoideanos (BAN *et al.*, 2012; CORDAS *et al.*, 2012; LINKOV *et al.*, 2008; SOFIADIS *et al.*, 2012).

O volume de dados é tão grande quanto a associação pouco clara entre os achados dos diferentes grupos. Síndromes e disfunções genéticas, podem apresentar o bócio como um de seus componentes, como na síndrome de Pendred, que cursa com surdez ou alterações do equilíbrio, e apresentam bases moleculares bem descritas, notadamente, relacionadas a proteínas de transporte transmembrana (MASINDOVA *et al.*, 2012), nesse grupo de proteínas destacam-se as conexinas, claudines e as aquaporinas (NG; KELLEY; FORREST, 2013; TWYFFELS *et al.*, 2011).

As conexinas, são um grupo hexamérico de proteínas transmembrana, que possuem quatro domínios alfa-helices transmembrana, conectados por duas voltas extracelulares e uma intracelular (DE MONTGOLFIER; AUDET; CYR, 2011; HILGERT; SMITH; VAN CAMP, 2009; PROST *et al.*, 2007).

Múltiplas isoformas das conexinas estão presentes no homem e todas desempenham importante papel da recirculação de K⁺ e outros metabólitos. Não só as doenças da tireoide auto-imunes, mas também nos bócios nodulares têm-se identificado mudanças na presença e distribuição de conexinas na imunohistoquímica dos tecidos doentes (DOMINGUEZ *et al.*, 2011).

Quando silenciada na tireoide de ratos, a conexina 32 (Cx32) esteve associada ao desenvolvimento de bócio após estimulação hormonal dos mesmos (PROST *et al.*, 2007).

A imunoreatividade à Cx43 também foi observada nos folículos tireoideanos, e é possível supor que mudanças na expressão da Cx43 afetem o tecido tireoideano incluindo variações da função da glândula (JIANG *et al.*, 2010).

Embora não haja consenso e a literatura seja vasta em relação ao papel das conexinas nas doenças da tireoide, há possibilidade de que um mecanismo fisiopatológico envolvendo as conexinas, seja um dos multi-fatores presentes nos bócio (LICHTENBERGER-GESLIN et al., 2013; WINTER et al., 2009).

As aquaporinas (AQP) são uma família de proteínas que controlam as trocas de água, glicerol e outros solutos entre as células. Além disso, a expressão das aquaporinas parece ser mediada por inúmeros fatores endócrinos, entre eles o fator de transcrição tireoideano (BELYANTSEVA et al., 2000; KIM et al., 2007).

No tecido tireoideano, a AQP3 foi encontrada expressa na membrana citoplasmática de células C normais, positivas para calcitonina à imuno-histoquímica, e completamente ausente nos tecidos com hiperplasia folicular por Graves e por bócio multinodular (NIU et al., 2012). Por outro lado, o desbalanço entre a expressão das isoformas AQP3 e AQP4, mostra AQP4 hiperexpresso em tecidos hiperplásicos e nodulares da tireoide (NIU et al., 2012).

As claudines são proteínas de membrana que contribuem para a formação da barreira paracelular, carga e tamanho seletivos. São proteínas expressas em diferentes tecidos, incluindo o sistema vestibular-auditivo, renal e hepático (HOU, 2012; NAKANO et al., 2009).

A associação com doenças ou disfunções tireoideanas ainda são pouco claras, mas alterações nos perfis de expressão das claudines tem sido reportados em diversas doenças nodulares humanas: cólon, pâncreas, próstata, esôfago, mama e tireoide (CHEN et al., 2013).

2.2 O ESTUDO MOLECULAR E O RNA-seq

O complexo processo celular transcricional, permite que a quase totalidade das células que contêm o mesmo DNA, num mesmo organismo, expressem em diferentes tecidos, e também no mesmo órgão, diferentes produtos transcricionais, fenotípica e funcionalmente distintas (CHU; COREY, 2012; MARTIN; WANG, 2011).

Caracterizar estes perfis de expressão, a partir da análise transcriptômica (conjuntos de RNA de um tecido), pode auxiliar na elucidação de processos e vias metabólicas celulares ainda obscuros (MUTZ et al., 2013).

A expressão gênica, o processamento do RNA mensageiro, as modificações transcricionais e a obtenção de moléculas de cDNA dependem de amostras de RNA em alto grau de pureza e integridade (NELSON; COX, 2011).

Uma biblioteca cDNA é um arranjo de cópias de DNA complementar que representa uma população de mRNAs. Basicamente são bibliotecas que permitem mostrar os genes expressos (induzidos ou reprimidos) sob uma determinada condição (tecido, célula ou estresse) (CHU; COREY, 2012; NELSON; COX, 2011).

Para tanto, é fundamental que tecido e genótipo permitam que os genes de interesse sejam expressos; e que a biblioteca possua qualidade e quantidade de clones suficiente para os mRNAs, incluindo aqueles com reduzido número de cópias (MUTZ et al., 2013).

Inúmeras foram as tecnologias desenvolvidas para o estudo transcriptômico de tecidos orgânicos incluindo técnicas de hibridização com uso de marcadores por fluorescência de cDNA (*microarray*) que embora possam ser utilizadas para análise de perfis de expressão, isoformas e quantificação de *splicing* alternativo, possuem a limitação da dependência de genomas conhecidos, a difícil comparação entre diferentes experimentos e a dificuldade de interpretação quantitativa dos dados (JIANG et al., 2015; TANG et al., 2009).

O advento de sequenciadores de nova geração (NGS) permitiu aumentar a fiabilidade dos perfis de expressão gênica, em especial permitindo o sequenciamento de transcriptomas inteiros a partir do RNA total, sem a necessidade de construções prévias de bibliotecas de cDNA com a tecnologia de RNA-seq, que permite o conhecimento de milhões de sequências, ou *reads*, de uma única célula ou tecido em um determinado momento (CHU; COREY, 2012).

O sequenciamento de RNA (RNA-seq) é uma modalidade de estudo molecular que permite a análise genômica quantitativa de moléculas de RNA de uma determinada amostra biológica com alta taxa de cobertura e acurácia de uma base nucleotídica (Figura 4, p. 28) (HAQUE et al., 2017; MARTIN; WANG, 2011; MUTZ et al., 2013).

A depender da disponibilidade e da qualidade do genoma de referência, classicamente tem-se três estratégias de análise (JIANG et al., 2015; MARTIN; WANG, 2011):

a) transcriptoma baseado em um genoma de referência, em que as sequências são alinhadas a um genoma usando *softwares* alinhadores, seguido de sobreposição das sequências similares e construção de possíveis isoformas;

b) estratégia “*de novo*” em que não utiliza um genoma de referência, mas sim a construção de um genoma por redundância de sobreposição de transcritos. Para muitos organismos que não possuem um genoma de referência de qualidade pode ser escolha válida, ou ainda para o melhoramento de genomas de referência pois não depende do bom alinhamento para análise, embora tenha a desvantagem de exigir um maior número de sequências, maior esforço de bioinformática e maior sensibilidade a erros do sequenciamento.

c) estratégia combinada.



Figura 4 –Visão geral do passo a passo de uma análise transcriptômica com RNA-seq. Adaptado de MARTIN; WANG, ZHONG, 2011.

A atual medicina é indissociável dos resultados de pesquisas celulares e moleculares. O estudo via RNA-seq de tecidos orgânicos gerou novos conhecimentos como o comportamento de diferentes doenças pulmonares, a estratificação do tratamento da leucemia mieloide e sua terapia personalizada, bem como a abordagem dos mielomas, hepatocarcinomas, câncer de mama, câncer colorretal, câncer do ovário, do pâncreas e da tireoide (KAN; SHUMYATCHER; HIMES, 2017; SHEN et al., 2015; TANG et al., 2009).

Na última década, frente ao forte e contínuo avanço dos grupos de pesquisa em RNA-seq com tecidos biológicos, a necessidade de se trabalhar com pequenos e escassos fragmentos de tecidos e questionamentos sobre o comportamento de uma única célula frente à heterogeneidade de padrões de expressão gênica entre as células de um mesmo tecido, possibilitou o desenvolvimento da análise transcriptômica de uma única célula (Figura 5, p.30): *single-cell RNA-seq (scRNA-seq)* (TANG et al., 2009; WANG, ZHONG; GERSTEIN; SNYDER, 2009).

O scRNA-seq deve ser considerado quando se busca comparar os transcriptomas de células individualmente, em especial na pesquisa de populações celulares raras que de outra forma passariam despercebidas em estudos de RNA-seq de tecidos, onde milhões a bilhões de células são estudadas de uma única vez (HAQUE et al., 2017).

Experiências exitosas têm sido descritas com populações de células periféricas como linfócitos T, células neuronais, nas primeiras linhagens de células embrionárias, na diferenciação de células cancerosas e imunes. Além disso, permite a melhor compreensão dos perfis de expressão gênica, em especial de RNAs não codificantes (ncRNA), de expressões monoalélicas e inúmeros fenômenos transcricionais como a flutuação da expressão de sequências não constitutivas (HAQUE et al., 2017).

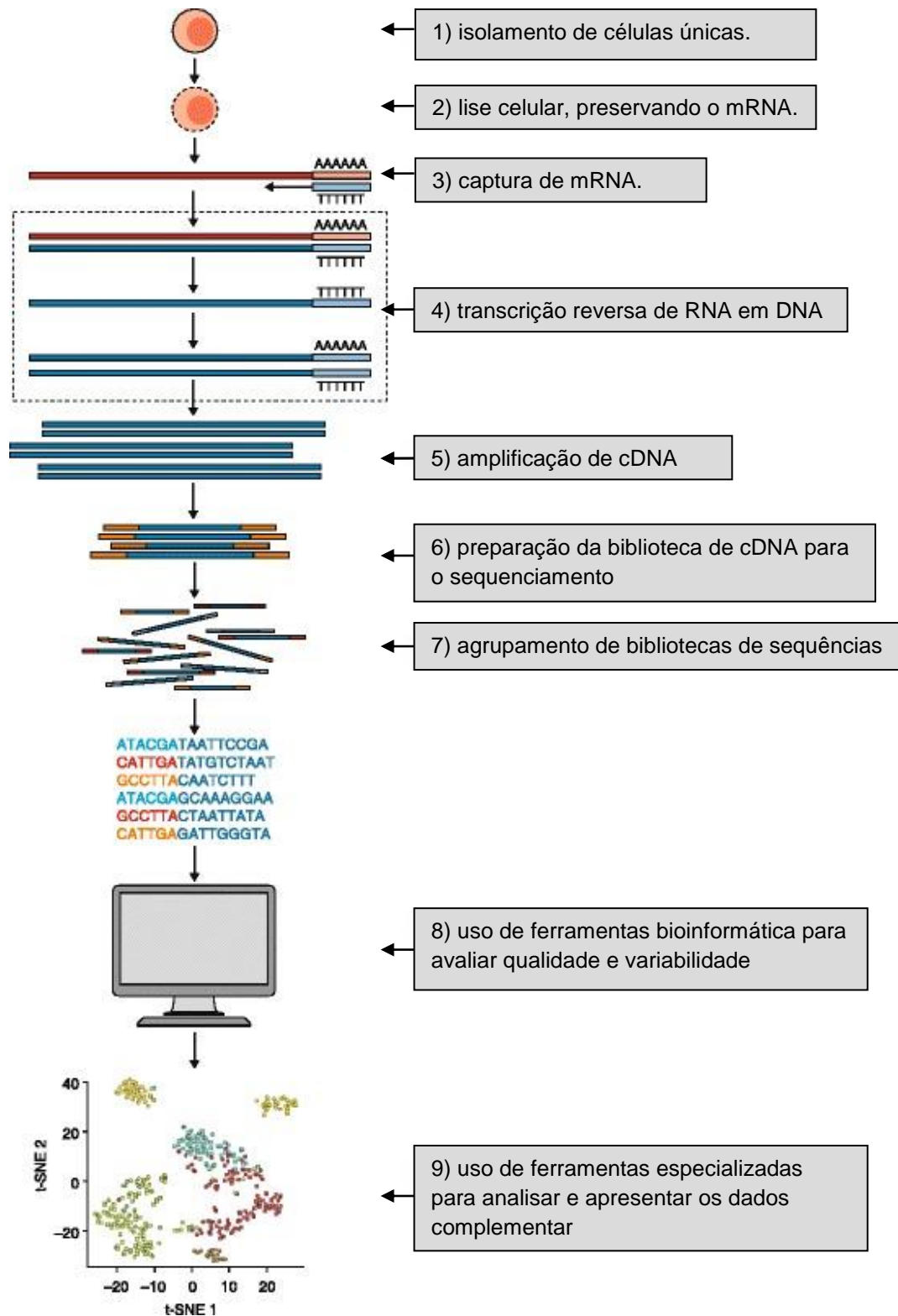


Figura 5 – Fluxo de trabalho geral de experiências de seqüenciamento de RNA de uma única célula (scRNA-seq). Um fluxo de trabalho típico de scRNA-seq inclui a maioria das seguintes etapas: 1) isolamento de células únicas, 2) lise celular, preservando o mRNA, 3) captura de mRNA, 4) transcrição reversa de RNA em DNA complementar (cDNA), 5) amplificação de cDNA, 6) preparação da biblioteca de cDNA para o seqüenciamento, 7) agrupamento de bibliotecas de seqüências, 8) uso de ferramentas bioinformática para avaliar qualidade e variabilidade, e 9) uso de ferramentas especializadas para analisar e apresentar os dados. Adaptado de HAQUE *et al*, 2017.

São estas experiências que tem tornado o estudo da expressão gênica muito menos concentrado em proteínas e permitido a descoberta ano após ano de novas funções para o RNA, dentre as quais as interações com proteína, no processo de splicing e mesmo na ativação ou repressão gênica de RNAs não codificadores (ncRNA) (NELSON; COX, 2011).

Em mamíferos, são maiores as concentrações celulares de ncRNA que de mRNA, e qualquer RNA que não codifique uma proteína; rRNA (ribossomais), tRNA (transportadores), sRNAi (interferentes), miRNA (micro), lncRNA (longos não codificantes), sRNAt (temporais) etc; é considerado ncRNA (EDDY, 2001).

A presença de RNAs não codificadores tem sido estudada em diferentes frentes da pesquisa em saúde. Acumulam-se dados tanto sobre lncRNA como miRNA que *in vitro* sugerem ser fundamentais no desenvolvimento muscular, no reparo de fibras musculares danificadas e na distrofia sindrômica como a distrofia de Duchene (SIMIONESCU-BANKSTON; KUMAR, 2016).

Em ratos experimentados com o silenciamento do lncRNA *H19* (uma sequência nucleotídica presente no cromossomo 11p15.5 que expressa um longo RNA não codificante e que comporta em seu *exon 1* o miRNA *mi-R-675*) (KURUKUTI et al., 2006), ficaram evidentes os defeitos na regeneração muscular, enquanto em ratos sem silenciamento de *H19*, havia uma hiper-expressão deste gene e de seus produtos (Figura 6, p. 32), com conseqüente regeneração muscular normal (SIMIONESCU-BANKSTON; KUMAR, 2016).

H19 é um monoalelo de origem materna, com o correspondente paterno imprintado pelo complexo telomérico *Igf2/H19* (*imprinting genômico*) e que não é expresso na maior parte dos tecidos humanos, exceto durante a fase embrionária, e tem sido demonstrado com atividades pró-tumorais em diferentes tipos e níveis de agressividade neoplásicos como mama, colorretal e renal (LIN et al., 2017; MEHDI; RIAZALHOSSEINI, 2017; POPE et al., 2017).

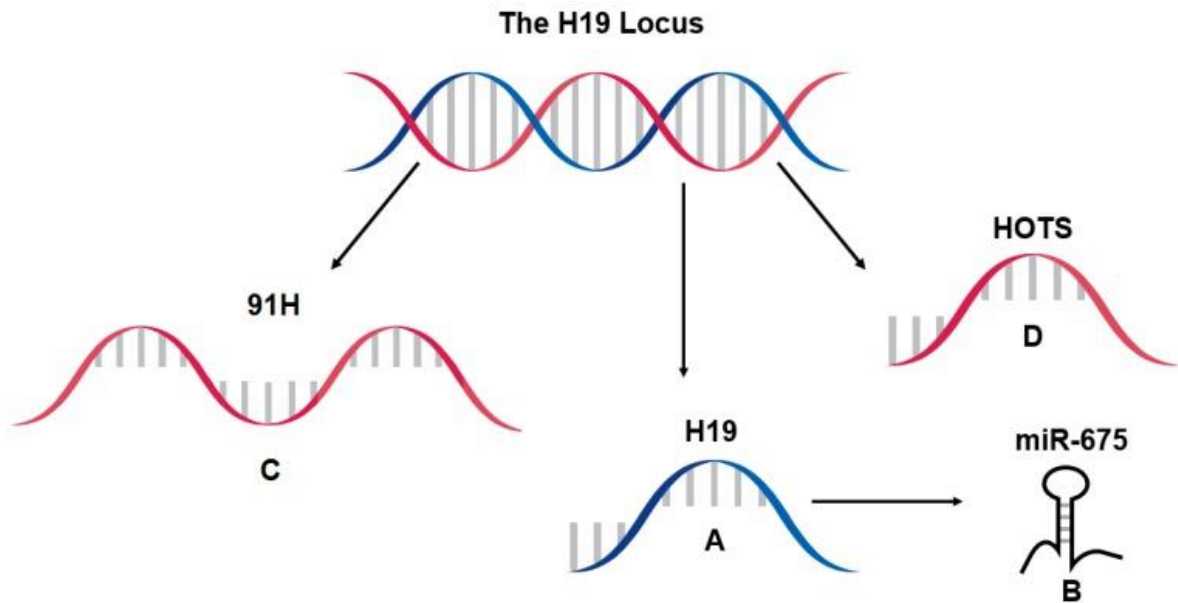


Figura 6 - A transcrição do *locus* de *H19* resulta no *H19* lncRNA (A) e seu microRNA *miR-675* (B) codificado a partir do *exon 1* de *H19*. Dois transcritos anti-sentido também são formados a partir do *locus*. O RNA *91H* (C) que é descrito em variados comprimentos, mas pode potencialmente ser codificado a partir da outra cadeia de DNA inteiramente abrangendo *H19* e outras porções de suas sequências reguladoras e o RNA *HOTS* (D) que é um transcrito a partir da maioria da sequência anti-sentido de *H19* e outras bases. *HOTS* pode ser traduzido em uma proteína nucleolar. Adaptado de POPE *et al.*, 2017.

A atuação dos transcritos do gene *H19* na iniciação tumoral não é clara. Contudo, alguns exemplos de teorias tentam unificar o papel de alguns deles, como na inibição do gene da proteína *p53*, importante fator de controle da ploidia celular e principal proteína associada à estabilidade genômica pré-tumoral (RAVEH *et al.*, 2015).

Fatores epigenéticos estariam envolvidos na hipóxia tecidual, na metilação de histonas H3K9 e H3K27 através do complexo metil-CpG, com consequente aumento da expressão de *H19* em células com *p53* deficiente, ativando uma cascata de oncogenes que culminaria com a poliploidia celular até um limite de não retorno em que a presença de *H19* tanto serviria como estabilizador de poliploidias e por um *feedback* negativo permitiria o desenvolvimento de aneuploidias e por consequente o desenvolvimento de cânceres, como o câncer gástrico. Portanto, *H19* estaria envolvido na instabilidade cromossômica de um estado celular pro-tumoral (RAVEH *et al.*, 2015).

É fato que a expressão de *H19* em tecidos onde normalmente não é expresso pode ser letal e pro-tumoral, mas de outro lado pode estar relacionada ao controle da de-diferenciação celular. Esta observação depende do estágio celular e merece

estudos posteriores para elucidação (Figura 7) (BRUNKOW; TILGHMAN, 1991; RAVEH et al., 2015).

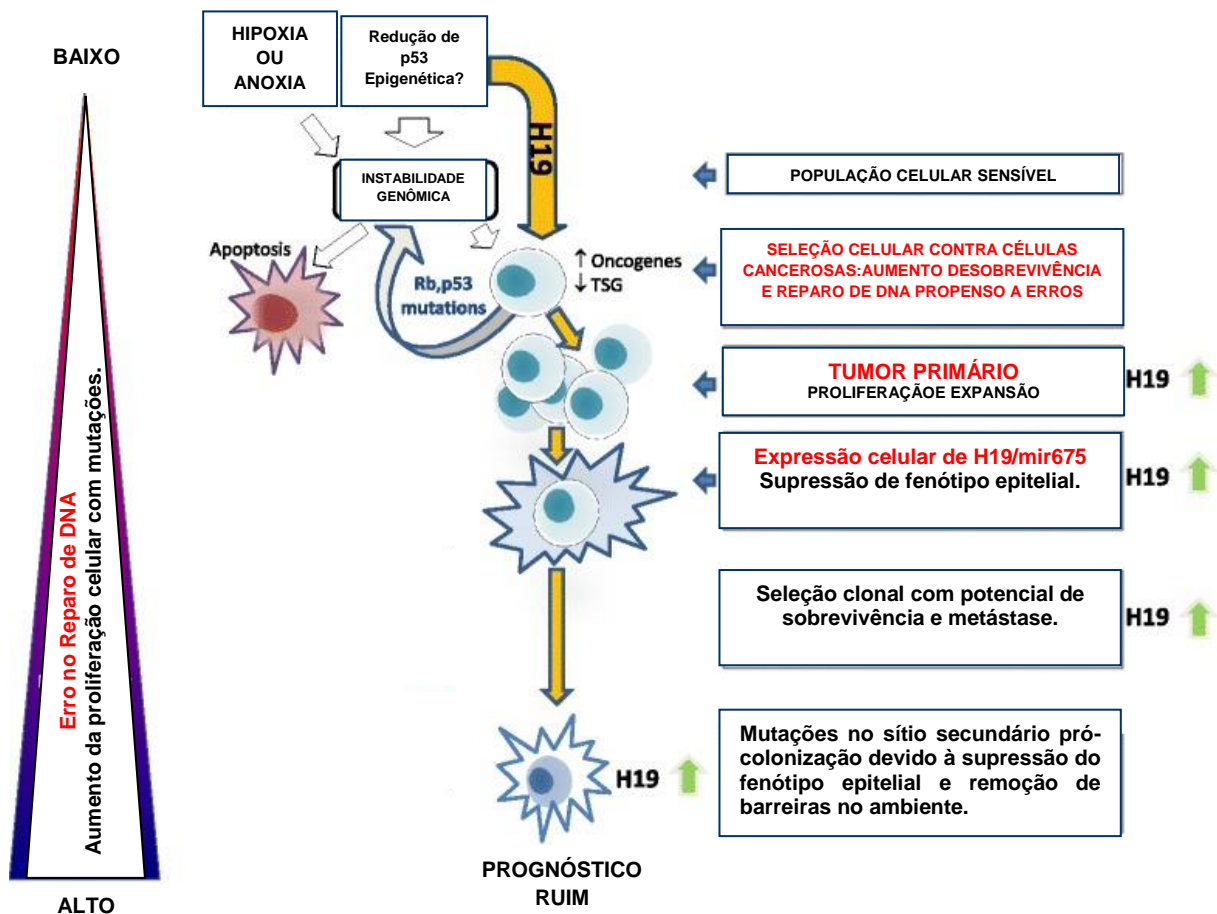


Figura 7 – Várias condições de estresse celular poderiam causar uma instabilidade genômica com redução da expressão de *p53*, aumento de *H19* em sub- clones celulares submetidos à estresse seguido de perda da reação celular ao estresse, culminando com proliferação desordenada. Adaptado de Raveh *et. al.* 2015.

Outros lncRNA têm demonstrado importância na regulação da expressão gênica e na homeostase tecidual, podendo inclusive ser classificados como onco-lncRNA (*Hotair, Malat1, Sox2-Ot, Hnf1a-As1, Anril, H19, Carlo-5, Mvih, Pvt1, Evadr*) e lncRNA supressores tumorais (*Meg3, Spry4-It1, Pandar, Gas5, Tug1, Bancr*). Este conhecimento é novo e merece melhores definições na literatura, mas não há dúvida quanto ao papel dos lncRNA em diversos processos biológicos, sua interação em cascatas moleculares pouco claras e cada vez mais em doenças proliferativas/neoplásicas (DUAN; ZHANG, C.; PENG, 2016).

2.3 O ESTUDO MOLECULAR DA TIREOIDE E O SEQUECIAMENTO DE NOVA GERAÇÃO (NGS)

O perfil molecular de expressão gênica, e a expressão diferencial entre tecidos tireoideanos normais e com bócio utilizando técnicas de *NGS* não é descrito na literatura. Os poucos estudos utilizando *NGS* para doenças tireoidianas são concentrados nos carcinomas bem diferenciados e não diferenciados da tireoide (CHA; KOO, 2016; RIA et al., 2016).

Com relação ao tecido sem doença/normal, um estudo com *microarray* mostrou que o tecido sem doença oriundo de tireoides com doença neoplásica quando comparado com tecidos sem doença oriundos de patologias não neoplásicas, possuem pelo menos 28 genes expressos diferencialmente (RIA et al., 2016).

Estiveram significativamente aumentados (*up-regulated*) no tecido sem doença (normal) oriundo de tireoides com carcinoma *versus* o tecido sem doença (normal) de tireoides operadas por outras doenças não neoplásicas os seguintes genes: *Zfp36l1*, *Tuft1*, *Slca2a3*, *Skp1*, *Ripk5*, *Rab7b*, *Ptgs2*, *Klf6*, *Jund*, *Irf1*, *Ier3*, *Hif1a*, *Hbegf*, *Gbp1*, *Gadd45a*, *Gadd45b*, *Eif4a3*, *Dusp5*, *Ccnl1*, *Cadm1*, *Bmp2*, *Bhlhb2*, *Auxd1*, e em contraste, a expressão reduzida de 5 genes (*down-regulated*): *Wdr48*, *Tef*, *Pnpla7*, *Accs* e *Klk-4* (RIA et al., 2016).

A lacuna deixada pela classificação citológica dos nódulos tireoideanos e o advento dos estudos moleculares destas doenças permitiu o desenvolvimento de testes moleculares para diagnóstico de doenças nodulares da tireoide, notadamente baseados na identificação das mutações *Braf*, *Ras* e nos rearranjos *RET/PTC* e *PAX8/PPAR γ* (NIKIFOROV et al., 2011).

Atualmente, os principais testes são o Afirma GEC®, baseado na tecnologia de *microarray* para análise de expressão de mRNA de 167 diferentes genes em produto de PAAF das categorias III e IV de Bethesda. Com alto valor preditivo negativo para a suspeição de malignidade (ZHANG; LIN, 2016).

De maneira semelhante, o ThyGenX TEST® desenvolvido para as PAAFs das categorias III e IV de Bethesda afim de detectar mutações e translocações de genes associados ao carcinoma bem diferenciado da tireoide (ZHANG; LIN, 2016).

Com a introdução do *NGS*, uma variante do ThyGenX TEST® está disponível para uso da plataforma *NGS* na identificação de 100 alterações genicas em 8 genes

associados a malignidade. A mesma empresa também passou a oferecer o ThyraMIR®, que inclui a análise de 10 diferentes miRNA (*miR-29b-1-5p*, *miR-31-5p*, *miR-138-1-3p*, *miR-139-5p*, *miR-146b-5p*, *miR-155*, *miR-204-5p*, *miR-222-3p*, *miR-375*, e *miR-551b-3p*), a ser usado quando o ThyGEnX TEST® é negativo para malignidade, associação que aumenta o valor preditivo negativo (ZHANG; LIN, 2016).

Também utilizando a plataforma NGS, o ThyroSeq TEST® propõe a análise de um painel de mutações e fusões de 12 genes com 284 *hot spots* mutacionais. Os primeiros testes mostraram valores preditivos discordantes, e por esta razão, carece de mais estudos e casos (ZHANG; LIN, 2016).

Os estudos a respeito do uso de NGS na prática clínica, vem de experiências realizadas com o tratamento oncológico personalizado. Há testes em fase II demonstrando que o uso de uma terapia personalizada, apresentou melhores resultados e menor toxicidade que aquela padronizada por protocolos não personalizados (HORAK; FRÖHLING; GLIMM, 2016).

Quer seja baseado no sequenciamento completo do genoma (*whole-genome-sequencing/WGS*) ou dos transcritos/exons alinhados à referência utilizando-se das ferramentas de RNA-Seq (*whole-exome-sequencing / WES*) e submetidos ao escrutínio estatístico de ferramentas de bioinformática, o NGS permite uma análise de múltiplas aberrações genéticas, tais como SNPs, inserções e deleções, número de cópias de variantes, *splicing* alternativo e outros rearranjos. Tais vantagens fazem frente aos painéis de grupos de genes (*microarray*), comumente aplicados na prática clínica para testes moleculares. Embora estes últimos já estejam padronizados, custem menos e sejam de fácil implementação, tem valor limitado para pesquisa de aberrações complexas e multigênicas (RIA et al., 2016; ZHANG; LIN, 2016).

A aplicabilidade clínica do NGS para doenças nodulares tireoidianas carece de validação por ensaios clínicos e em especial do conhecimento de fatores ainda obscuros da fisiopatogênese nodular e tumoral. Espera-se que além do diagnóstico pré-cirúrgico, os estudos com RNA-Seq por NGS contribuam fazendo parte de uma análise prognóstica, selecionando-se a terapia certa, para a doença certa na dose certa (CHA; KOO, 2016; HORAK; FRÖHLING; GLIMM, 2016).

