

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
QUÍMICA E BIOQUÍMICA



**Validação de genes alvo da via Rac1/PAK1-BCL6/STAT5
envolvidos na progressão tumoral**

Leonor Bárbara Freire Abrantes

Mestrado em Bioquímica
Especialização em Bioquímica Médica

Dissertação orientada por:
Doutor Paulo Matos
Doutora Patrícia Barros

Agradecimentos

Ao longo deste trabalho estive perto de pessoas que me acompanharam nesta etapa da minha vida, ao qual não poderia deixar de agradecer por todas as palavras e apoio que recebi.

Em primeiro lugar, gostaria de agradecer ao Doutor Paulo Matos por me ter dado a oportunidade de trabalhar consigo e com a sua equipa. A sua boa disposição sempre presente fez-me sentir em “casa”, muito obrigada por isso, por toda a compreensão e por me ter orientado e ajudado em tudo o que eu precisei. Agradeço ainda todo o tempo e trabalho que disponibilizou a corrigir esta tese e toda a ajuda que me deu na sua concretização.

Ao Doutor Peter Jordan, agradeço a amabilidade com que me recebeu no seu laboratório desde o primeiro dia e por me ter feito sentir à vontade desde o início. Obrigada ainda pela sua simpatia e otimismo ao longo da minha estadia no laboratório.

O maior dos obrigados à Patrícia, não consigo imaginar o meu percurso por esse laboratório sem ti. Foste sem dúvida alguma uma grande mentora, sempre disposta a ajudar e com muita paciência para me ensinar as coisas...sei que não foi fácil! Em todos os momentos estiveste lá e foi contigo que aprendi a não desesperar quando algo não corre como queremos! Muito obrigada por seres quem és, tive muita sorte em ter uma orientadora como tu! Espero um dia, poder vir a ser aquela pessoa a quem toda a gente pede uma 2ª opinião, aquela pessoa em quem toda a gente confia, uma pessoa como tu. Desejo-te a maior das sortes daqui para a frente e espero poder voltar a encontrar-te ao longo da minha carreira. Agradeço-te por tudo isto e muito mais!

Também tenho a agradecer às restantes meninas do laboratório como é óbvio! Obrigada por me terem recebido tão bem, nunca me vou esquecer do ambiente de boa disposição e diversão no laboratório. Vânia, Ana, Cláudia, Andreia, Joana e Márcia muito obrigada pelas conversas, pelos risos e por toda a vossa ajuda durante a minha estadia aí no laboratório. Não me vou esquecer daquele tenebroso “Dia das Centrífugas”, e do quanto se riram com isso! Gostei muito de vos conhecer e desejo-vos muita sorte para o futuro.

Aos meus amigos, Andreia, Patrícia, Filipe, Carlota, Reis, Daniel, Ricardo, Mini, Sofia, Miguel e afilhados, uns a percorrer o mesmo percurso, outros a trabalhar para lá chegar, muito obrigada por estarem sempre do meu lado. Não escolheria mais ninguém para percorrer este caminho a não ser vocês. Este ano não foi fácil para nenhum de nós, mas mesmo assim conseguimos chegar onde chegámos e sempre juntos! Onde quer que o Gerson esteja, sei que ele vai estar a olhar por nós e com certeza que está muito orgulhoso por conseguirmos chegar até aqui.

Muito obrigada a ti Diogo, por acreditares sempre em mim, mesmo quando eu própria não o fazia. Obrigada por toda a paciência, compreensão e apoio que me deste ao longo destes anos todos. Obrigada por teres aturado todas as conversas sobre PCR's e clonagens. Não tenho palavras que descrevam o quanto significas para mim.

Tenho a agradecer aos meus pais e ao meu irmão por terem proporcionado a minha chegada até aqui, por sempre me dizerem para seguir em frente e acreditarem sempre nas minhas capacidades. Aos meus tios e primo por estarem sempre lá para mim, em especial à minha tia-madrinha, por desde sempre ser o meu exemplo a seguir e me mostrar que temos que batalhar pelas coisas que queremos na vida. Aos meus avós por me acompanharem ao longo da vida e esperarem sempre grandes coisas de mim. Adoro-

vos a todos vós e muito obrigada por tudo, pois tudo o que sou hoje é resultado da minha convivência com vocês todos desde o início.

Um especial obrigado ao meu avô, apesar de já não estares aqui comigo espero que estejas orgulhoso de mim e que estejas a acompanhar todas as minhas vitórias. Todas essas batalhas eu venço-as por ti, porque sei que me vias como uma guerreira e quero que assim continue. Obrigada por tudo aquilo que me proporcionaste e ensinaste, és mesmo muito importante para mim!

Tentarei sempre alcançar os meus objetivos e fazer tudo ao meu alcance para que a minha contribuição na ciência torne este mundo um pouquinho melhor.

Resumo

O cancro colorretal é um dos tipos de cancro com maior incidência a nível mundial e também dos mais mortíferos, sendo o seu prognóstico tanto mais limitado quanto mais avançado for o estado da doença. A GTPase Rac1 encontra-se sobre-expressa em vários tipos de carcinoma, nomeadamente colorretal e a desregulação da sua sinalização celular tem sido associada à transformação maligna. Em particular, o eixo de sinalização Rac1/PAK1 encontra-se alterado em cerca de 60% dos tumores sólidos, estando esta alteração associada a tumores mais agressivos e invasivos, com prognósticos clínicos mais desfavoráveis, consequentes, muitas vezes, do desenvolvimento de resistência à quimioterapia. Para além disto, Rac1/PAK1 é também responsável pela ativação de várias vias de sinalização que conduzem à regulação da expressão génica, facto que tem adquirido cada vez mais destaque no estudo da progressão tumoral.

O laboratório de acolhimento descreveu, em células de carcinoma colorretal, uma nova via de sinalização em que a ativação de Rac1/PAK1 promove um *switch* transcricional entre o repressor BCL6 e o ativador STAT5, levando a um aumento da expressão génica. Assim, de forma a identificar todos os locais do genoma nos quais a expressão génica poderia ser modulada por esta via, o grupo de investigação utilizou uma abordagem inovadora de análise de dados de ChIP-seq, explorando a seletividade do *switch* transcricional, BCL6/STAT5. No presente estudo pretendeu-se validar experimentalmente esta nova abordagem de análise de dados de ChIP-seq, procurando, entre os múltiplos “hits” encontrados, um conjunto de genes cuja modulação da expressão por esta via pudesse elucidar algumas das consequências pro-oncogénicas da sobre-ativação de Rac1 e PAK1 observada em tumores agressivos com mau prognóstico.

Os nossos resultados identificaram 2402 genes que respondiam ao *switch* BCL6/STAT5 aquando da estimulação da via Rac1/PAK. Destes, selecionaram-se para validação experimental um conjunto de 15 genes, para os quais os picos detetados resumiam as características da totalidade dos picos identificados no que respeita a parâmetros como dimensão, amplitude, localização relativa ao respetivo gene, entre outros. Como era esperado, com base em estudos prévios, a variação na expressão dos genes identificados, em resposta à manipulação do estado de ativação da via Rac1/PAK1 é ligeira, não ultrapassando três vezes os valores basais. Curiosamente, alguns destes genes exibiam uma diminuição dos níveis de expressão aquando da ativação da via Rac1/PAK1 e o comportamento contrário na sua inibição. Uma análise mais detalhada das sequências delimitadas pelos picos correspondentes, permitiu identificar dois sub-motivos distintos para os genes que respondiam de forma positiva e negativa à ativação da via Rac1/PAK1, dentro do motivo *consensus* geral de ligação ao DNA dos fatores BCL6 e STAT5. Em paralelo, realizou-se uma análise de agrupamento funcional, tendo-se observado um enriquecimento dentro da lista de genes identificados de genes envolvidos na resposta e reparação de danos no DNA. Os picos correspondentes a estes genes continham o sub-domínio de resposta positiva à estimulação da via Rac1/PAK1 e a análise funcional da sua expressão demonstrou que os níveis de todos eles aumentavam em resposta à ativação da via e diminuía aquando da inibição da via. A avaliação do impacto fisiológico da ativação destes genes em células de carcinoma colorretal (DLD1) pelo ensaio do cometa, aquando da ativação da via Rac1/PAK1, revelou que esta via confere uma proteção parcial aos danos induzidos pelo tratamento com o agente alquilante sulfonato de etil-metano (EMS), acelerando o processo de reparação dos mesmos. De notar é o facto de que a inibição desta via com o inibidor seletivo de Rac1, EHT1846, bloqueia significativamente a reparação dos danos genómicos, promovendo mesmo o seu agravamento.

Este trabalho veio demonstrar que a nova estratégia de análise de ChIP-seq permite a identificação de pequenas variações transcricionais como a que advém da resposta transcricional ao *switch* BCL6/STAT5, modulada pela via Rac1/PAK1. Revelou, ainda, o papel da via de sinalização Rac1/PAK1-BCL6/STAT5 na resposta e reparação de danos no DNA, sugerindo que a sua inibição farmacológica poderá ter aplicação terapêutica no cancro, nomeadamente na potenciação dos efeitos de determinados agentes quimioterápicos com efeitos genotóxicos.

Palavras-chave: cancro colorretal; Rac1/PAK1; BCL6/STAT5; ChIP-seq; resposta e reparação de danos no DNA.

Abstract

Colorectal cancer is one of the most prevalent types of cancer worldwide and also one of the most deadly malignancies, with its prognosis being poorer the more advanced the state of the disease. The GTPase Rac1 is overexpressed in several types of carcinoma, namely in colorectal cancers, and the dysregulation of its cellular signaling has been tightly associated with malignant transformation. In particular, the Rac1/PAK1 signaling axis is altered in about 60% of solid tumors, this alteration being associated with more aggressive and invasive tumors, with more unfavorable clinical prognosis, often resulting from the development of resistance to chemotherapy. In addition, Rac1/PAK1 is also responsible for the activation of several signaling pathways that lead to the regulation of gene expression, an aspect that has become increasingly prominent in the study of tumor progression.

The host laboratory described, in colorectal carcinoma cells, a new signaling pathway in which activation of the Rac1/PAK1 pathway promotes a transcriptional switch between the BCL6 repressor and the STAT5 activator, leading to increased gene expression. Thus, in order to identify all sites in the genome in which gene expression could be modulated by this pathway, the research group used an innovative ChIP-seq data analysis approach, exploring the selectivity of the transcriptional switch, BCL6/STAT5. In the present study, we aimed to experimentally validate this new approach of ChIP-seq data analysis, searching among the multiple hits identified for a set of genes for which the modulation of their expression by this pathway could elucidate some of the pro-oncogenic consequences of the Rac1/PAK1 over-activation observed in aggressive tumors with poor prognosis.

Our results identified 2402 genes responding to the BCL6/STAT5 switch upon stimulation of the Rac1/PAK pathway. From these, a set of 15 genes were selected for experimental validation, since their corresponding ChIP-seq peaks summarized the characteristics of overall identified peaks with respect to parameters such as size, amplitude, location relative to the respective gene, among others. As expected, based on previous studies, the variation in expression levels of the selected genes was small in response to the manipulation of the Rac1/PAK1 pathway activity, not exceeding three times the baseline values. Interestingly, some of these genes exhibited a decrease in expression upon activation of the pathway and the opposite behavior in its inhibition. A more detailed analysis of the sequences delimited by the corresponding peaks allowed to identify two distinct DNA-binding sub-motifs for the genes that responded positively and negatively to the activation of the Rac1/PAK1 pathway, within the general DNA binding consensus factor for BCL6 and STAT5 factors. In parallel, a functional clustering analysis was performed, and enrichment in genes involved in DNA damage response and repair was observed within the identified 2402 gene list. Peaks corresponding to these genes contained the subdomain of positive response to Rac1/PAK1 pathway stimulation and the functional analysis of 4 of these for their expression levels demonstrated that all increased in response to Rac1/PAK1 pathway activation while inhibition of the pathway led to a decrease in their expression. The evaluation of the physiological impact of the activation of these genes in colorectal carcinoma cells (DLD1) by the comet assay, upon activation of the Rac1/PAK1 pathway, has shown that this pathway provides partial protection against damage induced by treatment with alkylating agents such as EMS, accelerating the process of DNA damage

repair. Of note is the fact that inhibition of this pathway with the selective Rac1 inhibitor, EHT1846, significantly blocks repair of genomic damage, even increasing its severity.

This work demonstrated that the new strategy for ChIP-seq data analysis allows the identification of small transcriptional variations such as that derived from the transcriptional response to the BCL6/STAT5 switch, modulated by the Rac1/PAK1 pathway. It also revealed a role of the Rac1/PAK1-BCL6/STAT5 pathway in the response to DNA damage and repair, suggesting that its pharmacological inhibition may have therapeutic application in cancer, namely in potentiating the effects of chemotherapeutic agents with genotoxic effects.

Keywords: colorectal cancer; Rac1/PAK1; BCL6/STAT5; ChIP-seq; Response to and repair of DNA damage.

Índice

Agradecimentos.....	i
Resumo.....	iii
Abstract	v
Lista de Figuras	ix
Lista de Tabelas.....	x
Lista de Abreviaturas.....	xi
1. Introdução	1
1.1. Cancro Colorretal	1
1.2. Rho-GTPases e a sua importância.....	1
1.3. Rac1 e sua sinalização.....	3
1.4. Via de sinalização Rac1/PAK1-BCL6/STAT5	4
1.5. ChIP-seq	5
1.6. Resposta a danos no DNA.....	8
2. Objetivos	10
3. Materiais e Métodos.....	11
3.1. Cultura de Células	11
3.2. Transfecção de células	12
3.3. Imunofluorescência	14
3.4. Lise de Células	15
3.5. SDS-PAGE e <i>Western-blot</i>	15
3.6. PCR e RT-PCR.....	17
3.6.1. Síntese de DNA complementar (cDNA)	18
3.6.2. Desenho de oligonucleótidos iniciadores (<i>primers</i>).....	18
3.6.3. Amplificação semi-quantitativa de fragmentos cDNA	20
3.6.4. Deteção de Mycoplasma por PCR.....	20
3.6.5. Eletroforese de produtos de PCR	21
3.6.6. Análise densitométrica	21
3.6.7. qPCR	21
3.7. Transformação de bactérias, culturas líquidas e extração de DNA plasmídico	22
3.8. Ensaio do Cometa.....	23
4. Resultados	27
4.1. Resumo da análise bioinformática dos dados de ChIP-seq	27
4.2. Seleção de genes para validação experimental.....	28
4.3. Ativação e inibição da via Rac1/PAK1 em células DLD1	31

4.4. Avaliação morfológica de células DLD1 nas condições de ativação e inibição da via Rac1/PAK1.....	32
4.5. Avaliação do papel da via Rac1/PAK1 na regulação da expressão dos genes selecionados.....	33
4.5.1. Cálculo do coeficiente de correlação de <i>Pearson</i>	34
4.5.2. Confirmação dos resultados por qPCR.....	36
4.5.3. Identificação de motivos consensus para a resposta não-canónica do <i>switch</i> BCL6/STAT5.....	37
4.6. Seleção de genes candidatos por agrupamento funcional	37
4.7. Identificação de motivos de ligação ao DNA nos genes do cluster.....	40
4.8. Envolvimento da via Rac1/PAK1 na suscetibilidade a agentes genotóxicos	41
5. Discussão e Conclusão	44
6. Referências Bibliográficas	48
7. Anexos	54

Lista de Figuras

Figura 1.1 – Modelo proposto para a via de sinalização Rac1/PAK1-BCL6/STAT5.....	5
Figura 1.2 – Metodologia típica de ChIP-seq.	6
Figura 1.3 – Representação esquemática dos resultados esperados aplicando o filtro biológico.	7
Figura 3.1 – Esquema representativo do procedimento experimental do ensaio do cometa.....	24
Figura 4.1 – Diagrama de Venn representativo das <i>features</i> identificadas por ChIP-seq.....	27
Figura 4.2 – Representação gráfica do motivo de <i>consensus</i> previsto pelo algoritmo “ <i>MEME Suite</i> ” para a ocorrência do <i>switch</i> BCL6/STAT5 em resposta à via Rac1/PAK1.....	28
Figura 4.3 – Resultado da análise de anotação e agrupamento funcional	29
Figura 4.4 – Histograma representativo da distribuição do número máximo de <i>reads</i> /pico pelos 3 intervalos de contagens indicados.	30
Figura 4.5 – Indução da ativação e inibição da via Rac1/PAK1 em células DLD1.	31
Figura 4.6 – Imagens de microscopia confocal representativas da imunofluorescência de células DLD1 em três condições distintas do estado de ativação da via Rac1/PAK1.	32
Figura 4.7 – Imagem representativa de RT-PCR semi-quantitativo	33
Figura 4.8 – Resultados dos cálculos do coeficiente de correlação de <i>Pearson</i>	35
Figura 4.9 – Representação gráfica do motivo de <i>consensus</i> previsto pelo algoritmo “ <i>MEME Suite</i> ” para a ocorrência do <i>switch</i> BCL6/STAT5 na resposta não-canônica à via Rac1/PAK1.	37
Figura 4.10 – Resultados obtidos da análise de agrupamento funcional via D.A.V.I.D. para as 2419 <i>features</i> identificadas em resposta à ativação da via Rac1/PAK1.	38
Figura 4.11 – Resultados obtidos na análise, via GeneMANIA, do cluster funcional de genes envolvidos na resposta a estímulos por danos no DNA.	38
Figura 4.12 – Gráfico representando as variações, apresentadas na Tabela 4.6., na expressão de 4 dos genes do cluster funcional, em resposta à manipulação da via Rac1/PAK1.	39
Figura 4.13 – Análise, por qPCR, das variações, em resposta à via Rac1/PAK1, na expressão de 4 dos genes selecionados a partir do cluster funcional.....	40
Figura 4.14 – Regiões de máxima homologia encontradas via “ <i>MEME Suite</i> ” nos picos correspondentes aos 4 genes estudados do cluster funcional.....	40
Figura 4.15 – Ensaio do cometa.....	41
Figura 4.16 – Resultados obtidos do ensaio do cometa.....	42

Lista de Tabelas

Tabela 3.1 – Condições das transfeções realizadas em células DLD1.....	13
Tabela 3.2 – Anticorpos primários e secundários utilizados na análise por <i>Western blot</i> das amostras obtidas por lise celular.....	17
Tabela 3.3 – <i>Primers</i> específicos das sequências humanas do mRNA dos genes selecionados para validação experimental.....	19
Tabela 3.4 – Composição da <i>mix</i> por amostra utilizada para cada PCR e definição dos programas de PCR.....	20
Tabela 4.1 – Resumo dos números máximos e mínimos de <i>reads</i> detetados por ChIP-seq mediante a ativação e inibição da via Rac1/PAK1.....	27
Tabela 4.2 – Genes selecionados com os critérios estabelecidos, separados em 3 grupos mediante o seu número de <i>reads</i>	30
Tabela 4.3 – Resultado da análise densitométrica semi-quantitativa das intensidades das bandas dos RT-PCRs para os 15 genes selecionados.....	34
Tabela 4.4 – Comparação da amplitude de variação na expressão dos genes candidatos com diversos parâmetros potencialmente relacionados.....	35
Tabela 4.5 – Resultado da análise por qPCR da variação nos níveis de expressão dos genes indicados em resposta à ativação e inibição da via Rac1/PAK1.....	35
Tabela 4.6 – Resultado da análise, por RT-PCR semi-quantitativo, da influência da via Rac1/PAK1 na expressão dos genes do cluster funcional relacionado com vias de reparação e recombinação de DNA.....	39

Lista de Abreviaturas

A260nm, A280nm	Absorvância 260nm e 280nm
APC	<i>Adenomatous Polyposis coli gene</i>
APS	Persulfato de Amônio
BCL6	<i>B-cell lymphoma-associated transcriptional factor 6</i>
Ca ²⁺	Catão Cálcio
CaCl ₂	Cloreto de cálcio
cDNA	DNA complementar
ChIP-Seq	<i>Chromatin Immunoprecipitation - Sequencing</i>
cm ²	Centímetro quadrado
CO ₂	Dióxido de carbono
CRC	Cancro Colorretal (<i>Colorectal Cancer</i>)
DAPI	dihidrocloreto 4',6'-diamidino-2-fenilindole
DDR	Resposta a Danos no DNA (<i>DNA Damage Response</i>)
DMEM	<i>Dulbecco's Modified Eagle's Medium</i>
DMSO	Sulfóxido de dimetilo
DNA	Ácido desoxiribonucleico
dsDNA	DNA de dupla cadeia (<i>double-stranded DNA</i>)
ECL	Reacção enzimática quimiluminescente (<i>Enhanced Chemiluminescence</i>)
EDTA	Tetra-acetato de etileno diamina
EGTA	<i>Ethylene Glycol-bis(β-aminoethyl ether)-N,N,N',N'-Tetraacetic acid</i>
EMS	Sulfonato de Etil-Metano
F	<i>primer Forward</i>
FBS	<i>Fetal Bovine Serum</i>
FDR	<i>False Discovery Rate</i>
g	Grama
G/C	Guanina/Citosina
GAPs	Proteínas estimuladoras da atividade catalítica da GTPase (<i>GTPase Activating Proteins</i>)
GDIIs	Inibidores de Dissociação de nucleótidos de Guanina (<i>Guanine nucleotide-Dissociation Inhibitors</i>)
GDP	Guanosina difosfato
GEFs	Fatores de Troca de nucleótidos de Guanina (<i>Guanine Nucleotide Exchange Factors</i>)
GFP	Proteína fluorescente verde (<i>Green Fluorescent Protein</i>)
GTP	Guanosina trifosfato
h	Hora
H ₂ O ₂	Peróxido de hidrogénio
HBSS	<i>Hank's Balanced Salt Solution</i>
HGF	<i>Hepatocyte Growth Factor</i>
HPR	<i>Horseradish Peroxidase</i>
HR	Recombinação Hómologa
IPs	Imunoprecipitados
IR	<i>Radiação Ionizante</i>
JNK	<i>c-Jun N-terminal Kinase</i>