3 OBJETIVOS

3.1 GERAL

Caracterizar o perfil de expressão gênica de tecidos tireoidianos com bócio, comparando com tecidos tireoidianos sem doença, oriundos de uma histoteca de tecidos tireoideanos de pacientes submetidos à tireoidectomia parcial ou total na Fundação Hospital Adriano Jorge (FHAJ).

3.2 ESPECÍFICOS

1. Analisar a coleção de RNAs mensageiros com a tecnologia RNA-Seq da plataforma Illumina HiSeq 2000 de amostras de tecidos tireoideanos com bócio e sem doença;
2. Estudar os perfis de expressão gênica diferencial entre os tecidos com bócio e os tecidos com ausência de doença.
3. Estudar as características biológicas, vias metabólicas e interação proteína-proteína dos genes diferencialmente expressos.

4 METODOLOGIA

Sequenciamento do transcriptoma de fragmentos de tecido tireoideano com bócio e sem doença utilizando a plataforma de sequenciamento de nova geração em Illumina HiSeq 2000 com protocolo de RNA-seq e análise de expressão diferencial.

4.1 ASPECTOS ÉTICOS

O presente estudo faz parte do projeto intitulado “Construção de uma biblioteca de tecidos tireoideanos com diagnóstico de bócio” avaliado pelo Comitê de Ética em Pesquisa com seres humanos da Fundação Hospital Adriano Jorge (FHAJ), tendo sido aprovado em 01 de junho de 2013 sob o CAAE 16463813.9.0000.0007.

Todos os sujeitos doadores de tecido para a histoteca utilizada neste estudo fizeram a leitura e a assinatura do Termo de Consentimento Livre e Esclarecido (TCLE) antes da tireoidectomia (Anexo1).

4.2 ORGANIZAÇÃO E SELEÇÃO DOS TECIDOS TIREOIDEANOS

As amostras utilizadas neste estudo foram obtidas diretamente da peça cirúrgica após sua retirada da região cervical por tireoidectomia parcial ou total indicadas pelo cirurgião assistente.

Os fragmentos mediam 1cm³, tendo sido retirados dos polos superiores e inferiores da tireoide pós-cirurgia, bem como do istmo, e quando presente, do lobo piramidal (Anexo 2), seguindo a orientação de um patologista especializado.

Imediatamente após a coleta, os fragmentos foram armazenados em microtúbulos contendo RNA *later* (conservante de RNA), sendo transportados à -20°C sob gelo seco até o Laboratório de Tecnologias do DNA do Centro de Apoio

Multidisciplinar (CAM) da Universidade Federal do Amazonas – UFAM e armazenados à -80°C até o seu uso.

4.3 AMOSTRAS DE TECIDO TIREOIDIANO

As amostras de tecido compreenderam: quatro amostras de bócio colóide e quatro amostras de tireoide com ausência de doença, selecionados na histoteca de tecidos tireoideanos do Prof. João Bosco Botelho e armazenados à -80°C e classificadas com o diagnóstico histopatológico de bócio ou com ausência de doença, conforme o resultado da análise histopatológica de cada região da peça cirúrgica (Anexo 3).

4.4 EXTRAÇÃO DE RNA TOTAL

Fizeram parte deste estudo oito amostras oriundas de tecido tireoideano, sendo quatro histologicamente sem doença e quatro com diagnóstico de bócio. Antes de seguir para a extração de RNA, foram previamente removidos o excedente de conservantes de RNA de cada amostra com o auxílio da micro-pipeta.

A amostra de tecido foi envolta em papel alumínio livre de RNase e foi banhada em um cadinho contendo $\text{N}_{2(l)}$ até o congelamento completo da amostra, seguida de trituração com pistilo e recuperação do tecido.

Após esta fase, seguiu-se o protocolo com TRIZOL® Reagent (lifetechnologies™) incluindo suas etapas opcionais conforme recomendação do fabricante (Figura 8).



Figura 8 – Fluxo de atividades em bancada para a extração de RNA.

4.5 CONTROLE DE QUALIDADE DAS AMOSTRAS

O controle de qualidade amostral foi realizado a partir da quantificação das amostras pelo espectrofotômetro Nanodrop® e pelo fluorômetro Qubit®, seguido do estudo por eletroforese em gel de Agarose e análise da integridade do RNA por Agilent 2100.

4.6 SEQUENCIAMENTO

As amostras de RNA íntegras, com concentração superior a 100ng/μl (Quibit®), bandas visíveis na eletroforese e razão de pureza por absorvância 260/280 > 2 (Nanodrop®) foram consideradas aprovadas ao controle de qualidade e sequenciadas utilizando o protocolo RNA-Seq da plataforma Illumina Hi-seq 2000, 150pb x 2 *paired end reads*. Este serviço foi realizado pela empresa GenONE (BRA).

4.6.1 Análise das sequências

Os dados recebidos do sequenciamento foram processados e armazenados em um servidor dedicado. Os conjuntos de sequencias obtidos foram inicialmente avaliadas quanto à qualidade, utilizando o programa FastQC seguido de filtragem pelo programa Trimmomatic 0.36 (BOLGER; LOHSE; USADEL, 2014) afim de remover adaptadores, bases do início com baixa qualidade, bases do fim das sequencias, e sequencias com qualidade abaixo de 30 (parâmetro PHRED) (BRUMATTI GONÇALVES; OTHERS, 2015).

As sequencias foram então alinhadas utilizando os *softwares* de mapeamento Bowtie2 e TopHat com o genoma de referência *Homo sapiens* GRCh38 disponível em: ftp://ussd-ftp.illumina.com/Homo_sapiens/NCBI/GRCh38/Homo_sapiens_NCBI_GRCh38.tar.gz (TRAPNELL et al., 2012).

Após isto, com a utilização dos softwares Cufflinks/Cuffdiff/Cuffmerge procedeu-se a análise da expressão diferencial gênica a partir da comparação com o genoma de referência utilizando o Protocolo Tuxedo. Para a visualização gráfica dos

genes diferencialmente expressos nas condições experimentais utilizaram-se as ferramentas R - CummeRbund, (Tabela 1) (TRAPNELL et al., 2012).

Tabela 1: Listagem de *softwares* utilizados no Protocolo Tuxedo.

Software	Finalidade	Versão	Endereço Eletrônico
1 FastQC	Controle de qualidade	0.11.5	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
2 Trimmomatic	Trimagem (Remover adaptadores e qualidade baixa).	0.36	http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic-0.36.zip
3 TopHat2	Mapeamento/Alinhamento de sequencias	2.1.1	https://ccb.jhu.edu/software/tophat/index.shtml
4 Cufflinks	Montagem do transcriptoma e análise de expressão diferencial de RNA-Seq	2.2.1	http://cole-trapnell-lab.github.io/cufflinks/
5 Cuffmerge	Fusão de diversas montagens geradas por Cufflinks.	1.0.0	http://cole-trapnell-lab.github.io/cufflinks/cuffmerge/
6 Cuffdiff	Encontra mudanças significativas na expressão do transcrito e splicing.	2.1.1	http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/
7 CummeRbund	Análise estatística e visualização de dados do Cuffdiff	3.5	https://bioconductor.org/packages/release/bioc/html/cummeRbund.html

Como o número de sequencias era limítrofe ao recomendado para a análise de bioinformática, para as análises posteriores foi utilizado o conjunto de sequencias original, em parceria com o grupo de Bioinformática do Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LaCTAD), parte da Universidade Estadual de Campinas (UNICAMP).

O mapeamento no genoma de referência (*Homo sapiens* - HG38) foi feito com o *software* Bowtie2 (LANGMEAD; SALZBERG, 2012) e a quantificação dos transcritos foi realizada com o *software* RSEM (LI; DEWEY, 2011).

Os genes diferencialmente expressos foram obtidos usando o *software* DESeq (ANDERS; HUBER, 2010). Devido à falta de réplicas em uma das condições, somente a condição com réplicas foi usada para estimar a dispersão.

4.6.2 Análise biológica

Os genes diferencialmente expressos foram analisados quanto à função biológica e classe proteínas utilizando a ferramenta *Panther* (MI et al., 2017) disponível na plataforma *Gene Ontology Consortium* (<http://www.geneontology.org/>) (WHEELER, 2004), seguido de análise de interação proteína-proteína com a ferramenta GeneMANIA (<http://www.genemania.org/>) (WARDE-FARLEY et al., 2010; ZUBERI et al., 2013) e revisão de literatura para compreensão dos achados (HUTCHINS, 2014; WONG et al., 2015).

5 RESULTADOS

As oito amostras coletadas foram submetidas à extração de RNA total, sendo quatro oriundas de tecido com ausência de doença (ao diagnóstico histopatológico) e quatro com diagnóstico histopatológico de bócio.

O controle de qualidade amostral realizado a partir da quantificação das amostras pelo método fluorimétrico (Qubit®) e por espectrofotometria (Nanodrop®), seguido do estudo por eletroforese em gel de agarose e checagem da integridade do RNA por Agilent 2100 está demonstrado no Quadro 2, Figura 9 (p. 43) e Anexos 4A e 4B.

Quadro 2 – Análise qualitativa e quantitativa de RNA total.

Amostra	Conc. (ng/μl) Qubit®	Volume (μl)	Quant (μg) Qubit®	260/280 Nanodrop®	260/230 Nanodrop®,	Conclusão
*B13	637	17	10.83	2.065	1.429	Aprovada
*B14	1500	17	25.5	2.064	2.139	Aprovada
**N6	321	17	5.457	2.061	1.7	Aprovada
**N17	326	16	5,216	1.94	1.44	RNA não qualificável
**N1	576	16	9.22	1.922	0.259	Degradação
**N5	2	16	0.03	1	0.006	Banda não visualizada
*B2	1682	16	26.91	1.965	1.261	Degradação
*B11	1598	16	25.57	1.954	1.455	RNA não qualificável

*B – tecido tireoideano com diagnóstico histológico de bócio;

**N - tecido tireoideano com ausência de doença.

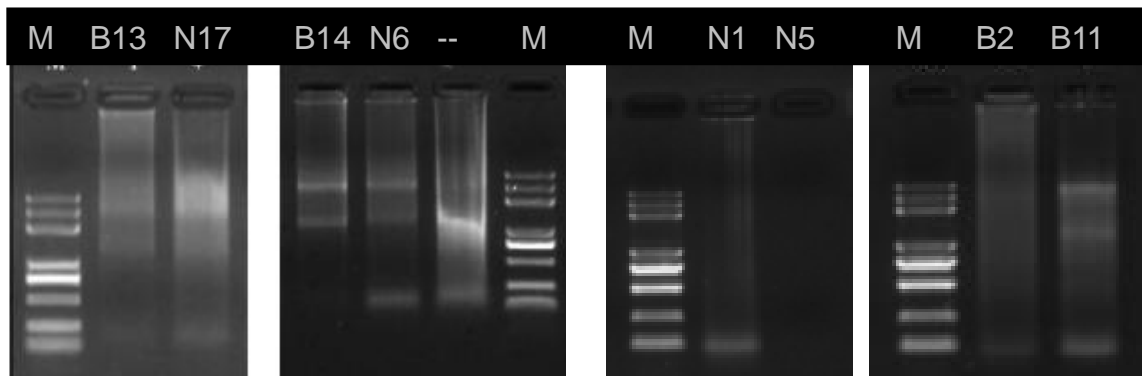


Figura 9 – Eletroforese em gel de Agarose das amostras indicadas no Quadro 2 (1.0%, 0.5xTBE; 180V em 16min). M: Trans 2K Plus DNA ladder.

Tendo em vista o controle de qualidade, seguiu-se com o sequenciamento das amostras de bócio B13, B14 e a amostra com ausência de doença N6, sendo as demais descartadas (Figura 10).

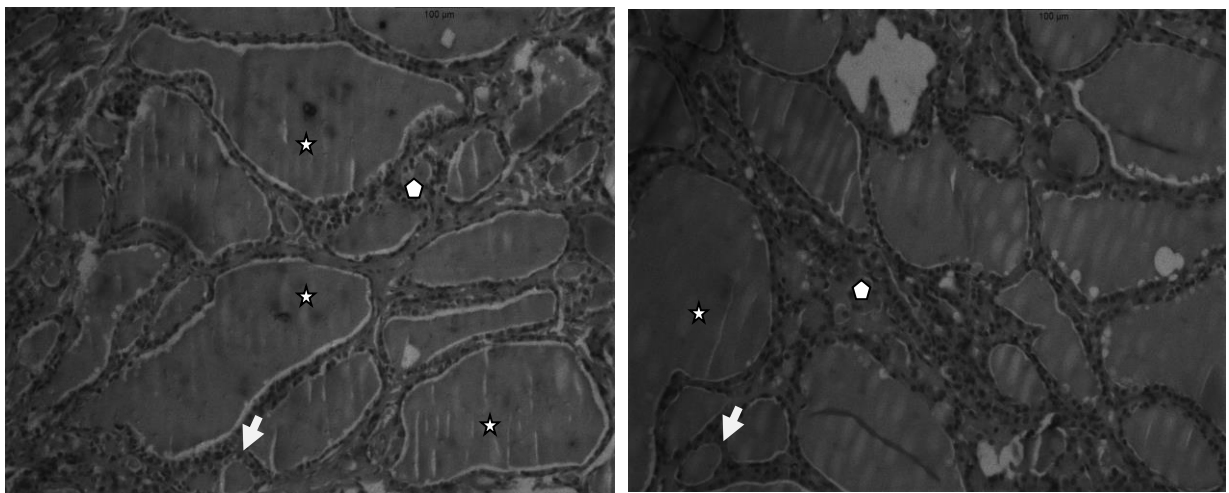


Figura 10 - Imagens histológicas dos tecidos com bócio das amostras B13 e B14 respectivamente. As imagens mostram a variabilidade de tamanho de pequenos (setas) e macro-folículos (estrela) colóides separados por densas bandas de fibrose (pentágono).

Todas as amostras sequenciadas apresentaram quantitativo maior que vinte milhões de sequências em pelo menos uma das fitas no controle de qualidade após a etapa de trimagem com o protocolo Trimmomatic utilizando os argumentos:

- HEADCROP 10 – Corte do início leitura com comprimento de 10 pb.
- SLIDINGWINDOW:4:20 – Remoção de bases de baixa qualidade neste intervalo.
- LEADING:20 - remoção de até 20pb com baixa qualidade a partir do início da sequência.

- TRAILING:20 - remoção de até 20pb com baixa qualidade a partir do final da sequência, ou seja, na direção contrária 3' - 5'.
- MINLEN:80 – Elimina as sequências abaixo de 80pb.

Dos resultados obtidos com o FastQC, mostramos a seguir a tabela de estatística básica de cada amostra e o gráfico de qualidade por base em cada sequência das amostras antes e depois da etapa de trimagem (Figuras 11-22, p. 45-50).

Em nosso experimento todas as amostras têm bases acima da escala 30 (escala *phred*), demonstradas no gráfico na camada de fundo verde (muito boa qualidade). A camada laranja mostra as bases de qualidade razoável e a vermelha com má qualidade.

Além disso, as tabelas mostram que todas as amostras contêm sequências com comprimento de 80 – 140pb (*Sequence length*) e, embora o total de sequências seja maior que 20 milhões em pelo menos um dos sentidos da fita de leitura (*Total Sequences*), nota-se grande perda de material em boa parte das sequências em N6 e B14 maior que 11% das sequências.

Os elementos gráficos são (BRUMATTI GONÇALVES; OTHERS, 2015):

- A linha vermelha central é o valor da mediana.
- A caixa amarela representa o intervalo inter-quartil (25-75%).
- Os traços superiores e inferiores representam os pontos 10 e 90%.
- Linha azul representando a qualidade média.

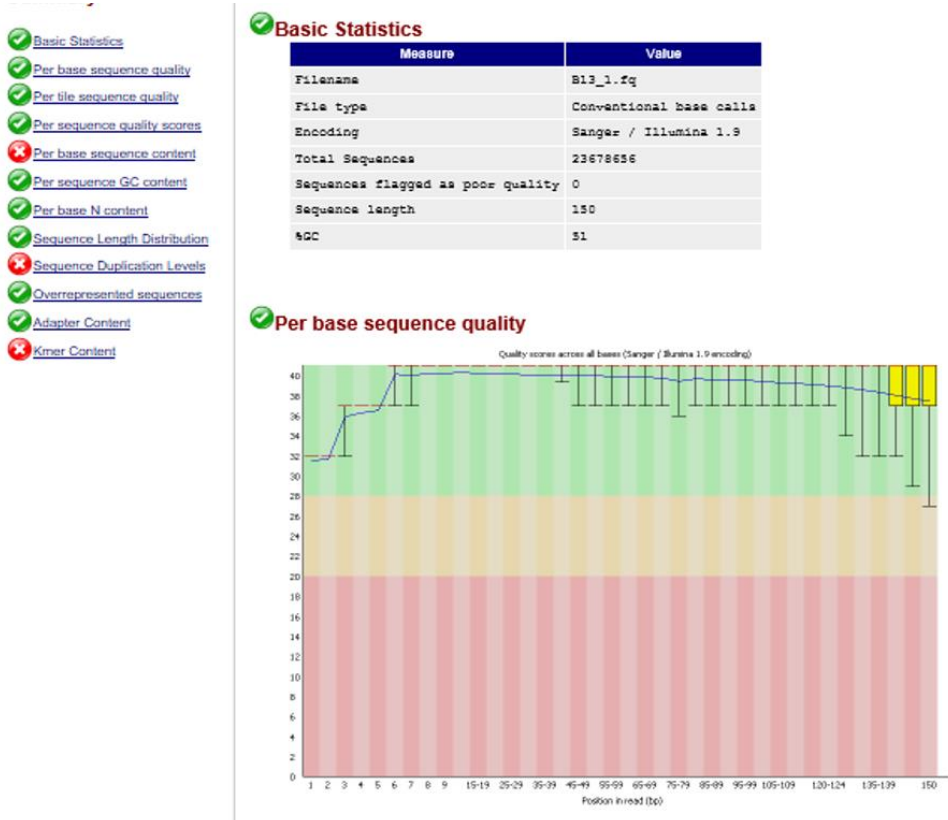


Figura 11 – Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.1.

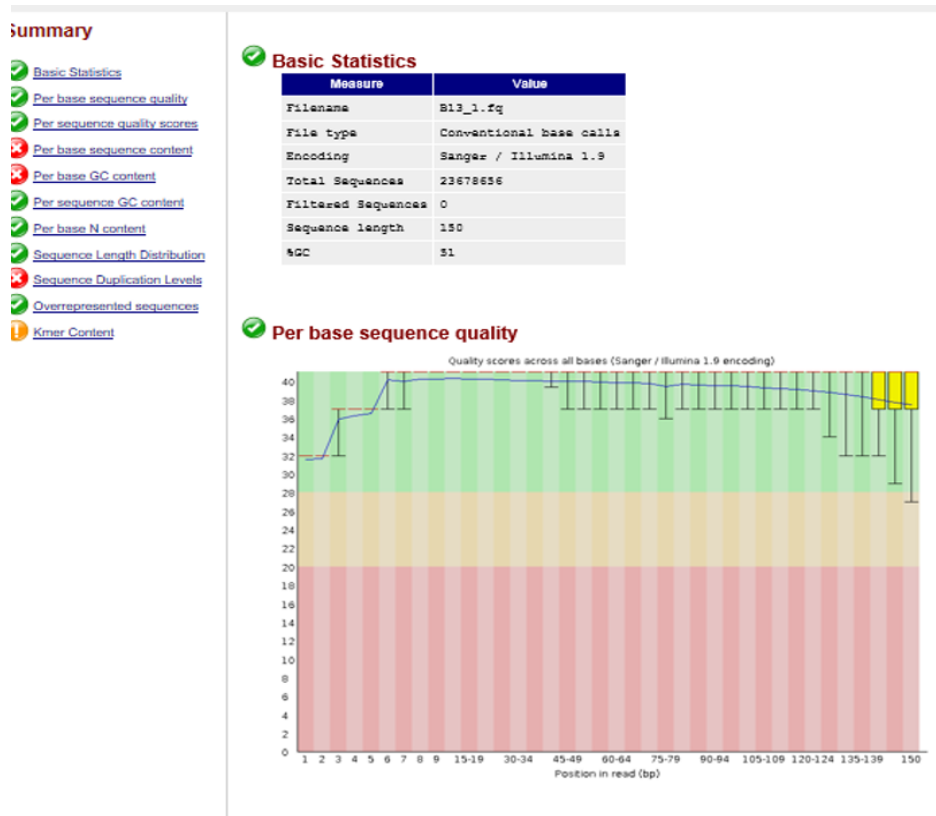


Figura 12– Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.1.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per file sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Adapter Content](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	B13_2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	23678656
Sequences flagged as poor quality	0
Sequence length	150
%GC	31

Per base sequence quality

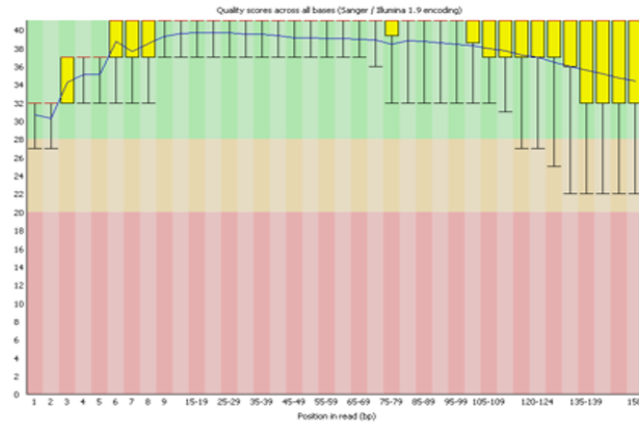


Figura 13 – Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.2.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per base GC content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	B13_2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	23678656
Filtered Sequences	0
Sequence length	150
%GC	31

Per base sequence quality

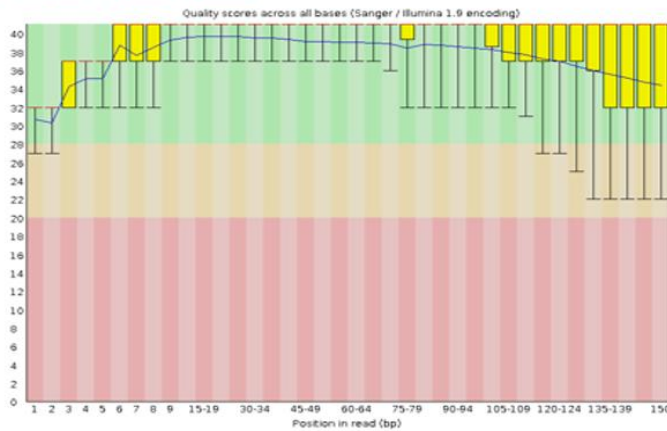


Figura 14 – Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B13.2.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per file sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Adapter Content](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	B14_1.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	27899802
Sequences flagged as poor quality	0
Sequence length	150
%GC	50

Per base sequence quality

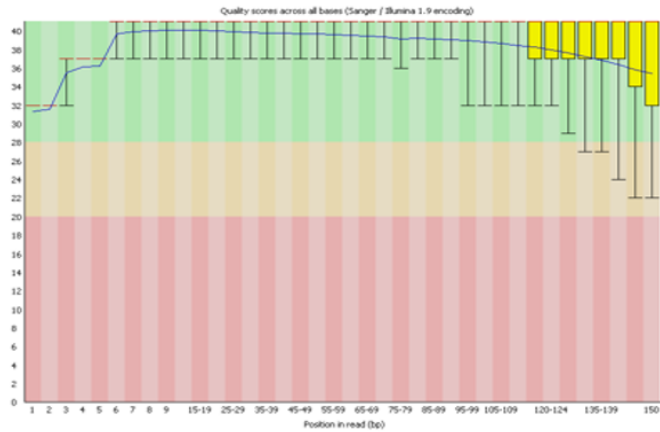


Figura 15 – Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.1.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per base GC content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	B14_1.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	20952062
Filtered Sequences	0
Sequence length	80-140
%GC	51

Per base sequence quality

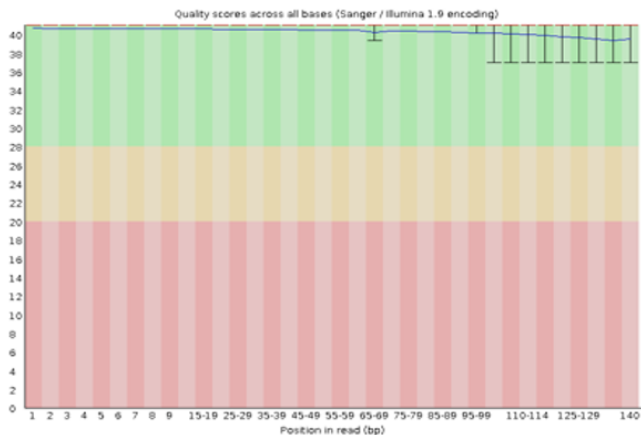


Figura 16 – Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.1.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per file sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Adapter Content](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	B14_2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	27899802
Sequences flagged as poor quality	0
Sequence length	150
%GC	50

Per base sequence quality

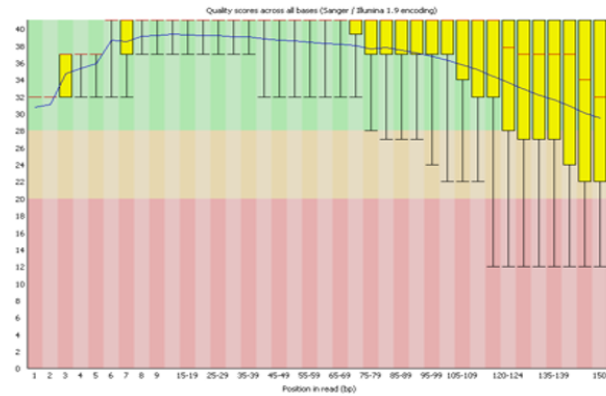


Figura 17 – Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.2.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per base GC content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	B14_2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1542461
Filtered Sequences	0
Sequence length	80-140
%GC	51

Per base sequence quality

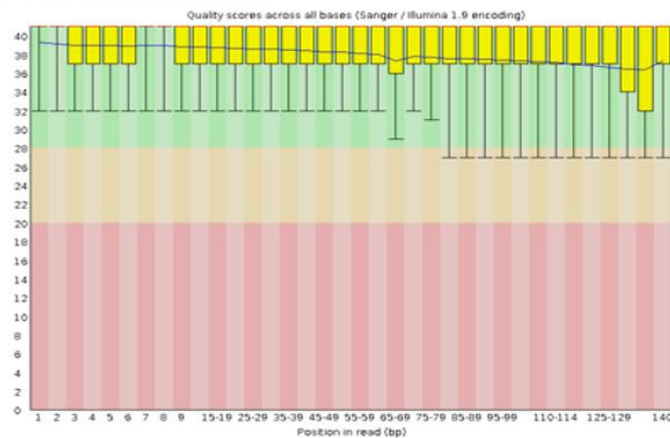


Figura 18 – Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra B14.2.

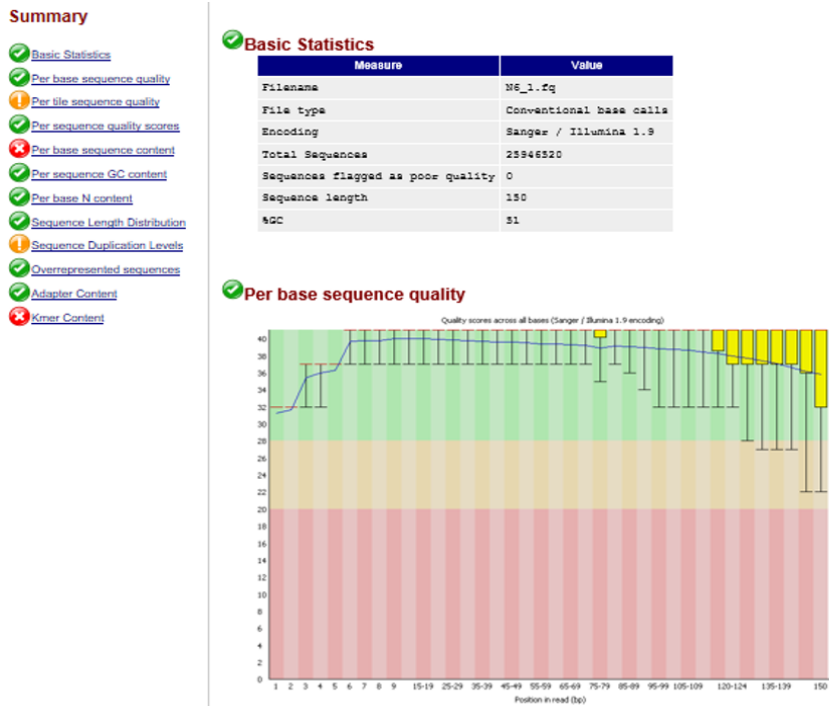


Figura 19 – Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.1.