Kan	Canamicina
L	Litro
LB	Meio de Cultura <i>Luria Broth</i>
LF2K	<i>Lipofectamine 2000</i>
M	Molar
m/v	Peso por volume
mA	Miliampere
MAPK	Proteínas cinase ativadas por mitogénios <i>Mitogen Activated Protein Kinase</i>
MgCl ₂	Cloreto de Magnésio
min	Minuto
mL	Mililitro
mm	Milímetro
mM	Milimolar
MMS	Sulfonato de Metil-Metano
mRNA	RNA mensageiro
NF-κB	<i>Nuclear factor κB</i>
NHEJ	<i>Nonhomologous End Joining</i>
NP-40	Nonidet P-40
nm	Nanómetro
O/N	<i>Overnight</i>
O ₂	Oxigénio
PAGE	Electroforese em gel de poliacrilamida (<i>Polyacrylamide Gel Electrophoresis</i>)
PAK	<i>p21-activated kinase</i>
pb	Pares de bases
PBS	Tampão fosfato salino (<i>Phosphate Buffered Saline</i>)
PCR	<i>Polymerase Chain Reaction</i>
PVDF	Bifluoreto de polivinilideno
qPCR	PCR em tempo Real (<i>real-time PCR</i>)
R	<i>Primer-Reverse</i>
<i>r</i>	coeficiente de correlação de <i>Pearson</i>
RNA	Ácido ribonucleico
ROS	Espécies reactivas de oxigénio (<i>Reactive Oxygen Species</i>)
rpm	Rotações por minuto
RT	Transcriptase Reversa
RT-PCR	<i>Reverse Transcription Polymerase Chain Reaction</i>
SB	<i>Laemmli Sample Buffer</i>
SDS	dodecilsufato de sódio
seg	Segundo
STAT	<i>Signal Transducers and Activators of Transcription</i>
Ta	Temperatura de emparelhamentos (Temperatura de <i>annealing</i>)
Tm	Temperatura de desnaturação
TBE	Tampão Tris-Borato-EDTA
TK	Tirosina cinase
TRITC	Isotiocianato de Tetrametilrodamina
TSS	<i>Transcription Start Site</i>
Tx-100	Triton X-100

UV	Radiação <u>U</u> ltrav <u>i</u> oleta
V	Volt
WB	<i>Western blot</i>
v/v	Volume por volume
%	Porcentagem
μg	Micrograma
μL	Microlitro
μM	Micromolar
°C	Grau centígrado

1. Introdução

1.1. Cancro Colorretal

O cancro colorretal (CRC) é o terceiro tipo de cancro com maior incidência a nível mundial, especialmente em países desenvolvidos, com mais de 1 milhão de novos casos todos os anos (De Rosa *et al.* 2015), e é também um dos mais mortíferos, encontrando-se em quarto lugar como principal causa de morte por cancro (Brenner, Kloor, and Pox 2014).

O prognóstico de doentes com cancro colorretal tem evoluído favoravelmente nas últimas décadas, apresentando taxas de sobrevivência ao fim de 5 anos de 50%, no entanto, varia bastante com o estadió em que a doença é detetada. Caso a deteção ocorra numa fase inicial da doença, em que o tumor se encontra confinado à parede do intestino, a ressecção cirúrgica é normalmente curativa, com taxas de sobrevivência ao fim de 5 anos de 90% ou mais. Todavia, no estado IV da doença, onde já existem metástases distantes, este valor desce para 15% (Watson 2006).

Os principais fatores de risco incluem a história familiar de cancro colorretal ou doença inflamatória do intestino e fatores relacionados com o estilo de vida, como dieta, obesidade e diabetes, sedentarismo, tabagismo e consumo excessivo de álcool (Chen *et al.* 2016).

A maioria dos tumores colorretais esporádicos decorre da acumulação progressiva de alterações genéticas e epigenéticas, em oncogenes específicos ou genes supressores de tumor, que conduzem à transformação do epitélio normal do cólon em adenocarcinoma do cólon (Grady 2004). Assim, tipicamente o CRC tem origem numa lesão precursora pré-maligna que hiperprolifera dando origem a um adenoma benigno, que evolui para carcinoma e metastiza ao longo de um período de 10 anos (sequência adenoma-carcinoma) (De Rosa *et al.* 2015). Ao nível molecular, o evento inicial do processo tumorigénico do CRC tem origem na mutação do gene APC (*adenomatous polyposis coli*), que se verifica em mais de 70% dos casos, seguida de mutações ativadoras do oncogene KRAS e mutações inibitórias do gene supressor de tumor TP53. Estas mutações oncogénicas são frequentemente acompanhadas de instabilidade cromossomal (Watson 2006; Brenner, Kloor, and Pox 2014). Os restantes CRC esporádicos, cerca de 15%, desenvolvem-se através de outros eventos moleculares, nomeadamente, instabilidade de microssatélites, caracterizada por defeitos nos mecanismos de metilação e reparação de DNA (Watson 2006; De Rosa *et al.* 2015).

Nos últimos anos, outras alterações genéticas e moleculares relevantes têm sido identificadas no processo tumorigénico colorretal, entre as quais se encontram as alterações na sinalização celular regulada pelas Rho-GTPases (Espina *et al.* 2008).

1.2. Rho-GTPases e a sua importância

A família Rho de GTPases de baixa massa molecular (~21kDa) constitui um importante subgrupo da superfamília Ras, cujo arquétipo é o proto-oncogene Ras. Tal como as proteínas Ras, as Rho-GTPases funcionam como “interruptores” moleculares que amplificam e convertem sinais extracelulares em respostas celulares, fazendo parte de uma vasta rede de vias de sinalização. Por esta razão, as Rho-

GTPases estão envolvidas na regulação de processos celulares tão diversos como a proliferação e apoptose, organização do citoesqueleto, adesão e polaridade celulares, tráfego membranar e transcrição gênica (Benitah *et al.* 2004).

No geral, as Rho-GTPases são ativadas na sequência da ligação de ligandos específicos a recetores de fatores de crescimento, de citocinas, de hormonas, ou recetores acoplados a proteína G (Fritz and Henninger 2015). O seu mecanismo de ação é regulado através de um ciclo de alteração conformacional entre uma conformação inativa e uma ativa, dependente da ligação a guanosina difosfato (GDP) ou a guanosina trifosfato (GTP), respetivamente (Rossman, Der, and Sondek 2005). Uma vez no estado ativo são capazes de se ligar e interagir com várias moléculas efetoras, como as proteínas cinase p21 (PAK do inglês, *p21-activated kinase*), iniciando cascatas de sinalização a jusante, que culminam na resposta celular correspondente, até que uma hidrólise de GTP a GDP induza a inativação da GTPase e, assim, termine a transdução do sinal (Manneville and Hall 2002). Esta transição entre o estado ativo e inativo desenrola-se na membrana plasmática e é controlada *in vivo* por três classes de proteínas reguladoras: os fatores de troca de nucleótidos de guanina (GEFs) que estimulam a troca de GDP por GTP e assim, promovem a ativação das Rho-GTPases, as proteínas estimuladoras da atividade catalítica da GTPase (GAPs) e os inibidores de dissociação de nucleótidos de guanina (GDIs) que participam da sua inativação (Rossman, Der, and Sondek 2005).

As Rho-GTPases contam com 22 membros (Rojas *et al.* 2012), dos quais RhoA, Rac1 e Cdc42 são os melhor caracterizados e principalmente conhecidos pelos seus efeitos no citoesqueleto de actina, com consequências ao nível da morfologia e mobilidade celular (Espina *et al.* 2008). Assim, a ativação de RhoA promove a formação de fibras de acto-miosina, conhecidas como *stress fibers*, e adesões focais, enquanto que Cdc42 ativo induz a formação de *filopodia*, isto é, protuberâncias filamentosas na membrana plasmática, ricas em actina. Já a ativação de Rac1 é conhecida por estimular a polimerização de uma rede de filamentos de actina na periferia celular, resultando na extensão da membrana plasmática e na formação de uma estrutura designada por *lamellipodia* (Hall 1998; Matos *et al.* 2000).

Embora não se conheçam mutações oncogénicas na família Rho, existem várias evidências que a desregulação da sua sinalização influencia uma grande variedade de processos celulares associados à progressão tumoral, nomeadamente a transformação celular, sobrevivência, migração, metastização e angiogénese (Sahai and Marshall 2002; Del Pulgar *et al.* 2005). Assim, o papel das Rho-GTPases na tumorigénese de vários tipos de cancro decorre da sobre-expressão de vários dos seus membros ou de alterações na sua atividade, promovida, por exemplo, por mutações ativadoras em alguns dos seus GEFs ou em outros reguladores a montante (Fritz, Just, and Kaina 1999; Sahai and Marshall 2002; Ellenbroek and Collard 2007). Por outro lado, também, a amplificação da sinalização das Rho-GTPases, através da alteração dos níveis de expressão dos seus efetores a jusante, tem sido descrita para alguns tumores (Benitah *et al.* 2004; Del Pulgar *et al.* 2005).

Em particular, existem várias evidências que indicam que a sinalização celular dependente de Rac1 possui um papel essencial na transformação maligna (Espina *et al.* 2008), tendo-se encontrado sobre-expressão de Rac1 em alguns carcinomas, nomeadamente da mama, gástrico, do pulmão (Del Pulgar *et al.* 2005) e colorretal (Espina *et al.* 2008). Curiosamente, para além dos seus efeitos na aquisição de um comportamento metastático pelas células tumorais (Malliri, Kammen, and Clark 2002), o papel específico da sobre-expressão de Rac1 no desenvolvimento tumoral tem vindo a ser cada vez mais relacionado com a ativação de vias de sinalização que levam à regulação da expressão gênica (Del Pulgar *et al.* 2007).

1.3. Rac1 e sua sinalização

A GTPase Rac1 é codificada pelo gene RAC1, cuja deleção é letal para os embriões, evidenciando a importância das vias de transdução de sinal em que participa, apresentando uma expressão ubíqua (Matos *et al.* 2000). Para além de regular a reorganização do citoesqueleto de actina, controla também processos celulares como a proliferação celular, adesão e transcrição génica, interagindo com múltiplas moléculas efetoras (Jaffe and Hall 2005). A proteína cinase PAK1, por exemplo, foi das primeiras a ser descrita e representa um dos seus principais efetores, estando envolvida na regulação do citoesqueleto, adesão celular e transcrição (Bosco, Mulloy, and Zheng 2009). À semelhança de Rac1, esta cinase tem sido detetada sobre-expressa e/ou hiperativada em vários tumores humanos, sendo que a desregulação da sua sinalização tem sido associada a um aumento da capacidade de invasibilidade e de sobrevivência das células tumorais (Rakesh Kumar, Gururaj, and Barnes 2006). De facto, estima-se que o eixo de sinalização Rac1/PAK1 se encontre alterado em cerca de 60% dos tumores sólidos, estando esta alteração frequentemente associada à progressão tumoral, invasão e metastização e, por conseguinte, a tumores mais agressivos, com prognósticos clínicos mais desfavoráveis (Espina *et al.* 2008; Zhu *et al.* 2012), geralmente decorrentes do desenvolvimento de resistência à quimioterapia (Bostner, Ahnstro, and Sta 2007; Dokmanovic, Hirsch, and Shen 2009).