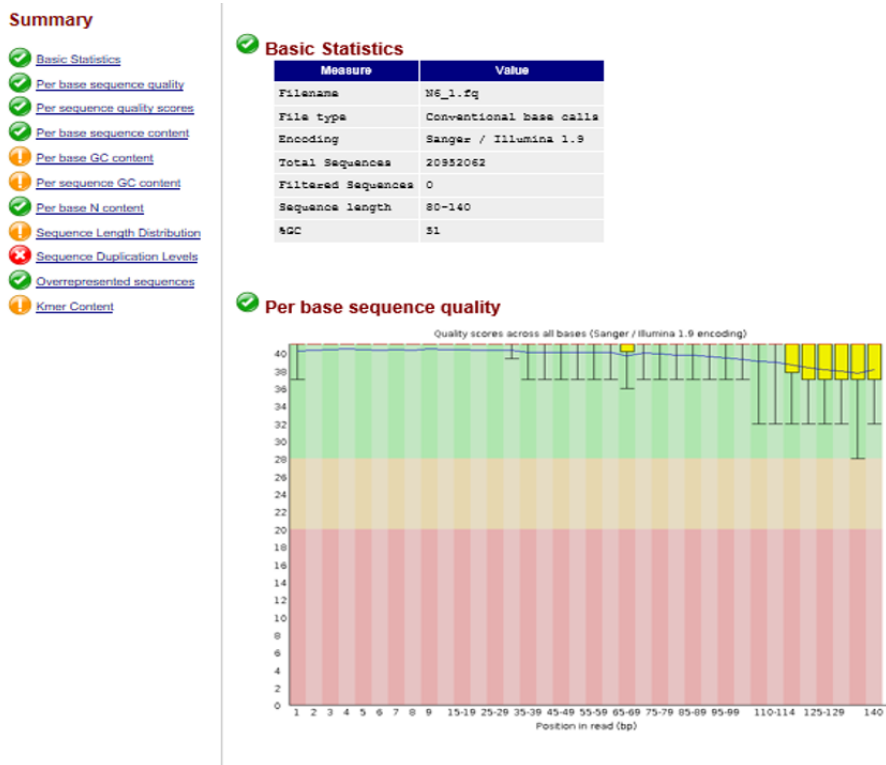


Figura 20 – Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.1.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per file sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Adapter Content](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	N6_2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	25946520
Sequences flagged as poor quality	0
Sequence length	150
%GC	51

Per base sequence quality

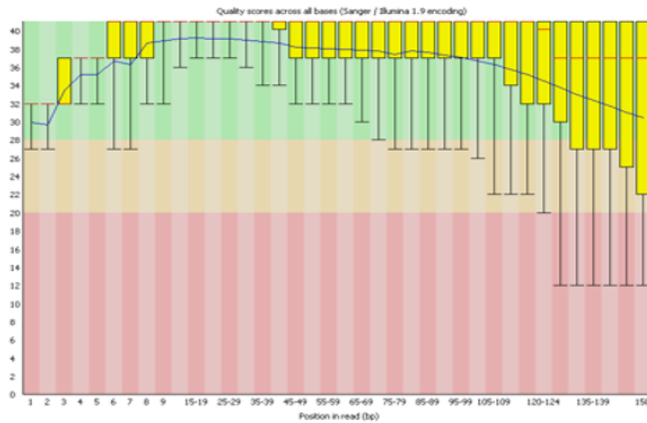


Figura 21 – Análise do FastQC pré etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.2.

Summary

- [Basic Statistics](#)
- [Per base sequence quality](#)
- [Per sequence quality scores](#)
- [Per base sequence content](#)
- [Per base GC content](#)
- [Per sequence GC content](#)
- [Per base N content](#)
- [Sequence Length Distribution](#)
- [Sequence Duplication Levels](#)
- [Overrepresented sequences](#)
- [Kmer Content](#)

Basic Statistics

Measure	Value
Filename	N6_2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	547124
Filtered Sequences	0
Sequence length	80-140
%GC	52

Per base sequence quality

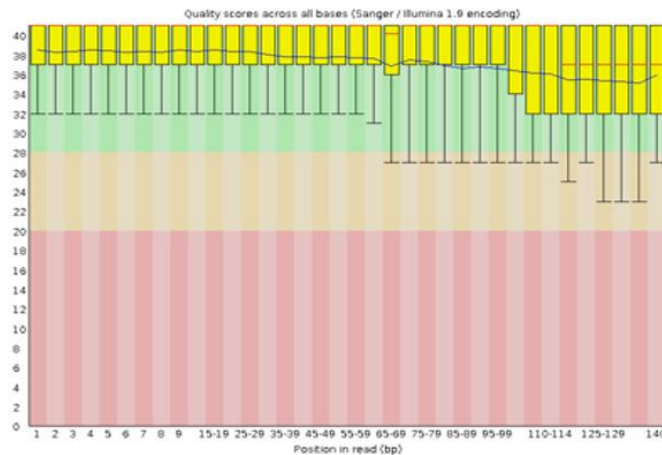


Figura 22 – Análise do FastQC pós etapa de trimagem (remoção de adaptadores) mostrando a estatística básica e a qualidade das sequencias por base da amostra N6.2.

A análise de expressão diferencial pelo protocolo Tuxedo demonstrou genes diferencialmente expressos utilizando a análise do Cuffdiff (TRAPNELL et al., 2012) onde houve diferença estatisticamente significativa entre os perfis de expressões gênicas de 6.102 genes de um total de 34.786 para B13 versus N6, mas não houve diferença estatística significativa da expressão 34.786 genes na comparação B14 versus N6 (Figuras 23 e 24).

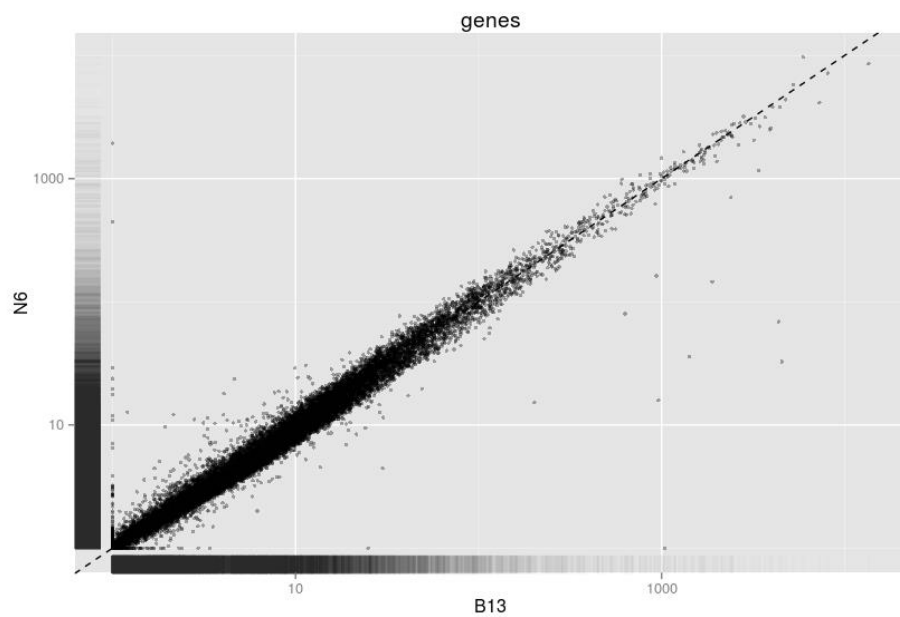


Figura 23 – Dispersão da expressão gênica entre as amostras B13 x N6.

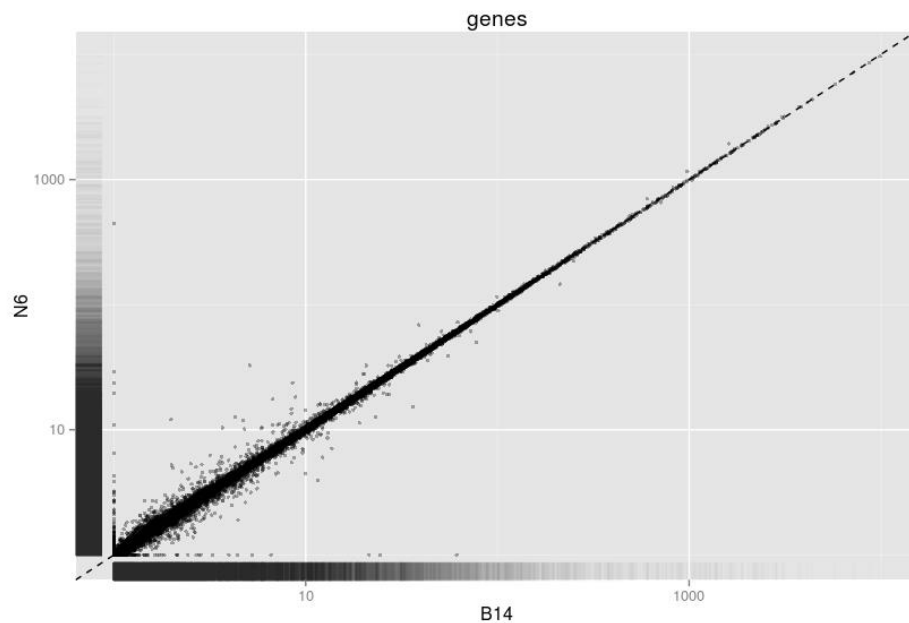


Figura 24 – Dispersão da expressão gênica entre as amostras B14 x N6.

A densidade da expressão gênica entre as amostras mostrou diferenças entre B13 e N6 mas não entre B14 e N6 (Figuras 25 e 26).

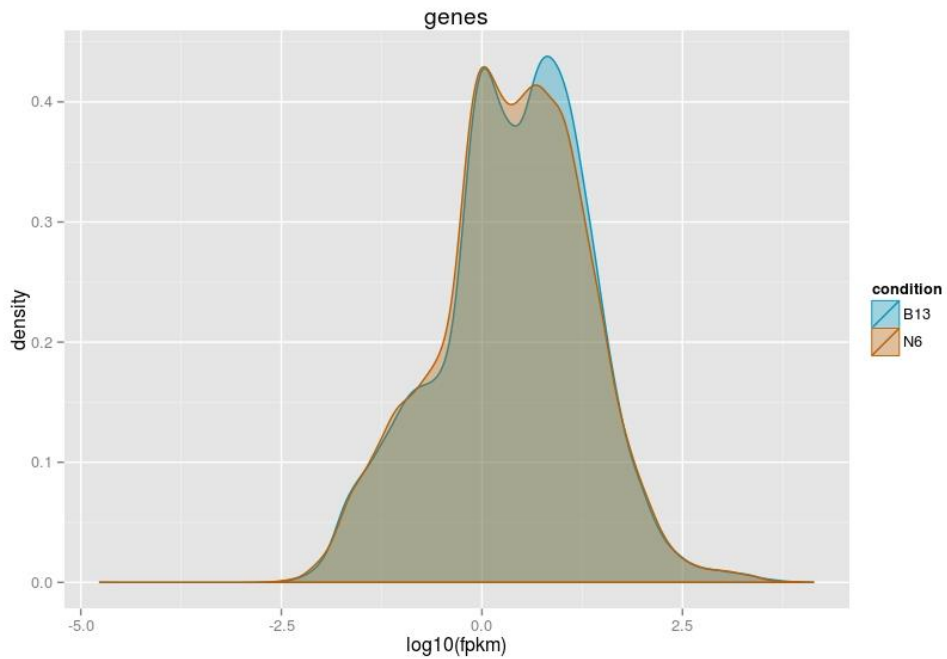


Figura 25 – Densidade da expressão gênica entre as amostras B13 x N6.

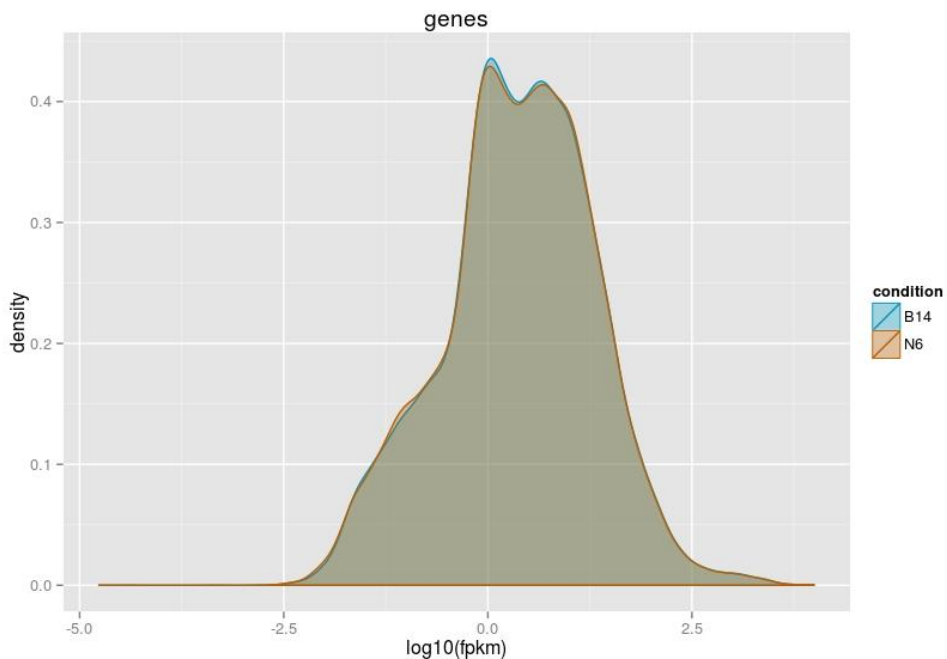


Figura 26 – Densidade da expressão gênica entre as amostras B14 x N6.

A partir dos parâmetros de análise do Cuffdiff, 24 genes expressos diferencialmente, podem ser novos transcritos já que não tem identificação e por

isso aparecem na tabela com “-“ , isso necessita de melhor investigação para confirmação (tabela 2).

O campo da tabela de genes diferencialmente expressos mostra:

- Transcrito: nome do transcrito;
- Status: Ok para o teste bem sucedido;
- Bócio / Controle : valores de FPKM de cada gene dentro da amostra;
- Log2FC : valor da mudança na base 2 (Bócio/Controle);
- P_value : O valor de p não corrigido;
- Q_value: O valor de p corrigido em FDR;
- Significant: Quando o valor de p < que 0,05.

Tabela 2- Tabela de saída utilizando o Protocolo Tuxedo das amostras B13 versus N6.

Transcrito	status	Bócio	Controle	log2FC	p_value	q_value	Significant
-	OK	269.621	188.165	-0.518938	5,00E-05	0.000265936	Yes
-	OK	404.011	253.277	-0.673677	5,00E-05	0.000265936	Yes
-	OK	738.238	517.126	-0.513569	5,00E-05	0.000265936	Yes
-	OK	535.805	375.929	-0.511248	5,00E-05	0.000265936	Yes
-	OK	287.997	350.392	0.282916	5,00E-05	0.000265936	Yes
-	OK	332.125	414.637	0.320122	5,00E-05	0.000265936	Yes
-	OK	271.143	216.045	-0.327727	5,00E-05	0.000265936	Yes
-	OK	936.36	161.295	-253.736	5,00E-05	0.000265936	Yes
-	OK	479.985	187.456	-135.644	5,00E-05	0.000265936	Yes
-	OK	65.843	391.219	-0.751054	5,00E-05	0.000265936	Yes
-	OK	225.608	184.376	-0.291166	5,00E-05	0.000265936	Yes
-	OK	770.415	573.124	-0.426787	5,00E-05	0.000265936	Yes
-	OK	302.304	220.222	-0.457042	5,00E-05	0.000265936	Yes
-	OK	957.059	603.057	-0.666313	5,00E-05	0.000265936	Yes
-	OK	788.102	431.867	-0.867795	5,00E-05	0.000265936	Yes
-	OK	1038.96	0	#NOME?	5,00E-05	0.000265936	Yes
-	OK	571.712	457.078	-0.322849	5,00E-05	0.000265936	Yes
-	OK	982.858	640.487	-0.617815	5,00E-05	0.000265936	Yes
-	OK	126.321	713.271	-0.824567	5,00E-05	0.000265936	Yes
-	OK	106.441	490.408	-1.118	5,00E-05	0.000265936	Yes
-	OK	192.04	145.442	-0.400962	5,00E-05	0.000265936	Yes
-	OK	708.541	177.141	132.197	5,00E-05	0.000265936	Yes
-	OK	4344.77	679.597	-599.846	5,00E-05	0.000265936	Yes
-	OK	1889.49	144.942	-370.445	5,00E-05	0.000265936	Yes

Fonte: Carvalho,DM, 2017.

Os resultados utilizando o Protocolo Tuxedo (TRAPNELL et al., 2012) não permitiram a continuidade das análises em R, somente a obtenção da densidade e dispersão gênicas.

O protocolo RSEM (LI; DEWEY, 2011) utilizando as sequencias originais considerando a importante perda de sequencias na etapa de trimagem (em torno de 11%), possibilitou a identificação de 70 sequencias de genes diferencialmente expressos (Anexo 5), com o software DESeq (ANDERS; HUBER, 2010), sendo que 61 deles estavam *down-regulated* no tecido tireoideano com bócio e nove *up-regulated* em comparação com o tecido sem doença (Figura 27).

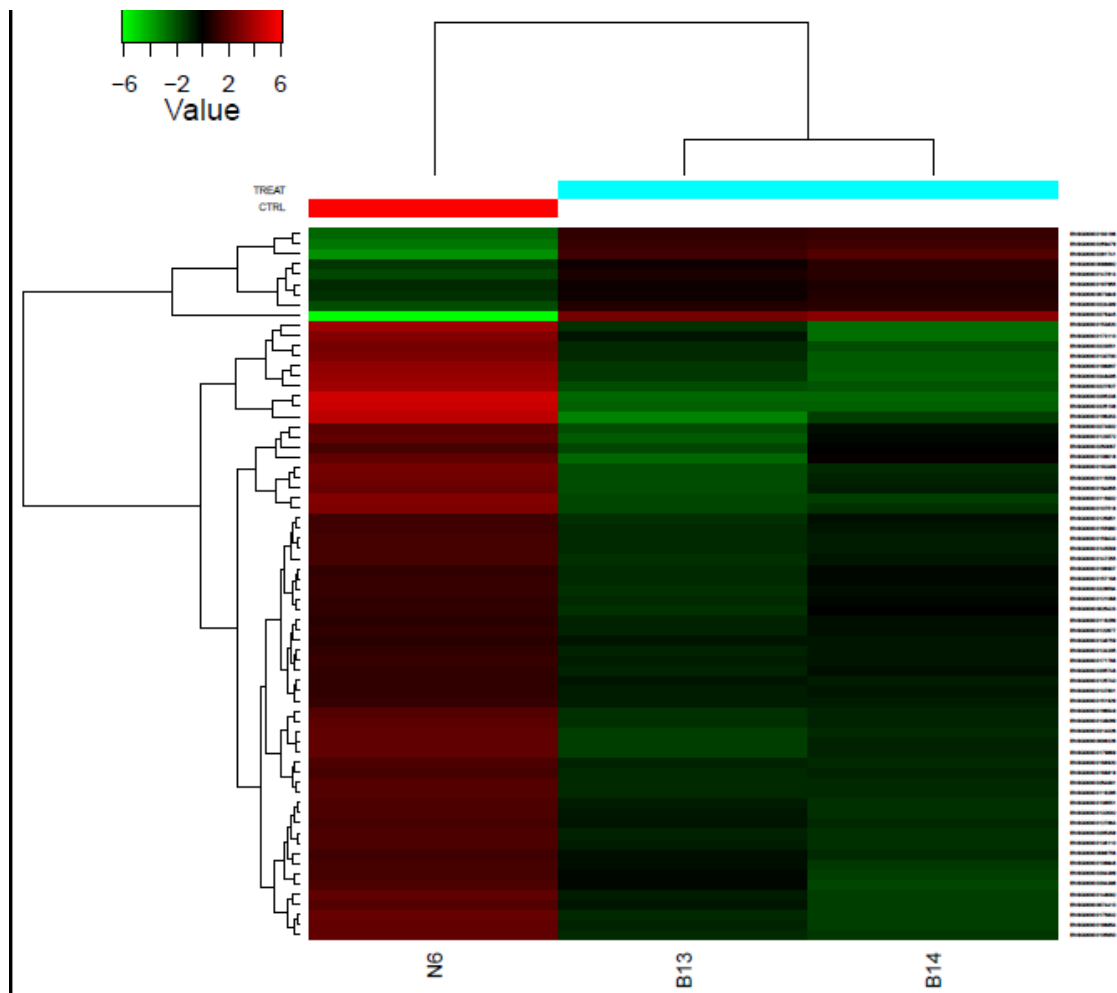
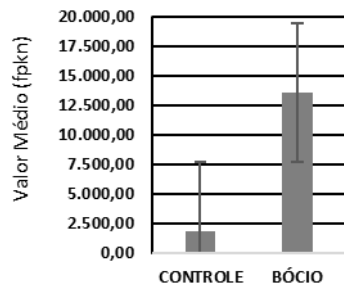
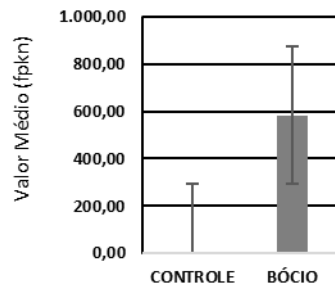
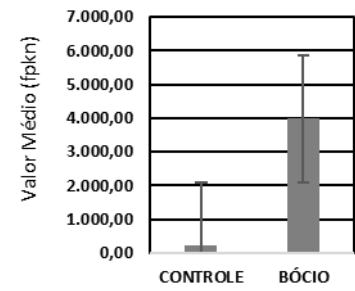
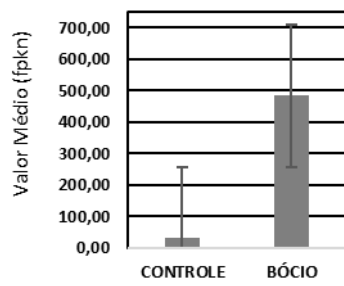
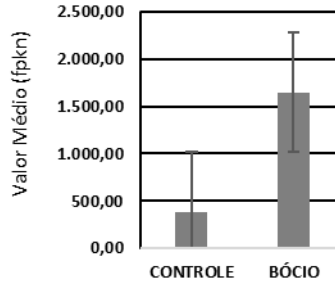
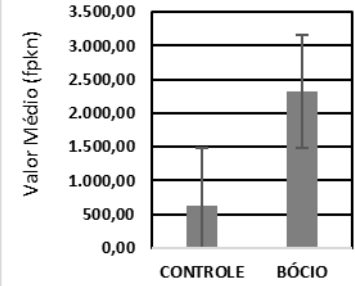
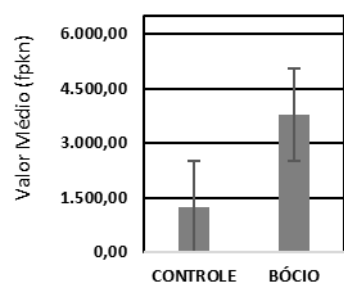
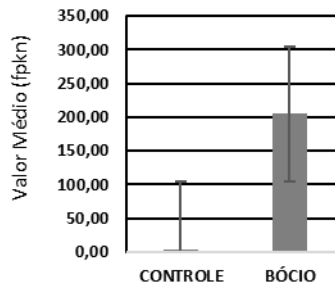
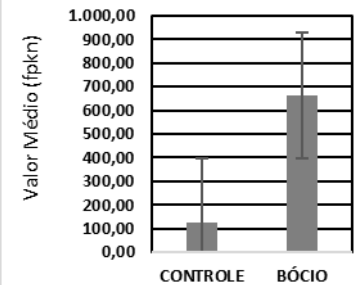


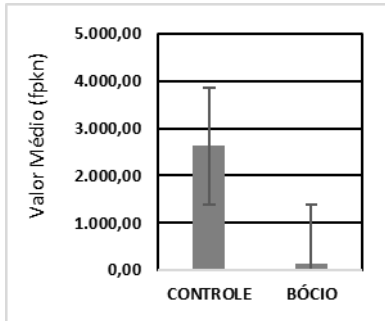
Figura 27 – Gráfico *Heat map* comparando os “*fold values*” dos 70 genes diferencialmente expressos em cada uma das amostras. Em verde representa-se os valores acima do valor de referência normalizada (Z score), em vermelho abaixo do valor referência e em tom escuro os valores próximo da referência. Cada coluna representa uma amostra e cada linha representa um gene.

Os resultados mostraram que há diferença entre os perfis de expressão gênica dos tecidos com doença em comparação com um controle com ausência de doença, em que 70 sequencias gênicas estiveram diferencialmente expressas, sendo 61 delas codificadoras de proteínas segundo a ferramenta *Panther*, 5 pseudogenes (*Sord2p*, *Pi4kap1*, *Zbtb45p1*, *Tmsb4xp4* e *Pkd1p5*), 3 sequencias relacionadas ao pré-mRNA do fator 31 de transcrição *PRPF31* (*Rp11-1212a22.1*, *Rp11-514p8.6* e *Rp11-958n24.2*), e 1 sequência relacionada à leitura anti-senso do locus *H19* (*Ac004556.1* ou gene da proteína nucleolar HOTS) (Figura 28 A e B, p. 56-63).

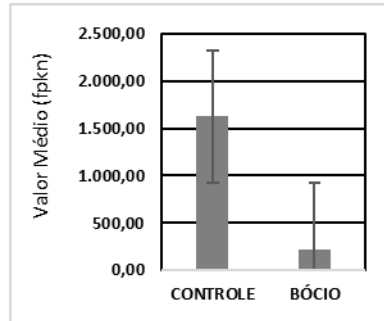
A**GENES UP-REGULATED****C4B****AC004556.1****SORD2P****C2CD4C****CPXM1****ST6GAL1****COL14A1****AC241377.2****NAPRT**