Adicionalmente, a sinalização de Rac1 induz a ativação das proteínas cinase ativadas por mitogénios (MAPK do inglês, *Mitogen Activated Protein Kinase*), p38 e JNK (do inglês, *c-Jun N-terminal Kinase*), que estimulam a expressão génica ao aumentarem a fosforilação e ativação dos seus fatores de transcrição alvo, c-Jun, ATF, ELK e AP-1 (Bosco, Mulloy, and Zheng 2009). Outro exemplo do papel de Rac1 na regulação da expressão génica é a estimulação do fator de transcrição NF- κ B (do inglês, *nuclear factor kappa B*), através de uma via de sinalização dependente da produção de espécies reativas de oxigénio (ROS do inglês, *reactive oxygen species*), que envolve a fosforilação e a degradação proteolítica das proteínas citoplasmáticas inibidoras I κ B α e NF- κ B2/p100 (Benitah *et al.* 2004). É de notar que a ativação do fator NF- κ B é um elo de ligação entre a resposta inflamatória e os processos que conduzem ao desenvolvimento tumoral, estando implicado em vários tipos de neoplasias (Sahai and Marshall 2002; Benitah *et al.* 2004).

Encontra-se também descrito que a sinalização de Rac1 modula a atividade dos fatores de transcrição da família STAT (do inglês, *Signal Transducers and Activators of Transcription*), nomeadamente, STAT3 e STAT5. Esta família, composta por sete proteínas de estrutura similar – STAT1, STAT2, STAT3, STAT4, STAT5a, STAT5b e STAT6, atua como mensageiro de sinais entre o citoplasma, onde se encontra latente, e o núcleo, para onde se desloca após ser ativada por fosforilação. Uma vez no núcleo agem como fatores de transcrição, com capacidade de ligação específica ao DNA, participando em processos fisiológicos como a diferenciação, angiogénese, proliferação e a apoptose. Para além da diversidade de funções sob condições fisiológicas, também a sua importância patogénica tem vindo a ser destacada ao longo do tempo, uma vez que STAT3 e STAT5 se encontram ativados num largo espectro de doenças hematológicas malignas, bem como em tumores sólidos, estando mesmo documentado uma ativação aberrante de STAT5 em carcinomas da próstata e do cólon (Kawashima *et al.* 2006).

Por último, destaca-se outro exemplo da relação de Rac1 com a regulação transcricional descrito pelo laboratório de acolhimento. O grupo observou que o repressor transcricional BCL6 (do inglês, *B-cell lymphoma-associated transcriptional factor 6*) é inibido em células de tumor colorretal aquando da ativação de Rac1, levando ao aumento da expressão de genes endógenos regulados negativamente por

este repressor. Este mecanismo de inibição requer a fosforilação de BCL6 pela cinase PAK1 que, por sua vez, é ativada a montante por Rac1 (Barros, Jordan, and Matos 2009).

O repressor BCL6 foi identificado como um dos genes mais frequentemente translocado em células B de linfoma não-Hodgkin, estando os genes que regula envolvidos na ativação e diferenciação de linfócitos, troca de isotipos de imunoglobulina (Ig) e implicados na regulação da inflamação e da progressão do ciclo celular (Dent, Vasanwala, and Toney 2002).

Curiosamente, o grupo de investigação do laboratório de acolhimento observou também que o motivo de DNA reconhecido pelo repressor BCL6 apresentava uma elevada homologia com a sequência de ligação do fator STAT5, levando à descrição de uma nova via de sinalização que correlaciona a atividade de Rac1 com estes dois fatores de transcrição em células de carcinoma colorretal (Barros *et al.* 2012).

1.4. Via de sinalização Rac1/PAK1-BCL6/STAT5

Dado que o motivo de DNA reconhecido pelos dois fatores de transcrição era semelhante, e ambos eram estimulados através da sinalização de Rac1, o grupo de investigação interrogou-se acerca da existência de papéis opostos na regulação transcricional e se a sinalização de Rac1 coordenava de alguma forma uma troca entre estes dois fatores de transcrição.

Em Barros *et al.* (2012), os autores descrevem em células colorretais que de facto a sinalização de Rac1 estimula a transcrição génica induzindo uma troca entre o repressor BCL6 e o ativador STAT5 no promotor de certos genes alvo. Os estudos efetuados, permitiram concluir que a ativação de Rac1, para além de estimular a inibição de BCL6 já referida acima, leva também ao aumento da fosforilação de STAT5 e à sua translocação nuclear. Mais importante ainda, a troca na ocupação do promotor entre BCL6 e STAT5, correlaciona-se diretamente com as alterações na expressão génica quer de um vetor repórter com o gene da luciferase de pirilampo, cujo promotor é controlado (reprimido) por BCL6, quer de três promotores endógenos de genes envolvidos no controlo do ciclo celular.

Com base nestes resultados, o grupo propôs um modelo para o papel da sinalização de Rac1 na troca transcricional observada entre os dois fatores de transcrição (Fig. 1.1). Neste modelo, um GEF específico de Rac1 promove a transição do estado inativo ao estado ativo da GTPase. Uma vez ativo, Rac1 liga-se e ativa a proteína cinase PAK1, que migra para o núcleo e fosforila o repressor BCL6 que se encontra ligado à cromatina, mais propriamente, ao promotor de um gene alvo, levando, assim, à sua inativação e à desocupação da ligação ao DNA. Em paralelo, o complexo formado entre Rac1 ativo e um GAP específico de Rac1 – MgcRacGAP – e STAT5, é fosforilado por uma tirosina cinase (TK) e posteriormente translocado para o núcleo. Aqui, estimulado pela ação de MgcRacGAP, a GTPase Rac1 volta ao estado inativo, libertando STAT5 fosforilado que reconhece o motivo de ligação no promotor deixado livre por BCL6 e se liga, ativando a transcrição desse gene, anteriormente reprimido (Barros *et al.* 2012).

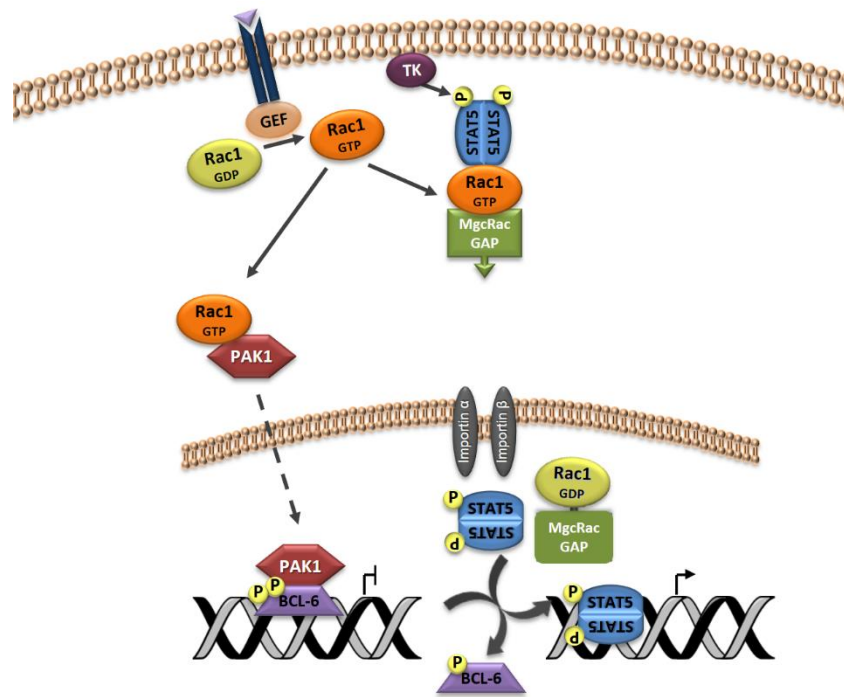


Figura 1.1- Modelo proposto para a via de sinalização Rac1/PAK1-BCL6/STAT5 (adaptado de Barros *et al.* 2012).

1.5. ChIP-seq

Os fatores de transcrição desempenham um papel crucial na regulação da expressão gênica (Barski and Zhao 2008). Estes reconhecem pequenas sequências específicas de DNA (designadas de motivos ou sequências *consensus*) no promotor de genes, às quais se podem ligar, após serem ativados por eventos de sinalização, e assim ativar ou reprimir a transcrição desses genes. Deste modo, a identificação e mapeamento do genoma dos locais de ligação dos fatores de transcrição, de outras proteínas que se ligam ao DNA e de marcas epigenéticas tornou-se essencial para melhor compreender a regulação transcricional. A técnica de imunoprecipitação da cromatina (ChIP do inglês, *Chromatin ImmunoPrecipitation*) é a ferramenta mais adequada e mais comumente usada para detectar interações entre proteínas e o DNA, pelo que a sua recente associação à sequenciação de nova geração, designada de ChIP-seq (*Chromatin ImmunoPrecipitation-sequencing*), tem permitido um progresso notável nesta área (Park 2009).

O método de ChIP-seq consiste na ligação covalente (*cross-linking*) entre o DNA e proteínas associadas com a cromatina no núcleo das células, levando à formação de complexos DNA-proteína que são posteriormente submetidos a sonicação, um método de fragmentação comumente utilizado neste procedimento que visa obter fragmentos de DNA com aproximadamente 200 pb nos complexos formados. Os fragmentos obtidos são, mais tarde submetidos a imunopurificação com um anticorpo específico da proteína de interesse presente nos complexos DNA-proteína. Por fim, os fragmentos de DNA isolados são purificados e em seguida sequenciados. Nesta fase, os milhões de fragmentos de DNA imunoprecipitados (IPs) são sequenciados (*reads*) e mapeados num genoma de referência, permitindo assim identificar os locais de ligação putativos da proteína de interesse no genoma (Liang and Kele 2012). O objetivo da análise de resultados obtidos por ChIP-seq cinge-se a encontrar regiões genômicas que se encontrem enriquecidas num conjunto de fragmentos específicos de DNA precipitado. Regiões de densidade elevada de sequenciação de *reads* são referidas como “picos”, e o *output* do *software* que

executa a metodologia de identificação de picos (*peak-calling*) é uma lista de “*peak calls*” que compreende as localizações genômicas dos locais inferidos a serem ocupados pela proteína de interesse (Valouev *et al.* 2008) .