B**GENES DOWN-REGULATED**

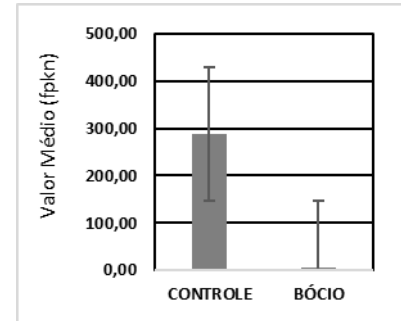
HSPA6



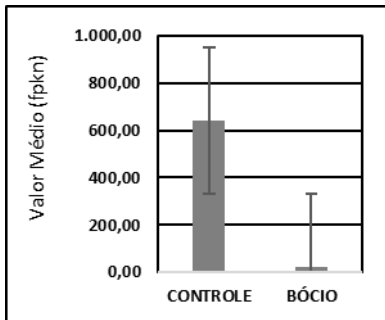
PKD1P5



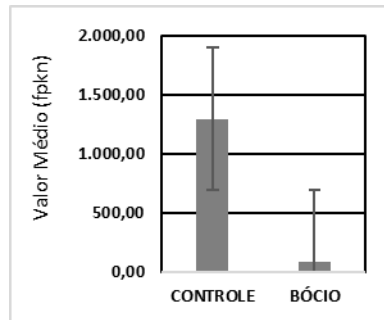
RP11-958N24.2



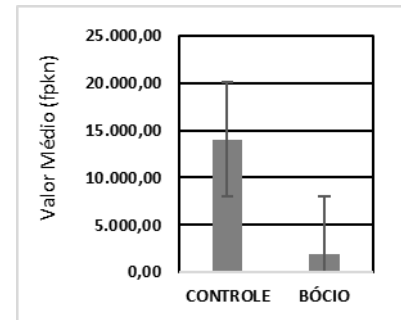
C1QL4



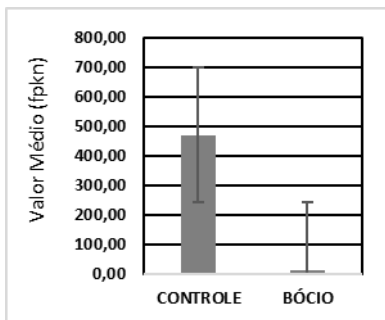
PLCD4



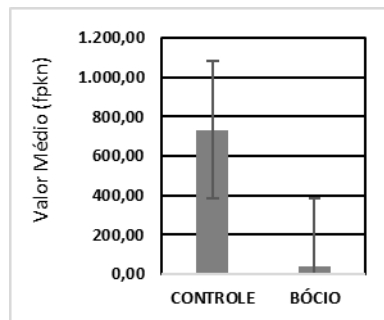
ERRFI1



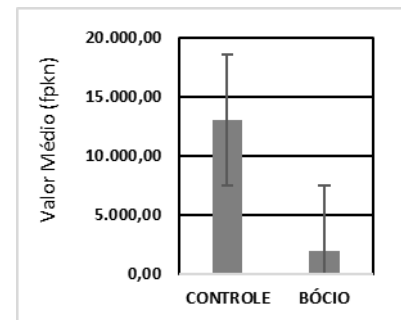
PCP4L1



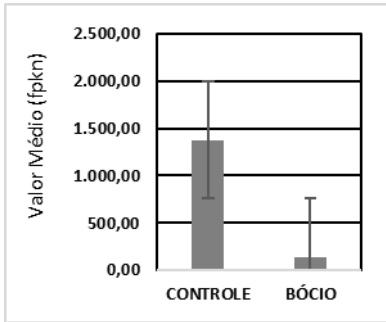
ATRNL1



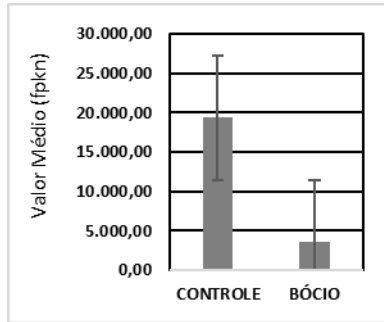
MT1H



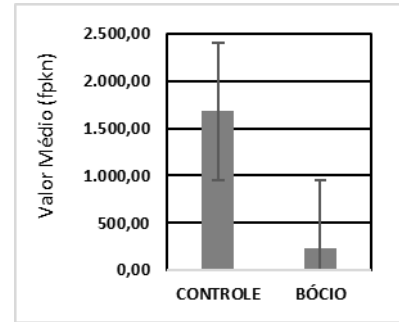
ABCA13



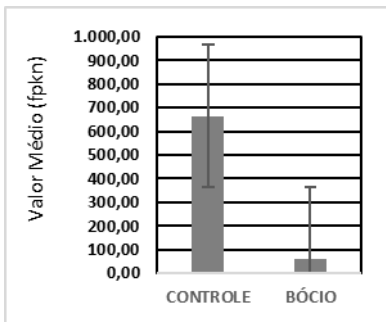
DNAJB1



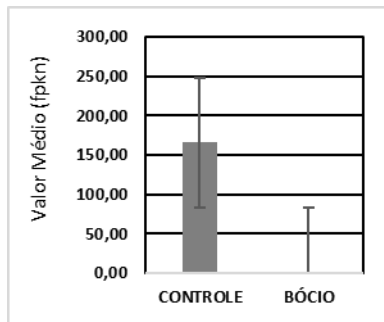
CA12



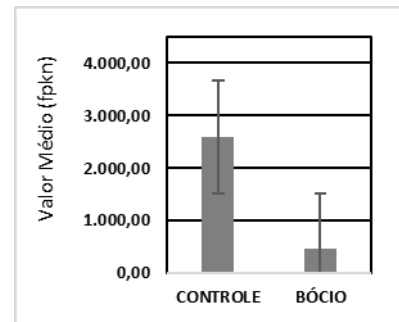
ETV4



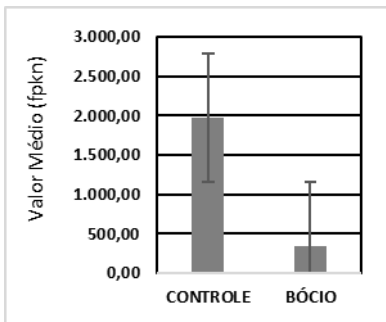
RP11-514P8.6



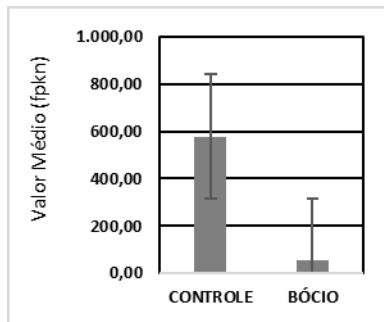
IGSF1



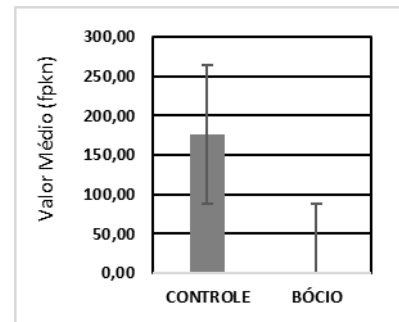
RASD1



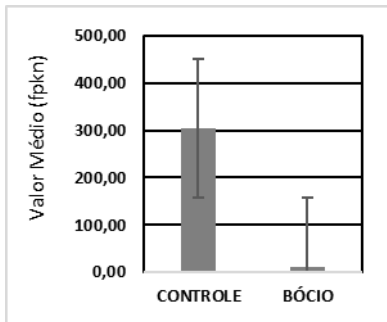
MRPL23



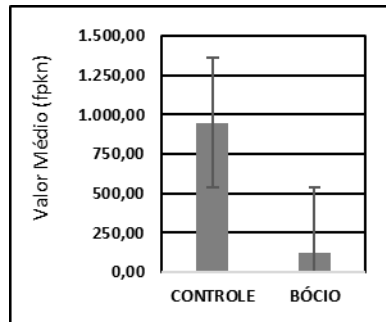
CPNE4



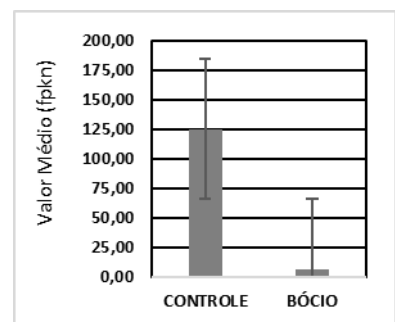
IL1RL1



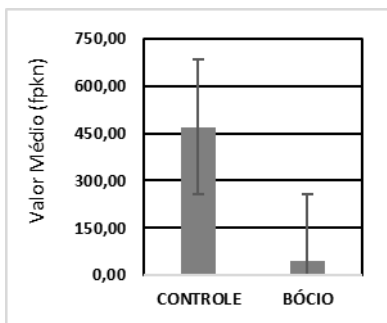
PI4KAP1



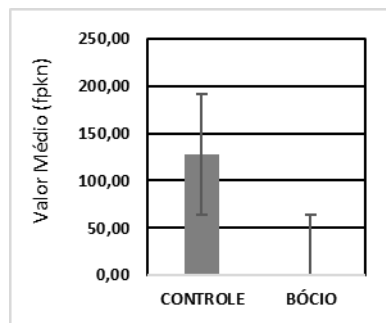
XKRX



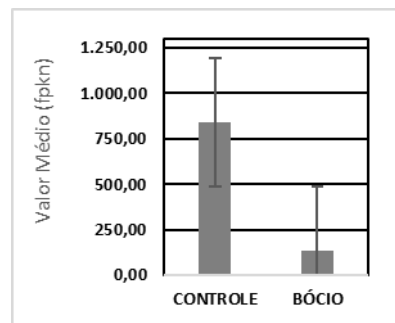
DLEC1



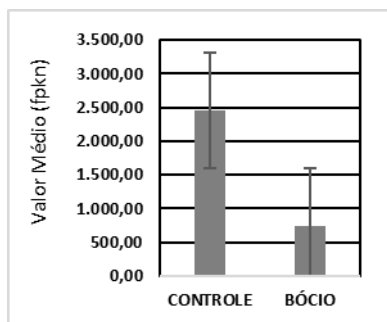
ZBTB45P1



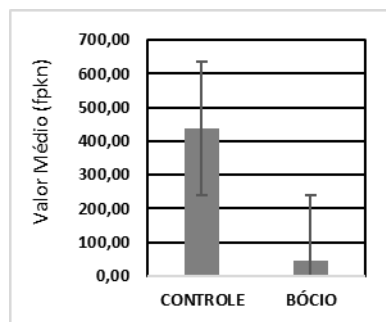
LECT1



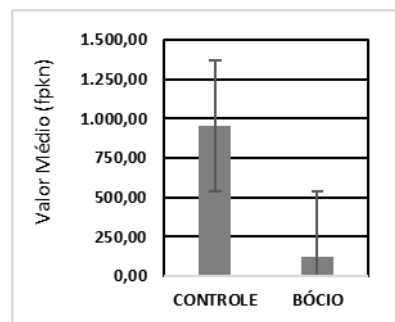
HSD17B6



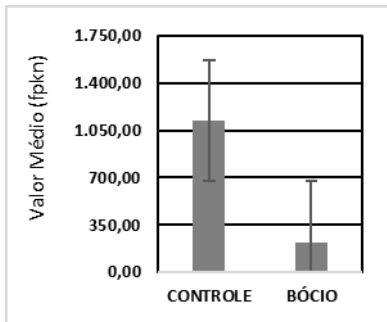
SHC3



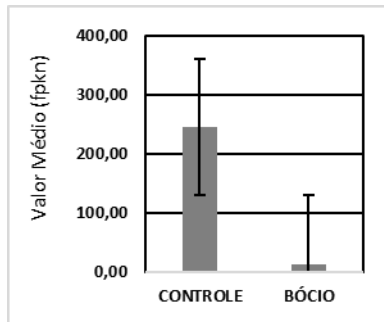
SORCS1



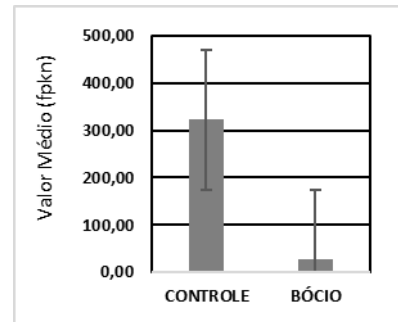
STARD9



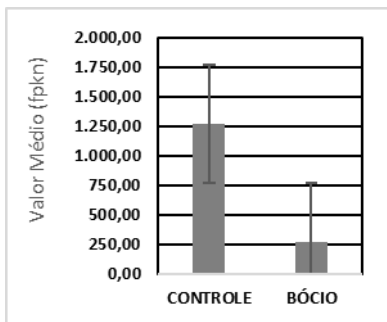
TMSB4XP4



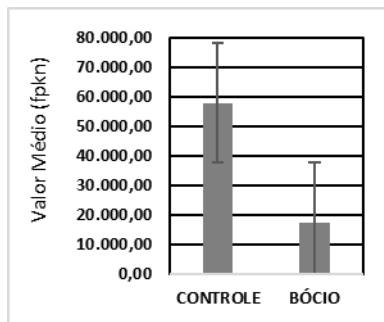
TMEM184A



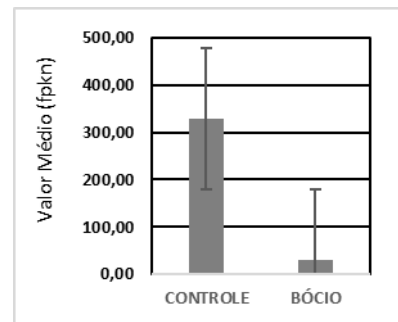
PCSK2



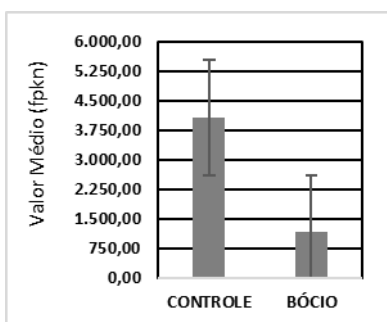
FOSB



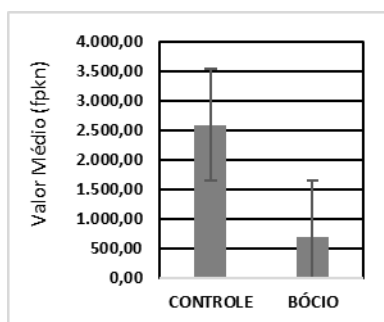
TRABD2A



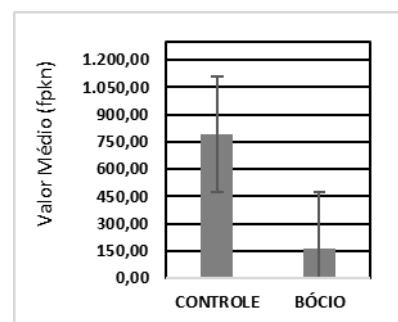
GATM



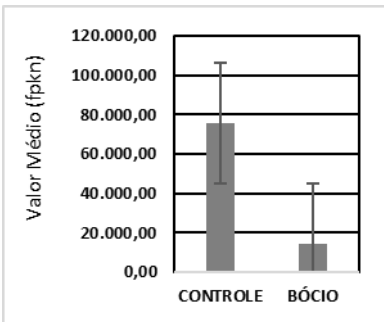
BAG3



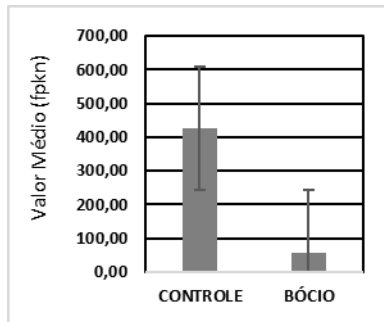
FAM105A



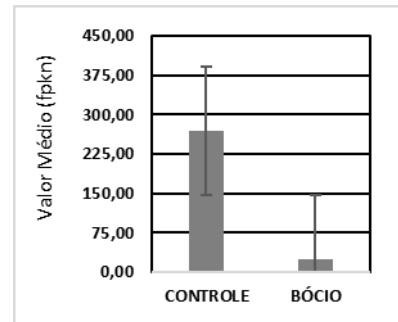
HSPA1B



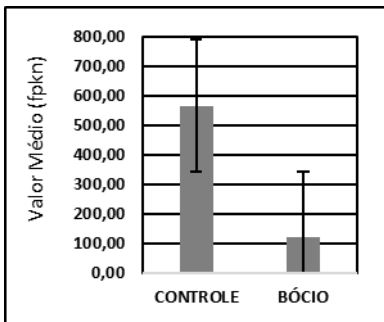
MME



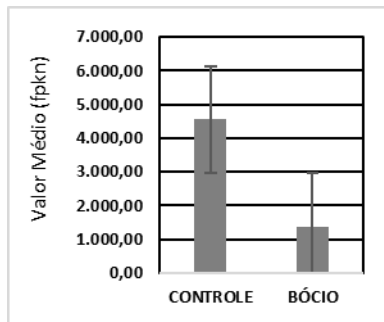
PDE4C



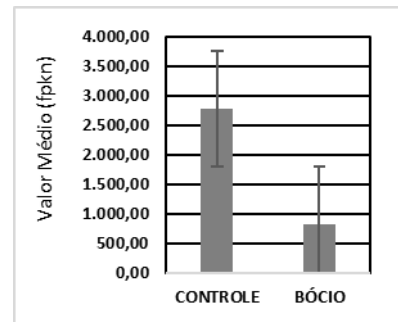
ABCC3



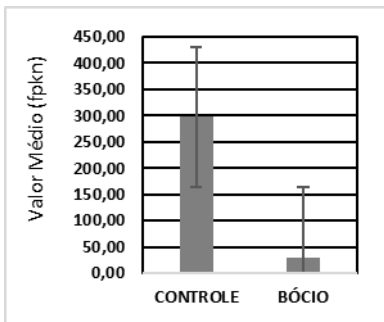
RTN4RL2



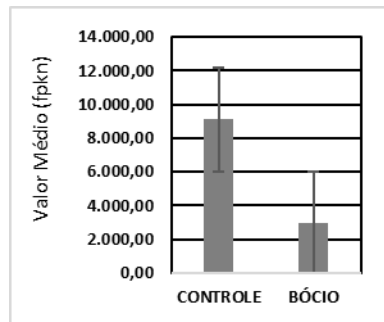
EDN3



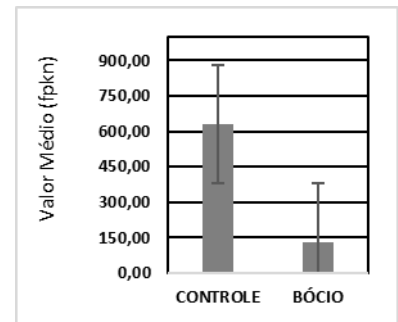
NRK



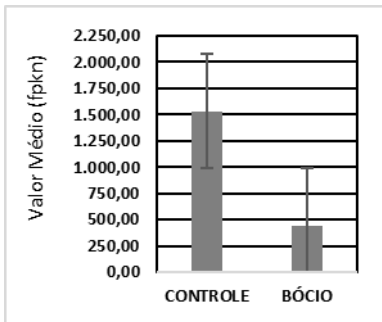
EGR2



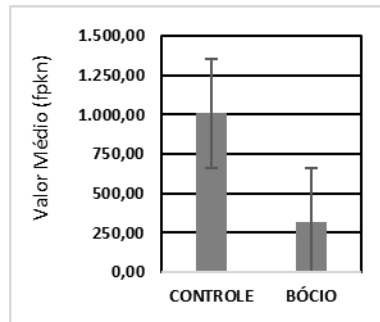
TNFRSF19



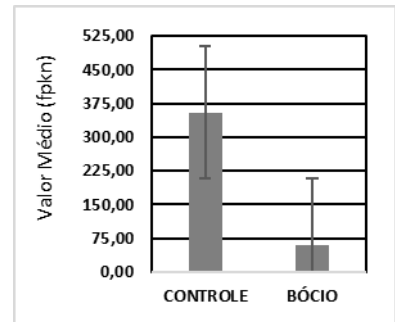
RP11-1212A22.1



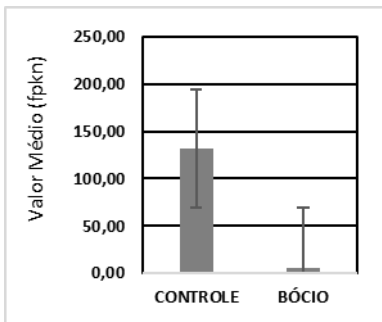
THBS1



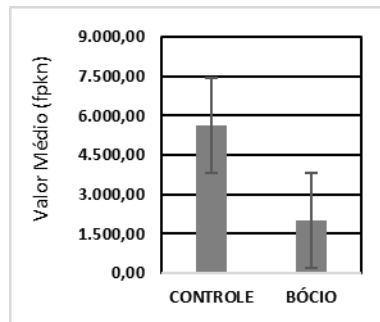
C15orf48



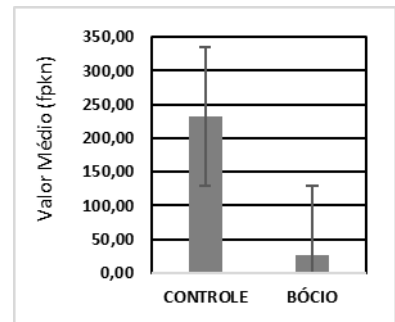
GATA5



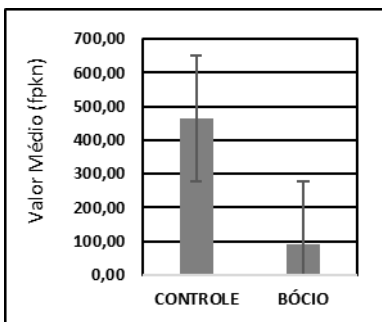
FRAS1



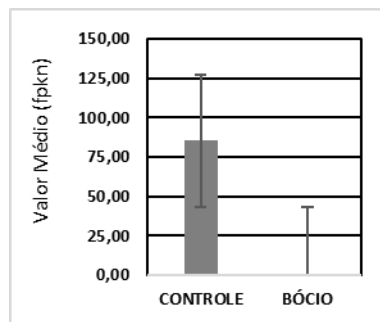
PCDH8



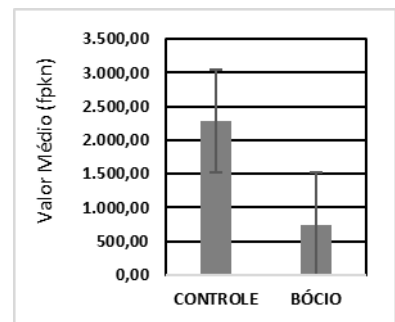
YJEFN3



SPHKAP



TBX2



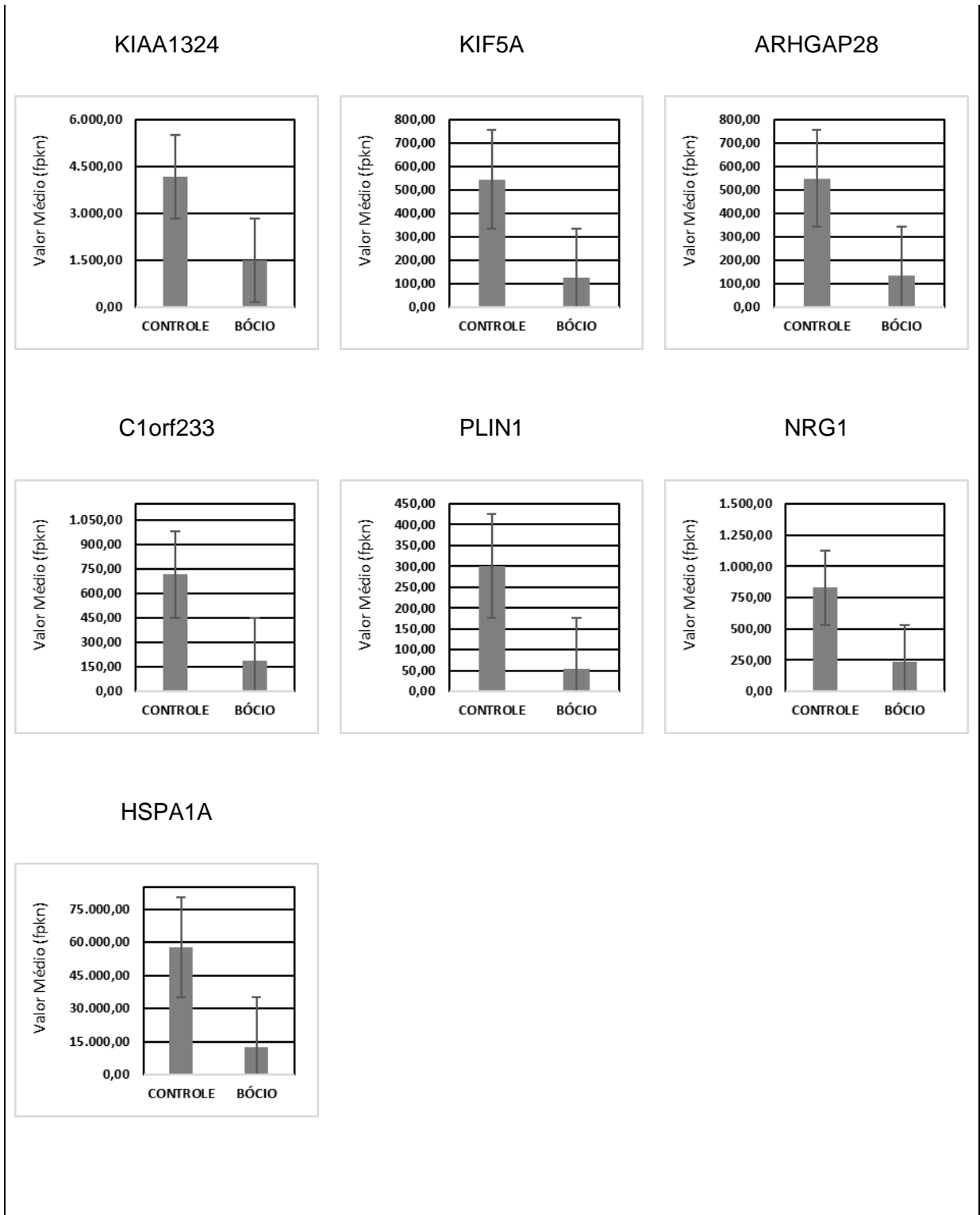


Figura 28 – Conjunto de gráficos das medidas da expressão gênica entre os genes (A) *up-regulated* e (B) *down-regulated* diferencialmente expressos no tecido com bócio vs o tecido sem doença considerando $p < 0,05$ (5%).

A análise ontológica realizada pela ferramenta *Panther* (MI et al., 2017) do *Gene Ontology Consortium* identificou a função molecular e classe proteica (Figuras 29 - 32) de 61 dos genes funcionais diferencialmente expressos (Anexo 6).

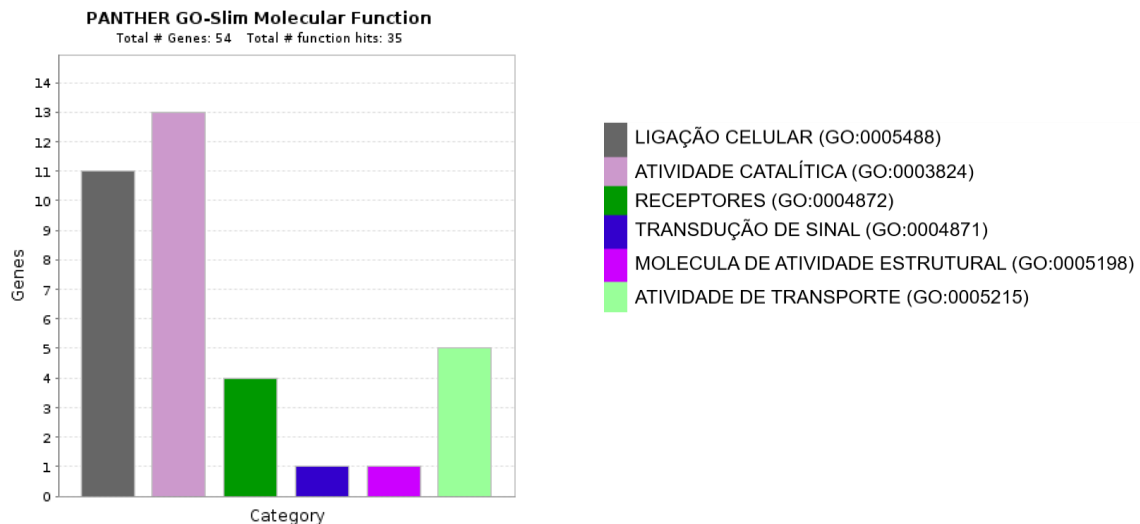


Figura 29- Função molecular dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos genes *down-regulated*: *Hspa6*, *C1ql4*, *Plcd4*, *Errfi1*, *Pcp4l1*, *Atrnl1*, *Mt1h*, *Abca13*, *Dnajb1*, *Ca12*, *Etv4*, *Igsf1*, *Rasd1*, *Mrpl23*, *Cpne4*, *Il1rl1*, *Dlec1*, *Lect1*, *Shc3*, *Sorcs1*, *Stard9*, *Tmem184a*, *Pcsk2*, *Fosb*, *Trabd2a*, *Gatm*, *Bag3*, *Fam105a*, *Hspa1b*, *Mme*, *Pde4c*, *Rtn4rl2*, *Edn3*, *Nrk*, *Egr2*, *Tnfrsf19*, *Abcc3*, *Hsd17b6*, *C15orf48*, *Gata5*, *Fras1*, *Pcdh8*, *Yjefn3*, *Sphkap*, *Tbx2*, *Kiaa1324*, *Kif5a*, *Arhgap28*, *C1orf233*, *Nrg1*, *Hspa1a*, *Thbs1*, *Plin1*, *Xkrx*

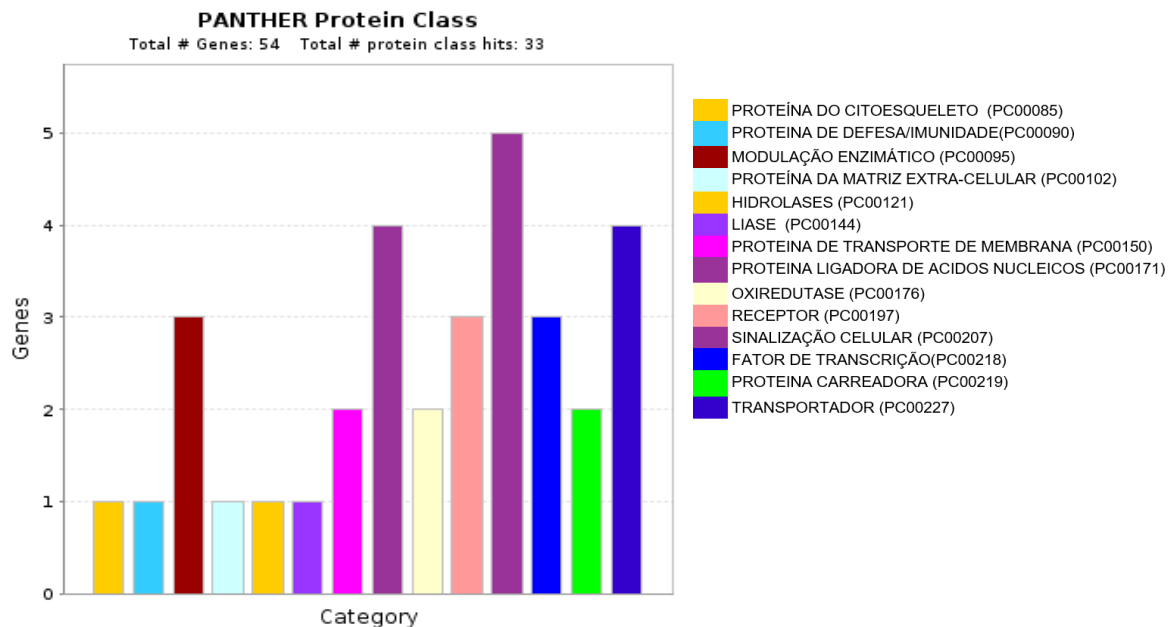


Figura 30- Classe proteica dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos genes *down-regulated*: *Hspa6*, *C1ql4*, *Plcd4*, *Errfi1*, *Pcp4l1*, *Atrnl1*, *Mt1h*, *Abca13*, *Dnajb1*, *Ca12*, *Etv4*, *Igsf1*, *Rasd1*, *Mrpl23*, *Cpne4*, *Il1rl1*, *Dlec1*, *Lect1*, *Shc3*, *Sorcs1*, *Stard9*, *Tmem184a*, *Pcsk2*, *Fosb*, *Trabd2a*, *Gatm*, *Bag3*, *Fam105a*, *Hspa1b*, *Mme*, *Pde4c*, *Rtn4rl2*, *Edn3*, *Nrk*, *Egr2*, *Tnfrsf19*, *Abcc3*, *Hsd17b6*, *C15orf48*, *Gata5*, *Fras1*, *Pcdh8*, *Yjefn3*, *Sphkap*, *Tbx2*, *Kiaa1324*, *Kif5a*, *Arhgap28*, *C1orf233*, *Nrg1*, *Hspa1a*, *Thbs1*, *Plin1*, *Xkrx*

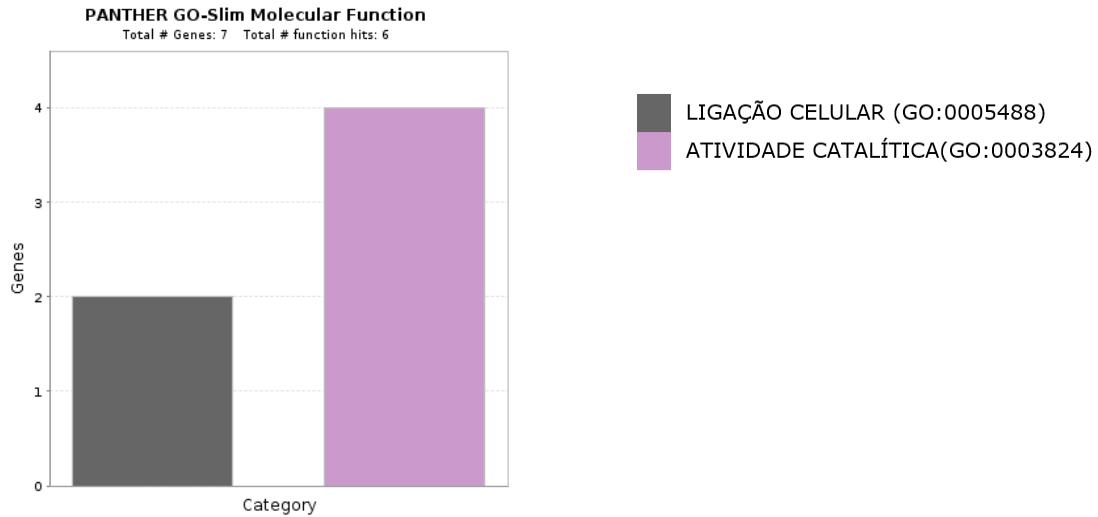


Figura 31 Função molecular dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos genes *up-regulated*: *C4b*, *C2cd4c*, *Ac241377.2*, *Cpxm1*, *Naprt*, *St6gal1*, *Col14a1*.