Na abordagem tradicional de ChIP-seq (Fig. 1.2) sequenciam-se os IPs em paralelo com uma amostra do *input*, i.e., uma amostra recolhida antes do passo de imunoprecipitação, que após purificação servirá de controlo para a frequência com que as sequências detetadas nos IPs ocorrem espontaneamente no genoma em estudo. Isto permite utilizar um filtro estatístico para eliminar o “ruído de fundo” gerado por sequências sobre-representadas no DNA que estamos a analisar, o que permitirá evitar falsos positivos e distinguir locais de ligação específicos das proteínas (neste caso, fatores de transcrição) de interesse. Na prática, o nível desejado de significância estatística é geralmente dado em termos da taxa de falsos positivos (FDR do inglês, *False Discovery Rate*) ou pelo número de posições falsas-positivas esperado (*E-value*). Os algoritmos de mapeamento podem então usar a distribuição calculada do *background* para determinar o *score* mínimo que um pico mapeado deve ter para ser considerado significativo. Muitos falsos-positivos resultantes, por exemplo, de grandes regiões de DNA repetitivo, podem, assim, ser sistematicamente filtrados pela determinação da sua significância estatística. Na ausência de uma estimativa empírica da distribuição do *background* na amostra, é ainda possível usar um modelo analítico para o estimar matematicamente. O modelo mais simples é o modelo de *Poisson*, que considera que a distribuição de *reads* ao longo das regiões acessíveis do genoma segue uma distribuição de probabilidade de variável aleatória discreta (Kharchenko, Tolstorukov, and Park 2008). Esta distribuição permite estimar um valor limite para o *E-value*, possibilitando, em muitos casos, a filtragem da maioria dos falsos-positivos no tratamento dos resultados (Barski and Zhao 2008).

O problema destas abordagens é que, devido à complexidade e heterogeneidade dos sistemas biológicos, a utilização deste tipo de filtros estatísticos, baseado em dados empíricos (*input*) ou em modelos probabilísticos, seleciona preferencialmente sinais fortemente enriquecidos nas IPs (retângulo verde, Fig. 1.2), excluindo pequenos sinais que, embora positivos, se encontram próximos dos níveis do “ruído de fundo” (retângulos a laranja, Fig. 1.2).

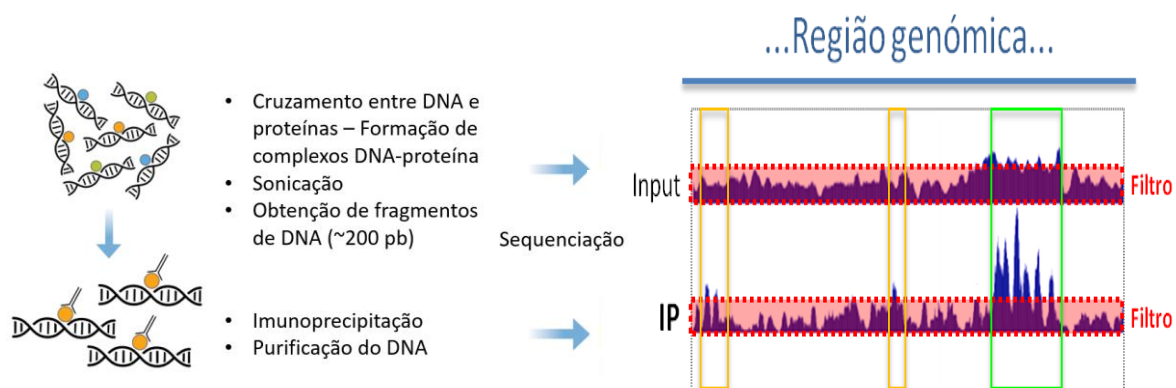


Figura 1.2 – Metodologia típica de ChIP-seq. Após a formação dos complexos DNA-proteína, estes são fragmentados em pequenas porções de DNA, retirando-se uma amostra (*input*) para determinar o nível de “ruído de fundo” (*background*) que servirá para estabelecer o limite (caixa pontilhada a vermelho - filtro) acima do qual os sinais detetados nas amostras de interesse (IP) serão considerados positivos. Sinais positivos, mas muito fracos (retângulos laranja), são filtrados por este método por não terem representatividade estatística.

Uma das linhas de investigação do grupo de acolhimento onde este trabalho foi realizado procurou identificar, através de ChIP-seq, todos os locais do genoma nos quais a expressão génica poderá ser

modulada pela via de sinalização Rac1/PAK1, através do *switch* dos fatores de transcrição BCL6/STAT5 caracterizado pelo mesmo grupo e descrito acima (Barros *et al.* 2012).

Acontece que, nos estudos anteriores, em que o grupo de acolhimento descobriu que a via Rac1/PAK1-BCL6/STAT5 modulava a expressão de um conjunto restrito de genes envolvidos no controlo do ciclo celular, também verificou que a amplitude de variação na expressão destes genes, a jusante desta via, não ultrapassava a de duas vezes a condição controlo (Barros *et al.* 2012). Assim sendo, a utilização da metodologia clássica na análise dos dados de ChIP-seq levantava um problema, pois variações desta ordem de grandeza são geralmente filtradas por esta metodologia juntamente com o “ruído de fundo”. Por essa razão, foi necessário criar uma nova estratégia que permitisse não excluir essas pequenas variações. A forma encontrada foi aplicar um filtro biológico antes do filtro estatístico. Este filtro biológico passa por explorar a seletividade acrescida conferida pelo *switch* transcricional que ocorre entre BCL6 e STAT5, aquando da inibição e ativação da via Rac1/PAK1-BCL6/STAT5, respetivamente, quando comparadas com condições controlo (i.e., condições basais - sem ativação nem inibição).

No sentido de validar este filtro biológico, realizaram-se ChIPs de BCL6 e STAT5 nestas três condições e desenvolveu-se um algoritmo em *R* para identificar entre os resultados da sequenciação destes, os picos que se manifestavam exclusivamente por ligação de STAT5 e saída de BCL6 do *locus* no DNA, aquando da ativação da via, e vice-versa aquando da inibição da mesma (Fig. 1.3).

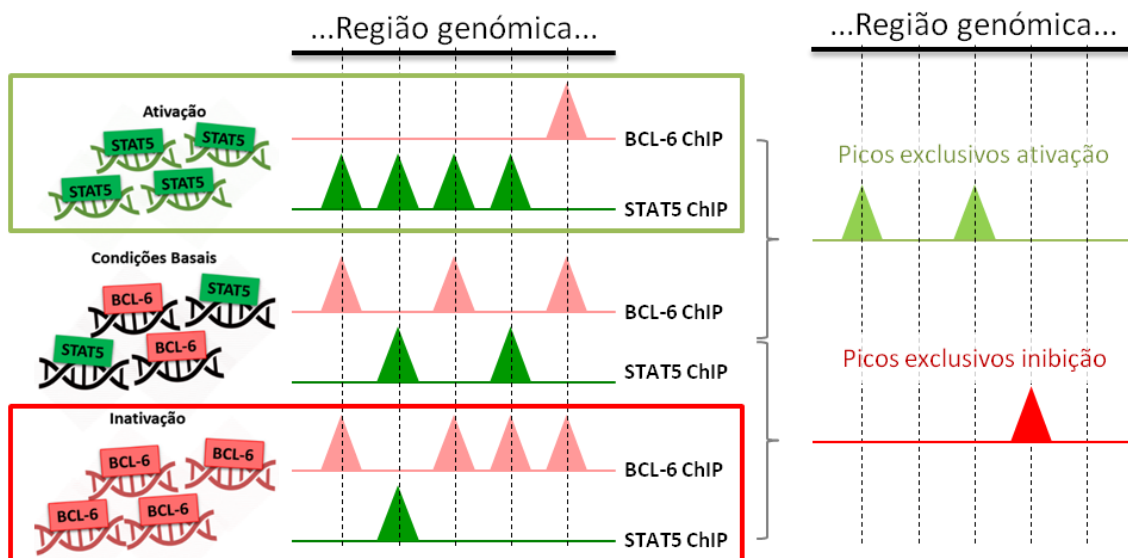


Figura 1.3 – Representação esquemática dos resultados esperados aplicando o filtro biológico. A aplicação do filtro biológico implica a ativação e inativação da via e permite a identificação dos picos presentes em cada uma das situações. A comparação entre os picos identificados quer na ativação quer na inativação, e as condições basais permite a identificação dos picos exclusivos a cada situação.

De uma forma muito simplificada, pode ver-se no esquema da Figura 1.3 um exemplo de picos exclusivos: assim quando se ativa a via, estimulando-se a ativação de Rac1 e PAK1, promove-se a troca de BCL6 por STAT5 em certos *loci*, ou seja, a ocorrência do *switch* entre os dois fatores de transcrição, que pode ser detetado por comparação com a ocupação dos mesmos *loci* nas condições basais. Por outro lado, a inibição da via deverá resultar no inverso, ou seja, a perda de ligação de STAT5 e ligação de BCL6 aos *loci* deixados vagos pelo primeiro.

A construção do algoritmo que permitiu fazer esta análise comparativa dos dados de ChIP-seq constituiu o foco da dissertação de mestrado do Licenciado Tiago Oliveira, orientado pelo Doutor Paulo Matos e pela Doutora Andreia Amaral, do grupo de bioinformática (GER) do centro de investigação BioISI da FCUL, liderado pela Professora Margarida Gama Carvalho. Neste sentido, considera-se que a sua descrição em maior detalhe se encontra fora do âmbito desta tese.

1.6. Resposta a danos no DNA

O DNA contido em cada uma das nossas células está sujeito a milhares de lesões por dia, geradas de forma espontânea durante o seu metabolismo ou induzidas por agentes ambientais, que caso não sejam corrigidas ou devidamente reparadas poderão dar origem a mutações ou aberrações genómicas.

As alterações espontâneas podem ocorrer durante a replicação do DNA pela inserção ou exclusão indevida de bases, ou através de modificações químicas na própria molécula de DNA, por desaminação (substituição do grupo amina (NH₂) por um átomo de oxigénio, dando origem a bases modificadas), despurinação (quebra da ligação entre as bases púricas e a desoxirribose deixando um local apurínico no DNA, ou seja, um local onde falta um nucleótido de adenina ou de guanina) ou alquilação (adição de grupos alquilo – metil ou etil – a bases de DNA). O DNA também pode ser danificado por ROS decorrentes do metabolismo normal da célula (Ciccia and Elledge 2010).