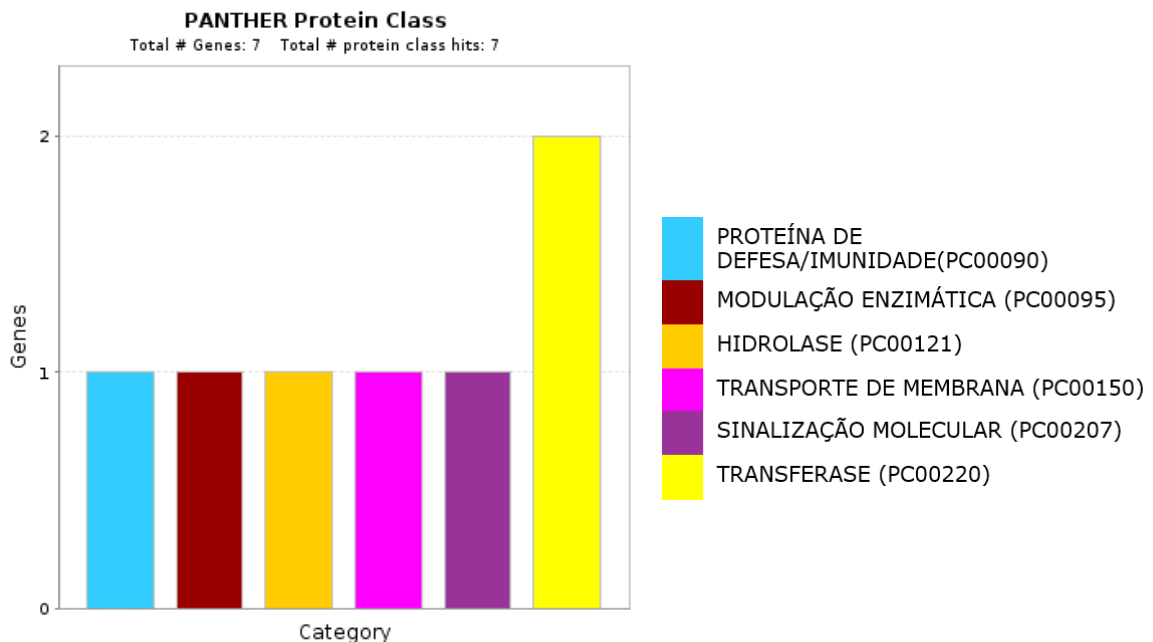


Figura 32 Classe proteica dos genes diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença) segundo *Panther* referente aos genes *up-regulated*: *C4b*, *C2cd4c*, *Ac241377.2*, *Cpxm1*, *Naprt*, *St6gal1*, *Col14a1*.

A análise de interação biológica entre estes genes (Anexo 7), identificou 423 interações possíveis utilizando a ferramenta Gene MANIA (Figura 33).

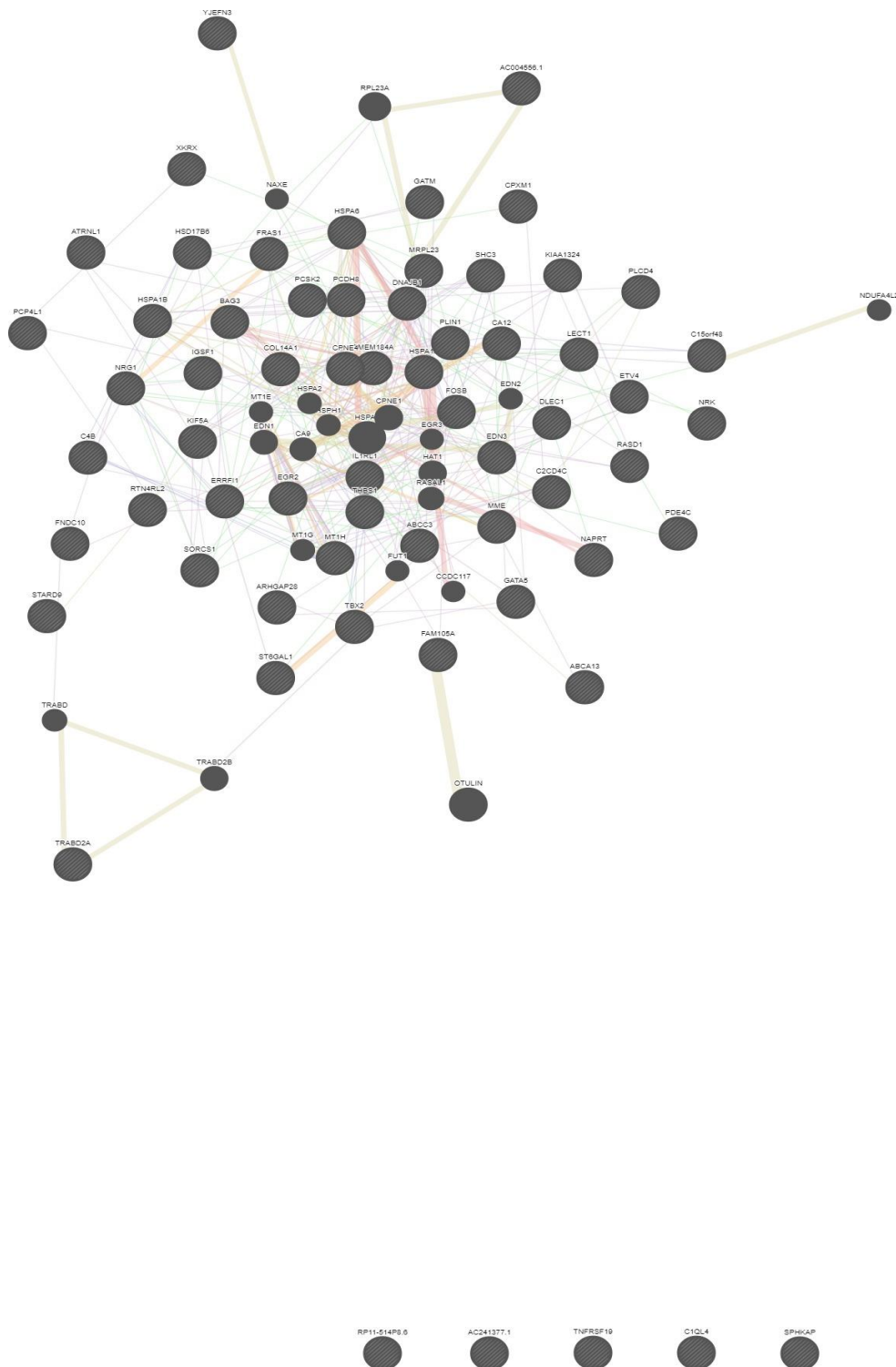


Figura 33 - Análise de interação proteína-proteína de 63 genes funcionais diferencialmente expressos em tecidos tireoideanos com bócio frente ao controle (tecido tireoideano livre de doença), utilizando o software GeneMANIA. Ao todo 423 interações gênicas (mais de 50% de co-expressão) 5 – genes: Tnf-Rsf19, Ac2413771, Rp11-514p8.6, C1ql4 E Sphkap não apresentavam interações identificadas com os demais.

6 DISCUSSÃO

Este estudo é inédito com pacientes de área endêmica para o bócio, e também quanto à metodologia utilizada de RNA-seq para o estudo comparativo entre tecidos tireoideanos doentes com bócio e com ausência de doença.

A vantagem dos estudos com NGS, são sua possível aplicação no diagnóstico, no prognóstico e no monitoramento e predição de respostas às terapias empregadas a partir do conhecimento da expressão gênica do tecido em condição de doença e em condição normal (KOTELNIKOVA et al., 2016).

A principal dificuldade deste desenho experimental é a necessidade de réplicas e um controle normal para análise de expressão gênica diferencial, preferencialmente com as mesmas origens para ambas as condições. Nesta tese em especial por empecilhos éticos na coleta de tireoides sem a indicação de tireoidectomias e a impossibilidade de garantir a normalidade de tecidos oriundos de tireoides removidas por doença, utilizou-se um controle negativo livre ou ausente de doença identificado no histopatológico pós-operatório, semelhante ao que fazem trabalhos similares com este mesmo desenho (RIA et al., 2016).

As publicações médicas para as doenças cirúrgicas da tireoide são concentradas na busca de biomarcadores para os carcinomas tireoideanos (LINKOV et al., 2008; NAGAR et al., 2014). Trabalhos iniciais com imunohistoquímica e *microarray* comparando os perfis de expressão entre amostras de tecido normal, com bócio, com adenoma e carcinoma identificou padrões diferentes entre as doenças, mas identificou similaridade entre os grupos de genes do tecido com bócio e os com carcinoma papilífero, o que poderia explicar o aumento da incidência de carcinomas em áreas bociogênicas ou o entendimento de um fator inicial comum na tumorigênese (STOLF et al., 2005).

A literatura é vasta e ao mesmo tempo rica em lacunas sobre a origem do bócio, certamente envolvendo fatores epigenéticos, hereditariedade e classicamente a deficiência de iodo, além das deficiências de ferro e selênio na dieta, a exposição a tipos alimentares ricos em flavanóides e cianogênicos como a mandioca, condições

que a partir da cronicidade levariam a mutações e ao aparecimento de nódulos na glândula (KNOBEL, 2016; KNOBEL; MEDEIROS-NETO, 2004).

Ocorre que na prática médica, diante de nodulações na tireoide, resta ao médico e ao doente o questionamento da presença ou não de lesão maligna (CHOW et al., 2001), e atualmente o padrão ouro de investigação pré-operatória são as punções com agulha fina e a classificação citológica de Bethesda, muitas vezes com necessidade de repetições e com sensibilidade e concordância com o histopatológico variáveis, além de ser inconclusiva em até 30 % dos casos (KIHARA et al., 2013; LINKOV et al., 2008).

Entender quem é o paciente com bócio que deve ser submetido à cirurgia ou a quais deles deve-se oferecer novas estratégias de terapia ou acompanhamento clínico justifica a necessidade de aprofundamento das características moleculares em especial da expressão gênica de bócios.

Os resultados desta tese mostraram haver diferenças na expressão gênica em pelo menos 70 sequencias. Os genes *down-regulated* no bócio relacionam-se a proteínas de diversas vias moleculares, com destaque as da fosfolipase C (PLCD4), vias de apoptose (TNFRSF19), proteínas do choque térmico (HSPA1A, HSPA6), fatores de crescimento (SHC3, NRG1), vias do protooncogene p53 (THBS1) e vias de reparo celular das chaperonas (BAG3), por outro lado, *up-regulated* nos tecidos com bócio as vias inflamatórias (COL14A) e do sistema complemento (C4B), além da presença exclusivamente no bócio de um transcrito anti-senso do *locus H19*, que codifica a proteína nucleolar HOTS (ONYANGO; FEINBERG, 2011).

Esses achados assemelham-se a características de doenças tumorais com a redução de sistemas de apoptose e de reparo celular, com aumento de atividade inflamatória na presença de produtos de um *locus* pró-tumoral, neste caso a proteína nucleolar HOTS, que junto do lncRNA H19 seriam possíveis indutores de lesões cancerosas de mama, fígado, rim e pulmão (DUAN; ZHANG; PENG, 2016; ONYANGO; FEINBERG, 2011; POPE et al., 2017).

Não é clara a ação do gene *H19* e seus produtos na iniciação de lesões tumorais e hiperplásicas, mas o antagonismo do lncRNA *H19* com *p53*, e a atividade de um de seus produtos gênicos o *miR-675* na promoção de instabilidade celular e cromossômica são bem descritos, bem como sua hiperexpressão na presença de fatores externos como hipóxia (POPE et al., 2017).

O papel de HOTS é pouco conhecido na literatura atual, a própria existência de uma proteína funcional do gene *H19* era ignorada até 2011, ocasião da publicação de Onyango (2011) que sugeriu HOTS como supressor tumoral no tumor de Wilms, em que a hiperexpressão nucleolar de HOTS desempenharia atividade de reforço no *imprinting* dos demais produtos conhecidamente oncogênicos do gene *H19*.

Cada vez mais presentes nas publicações dos últimos anos, os produtos senso do gene *H19* como o lncRNA H19 parecem desempenhar um papel central na tumorigenese de lesões neoplásicas gástricas, da mama, da vesícula, do cólon e da tireoide embora ainda sejam pouco compreendidas suas múltiplas funções e mecanismos de atividade (DUAN; ZHANG; PENG, 2016; LI; WANG, 2012; LIN et al., 2017).

Possivelmente um balanço entre os produtos dos transcritos senso (lncRNA) e anti-senso (HOTS) do gene *H19* estariam relacionados à regulação da homeostase celular (RAVEH et al., 2015), no caso do bócio, a hiperexpressão de HOTS sugeriria um ambiente tireoideano pró-tumoral.

Por tratar-se de uma doença negligenciada molecularmente em comparação aos carcinomas tireoideanos, a literatura atual não apresenta dados referentes ao perfil de expressão gênica dos bócios. Além disso, sobre a proteína HOTS, até dezembro de 2017, todas as 27 publicações indexadas no *Medline* utilizando os descritores “HOTS” e “H19” versam exclusivamente sobre carcinomas e os produtos senso do gene *H19*.

Não existem trabalhos com NGS de bócio com resultados semelhantes aos descritos nesta tese. Em carcinomas tireoideanos, além do lncRNA H19 outros lncRNA são discutidos na regulação gênica da progressão da doença, como o *Ptcsc3* (FAN et al., 2013) , também na metástase local o *Xloc051122* e o *Xloc006074* (LIYANARACHCHI et al., 2016) e ainda como possível alvo para terapias pró-apoptóticas de carcinomas o *Pandar* (LI et al., 2017).

A presença do bócio como fator de risco para o carcinoma da tireoide permanece incerta. A fisiopatologia hipotética do bócio uni ou multinodular depende de um estímulo externo como a deficiência de iodo, promovendo o desenvolvimento de uma lesão não tóxica, nodular, que pela heterogeneidade natural entre tireócitos resultaria em folículos tireoidianos de diferentes formas e tamanhos, bem como diferente sensibilidade ao estímulo do TSH. Contudo, com o passar do tempo a persistência do crescimento do bócio passa a ser independente do TSH por razões

ainda pouco claras e a replicação celular leva ao aumento da mutagênese com posterior hiperplasia (HEGEDÜS; BRIX; PASCHKE, 2009; KNOBEL, 2016).

Recentemente, foi demonstrado que uma mesma lesão tireoideana papilifera diagnosticada histopatologicamente tem comportamento de expressão proteica diferente se o doente tiver história de bócio multinodular previamente ao diagnóstico da neoplasia (ABDULLAH et al., 2016).

Outros trabalhos mostraram a importância das proteínas de membrana no desenvolvimento de doenças hiperplásicas e neoplásicas da tireoide, em especial destaque as conexinas e às aquaporinas (DE MONTGOLFIER; AUDET; CYR, 2011; JIANG et al., 2010; NIU et al., 2012). Neste estudo, não identificamos diferença no perfil de expressão de conexinas nem aquaporinas, mas estiveram *down-regulated* no bócio as proteínas de membrana STARD9 (apolipoproteína) e CPNE4.

Mais trabalhos devem prosseguir na elucidação molecular do bócio, e ainda, no melhor entendimento das funções de HOTS e dos RNA não codificantes nas doenças nodulares malignas ou não da tireoide.

7 CONCLUSÃO

Entender a doença nas múltiplas dimensões da matéria tem exigido das ciências da saúde a adoção e a criação de diferentes estratégias de investigação e entendimento das doenças.

Sendo assim, não é difícil entender que nas doenças cirúrgicas a proximidade entre o bisturi e a molécula nunca esteve tão encurtada. Isso exige de cirurgiões e clínicos o conhecimento das características moleculares das doenças e suas implicações na prática médica.

Pretendeu-se com esta tese analisar a diferença de expressão gênica entre tecidos com bócio e tecido tireoideano com ausência de doença, e foi possível demonstrar 70 sequencias diferencialmente expressas entre as duas condições estudadas, com níveis de co-expressão superior a 59% e 423 possibilidades de interação proteína-proteína entre os 63 genes funcionais diferencialmente expressos.

Os resultados ontológicos permitiram ainda caracterizar o bócio molecular como similar aos achados na gênese de lesões tumorais neoplásicas: 1) redução de atividade de reparo celular; 2) redução de atividade de vias apoptóticas; 3) aumento de atividade inflamatória; 4) expressão de HOTS, produto do gene imprintado *H19* com possível atividade supressora tumoral.

A presença de produtos do gene *H19*, hiperexpressos no bócio, uma doença não maligna com diferentes formas de apresentação em regiões endêmicas (bócios de pequeno e de grande volume), contribui para o entendimento da tumorigenese do bócio e seu possível papel como fator de risco para lesões malignas em regiões endêmicas, bem como um possível candidato para marcador molecular, a proteína HOTS.

Para a confirmação da hipótese do bócio como um estado pró-tumoral da tireoide serão necessários mais estudos em áreas endêmicas, com mais réplicas para análise em *NGS* e melhor entendimento da função dos ncRNAs para o desenvolvimento da doença.

8 REFERÊNCIAS BIBLIOGRÁFICAS

ABDULLAH, M. I. *et al.* Tissue and serum samples of patients with papillary thyroid cancer with and without benign background demonstrate different altered expression of proteins. **PeerJ**, 13 set. 2016. v. 4, p. 2450.

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome biology**, 2010. v. 11, n. 10, p. 106.

BAN, Y. *et al.* Proteomic Profiling of Thyroid Papillary Carcinoma. **Journal of Thyroid Research**, 2012. v. 2012, p. 1–7.

BELYANTSEVA, I. A. *et al.* Water permeability of cochlear outer hair cells: characterization and relationship to electromotility. **The Journal of Neuroscience**, 2000. v. 20, n. 24, p. 8996–9003.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, 1 ago. 2014. v. 30, n. 15, p. 2114–2120.

BOTELHO, J. B. **Tireoide: história, embriologia, anatomia e cirurgia**. 1. ed. Manaus: EDUA, 2012. p. 99-113.

_____. Le goitre endémique amazonien : rapports anatomiques de la glande thyroïde avec les structures vasculo-nerveuses contigües et les parathyroïdes. 2013. v. 4, n. 12.p. 4-12.

BRENNER, M.; HOISTAD, D. L.; HAIN, T. C. Prevalence of thyroid dysfunction in patients with Meniere's disease. **Archives of Otolaryngology–Head & Neck Surgery**, 2004. v. 130, n. 2, p. 226–228.

BRUMATTI GONÇALVES, K. C.; OTHERS. **Estudo de bioinformática aplicado à análise de expressão gênica utilizando dados oriundos de sequenciamento por tecnologia de “ Next-Generation” em animais controle e em modelos de epilepsia do lobo temporal mesial**. 2015. Dissertação de Mestrado (UNICAMP). Disponível em: <<http://repositorio.unicamp.br/handle/REPOSIP/312727>>. Acesso em: 27 jun. 2017.

BRUNKOW, M. E.; TILGHMAN, S. M. Ectopic expression of the H19 gene in mice causes prenatal lethality. **Genes & Development**, 1991. v. 5, n. 6, p. 1092–1101.

CHA, Y. J.; KOO, J. S. Next-generation sequencing in thyroid cancer. **Journal of Translational Medicine**, dez. 2016. v. 14, n. 1. Disponível em: <<http://translational-medicine.biomedcentral.com/articles/10.1186/s12967-016-1074-7>>. Acesso em: 1º abr. 2017.

CHEN, Y.-H. *et al.* The claudin family of proteins in human malignancy: a clinical perspective. **Cancer Management and Research**, nov. 2013. p. 367.

CHOI, Y. J. *et al.* Thyroid Nodule with Benign Cytology: Is Clinical Follow-Up Enough? **PLoS ONE**, 24 maio. 2013. v. 8, n. 5, p. e63834.

CHOW, L. S. *et al.* Nondiagnostic thyroid fine-needle aspiration cytology: management dilemmas. **Thyroid**, 2001. v. 11, n. 12, p. 1147–1151.

CHU, Y.; COREY, D. R. RNA sequencing: platform selection, experimental design, and data interpretation. **Nucleic acid therapeutics**, 2012. v. 22, n. 4, p. 271–274.

COLIN, I. M. *et al.* Recent Insights into the Cell Biology of Thyroid Angiofollicular Units. **Endocrine Reviews**, abr. 2013. v. 34, n. 2, p. 209–238.

COOPER, D.; ROSS, D. Diagnostic approach to and treatment of goiter in adults.pdf. **up to date**, [S.l.], 2013. Pesquisa Médica. Disponível em: <www.uptodate.org>. Acesso em: 25 jun. 2013.

CORDAS, E. A. *et al.* Thyroid Hormone Receptors Control Developmental Maturation of the Middle Ear and the Size of the Ossicular Bones. **Endocrinology**, mar. 2012. v. 153, n. 3, p. 1548–1560.

CORRÊA FILHO, H. R. *et al.* Inquérito sobre a prevalência de bócio endêmico no Brasil em escolares de 6 a 14 anos: 1994 a 1996. **Rev Panam Salud Pública**, 2002. v. 12, p. 317–26.

DEGROOT, L. J.; STANBURY, J. B. **The thyroid and its diseases**. 4. ed. [S.l.]: Wiley Biomedical Division, 1975. v. 1.

DOMINGUEZ, C. *et al.* Altered Expression of the Gap Junction Protein Connexin43 Is Associated with Papillary Thyroid Carcinomas When Compared with Other Noncancer Pathologies of the Thyroid. **Thyroid**, out. 2011. v. 21, n. 10, p. 1057–1066.

DUAN, C.; ZHANG, C.; PENG, Z. Functions and mechanisms of long noncoding RNAs in lung cancer. **OncoTargets and Therapy**, jul. 2016. v. Volume 9, p. 4411–4424.

EDDY, S. R. Non-coding RNA genes and the modern RNA world. **Nature Reviews Genetics**, 2001. v. 2, n. 12.

FAN, M. *et al.* A long non-coding RNA, PTCSC3, as a tumor suppressor and a target of miRNAs in thyroid cancer cells. **Experimental and Therapeutic Medicine**, abr. 2013. v. 5, n. 4, p. 1143–1146.

GARREL, R.; SOPHIE, P. **Pathologies chirurgicales de la glande thiroïd**. Paris: Société Française d'oto-rhino-laryngologie et chirurgie de la face et du cou, 2012. v. único.p.13 - 221.

GRAF, H. Therapy of Endocrine Disease: Recombinant human TSH and radioactive iodine therapy in the management of benign multinodular goiter. **European Journal of Endocrinology**, 30 dez. 2014. v. 172, n. 2, p. 47–52.

HAQUE, A. *et al.* A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. **Genome Medicine**, dez. 2017. v. 9, n. 1. Disponível em: <<http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-017-0467-4>>. Acesso em: 24 set. 2017.

HEGEDÜS, L.; BRIX, T. H.; PASCHKE, R. Etiology of simple goiter. **Thyroid**, 2009. v. 19, n. 3, p. 209–211.

HILGERT, N.; SMITH, R. J.; VAN CAMP, G. Function and expression pattern of nonsyndromic deafness genes. **Current molecular medicine**, 2009. v. 9, n. 5, p. 546.

HORAK, P.; FRÖHLING, S.; GLIMM, H. Integrating next-generation sequencing into clinical oncology: strategies, promises and pitfalls. **ESMO open**, 2016. v. 1, n. 5, p. 94.

HOU, J. The yin and yang of claudin-14 function in human diseases: Claudin-14 function. **Annals of the New York Academy of Sciences**, jul. 2012. v. 1258, n. 1, p. 185–190.

HUTCHINS, J. R. A. What's that gene (or protein)? Online resources for exploring functions of genes, transcripts, and proteins. **Molecular Biology of the Cell**, 15 abr. 2014. v. 25, n. 8, p. 1187–1201.

JIANG, X. *et al.* Differential expression of connexin 43 in human autoimmune thyroid disease. **Acta Histochemica**, maio. 2010. v. 112, n. 3, p. 278–283.

JIANG, Z. *et al.* Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. **Cellular and Molecular Life Sciences**, set. 2015. v. 72, n. 18, p. 3425–3439.

KAN, M.; SHUMYATCHER, M.; HIMES, B. E. Using omics approaches to understand pulmonary diseases. **Respiratory Research**, dez. 2017. v. 18, n. 1. Disponível em: <<http://respiratory-research.biomedcentral.com/articles/10.1186/s12931-017-0631-9>>. Acesso em: 24 set. 2017.

KANT, R.; RAVI, B. The Surgeon and Molecular Knowledge. **Indian Journal of Surgery**, jun. 2013. v. 75, n. 3, p. 167–169.

KIHARA, M. *et al.* Evaluation of cytologically benign solitary thyroid nodules by ultrasonography: A retrospective analysis of 1877 cases. **Auris Nasus Larynx**, jun. 2013. v. 40, n. 3, p. 308–311.

KIM, J. G. *et al.* Thyroid Transcription Factor-1 Facilitates Cerebrospinal Fluid Formation by Regulating Aquaporin-1 Synthesis in the Brain. **Journal of Biological Chemistry**, 18 maio. 2007. v. 282, n. 20, p. 14923–14931.

KNOBEL, M. Etiopathology, clinical features, and treatment of diffuse and multinodular nontoxic goiters. **Journal of Endocrinological Investigation**, abr. 2016. v. 39, n. 4, p. 357–373.

KNOBEL, M.; MEDEIROS-NETO, G. Disorders associated to chronic iodine deficiency. **Arquivos Brasileiros de Endocrinologia & Metabologia**, 2004. v. 48, n. 1, p. 53–61.

KOTELNIKOVA, E. A. *et al.* Practical aspects of NGS-based pathways analysis for personalized cancer science and medicine. **Practical aspects of NGS-based pathways analysis for personalized cancer science and medicine**, 2016. v. 7, n. 32, p. 52493–52509.

KURUKUTI, S. *et al.* CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. **Proceedings of the national academy of sciences**, 2006. v. 103, n. 28, p. 10684–10689.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, 4 mar. 2012. v. 9, n. 4, p. 357–359.

LI, B., DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC bioinformatics**, 2011. v. 12, n. 1, p. 323.

LI, X., WANG, Z. The role of noncoding RNA in thyroid cancer. **Gland surgery**, 2012. v. 1, n. 3, p. 146.

LI, Z. *et al.* Knockdown of lncRNA-PANDAR suppresses the proliferation, cell cycle and promotes apoptosis in thyroid cancer cells. **EXCLI journal**, 2017. v. 16, p. 354.

LICHTENBERGER-GESLIN, L. *et al.* Factors Associated With Hearing Impairment in Patients With Congenital Hypothyroidism Treated Since the Neonatal Period: A National Population-Based Study. **The Journal of Clinical Endocrinology & Metabolism**, set. 2013. v. 98, n. 9, p. 3644–3652.

LIN, Y. *et al.* Genetic variants in long noncoding RNA H19 contribute to the risk of breast cancer in a southeast China Han population. **OncoTargets and Therapy**, set. 2017. v. Volume 10, p. 4369–4378.

LINKOV, F. *et al.* Multiplex analysis of cytokines as biomarkers that differentiate benign and malignant thyroid diseases. **Proteomics - Clinical Applications**, dez. 2008. v. 2, n. 12, p. 1575–1585.

LIYANARACHCHI, S. *et al.* Genome-Wide Expression Screening Discloses Long Noncoding RNAs Involved in Thyroid Carcinogenesis. **The Journal of Clinical Endocrinology & Metabolism**, nov. 2016. v. 101, n. 11, p. 4005–4013.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature Reviews Genetics**, 7 set. 2011. v. 12, n. 10, p. 671–682.

MASINDOVA, I. *et al.* Molecular and hereditary mechanisms of sensorineural hearing loss with focus on selected endocrinopathies. **Endocrine Regulations**, 2012. v. 46, n. 3, p. 167–186.

MAZEH, H. *et al.* Differentiating benign from malignant thyroid nodules using micro ribonucleic acid amplification in residual cells obtained by fine needle aspiration biopsy. **Journal of Surgical Research**, abr. 2013. v. 180, n. 2, p. 216–221.

MEHDI, A.; RIAZALHOSSEINI, Y. Epigenome Aberrations: Emerging Driving Factors of the Clear Cell Renal Cell Carcinoma. **International Journal of Molecular Sciences**, 16 ago. 2017. v. 18, n. 8, p. 1774.

MI, H. *et al.* PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. **Nucleic Acids Research**, 4 jan. 2017. v. 45, n. D1, p. 183–189.

MONTGOLFIER, B. DE; AUDET, C.; CYR, D. G. Regulation of the connexin 43 promoter in the brook trout testis: Role of the thyroid hormones and cAMP. **General and Comparative Endocrinology**, jan. 2011. v. 170, n. 1, p. 110–118.

MORAN, C.; CHATTERJEE, K. Resistance to thyroid hormone due to defective thyroid receptor alpha. **Best Practice & Research Clinical Endocrinology & Metabolism**, ago. 2015. v. 29, n. 4, p. 647–657.

MUTZ, K.-O. *et al.* Transcriptome analysis using next-generation sequencing. **Current Opinion in Biotechnology**, fev. 2013. v. 24, n. 1, p. 22–30.

NAGAR, S. *et al.* Evaluation of genetic biomarkers for distinguishing benign from malignant thyroid neoplasms. **The American Journal of Surgery**, abr. 2014. v. 207, n. 4, p. 596–601.

NAKANO, Y. *et al.* A Claudin-9–Based Ion Permeability Barrier Is Essential for Hearing. **PLoS Genetics**, 21 ago. 2009. v. 5, n. 8, p. 1000610.

NELSON, D. L.; COX, Mi. M. **Princípios de bioquímica de Lehninger**. 5. ed. Porto Alegre: Artmed, 2011p. 1021 - 1162.

NG, L.; KELLEY, M. W.; FORREST, D. Making sense with thyroid hormone—the role of T3 in auditory development. **Nature Reviews Endocrinology**, 26 mar. 2013. v. 9, n. 5, p. 296–307.

NIKIFOROV, Y. E. *et al.* Impact of Mutational Testing on the Diagnosis and Management of Patients with Cytologically Indeterminate Thyroid Nodules: A Prospective Analysis of 1056 FNA Samples. **The Journal of Clinical Endocrinology & Metabolism**, nov. 2011. v. 96, n. 11, p. 3390–3397.

NIU, D. *et al.* Differential Expression of Aquaporins and Its Diagnostic Utility in Thyroid Cancer. **PLoS ONE**, 10 jul. 2012. v. 7, n. 7, p. 40770.

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. **Proceedings of the National Academy of Sciences**, 4 out. 2011. v. 108, n. 40, p. 16759–16764.

POPE, C. *et al.* Targeting H19, an Imprinted Long Non-Coding RNA, in Hepatic Functions and Liver Diseases. **Diseases**, 8 mar. 2017. v. 5, n. 1, p. 11.

PROST, G. *et al.* Connexin-32 acts as a downregulator of growth of thyroid gland. **AJP: Endocrinology and Metabolism**, 4 dez. 2007. v. 294, n. 2, p. 291–299.

RAVEH, E. *et al.* The H19 Long non-coding RNA in cancer initiation, progression and metastasis – a proposed unifying theory. **Molecular Cancer**, dez. 2015. v. 14, n. 1. Disponível em: <<http://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-015-0458-2>>. Acesso em: 15 out. 2017.

RIA, R. *et al.* Gene expression profiling of normal thyroid tissue from patients with thyroid carcinoma. **Oncotarget**, 2016. v. 7, n. 20, p. 29677.

ROSÁRIO, P. W. *et al.* Nódulo tireoidiano e câncer diferenciado de tireoide: atualização do consenso brasileiro. **Arquivos brasileiros de endocrinologia & metabologia**. São Paulo. Vol. 57, n. 4 (jul. 2013), p. 240-264, 2013. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/101528>>. Acesso em: 28 fev. 2016.

SHEN, T. *et al.* Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. **Frontiers in Genetics**, 17 jun. 2015. v. 6. Disponível em: <<http://journal.frontiersin.org/Article/10.3389/fgene.2015.00215/abstract>>. Acesso em: 2 out. 2017.

SIMIONESCU-BANKSTON, A.; KUMAR, A. Noncoding RNAs in the regulation of skeletal muscle biology in health and disease. **Journal of Molecular Medicine**, ago. 2016. v. 94, n. 8, p. 853–866.

SMITH, T. J. Neck and thyroid examination. 1990. Disponível em: <<http://www.ncbi.nlm.nih.gov/books/NBK244/>>. Acesso em: 28 fev. 2016.

SOFIADIS, A. *et al.* Proteomic profiling of follicular and papillary thyroid tumors. **European Journal of Endocrinology**, 1 abr. 2012. v. 166, n. 4, p. 657–667.

STOLF, B. S. *et al.* Expression profile of malignant and non-malignant diseases of the thyroid gland reveals altered expression of a common set of genes in goiter and papillary carcinomas. **Cancer Letters**, set. 2005. v. 227, n. 1, p. 59–73.

TANG, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. **Nature Methods**, maio. 2009. v. 6, n. 5, p. 377–382.

TRAPNELL, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, 1 mar. 2012. v. 7, n. 3, p. 562–578.

TWYFFELS, L. *et al.* Pendrin: the thyrocyte apical membrane iodide transporter? **Cellular physiology and biochemistry**, 2011. v. 28, n. 3, p. 491. ROSE

UGURLUOGLU, C. *et al.* Fine needle aspiration biopsy of thyroid nodules: cytologic and histopathologic correlation of 1096 patients. **International journal of clinical and experimental pathology**, 2015. v. 8, n. 11, p. 14800.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, jan. 2009. v. 10, n. 1, p. 57–63.

WARDE-FARLEY, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. **Nucleic Acids Research**, 1 jul. 2010. v. 38, n. suppl_2, p. 214–220.

WHEELER, D. L. Database resources of the National Center for Biotechnology Information: update. **Nucleic Acids Research**, 1 jan. 2004. v. 32, n. 90001, p. 35–40.

WINTER, H. *et al.* Deafness in TR Mutants Is Caused by Malformation of the Tectorial Membrane. **Journal of Neuroscience**, 25 fev. 2009. v. 29, n. 8, p. 2581–2587.

WONG, A. K. *et al.* IMP 2.0: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. **Nucleic Acids Research**, 1 jul. 2015. v. 43, n. 1, p. 128–133.

ZHANG, M.; LIN, O. Molecular testing of thyroid nodules: a review of current available tests for fine-needle aspiration specimens. **Archives of Pathology & Laboratory Medicine**, 2016. v. 140, n. 12, p. 1338–1344.

ZUBERI, K. *et al.* GeneMANIA Prediction Server 2013 Update. **Nucleic Acids Research**, 1 jul. 2013. v. 41, n. 1, p. 115–122.

9 ANEXOS

Anexo 1- Termo de Consentimento Livre e Esclarecido

Anexo 2 – Mapa de coleta de fragmentos tireoideanos.

Anexo 3 – Tabela de controle de amostras e histopatológicos.

Anexo 4A – Controle de qualidade das amostras sequenciadas B2 e B11 / N1 e N5

Anexo 4B – Controle de qualidade das amostras sequenciadas B13 e B14 / N6 e N17

Anexo 5 – Tabela de saída de genes diferencialmente expressos protocolo RSEM

Anexo 6 – Relatório do Panther.

Anexo 7 – Relatório da análise do GeneMANIA

ANEXO 1

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – TCLE

O (A) senhor (a), está sendo convidado a participar do projeto de: CONSTRUÇÃO DA BIBLIOTECA DE TECIDOS TIREOIDIANOS NORMAIS E COM BÓCIO. Tal pesquisa visa realizar o armazenamento de pequenos fragmentos, oriundos de amostras de tecido tireoidiano, submetidos a tireoidectomias na FHAJ por bócio para utilização em estudos posteriores de proteômica e transcriptoma.

A coleta e o armazenamento destes fragmentos tireoidianos, auxiliarão na execução de trabalhos científicos relacionados com bócios, possibilitando a criação de múltiplas linhas de pesquisa e fortalecimento das já existentes na área da genética em saúde humana.

Descrição do procedimento: Durante a tireoidectomia realizada na FHAJ, será acessada a região anterior do pescoço, através da mesma incisão realizada pelo cirurgião, expondo a glândula tireóide, seus ligamentos e nervos para posterior retirada exclusivamente do tecido tireoidiano, sem possibilidades de reconhecimento futuro de identidade.

Possíveis riscos e/ou desconfortos: A coleta de material será realizada da peça cirúrgica já retirada no ato operatório, assim, a coleta de material em si, não lhe trará riscos nem outros danos.

Para mais informações:

Prof. João Bosco Botelho Tel: (92) 9183-8680. Escola Superior de Ciência da Saúde da Universidade do Estado do Amazonas - Av. Carvalho Leal, s/n.

Declaro consentimento em participar deste Projeto, autorizando a retirada de material necessária segundo as condições a mim explicadas por este termo.

Declaro ainda estar ciente de que essa pesquisa tem finalidade científica, que a participação tem caráter voluntário e que as identidades dos voluntários serão mantidas em total sigilo por tempo indeterminado, tanto pelo executor como pela instituição onde será realizado, e que os resultados dos procedimentos executados na pesquisa serão analisados e alocados em tabelas, figuras ou gráficos e divulgados em palestras, conferências, periódico científico ou outra forma de divulgação que propicie o repasse dos conhecimentos para a sociedade e para autoridades normativas em saúde nacionais ou internacionais, de acordo com as normas/leis regulatórias de proteção nacional e internacional.

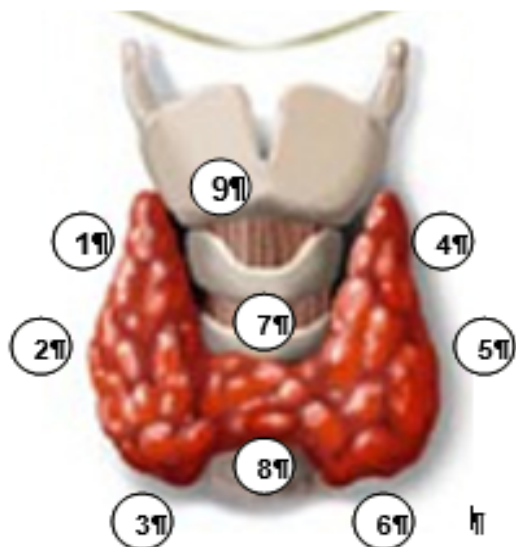
Também estou ciente de que a participação neste trabalho não trará danos, além dos já previstos neste termo, e que em qualquer momento o poderei obter esclarecimentos sobre todos os procedimentos utilizados na pesquisa e nas formas de divulgação dos resultados, bem como a liberdade e o direito de recusar sua participação ou retirar seu consentimento em qualquer fase da pesquisa, sem prejuízo físico, material ou moral.

Manaus, _____ de _____ de _____.

Assinatura do doente/responsável

Pesquisador Responsável

ANEXO 2



- 1- LOBO SUPERIOR DIREITO
- 2- LOBO MEDIO DIREITO
- 3- LOBO INFERIOR DIREITO
- 4- LOBO SUPERIOR ESQUERDO
- 5- LOBO MEDIO ESQUERDO
- 6- LOBO INFERIOR ESQUERDO
- 7- ISTIMO SUPERIOR
- 8- ISTIMO INFERIOR
- 9- PIRAMIDAL

DATA: _ / _ / _

LEITO:

NOME:

IDADE:

CIRURGIÃO:

TIPO DE CIRURGIA: TT TP: ()

REGISTRO DA PACIENTE NO LABORATÓRIO:

NÚMERO DE AMOSTRAS COLETADAS:

ANOTAÇÕES

ANEXO 3
**TRANSCRIPTOMA DE BÓCIO
CONTROLE DE AMOSTRAS RNA_t**

Número do paciente na biblioteca	Lobo tireoidiano da amostra	Nomenclatura adotada no estudo	Laudo Histopatológico da peça	Idade	Controle de qualidade de RNA GenOne	Lâminas Fotografadas
10	LMD	N1	BÓCIO ADENOMATOSO	50	Envio em JAN 2017	
11	LP	N2	BÓCIO ADENOMATOSO	39	X	X
	LMD	B1			----	
	LID	B11			Envio em JAN 2017	
13	LIE	B13	BÓCIO ADENOMATOSO	56	RNA contaminado DEZ 2016	X
14	LME	B2	BÓCIO ADENOMATOSO	?	Envio em JAN 2017	X
	LIE	B14			Aprovado DEZ 2016	
16	LMD	N3	ADENOMA FOLICULAR	56	-----	
Erro no cadastro	LID	N4	Amostra Teste	----	X	
17			BÓCIO ADENOMATOSO	64	----	X
	LP	N3			----	
	LSE	B3			----	
	LMD	N17			Reprovado DEZ 2016	
20	LMD	B4	BÓCIO ADENOMATOSO	48	-----	
21	P	N5	ADENOMA FOLICULAR	53	Envio em JAN 2017	
23	P	N6	BÓCIO ADENOMATOSO/ ADENOMA FOLICULAR	62	Aprovado DEZ 2016	X

LEGENDA

N – Amostra normal

B – Amostra com bócio Adenomatoso

L – LOBO + M (MÉDIO) OU I (INFERIOR) OU S (SUPEIROR)+ D (DIREITO) OU E(ESQUERDO)

OBSERVAÇÕES

- Todas as amostras de tecido com bócio foram selecionadas a partir do laudo histopatológico de diagnóstico de BÓCIO ADENOMATOSO. A amostras normais, no entanto, não seguiram a esse critério, sendo retiradas de peças com outros resultados de laudos.
- Com relação ao gênero, todos os pacientes pertencem ao **sexo feminino**.
- B2 é duplicata de B14 – B2 não deverá ser sequenciada pois B14 foi aprovado no controle de qualidade e já está com sequenciamento em curso.
- O paciente 17 é do sexo masculino, **NÃO DEVERÁ SER SEQUENCIADA**
- N2 N4 – erros no cadastro. Utilizadas para testes. Não devem ser sequenciadas

Sample Quality Control Report

Project General Info

Project name	NovoUS-Brazil-GenoneBiotech-4-human -RNAseq-WOBI	Contract NO.	C101HW16120007
Client	Daniel Carvalho	Institute/Company	GenOne Biotech
Sample Type	<input type="checkbox"/> DNA; <input checked="" type="checkbox"/> rRNA; <input type="checkbox"/> smRNA; <input type="checkbox"/> Tissue; <input type="checkbox"/> Library_____; <input type="checkbox"/> Others_____;		

Summary of RNA sample QC				
Test date	Sample number	Pass*	Hold*	Fail*
2017.02.24	4	0	0	4

Appendix: (For Hold and Fail level below, Novogene is not allowed to proceed with next move before permission from clients)

Pass*: Libraries preparation can be done regularly.

Hold*: Specific instructions from clients are required before library preparation by Novogene.

Fail*: It is highly recommended for clients to resend samples or you could choose to take high risk starting the library preparation (not recommended)

Sample Quality Control Report

Test Result

Sample NO.	Sample Name	Novogene ID	Conc (ng/ μ l)	Vol. * (μ L)	Amt. (μ g)	260/280	260/230	RIN	Conclusion	Note
1	N1	R17020719	576	16	9.22	1.922	0.259		Fail	Degraded
2	N5	R17020720	2	16	0.03	1	0.006		Fail	No banding
3	B2	R17020721	1682	16	26.91	1.965	1.261		Fail	Degraded
4	B11	R17020722	1598	16	25.57	1.954	1.455	2.2	Fail	Unqualified RIN

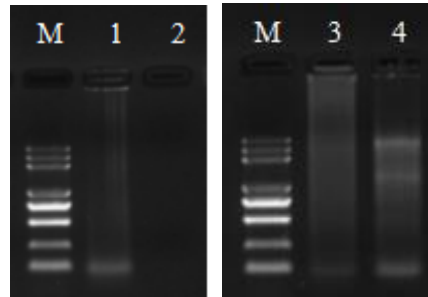
Note:

Vol.*: For samples above 15 μ L, 4 μ L for RNA samples will be taken to do sample QC, which makes the volume indicated above is less than that originally received.

RNA QC Method:

- (1) Nanodrop: Preliminary quantitation
- (2) Agarose Gel Electrophoresis: tests RNA degradation and potential contamination
- (3) Agilent 2100: checks RNA integrity and quantitation

The image of Agarose Gel Electrophoresis (1.0% agarose gel; 0.5× TBE solution; 180V voltage & 16 min)

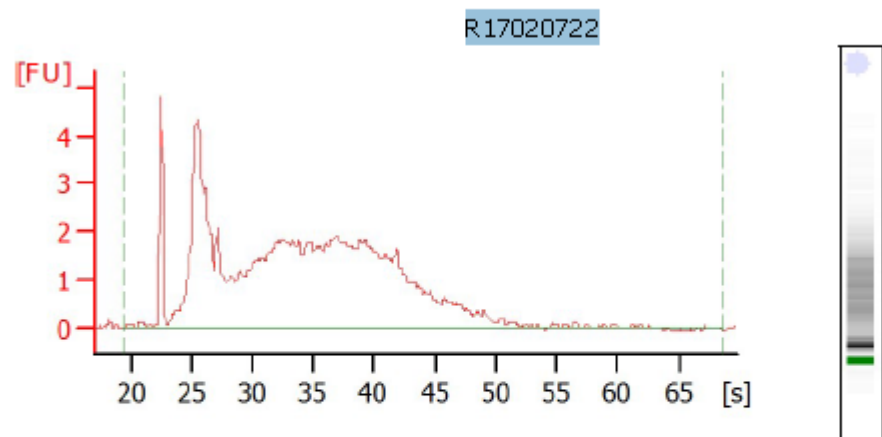


Note:

M: Trans 2K Plus DNA ladder; 1-4: samples ranged in the order of upper table (3-4 loaded 1 ul after 15 times diluted , 1 loaded 1 ul after 5 times diluted , 2 loaded 1 ul)

The result of Agilent 2100 analysis

4) B11



Overall Results for sample 1 : R17020722

RNA Area:	108.5
RNA Concentration:	138 ng/ μ l
rRNA Ratio [28s / 18s]:	0.0
RNA Integrity Number (RIN):	2.2 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 2.20

RNA Samples QC Criteria

Sequencing Type	Remarks	Amount (Qubit)	RIN	Volume	Concentration	Purity
Eukaryotic RNA-seq	Strongly Recommended*	≥2 µg	≥6.8 (Animals)	≥20 µL	≥50ng/µL	No degradation or DNA contamination
	Required*	≥1 µg	≥6.3 (Plants/Fungal)			
Prokaryotic RNA-seq	Strongly Recommended	≥6 µg	≥6.0	≥20 µL	≥50ng/µL	
	Required	≥3 µg				
Eukaryotic Long non-coding RNA-Seq	Strongly Recommended	≥4 µg	≥6.8 (Animals)	≥20 µL	≥50ng/µL	
	Required	≥2 µg	≥6.3 (Plants/Fungal)			
Small RNA-Seq	Strongly Recommended	≥6 µg	≥7.5(Animals)	≥20 µL	≥50ng/µL	
	Required	≥3 µg	≥7 (Plants/Fungal)			

Note:

1. Strongly Recommended*: In case unforeseen happening (lib prep failure, low quality, low amount, etc), the double amount requirements are highly recommended to avoid potential sample re-sending.
2. Required*: Samples amount required for one time library preparation
3. All the sample amount is determined based on Qubit. If client used Nanodrop/picogreen, the figure is expected to be times difference and this is normal due to the different working principle.

Sample Quality Control Report

Project General Info

Project name	GenoneBiotech-4-human-RNAseq-WOBI	Contract NO.	C101HW16120007
Client	Daniel Carvalho	Institute/Company	GenOne
Sample Type	<input type="checkbox"/> DNA; <input checked="" type="checkbox"/> RNA; <input type="checkbox"/> smRNA; <input type="checkbox"/> Tissue; <input type="checkbox"/> Library_____; <input type="checkbox"/> Others_____;		

Summary of RNA sample QC				
Test date	Sample number	Pass*	Hold*	Fail*
2016.12.12	4	2	1	1

Appendix: (For Hold and Fail level below, GenOne is not allowed to proceed with next move before permission from clients)

Pass*: Libraries preparation can be done regularly.

Hold*: Suggestions from client is needed before library preparation is proceed, due to the potential sample volume, total amount, degradation, contamination, extra fragment, etc. Libraries preparation can be done on risk and according to our experience, there will be no big problem.

Fail*: It is highly recommended for clients to resend samples or you could choose to take high risk starting the library preparation (not recommended)

Sample Quality Control Report

Test Result

Sample NO.	Sample Name	Novogene ID	Conc (ng/ μ l)	Vol. * (μ L)	Amt. (μ g)	260/280	260/230	RIN	Conclusion	Note
1	B13	R16120944	637	17	10.83	2.065	1.429	6.9	Hold	Severely impure
2	B14	R16120945	1500	17	25.5	2.064	2.139	8.1	Pass	Slightly impurity
3	N6	R16120946	321	17	5.457	2.061	1.7	8	Pass	Slightly impurity
4	N17	R16120947	326	16	5.216	1.94	1.442	3.7	Fail	Slightly impurity, Unqualified RIN

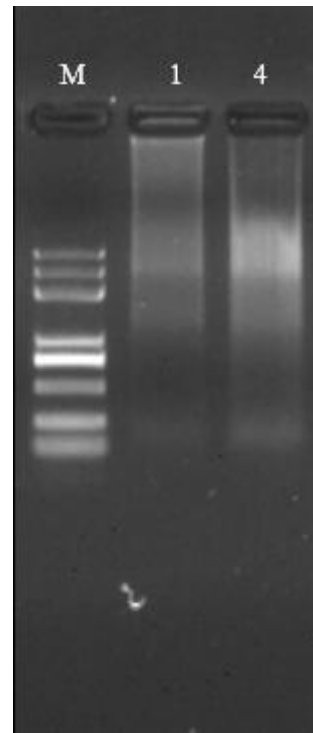
Note:

Vol.*: For samples above 15 μ L, 4 μ L for RNA samples will be taken to do sample QC, which makes the volume indicated above is less than that originally received.

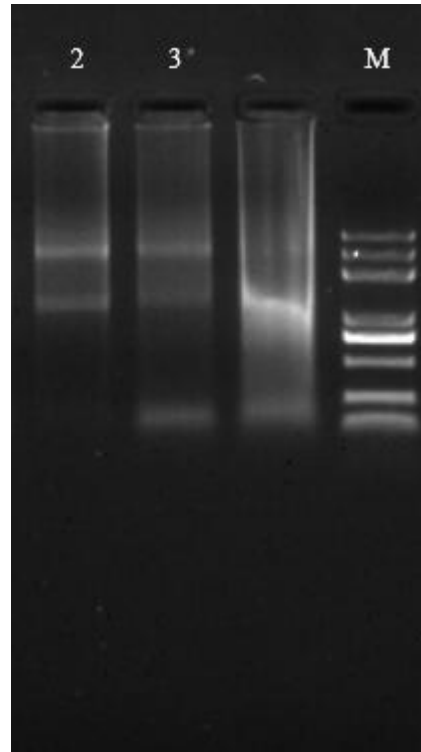
RNA QC Method:

- (1) Nanodrop: Preliminary quantitation
- (2) Agarose Gel Electrophoresis: tests RNA degradation and potential contamination
- (3) Agilent 2100: checks RNA integrity and quantitation

The image of Agarose Gel Electrophoresis (1.0% agarose gel; 0.5× TBE solution; 180V voltage & 16 min)



Sample Quality Control Report

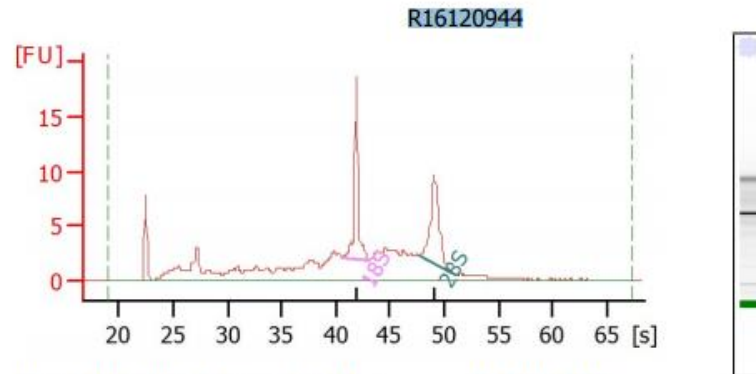


Note:

M: Trans 2K Plus DNA ladder; 1-4: samples ranged in the order of upper table (3-4 loaded 1 ul after 3 times diluted,2 loaded 1 ul after 20 times diluted,1 loaded 1 ul after 7 times diluted)

The result of Agilent 2100 analysis

1) B13



Overall Results for sample 11 : **R16120944**

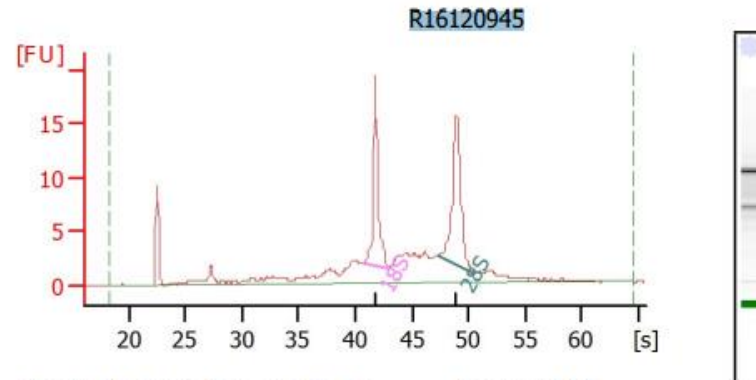
RNA Area: 150.4
 RNA Concentration: 91 ng/μl
 rRNA Ratio [28s / 18s]: 1.0
 RNA Integrity Number (RIN): 6.9 (B.02.08)
 Result Flagging Color:
 Result Flagging Label: RIN: 6.90

Fragment table for sample 11 : **R16120944**

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.66	43.05	16.0	10.7
28S	47.59	50.56	15.5	10.3

Sample Quality Control Report

2) B14



Overall Results for sample 8 : R16120945

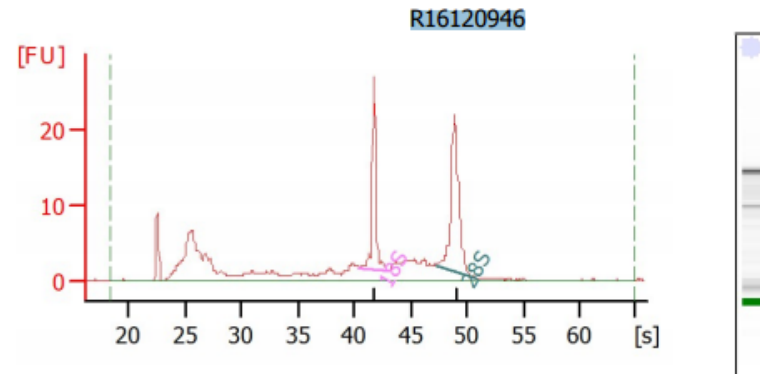
RNA Area: 129.8
RNA Concentration: 75 ng/ μ l
rRNA Ratio [28s / 18s]: 1.2
RNA Integrity Number (RIN): 8.1 (B.02.08)
Result Flagging Color:
Result Flagging Label: RIN: 8.10

Fragment table for sample 8 : R16120945

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.73	42.98	20.3	15.6
28S	47.39	50.48	24.6	18.9

Sample Quality Control Report

3) N6



Overall Results for sample 9 : **R16120946**

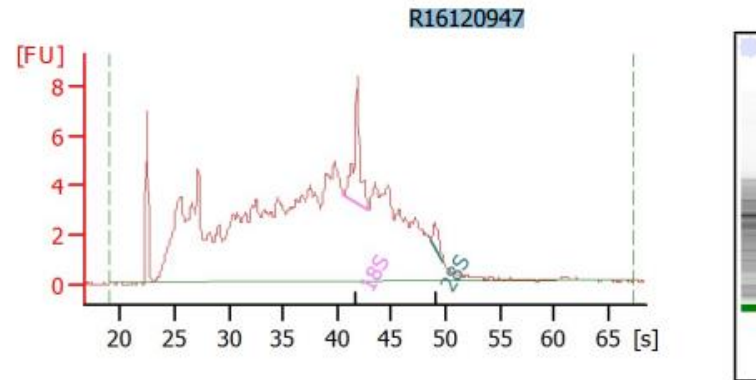
RNA Area: 185.4
RNA Concentration: 107 ng/ μ l
rRNA Ratio [28s / 18s]: 1.6
RNA Integrity Number (RIN): 8 (B.02.08)
Result Flagging Color:
Result Flagging Label: RIN:8

Fragment table for sample 9 : **R16120946**

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.48	43.02	21.8	11.8
28S	47.21	50.79	34.4	18.6

Sample Quality Control Report

4) N17



Overall Results for sample 12 : **R16120947**

RNA Area: 215.9
RNA Concentration: 131 ng/ μ l
rRNA Ratio [28s / 18s]: 0.2
RNA Integrity Number (RIN): 3.7 (B.02.08)
Result Flagging Color:
Result Flagging Label: RIN: 3.70

Fragment table for sample 12 : **R16120947**

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.61	42.85	7.6	3.5
28S	48.61	49.78	1.3	0.6

Sample Quality Control Report

RNA Samples QC Criteria

Sequencing Type	Remarks	Amount (Qubit)	RIN	Volume	Concentration	Purity
Eukaryotic RNA-seq	Strongly Recommended*	≥2 µg	≥6.8 (Animals)	≥20 µL	≥50ng/µL	No degradation or DNA contamination
	Required*	≥1 µg	≥6.3 (Plants/Fungal)			
Prokaryotic RNA-seq	Strongly Recommended	≥6 µg	≥6.0	≥20 µL	≥50ng/µL	
	Required	≥3 µg				
Eukaryotic Long non-coding RNA-Seq	Strongly Recommended	≥4 µg	≥6.8 (Animals)	≥20 µL	≥50ng/µL	
	Required	≥2 µg	≥6.3 (Plants/Fungal)			
Small RNA-Seq	Strongly Recommended	≥6 µg	≥7.5(Animals)	≥20 µL	≥50ng/µL	
	Required	≥3 µg	≥7 (Plants/Fungal)			

Note:

1. Strongly Recommended*: In case unforeseen happening (lib prep failure, low quality, low amount, etc), the double amount requirements are highly recommended to avoid potential sample re-sending.
2. Required*: Samples amount required for one time library preparation
3. All the sample amount is determined based on Qubit. If client used Nanodrop/picogreen, the figure is expected to be times difference and this is normal due to the different working principle.