Relativamente aos danos no DNA de origem ambiental, estes podem ser físicos ou químicos. A radiação ultravioleta proveniente do sol (UV) será provavelmente o agente genotóxico ambiental mais conhecido, sendo também o mais penetrante (Jackson and Bartek 2009), podendo causar danos estruturais e oxidativos no DNA que podem conduzir ao desenvolvimento de cancro da pele (T Helleday *et al.* 2008). Outro exemplo de agente físico é a radiação ionizante (IR) que pode induzir a oxidação de bases do DNA e gerar quebras no DNA de cadeia simples ou dupla. Os compostos químicos utilizados em quimioterapia são exemplo de agentes químicos mutagénicos, causando uma grande variedade de lesões no DNA. Alguns atuam por alquilação, adicionando grupos alquilo a bases do DNA modificando-as, como é o caso do quimioterápico *metil-metanosulfonato* (MMS) ou do sulfonato de etil-metano (EMS), um composto com propriedades equivalentes utilizado com o mesmo propósito em laboratório. Outros compostos, como a cisplatina, intercalam-se no DNA, ligando-se a bases da mesma cadeia ou entre cadeias diferentes, formando aductos que alteram a estrutura da molécula e assim impedem a transcrição e replicação da mesma (Ciccia and Elledge 2010; Thomas Helleday, Eshtad, and Nik-zainal 2014).

Esta variedade e frequência de lesões no DNA são acompanhadas por uma complexidade de mecanismos de deteção e reparação de danos no DNA que combatem estas ameaças à integridade genómica, tentando assegurar a transmissão correta da informação genética à descendência e impedir o desenvolvimento de doenças, a tumorigénese e a transformação celular (Jackson and Bartek 2009; Ciccia and Elledge 2010). Este conjunto de mecanismos é conhecido como a resposta aos danos no DNA (DDR) do organismo e atua de uma forma concertada e coordenada por uma série de vias de sinalização, que interagem funcionalmente, e são definidas pelo tipo de lesão que processam. Assim, existem mecanismos que reparam por excisão de base, isto é, que removem bases de DNA danificadas, que reparam por excisão de nucleótidos, removendo aductos de DNA e aqueles que reparam *mismatches*, reconhecendo e corrigindo erros de introdução de bases não complementares durante a replicação. A reparação de quebras no DNA é feita essencialmente por recombinação homóloga (HR) ou por junção de extremidades não homólogas (NHEJ do inglês, *nonhomologous end joining*) (Lord and Ashworth 2012).

A maioria das vias DDR abrangem um conjunto semelhante de processos bem coordenados como a detecção de danos no DNA, a acumulação de fatores de reparação no local onde ocorreu a lesão no DNA, e finalmente a reparação física da lesão originada (Lord and Ashworth 2012). Existem evidências da interferência da sinalização de Rac1 na regulação da resposta a danos no DNA. Tendo sido demonstrado, que Rac1 ativo se encontra no núcleo aquando de lesões no DNA, enquanto que Rac1 inativo permanece no citoplasma, sustentando a hipótese da existência de um envolvimento direto de Rac1 nuclear na reparação do DNA. Estes dados são também suportados, pela existência de uma interação física entre uma proteína de reparação de excisão de base com a forma inativa desta GTPase (ligada a GDP). De facto, estas e outras evidências experimentais têm vindo a sugerir que a GTPase Rac1 pode ser uma coordenadora das respostas celulares ao stress genotóxico originado quer por danos causados no DNA, quer por outro tipo de danos a macromoléculas celulares (Fritz and Henninger 2015).

2. Objetivos

O principal objetivo deste projeto foi validar experimentalmente os resultados de uma nova estratégia de análise de dados de ChIP-seq, desenvolvida pelo laboratório de acolhimento, para caracterizar, a nível genómico, os genes alvo da via de sinalização Rac1/PAK1-BCL6/STAT5 (ver *Introdução*). Com esta validação, pretendia-se ainda selecionar, entre os múltiplos “*hits*” encontrados, genes, ou conjuntos de genes, cuja modulação da expressão por esta via de sinalização pudesse elucidar algumas das consequências pro-oncogénicas da sobre-ativação de Rac1 e PAK1 observada em tumores agressivos com mau prognóstico (ver *Introdução*).

3. Materiais e Métodos

3.1. Cultura de Células

A cultura de células é um processo no qual células são retiradas de um animal ou planta e mantidas em crescimento num ambiente artificial favorável. Um fator muito importante a ter em conta são as condições de cultura, uma vez que cada tipo de células tem as suas próprias condições favoráveis ao crescimento, pelo que o ambiente artificial a que são expostas tem que conter um substrato ou meio apropriado com os nutrientes essenciais, fatores de crescimento, hormonas, gases (como O₂, CO₂) e um ambiente físico-químico regulado (pH, pressão osmótica, temperatura) (Freshney 2000).

Este método revelou ser de grande importância para o estudo de tecidos humanos, dando provas de que era possível a formação de linhas celulares contínuas de tumores humanos, possibilitando, entre outros, a produção de vacinas antivirais e o estudo do desenvolvimento tumoral (Freshney 2000). Assim sendo, a cultura de células fornece excelentes sistemas modelo que permitem analisar as células a nível bioquímico e morfológico, bem como o efeito de fármacos ou compostos tóxicos, fazendo das linhas celulares um dos modelos mais utilizados em biologia celular e molecular.

Apesar de essenciais, as culturas celulares não podem estar sistematicamente em cultura, na medida em que após várias passagens podem ocorrer alterações genéticas e funcionais que modificam as características originais das células e podem desvirtuar resultados. A criopreservação é por isso, um procedimento da máxima importância em cultura celular, permitindo manter *stocks* de células num estado próximo ao qual são originários, livres de contaminação e prontos a ser utilizados. O melhor método de preservação é armazenar os *stocks* celulares em azoto líquido (-196°C) na presença de um de um criopreservante hidrofílico (os mais usados são o glicerol e o sulfóxido de dimetilo – DMSO). Os agentes criopreservantes reduzem o ponto de congelação e permitem uma congelação lenta e progressiva, reduzindo o risco de formação de cristais de gelo que pode causar morte celular. O DMSO, apesar da sua citotoxicidade, apresenta uma maior eficiência de criopreservação que, por exemplo, o glicerol, possivelmente por conseguir penetrar melhor a membrana celular (Freshney 2000).

Neste estudo utilizou-se a linha celular de cancro colorectal DLD1, que foi estabelecida a partir de um adenocarcinoma colorectal (ATCC® CCL-221™). Estas células apresentam uma morfologia epitelial aderente, desenvolvendo-se em monocamada.

Para começar a cultura desta linha celular, foi necessário proceder à descongelação rápida a 37°C de uma ampola com ~10⁷ de células, criopreservadas em azoto líquido, que após uma suave homogeneização se semearam num frasco de cultura com 75 cm² (T75) (VWR, <https://pt.vwr.com>). Este T75 continha 13 ml de meio Eagle's com a modificação de Dulbecco's (DMEM – *Dulbecco's Modified Eagle's Medium*, Gibco®, Thermo Fisher Scientific, www.thermofisher.com), suplementado com 10% (v/v) de soro fetal bovino (FBS – *Fetal Bovine Serum*, Gibco®, Thermo Fisher Scientific) e pré-aquecido a 37°C. Em seguida, colocou-se o frasco de cultura numa estufa de incubação a 37°C, com 5% de CO₂, sem estar completamente fechado.

Tendo em conta que as células são congeladas na presença de um agente criopreservante, o DMSO, que é tóxico para as células, o meio foi mudado no dia seguinte para o remover e assim melhorar a viabilidade celular. Quarenta e oito horas após terem sido semeadas, as células apresentavam uma confluência igual ou superior a 90%, reunindo assim as condições necessárias à sua sub-cultura. Estas

foram então submetidas a um processo de tripsinização (Trypsin + 0,05% EDTA (do inglês, *ethylenediaminetetraacetic acid*), Gibco®, Thermo Fisher Scientific)) e dependendo da sua confluência foi estabelecida uma diluição (geralmente de 1:3 a 1:8) a utilizar na nova cultura. Antes de cada tripsinização o meio foi retirado e foram feitas 2 lavagens com tampão fosfato salino – PBS (do inglês, *Phosphate buffered saline* – Gibco®, Thermo Fisher Scientific) + EGTA (do inglês, *ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid*) 0,05%, de forma a remover por completo o soro das células e, assim, a permitir a atuação da tripsina. O processo de tripsinização ocorre com a ação combinada entre a tripsina, que promove o isolamento das células ao hidrolizar as proteínas de membrana responsáveis pelas interações célula-célula e célula-substrato, e o agente quelante EDTA, que captura íons divalentes, como o Ca^{2+} , necessário para a adesão entre células mediada por caderinas, promovendo a dissociação celular e facilitando a ação da tripsina na destruição das ligações célula-substrato, o que agiliza o processo de isolamento da suspensão celular. A hidrólise de proteínas membranares, aumenta o tempo necessário a que, após passagem, as células recuperem a adesão, quer intercelular quer ao substrato. Se esta ação for demasiado agressiva corre-se o risco de induzir morte celular (*anoikis*). Assim, em culturas de monocamadas epiteliais, em que existem fortes ligações célula-célula, para reduzir o tempo de ação da tripsina e atenuar o seu efeito deletério é conveniente efetuar-se primeiro lavagens da cultura celular com PBS+EGTA, dado que o EGTA é também um agente quelante, mas mais específico para o Ca^{2+} , pelo que é mais eficiente na dissociação das ligações intercelulares, o que potencia a ação da tripsina e reduz o seu tempo de atuação.

Para prevenir possíveis infeções bacterianas ou fúngicas e monitorizar a proliferação e consequente sub-cultura, as células foram vigiadas regularmente. Possíveis contaminações com *Mycoplasma* são periodicamente verificadas por PCR (ver 3.6.4. Detecção de *Mycoplasma* por PCR) em todas as culturas celulares do laboratório de acolhimento.

No início da sub-cultura prepararam-se *pellets* de células para posterior congelação. Este processo iniciou-se com a centrifugação de uma suspensão celular (10 min a 200 g) provinda da tripsinização de dois T75, em que o *pellet* resultante foi ressuscitado suavemente em FBS com 10% de DMSO. Finalizando com a distribuição do volume final por várias ampolas de criopreservação (10^7 células/ml/ampola) que são congeladas a -80°C e armazenadas, após 48 h, em contentores de azoto líquido (-196°C).

A cultura de células foi executada em câmaras de fluxo laminar, sendo o material não estéril utilizado devidamente desinfetado, com álcool a 70%, antes de ser colocado a uso na câmara, para minimizar eventuais contaminações.

3.2. Transfeção de células

Transfeção consiste na introdução de DNA ou RNA exógeno em células com o objetivo de induzir nestas a expressão ou sobre-expressão ectópica de RNAs ou proteínas de interesse. Este é um poderoso método analítico que permite estudar não só a função génica como também a função e regulação proteica. O principal objetivo do uso desta técnica é o estudo da função de genes ou produto dos mesmos, através da ativação ou inibição da expressão de genes específicos nas células, ou a produção de proteínas recombinantes (Kim and Eberwine 2010).

As transfeções podem ser estáveis ou transientes. Uma transfeção estável consiste na introdução de material genético (transgene), juntamente com um gene marcador para seleção (e.g., um gene que confere resistência a um determinado antibiótico), no genoma hospedeiro, permitindo a expressão

continuada do transgene mesmo após a replicação das células hospedeiras. Por outro lado, numa transfeção transiente, o material genético transfetado apenas é expresso por um período limitado de tempo, não sendo integrado no genoma (Kim and Eberwine 2010).

Ao longo do tempo têm vindo a ser desenvolvidos vários métodos de transfeção, cada um deles com diferentes abordagens que variam com o tipo de células e o objetivo final, procurando todos maximizar a eficiência de transfeção e minimizar a toxicidade celular, por forma a reduzir os efeitos adversos na fisiologia normal das células mas mantendo a reprodutibilidade experimental e a facilidade de utilização (Kim and Eberwine 2010).

O uso de lipossomas catiónicos como mediadores da transfeção, acarreta algumas vantagens sobre os restantes métodos, nomeadamente, a baixa toxicidade para as células e uma boa eficiência de transfeção (Kamata *et al.* 1994). Isto porque as cargas positivas na superfície do lipossoma geram uma interação eletrostática com os ácidos nucleicos negativamente carregados, permitindo a formação de complexos lipossoma/ácido nucleico estáveis que, assim, são atraídos para a membrana plasmática (com carga negativa) e integrados no citoplasma da célula por endocitose (Dalby *et al.* 2004).

Pelas razões apresentadas, o reagente de transfeção utilizado neste trabalho foi o *Lipofectamine™ 2000* (LF2K – Invitrogen, Thermo Fisher Scientific) que opera segundo o método descrito acima e apresenta uma elevada eficiência de transfeção para a linha celular DLD1.

As transfeções realizaram-se em células DLD1 com uma confluência de aproximadamente 80% e iniciaram-se com a diluição dos DNAs plasmídicos a transfetar e de LF2K em Opti-MEM® I *Reduced Serum Medium* (Gibco, Thermo Fisher Scientific), de acordo com as instruções do fabricante. Após 5 min em repouso à temperatura ambiente, as preparações anteriores foram combinadas, seguindo-se 20 min de incubação para permitir a formação dos complexos lipossoma/DNA estáveis. Estes foram então adicionados gota-a-gota às células, tendo-se agitado suavemente o meio de cultura para assegurar uma distribuição uniforme dos complexos pela superfície da monocultura. Para aumentar a viabilidade celular o meio das células foi mudado antes da transfeção.

As transfeções foram efetuadas em placas de 10 cm² (35 mm – Thermo Fisher Scientific), segundo as condições descritas na Tabela 3.1 e repetidas por seis vezes. Os vetores plasmídicos para expressão em células de mamífero utilizados foram o vetor pEGFP vazio (GFP-ctrl) que serviu de controlo e o vetor pEGFP-Tiam1C1199 (GFP-Tiam1) para estimular a ativação da via Rac1/PAK1. 16 h após a transfeção, um dos replicados celulares transfetados com GFP-Tiam1 foi tratado, durante 4 h, com 20 µM de EHT 1864 (EHT – inibidor específico da atividade de Rac1) para mimetizar a inibição da via.

Tabela 3.1 - Condições das transfeções realizadas em células DLD1. Inclui a identificação e as quantidades respetivas de cada DNA plasmídico transfetado, assim como, a existência de tratamentos.

Condições	Plasmídeos	Quantidades (µg)	Tratamento com EHT
- 1 x 10 ⁶ cél/placa 10 cm ² - 2,5 µg DNA - 10 µl LF2K - 125 µl OptiMEM-DNA - 125 µl OptiMEM-LF2K	GFP-ctrl	2,5	-
	GFP-Tiam1	2,5	
	GFP-Tiam1	2,5	20µM EHT durante 4h

3.3. Imunofluorescência

Imunofluorescência é uma técnica de coloração histoquímica usada para demonstrar a presença e localização sub-celular de antígenos de interesse. Nesta técnica, os anticorpos são quimicamente conjugados a um marcador fluorescente (fluorocromo) e, uma vez marcados, ligam-se ao antígeno de interesse, permitindo assim a sua detecção por técnicas de microscopia de fluorescência (Robinson, Bs, and Kumar 2013).

Este método pode ser direto ou indireto, sendo este último mais sensível. A imunofluorescência direta, onde o anticorpo dirigido para o antígeno de interesse é diretamente conjugado ao fluorocromo, é menos utilizada. Na imunofluorescência indireta, o anticorpo que se liga especificamente ao antígeno de interesse (designado de anticorpo primário) não é marcado, sendo a detecção feita através de um anticorpo secundário, este sim conjugado com um fluorocromo, direcionado ao anticorpo primário (Robinson, Bs, and Kumar 2013). Nesta última abordagem, há uma amplificação do sinal de detecção, uma vez que várias moléculas do anticorpo secundário se podem ligar ao complexo formado pelo anticorpo primário e o antígeno de interesse.

Cada fluorocromo apresenta propriedades espectrais diferentes, como a intensidade e o padrão espectral de absorção e emissão, podendo, por isso, ser separadas com filtros seletivos recorrendo a um microscópio de fluorescência, permitindo o estudo em simultâneo de vários antígenos.

Neste estudo marcaram-se, como indicador da morfologia geral das células, os microfilamentos de actina, constituintes do citoesqueleto, com uma solução de faloidina (uma micotoxina produzida pelo fungo *Amanita phalloides* que se liga com alta afinidade aos microfilamentos de actina, impedindo a sua despolimerização) conjugada com o fluoróforo TRITC (isotiocianato de tetrametilrodamina), que emite na zona vermelha do espectro visível (595 nm). Para a marcação dos núcleos utilizou-se o dihidrocloreto 4',6'-diamidino-2-fenilindole (DAPI), dado que se intercala no DNA, ligando-se fortemente, e emitindo, quando excitado por luz ultravioleta, comprimentos de onda na gama do azul. E como as construções transfetadas, GFP-ctrl e GFP-Tiam1, contêm a proteína verde fluorescente (GFP-*Green Fluorescent Protein*) que apresenta capacidades intrínsecas de emissão de fluorescência (zona verde do espectro visível, 509 nm) foram detetadas sem recorrer a qualquer anticorpo específico.

Assim, células DLD1 foram crescidas em lamelas 10 x 10 mm, transfetadas e tratadas de acordo com as condições indicadas no ponto anterior (Tabela 3.1). Após 20 h de expressão, as lamelas foram lavadas, por três vezes, com PBS⁺⁺ (1x PBS + 0,5 mM MgCl₂ + 0,1 mM CaCl₂), seguida de fixação em 1x PBS contendo 4% de formaldeído (v/v), durante 30 min. O formaldeído é um agente fixador capaz de formar ligações cruzadas entre as proteínas intracelulares, “imobilizando” os processos metabólicos. Deste modo, a sua aplicação permite preservar os organelos, estruturas celulares e complexos proteicos que de outra forma seriam degradados ou dissociados, afastando a amostra do seu estado original e inviabilizando uma correta avaliação da distribuição sub-celular das proteínas de interesse.

Em seguida fizeram-se três lavagens, em agitação, com PBST (1x PBS, 0,05% Triton X-100 (Tx-100) (v/v)), durante 5 min. Para o passo de permeabilização, as células foram colocadas em PBS + 0,2% Tx-100 (1x PBS, 0,2% Tx-100 (v/v)) por um período de 10 min. Esta fase permite a entrada e ligação dos anticorpos às proteínas intracelulares, uma vez que esta solução dissolve as membranas lipídicas. Seguidamente realizaram-se outras três lavagens com PBST e prosseguiu-se com a incubação em câmara húmida, durante 30 min, com faloidina (dilução 1:500 em PBST (ECM Biosciences, www.ecmbiosciences.com)). No final desta incubação seguiram-se três lavagens, em agitação, com PBST cada uma de 5 min. Seguiu-se a coloração dos núcleos com DAPI (Sigma-Aldrich) – diluído

1:200000 em PBST), com leve agitação manual durante 30 seg. Realizaram-se outras três lavagens com PBST e fixou-se a imunomarcação com 1x PBS contendo 4% de formaldeído (v/v) durante 10 min. Após três lavagens com PBST, as lamelas foram montadas em lâminas com meio de montagem *VectaShield*[®] (Vector Laboratories, VWR, pt.vwr.com), seladas com verniz e observadas por microscopia confocal (Leica TCS-SPE, www.leica-microsystems.com).

3.4. Lise de Células

A lise de células é fulcral no tratamento de amostras para o estudo de processos intracelulares como a proteômica, metabolômica ou genômica. Consoante o objetivo e tipo de células podem ser usados métodos mecânicos, físicos, químicos e/ou biológicos na desagregação celular. Em células de mamífero, recorre-se frequentemente a procedimentos suaves como variações osmóticas, congelamento e descongelamento e, em particular, diferentes tipos de procedimentos de lise química (Tran *et al.* 2010; Mernier *et al.* 2012).

Neste trabalho, a lise de células foi feita em simultâneo para a obtenção de extratos proteicos e extração do RNA total.

A lise celular foi efetuada 20 h após a transfeção, em gelo, e iniciou-se com a remoção do meio das placas e lavagem com meio de lavagem frio (HBSS do inglês, *Hank's Balanced Salt Solution*). Em seguida, adicionaram-se 100 µL de tampão de lise com 1% de NP40 (composição no Anexo VI) a cada poço e após uma incubação de 5 min, raspou-se gentilmente o conteúdo celular e recolheu-se para tubos devidamente identificados. Os lisados foram depois centrifugados a 6000 rpm, a 4°C, durante 5 min. Para a preparação dos lisados proteicos retiraram-se 40 µL do sobrenadante e homogeneizaram-se em 10 µL de tampão de aplicação 5x concentrado (*Laemmli sample buffer* 5x – composição no Anexo VI), sendo de seguida desnaturados (95°C durante 10 min) e analisados por electroforese em gel de poliacrilamida de dodecilsulfato de sódio (SDS-PAGE) e *Western blot*. Para o isolamento do RNA total foi utilizado o sobrenadante remanescente e o *pellet* que se homogeneizaram em 350 µL de tampão de lise RA1 do kit Nucleospin RNA II (Macherey-Nagel, www.mn-net.com). Este kit foi também utilizado para purificar o RNA no lisado.

3.5. SDS-PAGE e *Western-blot*

SDS-PAGE é uma das técnicas mais comuns em ciências biológicas e biomédicas, sendo utilizada para separar proteínas em amostras biológicas, de acordo com o seu tamanho via eletroforese em gel.