Sample Quality Control Report

ANEXO 5

id	Gene symb	baseMean	baseMean		foldChange	log2FoldCh	pval	padj
			A	B				
			CTRL	TREAT				
ENSG00000105421	(HSPA6	969,6475	2624,402	142,2701	0,05421	-4,20528	5,58E-25	1,52E-20
ENSG00000105422	(AC004556.	389,2367	0	583,8551	Inf	Inf	8,75E-22	1,19E-17
ENSG00000105423	(SORD2P	2724,63	219,5988	3977,146	18,11096	4,178791	3,21E-18	2,92E-14
ENSG00000105424	(C1QL4	226,2702	640,1698	19,32045	0,03018	-5,05025	2,94E-17	2,01E-13
ENSG00000105425	(PLCD4	492,752	1295,045	91,60554	0,070735	-3,82142	2,70E-16	1,47E-12
ENSG00000105426	(ERRFI1	5980,52	14025,9	1957,832	0,139587	-2,84076	5,14E-16	2,34E-12
ENSG00000105427	(PCP4L1	164,9995	470,5689	12,21474	0,025957	-5,26771	4,46E-15	1,52E-11
ENSG00000105428	(ATRNL1	268,6142	733,3033	36,26969	0,049461	-4,33757	4,46E-15	1,52E-11
ENSG00000105429	(MT1H	5698,956	13089,66	2003,605	0,153068	-2,70776	6,68E-15	2,02E-11
ENSG00000105430	(ABCA13	550,8222	1376,414	138,0262	0,10028	-3,3179	1,38E-13	3,77E-10
ENSG00000105431	(DNAJB1	8794,016	19352,15	3514,95	0,181631	-2,46092	4,99E-13	1,24E-09
ENSG00000105432	(CA12	711,3262	1679,343	227,3179	0,135361	-2,88511	2,32E-12	5,28E-09
ENSG00000105433	(PKD1P5	687,5309	1624,443	219,0748	0,134861	-2,89045	7,23E-12	1,52E-08
ENSG00000105434	(RP11-958N	99,45817	288,2235	5,07551	0,01761	-5,82749	8,52E-12	1,66E-08
ENSG00000105435	(C4B	9662,877	1870,512	13559,06	7,248851	2,857752	1,54E-11	2,80E-08
ENSG00000105436	(ETV4	261,3134	664,6786	59,63085	0,089714	-3,47853	5,86E-11	9,98E-08
ENSG00000105437	(RP11-514P	55,22649	165,6795	0	0	#NOME?	9,92E-11	1,59E-07
ENSG00000105438	(IGSF1	1171,612	2594,011	460,4127	0,177491	-2,49419	2,11E-10	3,20E-07
ENSG00000105439	(RASD1	888,867	1972,468	347,0664	0,175955	-2,50672	2,49E-10	3,57E-07
ENSG00000105440	(MRPL23	227,2428	577,4273	52,15061	0,090315	-3,46888	7,01E-10	9,55E-07
ENSG00000105441	(CPNE4	59,4177	176,4634	0,894866	0,005071	-7,62348	8,31E-10	1,04E-06
ENSG00000105442	(IL1RL1	109,0454	303,9091	11,61355	0,038214	-4,70976	8,42E-10	1,04E-06
ENSG00000105443	(PI4KAP1	397,4859	947,02	122,7189	0,129584	-2,94804	1,09E-09	1,29E-06
ENSG00000105444	(C2CD4C	333,7567	33,33197	483,9691	14,51967	3,859936	3,69E-09	4,20E-06
ENSG00000105445	(DLEC1	185,206	469,5886	43,01469	0,091601	-3,4485	9,84E-09	1,07E-05
ENSG00000105446	(ZBTB45P1	42,48192	127,4458	0	0	#NOME?	1,35E-08	1,42E-05
ENSG00000105447	(LECT1	368,9326	842,1223	132,3378	0,157148	-2,6698	1,57E-08	1,58E-05
ENSG00000105448	(AC241377.	137,7933	3,921408	204,7293	52,20811	5,706202	3,20E-08	3,12E-05
ENSG00000105449	(SHC3	175,8071	437,237	45,09222	0,10313	-3,27746	3,67E-08	3,45E-05
ENSG00000105450	(SORCS1	400,9828	955,8432	123,5526	0,12926	-2,95165	3,96E-08	3,59E-05
ENSG00000105451	(STARD9	521,5333	1121,523	221,5386	0,197534	-2,33983	1,20E-07	0,000106
ENSG00000105452	(TMSB4XP4	90,81418	245,088	13,67727	0,055806	-4,16345	1,58E-07	0,000135
ENSG00000105453	(TMEM184/	125,2003	322,5358	26,53259	0,082262	-3,60362	2,74E-07	0,000226
ENSG00000105454	(PCSK2	606,0039	1271,517	273,2477	0,214899	-2,21827	2,93E-07	0,000235
ENSG00000105455	(FOSB	31029,91	57969,19	17560,27	0,302924	-1,72297	3,08E-07	0,000238
ENSG00000105456	(TRABD2A	129,9573	328,4179	30,72705	0,093561	-3,41795	3,15E-07	0,000238
ENSG00000105457	(GATM	2134,98	4088,068	1158,436	0,28337	-1,81924	6,37E-07	0,000464
ENSG00000105458	(BAG3	1334,77	2594,011	705,149	0,271837	-1,87918	6,47E-07	0,000464
ENSG00000105459	(FAM105A	370,7952	794,0851	159,1502	0,20042	-2,3189	1,10E-06	0,000765
ENSG00000105460	(HSPA1B	34641,24	75642	14140,87	0,186945	-2,41932	1,13E-06	0,000771
ENSG00000105461	(MME	181,4371	426,4531	58,92913	0,138184	-2,85533	1,47E-06	0,000979
ENSG00000105462	(PDE4C	106,0494	269,5968	24,27572	0,090045	-3,47322	1,73E-06	0,00112
ENSG00000105463	(CPXM1	1226,196	381,3569	1648,616	4,323026	2,112042	2,63E-06	0,00167
ENSG00000105464	(RTN4RL2	2433,253	4547,853	1375,954	0,30255	-1,72475	3,11E-06	0,001926
ENSG00000105465	(EDN3	1481,034	2787,141	827,9802	0,297072	-1,75112	3,62E-06	0,002193
ENSG00000105466	(NRK	119,1481	297,0466	30,19877	0,101663	-3,29813	3,82E-06	0,002265

ENSG0000(EGR2	4993,615	9088,843	2946,001	0,324134	-1,62534	4,19E-06	0,002432
ENSG0000(TNFRSF19	298,6224	632,327	131,7701	0,208389	-2,26265	4,65E-06	0,002643
ENSG0000(ABCC3	269,1327	566,6434	120,3773	0,212439	-2,23488	8,87E-06	0,004937
ENSG0000(NAPRT	482,7944	127,4458	660,4686	5,182351	2,373607	1,01E-05	0,005499
ENSG0000(HSD17B6	1313,127	2453,821	742,7804	0,302704	-1,72402	1,09E-05	0,005845
ENSG0000(RP11-1212	806,2261	1531,31	443,6843	0,289742	-1,78716	1,30E-05	0,006824
ENSG0000(ST6GAL1	1753,427	635,2681	2312,506	3,640206	1,86402	1,42E-05	0,007316
ENSG0000(C15orf48	157,7584	353,9071	59,68406	0,168643	-2,56795	2,65E-05	0,013176
ENSG0000(GATA5	47,71149	131,3672	5,883652	0,044788	-4,48075	2,66E-05	0,013176
ENSG0000(FRAS1	3217,555	5631,142	2010,761	0,357079	-1,48569	2,89E-05	0,01406
ENSG0000(PCDH8	94,80243	232,3434	26,03194	0,112041	-3,1579	3,19E-05	0,015234
ENSG0000(YJEFN3	214,4661	462,7261	90,33615	0,195226	-2,35678	3,67E-05	0,017243
ENSG0000(SPHKAP	29,1871	85,29062	1,135338	0,013311	-6,23119	4,51E-05	0,020827
ENSG0000(TBX2	1259,005	2275,397	750,8086	0,329968	-1,5996	4,61E-05	0,02095
ENSG0000(KIAA1324	2393,249	4179,24	1500,253	0,358978	-1,47803	5,64E-05	0,024991
ENSG0000(KIF5A	266,3902	545,0757	127,0474	0,233082	-2,10109	5,69E-05	0,024991
ENSG0000(ARHGAP28	273,7577	548,9971	136,138	0,247976	-2,01173	5,78E-05	0,024991
ENSG0000(C1orf233	362,9637	715,6569	186,6171	0,260763	-1,93919	6,19E-05	0,02636
ENSG0000(COL14A1	2917,666	1222,499	3765,249	3,079961	1,622912	7,66E-05	0,032131
ENSG0000(NRG1	431,9334	829,3778	233,2112	0,281188	-1,83039	8,27E-05	0,034149
ENSG0000(HSPA1A	27717,93	57801,55	12676,11	0,219304	-2,189	0,000102	0,041454
ENSG0000(THBS1	547,7865	1009,763	316,7986	0,313736	-1,67238	0,000104	0,041727
ENSG0000(PLIN1	135,6509	300,9681	52,99226	0,176073	-2,50576	0,000109	0,043242
ENSG0000(XKRX	46,40518	125,4851	6,865241	0,05471	-4,19206	0,000113	0,044048



PANTHER™ Version 12.0 (released 2017-07-10). [View release notes here.](#)

PANTHER GENE LIST [?](#) [Customize Gene list](#)

 Convert List to: Send list to:

 Display: items per page [Refine Search](#)

 Hits 1-30 of 54 [page: (1) 2] Number of mapped ids found 54 [IDs not found \(5\)](#)

<input type="checkbox"/>	<input type="checkbox"/>	Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	<input checked="" type="checkbox"/> PANTHER Family/Subfamily	<input checked="" type="checkbox"/> PANTHER Protein Class	<input checked="" type="checkbox"/> Species
<input type="checkbox"/>	<input type="checkbox"/>	1. HUMAN HGNC=9062 UniProtKB=Q9BRC7	ENSG00000115556	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4 PLCD4 ortholog	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4 (PTHR10336:SF31)	guanyl-nucleotide exchange factor signaling molecule	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	2. HUMAN HGNC=11597 UniProtKB=O13207	ENSG00000121068	T-box transcription factor TBX2 TBX2 ortholog	T-box transcription factor TBX2 (PTHR11267:SF82)	nucleic acid binding transcription factor	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	3. HUMAN HGNC=5948 UniProtKB=Q8N6C5	ENSG00000147255	Immunoglobulin superfamily member 1 IGSF1 ortholog	Immunoglobulin superfamily member 1 (PTHR11738:SF102)	immunoglobulin receptor superfamily	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	4. HUMAN HGNC=15802 UniProtKB=Q9BWX5	ENSG00000130700	Transcription factor GATA-5 GATA5 ortholog	Transcription factor GATA-5 (PTHR10071:SF289)	nucleic acid binding transcription factor	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	5. HUMAN HGNC=14638 UniProtKB=Q86UQ4	ENSG00000179869	ATP-binding cassette sub-family A member 13 ABCA13 ortholog	ATP-binding cassette sub-family A member 13 (PTHR19229:SF113)	ATP-binding cassette (ABC) transporter	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	6. HUMAN HGNC=25391 UniProtKB=Q7Z2Y5	ENSG00000123572	Nik-related protein kinase NRK ortholog	Nik-related protein kinase (PTHR24361:SF438)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	7. HUMAN HGNC=42951 UniProtKB=F2Z333	ENSG00000228594	Fibronectin type-III domain-containing transmembrane protein C1orf233 C1orf233 ortholog	Fibronectin type-III domain-containing transmembrane protein C1orf233 (PTHR39233:SF1)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	8. HUMAN HGNC=16697 UniProtKB=Q8WY21	ENSG00000108018	VPS10 domain-containing receptor SorCS1 SORCS1 ortholog	VPS10 domain-containing receptor SorCS1 (PTHR12106:SF8)	receptor transporter	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	9. HUMAN HGNC=29063 UniProtKB=Q5VV63	ENSG00000107518	Attractin-like protein 1 ATRNL1 ortholog	Attractin-like protein 1 (PTHR10574:SF224)	extracellular matrix linker protein receptor	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	10. HUMAN HGNC=15828 UniProtKB=Q9Y272	ENSG00000108551	Dexamethasone-induced Ras-related protein 1 RASD1 ortholog	Dexamethasone-induced Ras-related protein 1 (PTHR24070:SF255)	small GTPase	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	11. HUMAN HGNC=28797 UniProtKB=Q6ZMB5	ENSG00000164855	Transmembrane protein 184A TMEM184A ortholog	Transmembrane protein 184A (PTHR23423:SF59)	transporter	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	12. HUMAN HGNC=11915 UniProtKB=Q9NS68	ENSG00000127863	Tumor necrosis factor receptor superfamily member 19 TNFRSF19 ortholog	Tumor necrosis factor receptor superfamily member 19 (PTHR12120:SF1)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	13. HUMAN HGNC=20448 UniProtKB=A6NKN8	ENSG00000248485	Purkinje cell protein 4-like protein 1 PCP4L1 ortholog	Purkinje cell protein 4-like protein 1 (PTHR15359:SF5)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	14. HUMAN HGNC=1371 UniProtKB=O43570	ENSG00000074410	Carbonic anhydrase 12 CA12 ortholog	Carbonic anhydrase 12 (PTHR18952:SF19)	dehydratase	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	15. HUMAN HGNC=5232 UniProtKB=P0DMV8	ENSG00000204389	Heat shock 70 kDa protein 1A HSPA1A ortholog	Heat shock 70 kDa protein 1A-related (PTHR19375:SF223)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	16. HUMAN HGNC=18181 UniProtKB=Q92529	ENSG00000148082	SHC-transforming	SHC-transforming	signaling molecule	Homo

			protein 3 SHC3 ortholog	protein 3 (PTHR10337:SF4)		sapiens	
<input type="checkbox"/>	17.	HUMAN HGNC=5270 UniProtKB=P25685	ENSG00000132002	DnaJ homolog subfamily B member 1 DNAJB1 ortholog	DnaJ homolog subfamily B member 1 (PTHR24078:SF306)	-	Homo sapiens
<input type="checkbox"/>	18.	HUMAN HGNC=11785 UniProtKB=P07996	ENSG00000137801	Thrombospondin-1 THBS1 ortholog	Thrombospondin-1 (PTHR10199:SF78)	-	Homo sapiens
<input type="checkbox"/>	19.	HUMAN HGNC=3239 UniProtKB=P11161	ENSG00000122877	E3 SUMO-protein ligase EGR2 EGR2 ortholog	E3 SUMO-protein ligase EGR2 (PTHR44234:SF2)	-	Homo sapiens
<input type="checkbox"/>	20.	HUMAN HGNC=5998 UniProtKB=Q01638	ENSG00000115602	Interleukin-1 receptor-like 1 IL1RL1 ortholog	Interleukin-1 receptor-like 1 (PTHR11890:SF7)	type I cytokine receptor	Homo sapiens
<input type="checkbox"/>	21.	HUMAN HGNC=25629 UniProtKB=Q9NUU6	ENSG00000145569	Inactive ubiquitin thioesterase FAM105A FAM105A ortholog	Inactive ubiquitin thioesterase FAM105A (PTHR33662:SF1)	-	Homo sapiens
<input type="checkbox"/>	22.	HUMAN HGNC=30619 UniProtKB=Q2M3C7	ENSG00000153820	A-kinase anchor protein SPHKAP SPHKAP ortholog	A-kinase anchor protein SPHKAP (PTHR10226:SF7)	-	Homo sapiens
<input type="checkbox"/>	23.	HUMAN HGNC=5233 UniProtKB=P0DMV9	ENSG00000204388	Heat shock 70 kDa protein 1B HSPA1B ortholog	Heat shock 70 kDa protein 1A-related (PTHR19375:SF223)	-	Homo sapiens
<input type="checkbox"/>	24.	HUMAN HGNC=8744 UniProtKB=P16519	ENSG00000125851	Neuroendocrine convertase 2 PCSK2 ortholog	Neuroendocrine convertase 2 (PTHR42884:SF13)	-	Homo sapiens
<input type="checkbox"/>	25.	HUMAN HGNC=23316 UniProtKB=O14756	ENSG00000025423	17-beta-hydroxysteroid dehydrogenase type 6 HSD17B6 ortholog	17-beta-hydroxysteroid dehydrogenase type 6 (PTHR43313:SF4)	dehydrogenase reductase	Homo sapiens
<input type="checkbox"/>	26.	HUMAN HGNC=10322 UniProtKB=Q16540	ENSG00000214026	39S ribosomal protein L23, mitochondrial MRPL23 ortholog	39S ribosomal protein L23, mitochondrial (PTHR12059:SF5)	ribosomal protein	Homo sapiens
<input type="checkbox"/>	27.	HUMAN HGNC=31416 UniProtKB=Q86Z23	ENSG00000186897	Complement C1q-like protein 4 C1QL4 ortholog	Complement C1q-like protein 4 (PTHR22923:SF67)	-	Homo sapiens
<input type="checkbox"/>	28.	HUMAN HGNC=3493 UniProtKB=P43268	ENSG00000175832	ETS translocation variant 4 ETV4 ortholog	ETS translocation variant 4 (PTHR11849:SF181)	nucleic acid binding signaling molecule transcription factor	Homo sapiens
<input type="checkbox"/>	29.	HUMAN HGNC=25509 UniProtKB=Q9P2N2	ENSG00000088756	Rho GTPase-activating protein 28 ARHGAP28 ortholog	Rho GTPase-activating protein 28 (PTHR14963:SF5)	G-protein modulator	Homo sapiens
<input type="checkbox"/>	30.	HUMAN HGNC=7400 UniProtKB=P80294	ENSG00000205358	Metallothionein-1H MT1H ortholog	Metallothionein 1H-like protein 1-related (PTHR23299:SF4)	-	Homo sapiens

Hits 1-30 of 54 [page: (1) 2]



PANTHER™ Version 12.0 (released 2017-07-10). [View release notes here.](#)

PANTHER GENE LIST [?](#) [Customize Gene list](#)

 Convert List to: Send list to: 

 Display: items per page [Refine Search](#)

 Hits 31-54 of 54 [page: [1](#) (2)] Number of mapped ids found 54 [IDs not found \(5\)](#)

<input type="checkbox"/>	<input type="checkbox"/>	Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	<input checked="" type="checkbox"/> PANTHER Family/Subfamily	<input checked="" type="checkbox"/> PANTHER Protein Class	<input checked="" type="checkbox"/> Species
<input type="checkbox"/>	<input type="checkbox"/>	31. HUMAN HGNC=23053 UniProtKB=O86UN3	ENSG00000186907	Reticulon-4 receptor-like 2 RTN4RL2 ortholog	Reticulon-4 receptor-like 2 (PTHR44065:SF5)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	32. HUMAN HGNC=54 UniProtKB=O15438	ENSG00000108846	Canalicular multispecific organic anion transporter 2 ABCC3 ortholog	Canalicular multispecific organic anion transporter 2 (PTHR24223:SF340)	ATP-binding cassette (ABC) transporter	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	33. HUMAN HGNC=17005 UniProtKB=O75829	ENSG00000136110	Leukocyte cell-derived chemotaxin 1 LECT1 ortholog	Leukocyte cell-derived chemotaxin 1 (PTHR14064:SF6)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	34. HUMAN HGNC=29618 UniProtKB=O6UXG2	ENSG00000116299	UPF0577 protein KIAA1324 KIAA1324 ortholog	UPF0577 protein KIAA1324 (PTHR22727:SF13)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	35. HUMAN HGNC=27013 UniProtKB=O86V40	ENSG00000186854	Metalloprotease TIK1 TRABD2A ortholog	Metalloprotease TIK1 (PTHR31120:SF7)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	36. HUMAN HGNC=24785 UniProtKB=A6XGL0	ENSG00000250067	YjeF N-terminal domain-containing protein 3 YJEFN3 ortholog	YjeF N-terminal domain-containing protein 3 (PTHR13232:SF12)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	37. HUMAN HGNC=4175 UniProtKB=P50440	ENSG00000171766	Glycine amidinotransferase, mitochondrial GATM ortholog	Glycine amidinotransferase, mitochondrial (PTHR10488:SF1)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	38. HUMAN HGNC=8782 UniProtKB=Q08493	ENSG00000105650	cAMP-specific 3',5'-cyclic phosphodiesterase 4C PDE4C ortholog	cAMP-specific 3',5'-cyclic phosphodiesterase 4C (PTHR11347:SF135)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	39. HUMAN HGNC=9076 UniProtKB=O60240	ENSG00000166819	Perilipin-1 PLIN1 ortholog	Perilipin-1 (PTHR14024:SF30)	transfer/carrier protein	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	40. HUMAN HGNC=7997 UniProtKB=Q02297	ENSG00000157168	Pro-neuregulin-1, membrane-bound isoform NRG1 ortholog	Pro-neuregulin-1, membrane-bound isoform (PTHR11100:SF7)	growth factor	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	41. HUMAN HGNC=8660 UniProtKB=O95206	ENSG00000136099	Protocadherin-8 PCDH8 ortholog	Protocadherin-8 (PTHR24028:SF46)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	42. HUMAN HGNC=18185 UniProtKB=Q9UJM3	ENSG00000116285	ERBB receptor feedback inhibitor 1 ERRF1 ortholog	ERBB receptor feedback inhibitor 1 (PTHR14254:SF5)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	43. HUMAN HGNC=7154 UniProtKB=P08473	ENSG00000196549	Nepriylsin MME ortholog	Nepriylsin (PTHR11733:SF114)	metalloprotease	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	44. HUMAN HGNC=19185 UniProtKB=O86XX4	ENSG00000138759	Extracellular matrix protein FRAS1 FRAS1 ortholog	Extracellular matrix protein FRAS1 (PTHR11878:SF29)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	45. HUMAN HGNC=3797 UniProtKB=P53539	ENSG00000125740	Protein fosB FOSB ortholog	Protein fosB (PTHR23351:SF3)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	46. HUMAN HGNC=939 UniProtKB=O95817	ENSG00000151929	BAG family molecular chaperone regulator 3 BAG3 ortholog	BAG family molecular chaperone regulator 3 (PTHR12329:SF12)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	47. HUMAN HGNC=3178 UniProtKB=P14138	ENSG00000124205	Endothelin-3 EDN3 ortholog	Endothelin-3 (PTHR13874:SF11)	peptide hormone	Homo sapiens

<input type="checkbox"/>	48. HUMAN HGNC=6323 UniProtKB=Q12840	ENSG00000155980	Kinesin heavy chain isoform 5A KIF5A ortholog	Kinesin heavy chain isoform 5A (PTHR24115:SF317)	microtubule binding motor protein	Homo sapiens
<input type="checkbox"/>	49. HUMAN HGNC=2899 UniProtKB=Q9Y238	ENSG00000008226	Deleted in lung and esophageal cancer protein 1 DLEC1 ortholog	Deleted in lung and esophageal cancer protein 1 (PTHR23053:SF12)	-	Homo sapiens
<input type="checkbox"/>	50. HUMAN HGNC=5239 UniProtKB=P17066	ENSG00000173110	Heat shock 70 kDa protein 6 HSPA6 ortholog	Heat shock 70 kDa protein 6-related (PTHR19375:SF168)	-	Homo sapiens
<input type="checkbox"/>	51. HUMAN HGNC=29845 UniProtKB=Q6PP77	ENSG00000182489	XK-related protein 2 XKRX ortholog	XK-related protein 2 (PTHR14297:SF4)	-	Homo sapiens
<input type="checkbox"/>	52. HUMAN HGNC=29898 UniProtKB=Q9C002	ENSG00000166920	Normal mucosa of esophagus-specific gene 1 protein NMES1 ortholog	Normal mucosa of esophagus-specific gene 1 protein (PTHR14256:SF3)	oxidoreductase	Homo sapiens
<input type="checkbox"/>	53. HUMAN HGNC=19162 UniProtKB=Q9P2P6	ENSG00000159433	StAR-related lipid transfer protein 9 STARD9 ortholog	StAR-related lipid transfer protein 9 (PTHR12136:SF67)	membrane traffic protein transfer/carrier protein	Homo sapiens
<input type="checkbox"/>	54. HUMAN HGNC=2317 UniProtKB=Q96A23	ENSG00000196353	Copine-4 CPNE4 ortholog	Copine-4 (PTHR10857:SF4)	membrane traffic protein	Homo sapiens

Hits 31-54 of 54 [page: [1](#) (2)]

PANTHER™ Version 12.0 (released 2017-07-10). [View release notes here.](#)
PANTHER GENE LIST [?](#) [Customize Gene list](#)

 Convert List to: Send list to: 

 Display: items per page [Refine Search](#)

 Hits 1-7 of 7 [page: (1)] Number of mapped ids found 7 [IDs not found \(1\)](#)

<input type="checkbox"/>	Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	PANTHER Family/Subfamily	<input checked="" type="checkbox"/> PANTHER Protein Class <input checked="" type="checkbox"/>
<input type="checkbox"/>	1. HUMAN HGNC=30450 UniProtKB=Q6XQN6	ENSG00000147813	Nicotinate phosphoribosyltransferase NAPRT ortholog	Nicotinate phosphoribosyltransferase (PTHR11098:SF1)	glycosyltransferase
<input type="checkbox"/>	2. HUMAN Ensembl=ENSG00000281741 UniProtKB=A0A0D9SF21	ENSG00000281741	Protein LOC100996720 LOC100996720 ortholog	-	-
<input type="checkbox"/>	3. HUMAN HGNC=10860 UniProtKB=P15907	ENSG00000073849	Beta-galactoside alpha-2,6-sialyltransferase 1 ST6GAL1 ortholog	Beta-galactoside alpha-2,6-sialyltransferase 1 (PTHR13713:SF55)	glycosyltransferase
<input type="checkbox"/>	4. HUMAN HGNC=1324 UniProtKB=P0C0L5	ENSG00000224389	Complement C4-B C4B ortholog	Complement C4-B (PTHR11412:SF86)	complement component cytokine serine protease inhibitor
<input type="checkbox"/>	5. HUMAN HGNC=2191 UniProtKB=Q0570Z	ENSG00000187955	Collagen alpha-1(XIV) chain COL14A1 ortholog	Collagen alpha-1(XIV) chain (PTHR44557:SF1)	-
<input type="checkbox"/>	6. HUMAN HGNC=29417 UniProtKB=Q8TF44	ENSG00000183186	C2 calcium-dependent domain-containing protein 4C C2CD4C ortholog	C2 calcium-dependent domain-containing protein 4C (PTHR10024:SF116)	membrane trafficking regulatory protein
<input type="checkbox"/>	7. HUMAN HGNC=15771 UniProtKB=Q96SM3	ENSG00000088882	Probable carboxypeptidase X1 CPXM1 ortholog	carboxypeptidase X1-related (PTHR11532:SF43)	metalloprotease

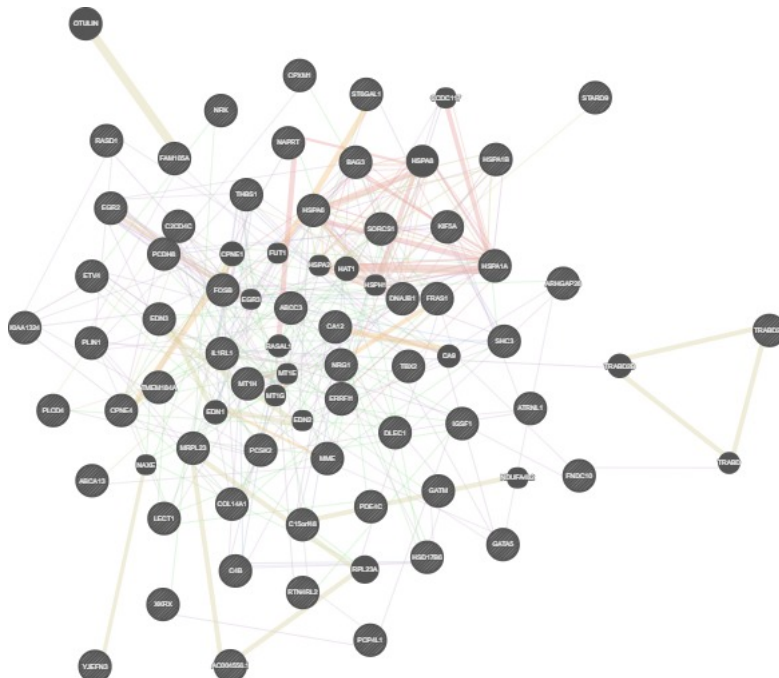
Hits 1-7 of 7 [page: (1)]

GeneMANIA report

Created on : 30 September 2017 19:27:28

Last database update : 13 March 2017 00:00:00

Application version : 3.5.0



Networks

- Co-expression
- Physical Interactions
- Shared protein domains
- Predicted
- Co-localization
- Genetic Interactions

Functions

N/A

Search parameters

Organism Homo sapiens (human)

Genes ENSG00000008226 , ENSG00000025423 , ENSG00000073849 , ENSG00000074410 , ENSG00000088756 , ENSG00000088882 , ENSG00000105650 , ENSG00000107518 , ENSG00000108018 , ENSG00000108551 , ENSG00000108846 , ENSG00000115556 , ENSG00000115602 , ENSG00000116285 , ENSG00000116299 , ENSG00000121068 , ENSG00000122877 , ENSG00000123572 , ENSG00000124205 , ENSG00000125740 , ENSG00000125851 , ENSG00000127863 , ENSG00000130700 , ENSG00000132002 , ENSG00000136099 , ENSG00000136110 , ENSG00000137801 , ENSG00000138759 , ENSG00000145569 , ENSG00000147255 , ENSG00000147813 , ENSG00000148082 , ENSG00000151929 , ENSG00000153820 , ENSG00000155980 , ENSG00000157168 , ENSG00000159433 , ENSG00000164855 , ENSG00000166819 , ENSG00000166920 , ENSG00000171766 , ENSG00000173110 , ENSG00000175832 , ENSG00000179869 , ENSG00000182489 , ENSG00000183186 , ENSG00000186854 , ENSG00000186897 , ENSG00000186907 , ENSG00000187955 , ENSG00000196353 , ENSG00000196549 , ENSG00000204388 , ENSG00000204389 , ENSG00000205236 , ENSG00000205358 , ENSG00000214026 , ENSG00000224389 , ENSG00000228594 , ENSG00000248485 , ENSG00000250067 , ENSG00000276345 , ENSG00000281741

Network weighting Automatically selected weighting method

Networks **A**

Abu-Odeh-Aqeilan-2014 , Agrawal-Sedivy-2010 , Aichem-Groettrup-2012 , Albers-Koegl-2005 , Alexandru-Deshaies-2008 , Alizadeh-Staudt-2000 , Andresen-Flores-Morales-2014 , Arbuckle-Grant-2010 , Arroyo-Aloy-2014 , Arroyo-Aloy-2015

B

Bahr-Bowler-2013 , Bailey-Hieter-2015 , Bandyopadhyay-Ideker-2010 , Bantscheff-Drewes-2011 , Barr-Knapp-2009 , Barrios-Rodiles-Wrana-2005 , Behrends-Harper-2010 , Behzadnia-Lührmann-2007 , Bennett-Harper-2010 , Benzinger-Hermeking-2005 , Berggård-James-2006 , Bett-Hay-2013 , Bhatnagar-Attie-2014 , Bild-Nevins-2006 B , BIOGRID-SMALL-SCALE-STUDIES , BIOGRID-SMALL-SCALE-STUDIES , Blandin-Richard-2013 , Blomen-Brummelkamp-2015 , Blomen-Brummelkamp-2015 , Bogachek-Weigel-2014 , Boldrick-Relman-2002 , Bonacci-Soubeyran-2014 , Bouwmeester-Supert-Furga-2004 , Brajenovic-Drewes-2004 , Brehme-Supert-Furga-2009 , Bruderer-Hay-2011 , Burington-Shaughnessy-2008 , Butland-Hayden-2014 , Byron-Humphries-2012

C

Cai-Conaway-2007 , Camargo-Brandon-2007 , Campos-Reinberg-2015 , Cao-Chinnaiyan-2014 , Carmon-Liu-2014 , CELL_MAP , Chen-Brown-2002 , Chen-Ge-2013 , Chen-Huang-2014 , Chen-Zhang-2013 , Christianson-Kopito-2011 , Cloutier-

C

Coulombe-2013 , Colland-Gauthier-2004 , Corominas-Iakoucheva-2014 , Couzens-Gingras-2013 , Cox-Rizzino-2013 , Coyaud-Raught-2015

D

Danielsen-Nielsen-2011 , Dart-Wells-2015 , de Hoog-Mann-2004 , Diner-Cristea-2015 , Dobbin-Giordano-2005 , Drissi-Boisvert-2015 , Dyer-Sobral-2010

E

Emanuele-Elledge-2011 , Emdal-Olsen-2015 , Ewing-Figeys-2007

F

Fenner-Prehn-2010 , Floyd-Pagliarini-2016 , Foerster-Ritter-2013 , Fogeron-Lange-2013 , Foster-Marshall-2013 , Freibaum-Taylor-2010

G

Gabriel-Baumgrass-2016 , Galligan-Howley-2015 , Gao-Reinberg-2012 , Gautier-Hall-2009 , Giannone-Liu-2010 , Glatter-Gstaiger-2009 , Gloeckner-Ueffing-2007 , Goehler-Wanker-2004 , Golebiowski-Hay-2009 , Goudreault-Gingras-2009 , Grant-2010 , Greco-Cristea-2011 , Grossmann-Stelzl-2015 , Guarani-Harper-2014 , Gupta-Pelletier-2015

H

Hanson-Clayton-2014 , Hauri-Gstaiger-2013 , Havrylov-Redowicz-2009 , Havugimana-Emili-2012 , Hayes-Urbé-2012 , Hegele-Stelzl-2012 A , Hegele-Stelzl-2012 B , Hein-Mann-2015 , Hill-Livingston-2014 , HUMANCYC , Humphries-Humphries-2009 , Hutchins-Peters-2010 , Huttlin-Gygi-2015

I

I2D-BIND-Fly2Human , I2D-BIND-Mouse2Human , I2D-BIND-Rat2Human , I2D-BIND-Worm2Human , I2D-BIND-Yeast2Human , I2D-BioGRID-Fly2Human , I2D-BioGRID-Mouse2Human , I2D-BioGRID-Rat2Human , I2D-BioGRID-Worm2Human , I2D-BioGRID-Yeast2Human , I2D-Chen-Pawson-2009-PiwiScreen-Mouse2Human , I2D-Formstecher-Daviet-2005-Embryo-Fly2Human , I2D-Giot-Rothbert-2003-Low-Fly2Human , I2D-INNATEDB-Mouse2Human , I2D-IntAct-Fly2Human , I2D-IntAct-Mouse2Human , I2D-IntAct-Rat2Human , I2D-IntAct-Worm2Human , I2D-IntAct-Yeast2Human , I2D-Krogan-Greenblatt-2006-Core-Yeast2Human , I2D-Krogan-Greenblatt-2006-NonCore-Yeast2Human , I2D-Li-Vidal-2004-CORE-1-Worm2Human , I2D-Li-Vidal-2004-non-core-Worm2Human , I2D-Manual-Mouse2Human , I2D-Manual-Rat2Human , I2D-MGI-Mouse2Human , I2D-MINT-Fly2Human , I2D-MINT-Mouse2Human , I2D-MINT-Rat2Human , I2D-MINT-Worm2Human , I2D-MINT-Yeast2Human , I2D-Ptacek-Snyder-2005-Yeast2Human , I2D-Tarassov-PCA-Yeast2Human , I2D-Tewari-Vidal-2004-TGFb-Worm2Human , I2D-vonMering-Bork-2002-High-Yeast2Human , I2D-vonMering-Bork-2002-Low-Yeast2Human , I2D-vonMering-Bork-2002-Medium-Yeast2Human , I2D-Wang-Orkin-2006-EScmplx-Mouse2Human , I2D-Wang-Orkin-2006-

I

EScmplxlow-Mouse2Human , I2D-Yu-Vidal-2008-GoldStd-Yeast2Human , IMID , Ingham-Pawson-2005 , Innocenti-Brown-2011 , INTERPRO , IREF-BIND , IREF-BIOGRID , IREF-DIP , IREF-HPRD , IREF-INTACT , IREF-MATRIXDB , IREF-MPPI , IREF-PUBMED , IREF-SMALL-SCALE-STUDIES , IREF-SMALL-SCALE-STUDIES

J

Jeronimo-Coulombe-2007 , Jin-Pawson-2004 , Johnson-Kerner-Wichterle-2015 , Johnson-Shoemaker-2003 , Jones-MacBeath-2006 , Joshi-Cristea-2013 , Jäger-Krogan-2011

K

Kahle-Zoghbi-2011 , Kaltenbach-Hughes-2007 , Katsogiannou-Rocchi-2014 , Kim-Gygi-2011 , Kim-Major-2015 , Kneissl-Grummt-2003 , Koch-Hermeking-2007 , Kotlyar-Jurisica-2015 , Kristensen-Foster-2012 , Kärblane-Sarmiento-2015 , Kırılı-Görlich-2015

L

Lambert-Gingras-2015 , Lamoliatte-Thibault-2014 , Lau-Ronai-2012 , Lee-Songyang-2011 , Lehner-Sanderson-2004 A , Lehner-Sanderson-2004 B , Leng-Wang-2014 , Leung-Jones-2014 , Li-Chen-2015 , Li-Dorf-2011 A , Li-Dorf-2011 B , Li-Dorf-2014 , Li-Haura-2013 , Lim-Zoghbi-2006 , Lin-Smith-2010 , Lipp-Guthrie-2015 , Liu-Wang-2012 , Llères-Lamond-2010 , Loch-Strickler-2012 , Low-Heck-2014 , Lu-Zhang-2013 , Luo-Elledge-2009

M

Mak-Moffat-2010 , Mallon-McKay-2013 , Malovannaya-Qin-2010 , Markson-Sanderson-2009 , Maréchal-Zou-2014 , Matsumoto-Nakayama-2005 , McCracken-Blencowe-2005 , McFarland-Nussbaum-2008 , Meek-Piwnica-Worms-2004 , Milev-Mouland-2012 , Miyamoto-Sato-Yanagawa-2010 , Murakawa-Landthaler-2015

N

Nakayama-Ohara-2002 , Nakayasu-Adkins-2013 , Napolitano-Meroni-2011 , Narayan-Bennett-2012 , Nathan-Goldberg-2013 , NCI_NATURE , Neganova-Lako-2011 , Newman-Keating-2003 , Nicholson-Hupp-2014 , Noble-Diehl-2008

O

Oliviero-Cagney-2015 , Olma-Pintard-2009 , Oláh-Ovádi-2011 , Oshikawa-Nakayama-2012 , Ouyang-Gill-2009

P

Panigrahi-Pati-2012 , Papp-Lamia-2015 , Perez-Hernandez-Yáñez-Mó-2013 , Perou-Botstein-1999 , Perou-Botstein-2000 , Persaud-Rotin-2009 , Petschnigg-Stagljar-2014 , PFAM , Phillips-Corn-2013 , Pichlmair-Superti-Furga-2011 , Pichlmair-Superti-Furga-2012 , Pilot-Storck-Goillot-2010 , Povlsen-Choudhary-2012

R

Ramachandran-LaBaer-2004 , Raman-Harper-2015 , Ramaswamy-Golub-2001 , Ravasi-Hayashizaki-2010 , REACTOME , Reinke-Keating-2013 , Reyniers-Taymans-2014 , Richter-Chrzanowska-Lightowlers-2010 , Rieger-Chu-2004 , Rolland-Vidal-2014 , Rosenwald-Staudt-2001 , Roth-Zlotnik-2006 , Roux-Burke-2012 , Rowbotham-Mermoud-2011 , Roy-Pardo-2014 , Roy-Parent-2013 , Rual-Vidal-2005 A , Rual-Vidal-2005 B

S

Sang-Jackson-2011 , Sato-Conaway-2004 , Schadt-Shoemaker-2004 , Scholz-Taylor-2016 , Singh-Moore-2012 , Smirnov-Cheung-2009 , So-Colwill-2015 , Soler-López-Aloy-2011 , Sowa-Harper-2009 , Stehling-Lill-2012 , Stehling-Lill-2013 , Stelzl-Wanker-2005 , Stes-Gevaert-2014 , Stuart-Kim-2003 , Suter-Wanker-2013

T

Taipale-Lindquist-2012 , Taipale-Lindquist-2014 , Takahashi-Conaway-2011 , Tarallo-Weisz-2011 , Tatham-Hay-2011 , Teixeira-Gomes-2010 , Thalappilly-Duseti-2008 , Thompson-Luchansky-2014 , Tong-Moran-2014 , Toyoshima-Grandori-2012 , Tsai-Cristea-2012

U

Udeshi-Carr-2012

V

van Wijk-Timmers-2009 , Vandamme-Angrand-2011 , Varjosalo-Gstaiger-2013 , Varjosalo-Supertti-Furga-2013 , Venkatesan-Vidal-2009 , Vermeulen-Mann-2010 , Vinayagam-Wanker-2011 , Virok-Fülöp-2011 , Vizeacoumar-Moffat-2013

W

Wagner-Choudhary-2011 , Wallach-Kramer-2013 , Wan-Emili-2015 , Wang-Balch-2006 , Wang-Cheung-2015 , Wang-He-2008 , Wang-Maris-2006 , Wang-Xu-2015 , Wang-Yang-2011 , Weimann-Stelzl-2013 A , Weimann-Stelzl-2013 B , Weinmann-Meister-2009 , Wen-Wu-2014 , Whisenant-Salomon-2015 , Wilker-Yaffe-2007 , Willingham-Muchowski-2003 , Witt-Labeit-2008 , Wong-O'Bryan-2012 , Woods-Monteiro-2012 , Woodsmith-Sanderson-2012 , Wu-Garvey-2007 , Wu-Li-2007 , Wu-Ma-2012 , Wu-Stein-2010 , Wu-Stein-2010

X

Xiao-Lefkowitz-2007 , Xie-Cong-2013 , Xie-Green-2012 , Xu-Ye-2012

Y

Yang-Chen-2010 , Yatim-Benkirane-2012 , Yu-Chow-2013 , Yu-Vidal-2011

Z

Zanon-Pichler-2013 , Zhang-Shang-2006 , Zhang-Zou-2011 , Zhao-Krug-2005 , Zhao-Yang-2011 , Zhou-Conrads-2004 , Zhou-Hanemann-2016

Genes

Gene	Description	Rank
RP11-514P8.6		N/A
AC241377.1		N/A
AC004556.1		N/A
C1QL4	complement component 1, q subcomponent-like 4 [Source:HGNC Symbol;Acc:HGNC:31416]	N/A
HSPA1B	heat shock protein family A (Hsp70) member 1B [Source:HGNC Symbol;Acc:HGNC:5233]	N/A
YJEFN3	YjeF N-terminal domain containing 3 [Source:HGNC Symbol;Acc:HGNC:24785]	N/A
NRK	Nik related kinase [Source:HGNC Symbol;Acc:HGNC:25391]	N/A
ABCA13	ATP binding cassette subfamily A member 13 [Source:HGNC Symbol;Acc:HGNC:14638]	N/A
PDE4C	phosphodiesterase 4C [Source:HGNC Symbol;Acc:HGNC:8782]	N/A
PCP4L1	Purkinje cell protein 4 like 1 [Source:HGNC Symbol;Acc:HGNC:20448]	N/A
TRABD2A	TraB domain containing 2A [Source:HGNC Symbol;Acc:HGNC:27013]	N/A
TMEM184A	transmembrane protein 184A [Source:HGNC Symbol;Acc:HGNC:28797]	N/A
TNFRSF19	tumor necrosis factor receptor superfamily member 19 [Source:HGNC Symbol;Acc:HGNC:11915]	N/A
FNDC10	fibronectin type III domain containing 10 [Source:HGNC Symbol;Acc:HGNC:42951]	N/A
XKRX	XK related, X-linked [Source:HGNC Symbol;Acc:HGNC:29845]	N/A
GATA5	GATA binding protein 5 [Source:HGNC Symbol;Acc:HGNC:15802]	N/A
STARD9	StAR related lipid transfer domain containing 9 [Source:HGNC Symbol;Acc:HGNC:19162]	N/A
KIAA1324	KIAA1324 [Source:HGNC Symbol;Acc:HGNC:29618]	N/A
C4B	complement component 4B (Chido blood group) [Source:HGNC Symbol;Acc:HGNC:1324]	N/A
ATRNL1	attractin like 1 [Source:HGNC Symbol;Acc:HGNC:29063]	N/A
C2CD4C	C2 calcium dependent domain containing 4C [Source:HGNC Symbol;Acc:HGNC:29417]	N/A
SPHKAP	SPHK1 interactor, AKAP domain containing [Source:HGNC Symbol;Acc:HGNC:30619]	N/A

Gene	Description	Rank
RASD1	ras related dexamethasone induced 1 [Source:HGNC Symbol;Acc:HGNC:15828]	N/A
NAPRT	nicotinate phosphoribosyltransferase [Source:HGNC Symbol;Acc:HGNC:30450]	N/A
FAM105A	family with sequence similarity 105 member A [Source:HGNC Symbol;Acc:HGNC:25629]	N/A
CPNE4	copine 4 [Source:HGNC Symbol;Acc:HGNC:2317]	N/A
FRAS1	Fraser extracellular matrix complex subunit 1 [Source:HGNC Symbol;Acc:HGNC:19185]	N/A
CPXM1	carboxypeptidase X (M14 family), member 1 [Source:HGNC Symbol;Acc:HGNC:15771]	N/A
ARHGAP28	Rho GTPase activating protein 28 [Source:HGNC Symbol;Acc:HGNC:25509]	N/A
PLCD4	phospholipase C delta 4 [Source:HGNC Symbol;Acc:HGNC:9062]	N/A
SORCS1	sortilin related VPS10 domain containing receptor 1 [Source:HGNC Symbol;Acc:HGNC:16697]	N/A
HSPA1A	heat shock protein family A (Hsp70) member 1A [Source:HGNC Symbol;Acc:HGNC:5232]	N/A
NRG1	neuregulin 1 [Source:HGNC Symbol;Acc:HGNC:7997]	N/A
GATM	glycine amidinotransferase [Source:HGNC Symbol;Acc:HGNC:4175]	N/A
COL14A1	collagen type XIV alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2191]	N/A
C15orf48	chromosome 15 open reading frame 48 [Source:HGNC Symbol;Acc:HGNC:29898]	N/A
EDN3	endothelin 3 [Source:HGNC Symbol;Acc:HGNC:3178]	N/A
PCDH8	protocadherin 8 [Source:HGNC Symbol;Acc:HGNC:8660]	N/A
ERRFI1	ERBB receptor feedback inhibitor 1 [Source:HGNC Symbol;Acc:HGNC:18185]	N/A
RTN4RL2	reticulon 4 receptor-like 2 [Source:HGNC Symbol;Acc:HGNC:23053]	N/A
ETV4	ETS variant 4 [Source:HGNC Symbol;Acc:HGNC:3493]	N/A
SHC3	SHC adaptor protein 3 [Source:HGNC Symbol;Acc:HGNC:18181]	N/A
DLEC1	deleted in lung and esophageal cancer 1 [Source:HGNC Symbol;Acc:HGNC:2899]	N/A
LECT1	leukocyte cell derived chemotaxin 1 [Source:HGNC Symbol;Acc:HGNC:17005]	N/A
ST6GAL1	ST6 beta-galactoside alpha-2,6-sialyltransferase 1 [Source:HGNC	N/A

Gene	Description	Rank
	Symbol;Acc:HGNC:10860]	
CA12	carbonic anhydrase 12 [Source:HGNC Symbol;Acc:HGNC:1371]	N/A
TBX2	T-box 2 [Source:HGNC Symbol;Acc:HGNC:11597]	N/A
IGSF1	immunoglobulin superfamily member 1 [Source:HGNC Symbol;Acc:HGNC:5948]	N/A
HSPA6	heat shock protein family A (Hsp70) member 6 [Source:HGNC Symbol;Acc:HGNC:5239]	N/A
PCSK2	proprotein convertase subtilisin/kexin type 2 [Source:HGNC Symbol;Acc:HGNC:8744]	N/A
MRPL23	mitochondrial ribosomal protein L23 [Source:HGNC Symbol;Acc:HGNC:10322]	N/A
IL1RL1	interleukin 1 receptor like 1 [Source:HGNC Symbol;Acc:HGNC:5998]	N/A
KIF5A	kinesin family member 5A [Source:HGNC Symbol;Acc:HGNC:6323]	N/A
ABCC3	ATP binding cassette subfamily C member 3 [Source:HGNC Symbol;Acc:HGNC:54]	N/A
PLIN1	perilipin 1 [Source:HGNC Symbol;Acc:HGNC:9076]	N/A
THBS1	thrombospondin 1 [Source:HGNC Symbol;Acc:HGNC:11785]	N/A
EGR2	early growth response 2 [Source:HGNC Symbol;Acc:HGNC:3239]	N/A
HSD17B6	hydroxysteroid 17-beta dehydrogenase 6 [Source:HGNC Symbol;Acc:HGNC:23316]	N/A
MT1H	metallothionein 1H [Source:HGNC Symbol;Acc:HGNC:7400]	N/A
MME	membrane metallo-endopeptidase [Source:HGNC Symbol;Acc:HGNC:7154]	N/A
FOSB	FosB proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:3797]	N/A
DNAJB1	DnaJ heat shock protein family (Hsp40) member B1 [Source:HGNC Symbol;Acc:HGNC:5270]	N/A
BAG3	BCL2 associated athanogene 3 [Source:HGNC Symbol;Acc:HGNC:939]	N/A
OTULIN	OTU deubiquitinase with linear linkage specificity [Source:HGNC Symbol;Acc:HGNC:25118]	1
HSPA8	heat shock protein family A (Hsp70) member 8 [Source:HGNC Symbol;Acc:HGNC:5241]	2
RPL23A	ribosomal protein L23a [Source:HGNC Symbol;Acc:HGNC:10317]	3
CPNE1	copine 1 [Source:HGNC Symbol;Acc:HGNC:2314]	4
HAT1	histone acetyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4821]	5
TRABD2B	TraB domain containing 2B [Source:HGNC Symbol;Acc:HGNC:44200]	6

Gene	Description	Rank
EDN1	endothelin 1 [Source:HGNC Symbol;Acc:HGNC:3176]	7
CA9	carbonic anhydrase 9 [Source:HGNC Symbol;Acc:HGNC:1383]	8
RASAL1	RAS protein activator like 1 [Source:HGNC Symbol;Acc:HGNC:9873]	9
TRABD	TraB domain containing [Source:HGNC Symbol;Acc:HGNC:28805]	10
MT1G	metallothionein 1G [Source:HGNC Symbol;Acc:HGNC:7399]	11
HSPH1	heat shock protein family H (Hsp110) member 1 [Source:HGNC Symbol;Acc:HGNC:16969]	12
HSPA2	heat shock protein family A (Hsp70) member 2 [Source:HGNC Symbol;Acc:HGNC:5235]	13
NDUFA4L2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2 [Source:HGNC Symbol;Acc:HGNC:29836]	14
EDN2	endothelin 2 [Source:HGNC Symbol;Acc:HGNC:3177]	15
MT1E	metallothionein 1E [Source:HGNC Symbol;Acc:HGNC:7397]	16
CCDC117	coiled-coil domain containing 117 [Source:HGNC Symbol;Acc:HGNC:26599]	17
FUT1	fucosyltransferase 1 (H blood group) [Source:HGNC Symbol;Acc:HGNC:4012]	18
EGR3	early growth response 3 [Source:HGNC Symbol;Acc:HGNC:3240]	19
NAXE	NAD(P)HX epimerase [Source:HGNC Symbol;Acc:HGNC:18453]	20

Networks

Co-expression	59.54%
Innocenti-Brown-2011	9.58%
Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. Innocenti et al (2011). <i>PLoS Genet</i> Co-expression with 603,765 interactions from GEO	
Ramaswamy-Golub-2001	9.18%
Multiclass cancer diagnosis using tumor gene expression signatures. Ramaswamy et al (2001). <i>Proc Natl Acad Sci U S A</i> Co-expression with 275,113 interactions from supplementary material	
Roth-Zlotnik-2006	8.50%
Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Roth et al (2006). <i>Neurogenetics</i> Co-expression with 669,062 interactions from GEO	
Mallon-McKay-2013	7.77%
StemCellDB: the human pluripotent stem cell database at the National Institutes of Health. Mallon et al (2013). <i>Stem Cell Res</i> Co-expression with 585,265 interactions from GEO	
Burington-Shaughnessy-2008	6.36%
Tumor cell gene expression changes following short-term in vivo exposure to single agent chemotherapeutics are related to survival in multiple myeloma. Burington et al (2008). <i>Clin Cancer Res</i> Co-expression with 290,538 interactions from GEO	
Chen-Brown-2002	5.16%
Gene expression patterns in human liver cancers. Chen et al (2002). <i>Mol Biol Cell</i> Co-expression with 282,241 interactions from supplementary material	
Noble-Diehl-2008	3.92%
Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. Noble et al (2008). <i>Gut</i> Co-expression with 661,539 interactions from GEO	
Wang-Maris-2006	2.74%
Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. Wang et al (2006). <i>Cancer Res</i> Co-expression with 264,023 interactions from GEO	
Wang-Cheung-2015	2.53%
Genetic variation in insulin-induced kinase signaling. Wang et al (2015). <i>Mol Syst Biol</i> Co-expression with 411,047 interactions from GEO	
Perou-Botstein-1999	1.56%
Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Perou et al (1999). <i>Proc Natl Acad Sci U S A</i> Co-expression with 65,069 interactions from supplementary material	
Bild-Nevins-2006 B	0.82%
Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Bild et al (2006). <i>Nature</i> Co-expression with 280,683 interactions from GEO	

Co-expression	59.54%
Perou-Botstein-2000	0.59%
Molecular portraits of human breast tumours. Perou et al (2000). <i>Nature</i> Co-expression with 185,068 interactions from supplementary material	
Alizadeh-Staudt-2000	0.53%
Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Alizadeh et al (2000). <i>Nature</i> Co-expression with 90,336 interactions from supplementary material	
Rosenwald-Staudt-2001	0.30%
Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. Rosenwald et al (2001). <i>J Exp Med</i> Co-expression with 114,694 interactions from supplementary material	
Physical Interactions	27.38%
Kristensen-Foster-2012	10.06%
A high-throughput approach for measuring temporal changes in the interactome. Kristensen et al (2012). <i>Nat Methods</i> Physical Interactions with 7,115 interactions from BioGRID	
Taipale-Lindquist-2014	5.88%
A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. Taipale et al (2014). <i>Cell</i> Physical Interactions with 1,362 interactions from iRefIndex	
Havugimana-Emili-2012	3.11%
A census of human soluble protein complexes. Havugimana et al (2012). <i>Cell</i> Physical Interactions with 13,716 interactions from BioGRID	
Li-Haura-2013	3.03%
Perturbation of the mutated EGFR interactome identifies vulnerabilities and resistance mechanisms. Li et al (2013). <i>Mol Syst Biol</i> Physical Interactions with 403 interactions from BioGRID	
Chen-Ge-2013	2.82%
Bcl2-associated athanogene 3 interactome analysis reveals a new role in modulating proteasome activity. Chen et al (2013). <i>Mol Cell Proteomics</i> Physical Interactions with 368 interactions from iRefIndex	
Hein-Mann-2015	2.25%
A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Hein et al (2015). <i>Cell</i> Physical Interactions with 27,044 interactions from BioGRID	
IREF-INTACT	0.23%
Physical Interactions with 56,297 interactions from iRefIndex	
Shared protein domains	5.18%
PFAM	5.18%
Shared protein domains with 457,054 interactions from Pfam	
Predicted	3.72%
Wu-Stein-2010	3.72%

Predicted 3.72%

Wu-Stein-2010

A human functional protein interaction network and its application to cancer data analysis. Wu et al (2010). *Genome Biol*
Predicted with 87,829 interactions from supplementary material

Co-localization 3.67%

Johnson-Shoemaker-2003

Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Johnson et al (2003). *Science*
Co-localization with 426,332 interactions from GEO

Genetic Interactions 0.52%

Lin-Smith-2010

A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. Lin et al (2010). *Genome Res*
Genetic Interactions with 4,820,370 interactions from supplementary material