Este procedimento consiste na preparação das proteínas a separar num tampão contendo dodecilsulfato de sódio (SDS), um detergente fortemente aniónico (Swank *et al.* 2006). Quando dissolvidas as moléculas de SDS têm uma carga líquida negativa dentro de uma larga gama de pHs e ligam-se às cadeias polipeptídicas na proporção da sua massa molecular relativa. Assim sendo, o rácio molar de SDS negativamente carregado para as proteínas é relativamente o mesmo para proteínas de diferentes pesos moleculares. Para além disto, as cargas negativas do SDS também destroem a maioria das conformações estruturais das proteínas, desnaturando-as. Este tratamento anula assim a carga elétrica intrínseca e a conformação tridimensional das proteínas, fazendo com que a principal força de condução para uma resolução vertical por eletroforese seja a taxa de migração dependente da sua massa molecular

(Jensen 2012). Adicionalmente, as propriedades porosas de um gel de poliacrilamida, aquando da passagem de corrente elétrica, completam o processo de separação, em que proteínas com baixo peso molecular migram mais depressa no gel, ao passo que proteínas com maior peso molecular migram mais lentamente (Swank *et al.* 2006). O tamanho das proteínas separadas é mais tarde estimado pela comparação com um marcador de peso molecular, com base nas suas distâncias percorridas em gel.

O SDS-PAGE é geralmente seguido de uma análise por *Western blot*, onde as proteínas que foram resolvidas em gel são electroforéticamente transferidas para uma membrana de nitrocelulose ou bifluoreto de polivinilideno (PVDF) que pode ser então marcada com anticorpos para as proteínas de interesse (Kurien and Scofield 2006).

Western blot é um método de deteção de proteínas específicas em amostras biológicas (Heidebrecht *et al.* 2009), sendo geralmente usado para identificar, estimar o tamanho e a abundância de proteínas específicas (Jensen 2012). A sua eficiência foi melhorada com a descoberta da quimiluminescência que veio permitir o aumento da sensibilidade e estabilidade na deteção do sinal.

Como já referido, este procedimento necessita primeiramente que haja a separação proteica mediante o seu peso molecular, em seguida a transferência das proteínas do gel para uma membrana, podendo estas ser então coradas com solução de *Ponceau* (PVDF e nitrocelulose) ou azul brilhante de *Coomassie* (PVDF), permitindo assim a visualização das proteínas separadas. Uma vez corada a membrana, prossegue-se com o bloqueio com uma solução proteica diluída (geralmente 2-5% de leite (p/v) em TBST (composição no Anexo VI)) para prevenir interações inespecíficas entre a membrana e os anticorpos escolhidos para detetar as proteínas de interesse (Jensen 2012). Após o bloqueio, segue-se a incubação da membrana com o anticorpo primário específico para as proteínas de interesse, seguida de algumas lavagens em TBST para remover o anticorpo não ligado, incubando-se novamente com o anticorpo secundário que irá reconhecer o complexo anticorpo primário-proteína. Normalmente, o anticorpo secundário encontra-se ligado a um enzima (tipicamente, o peroxidase de rábano (do inglês, *horseradish peroxidase* – HPR)), que em contato com um agente quimiluminescente produz luminescência nos locais onde complexo é reconhecido (Jensen 2012). A reação enzimática quimiluminescente (ECL) é um método sensível que pode ser usado para quantificar a abundância relativa das proteínas de interesse. Esta reação ocorre quando um agente quimiluminescente (como o luminol) é oxidado pelo peróxido de hidrogénio (H_2O_2), resultando na emissão de fotões nos locais de reconhecimento da proteína (local da reação), sendo esta emissão proporcional à quantidade de proteína presente no local e detetável mediante a sua exposição a chapas autorradiográficas (MacPhee 2010).

Procedeu-se, então, à separação dos lisados proteicos por SDS-PAGE em mini-géis de acrilamida com uma concentração de 10% (v/v) (composição no Anexo VI), durante cerca de 60 minutos, com uma amperagem constante de 20 mA por gel. Em simultâneo com as amostras foi aplicado um marcador de proteínas de peso molecular conhecido - *Precision Plus Protein™ Standards* (Bio-Rad, www.bio-rad.com).

Uma vez resolvidas no gel, transferiram-se as proteínas para membranas de PVDF (Bio-Rad), previamente equilibradas em metanol e hidratadas em H_2O bidestilada, utilizando um dispositivo de transferência da Bio-Rad, o *Mini Trans-Blot Cell*, que submerso em tampão de transferência 1x (composição no Anexo VI) foi submetido a uma diferença de potencial de 100 V, durante 60 min. Após a transferência, as membranas foram coradas com azul brilhante de *Coomassie* (composição no Anexo VI) e, em seguida, submersas numa solução descorante (solução *Destain* – composição no Anexo VI) para remover o corante em excesso, permitindo assim visualizar as proteínas transferidas e a eficiência da transferência. Em seguida, procedeu-se ao bloqueio das membranas com tampão de bloqueio (solução de leite magro em pó (5%) em tampão TBST – composição no Anexo VI) durante 16 a 18 h (*overnight*

– O/N) a 4°C. Concluído o bloqueio, realizaram-se três lavagens com TBST seguida de incubação de 1 h, em câmara húmida, com os anticorpos primários necessários (Tabela 2.2) (diluídos em tampão de bloqueio). Utilizaram-se dois anticorpos primários, o anti-GFP (Roche, www.lifescience.roche.com) que reconhece a proteína GFP presente em todas as construções transfetadas e o anti- α -tubulina que foi usado para servir de controlo da quantidade de proteína total carregada em cada poço do gel.

Após a fase de ligação do anticorpo primário às proteínas de interesse, realizaram-se três lavagens, em agitação, com TBST, uma de 10 min e duas de 5 min. Seguiu-se a incubação com o anticorpo secundário (Tabela 3.2) também diluído em tampão de bloqueio, durante 1 h, em agitação. A esta incubação, sucederam-se três lavagens de 10 min com TBST.

Tabela 3.2 - Anticorpos primários e secundários utilizados na análise por *Western blot* das amostras obtidas por lise celular. Apresenta informações relativas ao tipo de anticorpo, marca e diluições utilizadas.

Amostras	Anticorpos Primários			Anticorpos Secundários		
	Identificação	Marca	Diluição	Identificação	Marca	Diluição
Lisados celulares das transfeções com GFP-ctrl, GFP-Tiam1 (com ou sem EHT)	Monoclonal de murganho anti-GFP	Roche	1:500	Anticorpo de cabra anti-murganho-HRP	Bio-Rad	1:3000
	Monoclonal de murganho anti- α -tubulina	Sigma-Aldrich	1:10 000			

A fase de deteção iniciou-se com a preparação de duas soluções de ECL (composição no Anexo VI) que após se misturarem, na câmara escura, se utilizaram para incubar as membranas durante 1 min. Em seguida, as membranas foram colocadas e seladas em cassetes de autoradiografia e expostas a chapas autoradiográficas (Fujifilm, www.fujifilm.eu) por períodos variáveis de tempo de 1 seg a 15 min. Por fim, as chapas foram reveladas num revelador automático (Kodak Medical X-ray processor 102, www.kodak.com).

3.6. PCR e RT-PCR

A reação de PCR (*Polymerase Chain Reaction*) veio revolucionar as diversas áreas das ciências biológicas, trazendo enormes benefícios e desenvolvimentos científicos como a sequenciação do genoma, expressão génica em sistemas recombinantes, análises genéticas moleculares, incluindo a rápida determinação tanto da paternidade como do diagnóstico de doenças infecciosas. Esta técnica permite a amplificação *in vitro* de fragmentos específicos de DNA usando um enzima designado polimerase de DNA, que participa na replicação celular do material genético. Este enzima sintetiza uma sequência de DNA complementar, assim que um pequeno oligonucleótido (*primer*) se liga a uma das cadeias do DNA no local específico escolhido para o início da síntese. Os *primers* irão limitar a sequência a ser replicada e o resultado é a amplificação de uma sequência de DNA particular com biliões de cópias (Valones *et al.* 2009).

A *Reverse Transcription Polymerase Chain Reaction* (RT-PCR) é uma adaptação da técnica laboratorial de PCR, altamente sensível e bastante útil, por exemplo, para a detecção de transcritos específicos ou para a clonagem das sequências codificantes, completas ou parciais, de genes de interesse (Marone *et al.* 2001). Para que esta se realize, é necessária a conversão de fragmentos específicos de mRNA em DNA complementar (cDNA), por ação do enzima transcriptase reversa (RT), e em seguida a realização de técnicas convencionais de PCR, para se proceder à amplificação do cDNA sintetizado (Rio 2014).

3.6.1. Síntese de DNA complementar (cDNA)

O RNA total utilizado foi extraído e purificado com o kit Nucleospin RNA II (Macherey-Nagel), de acordo com as condições indicadas pelo fabricante. Em seguida, partindo de 2 µg de RNA, sintetizou-se cDNA para cada uma das amostras em estudo usando 0,5 µg de *random primers* (Invitrogen, Thermo Fisher Scientific) e *Ready-To-Go You-Prime First-Strand beads* (GE Healthcare, www.gelifesciences.com), de acordo com as instruções do fabricante. Este procedimento utiliza uma mistura de oligonucleótidos que representam todas as sequências possíveis de hexâmeros denominada por *random primers*, cuja função é ligarem-se ao RNA, servindo de molde para a reação da RT e iniciarem a síntese de cDNA (Nie and Singh 2001).

3.6.2. Desenho de oligonucleótidos iniciadores (*primers*)

Com o objetivo de amplificar as sequências correspondentes ao mRNA dos genes envolvidos na via em estudo desenharam-se pares de *primers* específicos para cada transcrito (Tabela 3.3). Para tal recorreu-se à ferramenta *PRIMER BLAST* (*NCBI database*, www.ncbi.nlm.nih.gov) e definiram-se os seguintes parâmetros: entre 18 a 23 pares de bases (pb), 40 a 60% de conteúdo G/C, tamanho do amplicão entre 100 a 200 pb, temperaturas de desnaturação (T_m) entre 58°C e 62°C e que não diferissem em mais de 3°C entre si, e de modo a limitar a amplificação apenas ao mRNA, pelo menos um dos *primers* deve estar contido entre uma junção exão-exão. Um dos elementos do par de *primers* é complementar à extremidade 5' da sequência direta (*forward* - F) e o outro complementar à extremidade 3' da sequência reversa (*reverse* - R).

Tabela 3.3 – Primers específicos das sequências humanas do mRNA dos genes selecionados para validação experimental. Na tabela constam o gene, a designação dos primers e a sua sequência, assim como a temperatura de emparelhamento (do inglês annealing, Ta) e o número de ciclos de PCR utilizado para cada reação.

Gene	Primer (F – Forward; R – Reverse)	Sequência (5'→3')	Ta (°C)	Ciclos
BMI1	BMI1-F	GTTACCTGGAGACCAGCAAGT	60°C	30
	BMI1-R	TTAGAGCCATTGGCAGCATCA		
CBX7	CBX7-F	TGCGGAAGGGTAAAGTCGAG	58°C	32
	CBX7-R	CTCTCCTCCTTCTCCTCGTAG		
H2AFX	H2AFX-F	GGCCTCCCAGGAGTACTAA	56°C	32
	H2AFX-R	CTCTTTCCATGAGGGCGGT		
HIPK1	HIPK1-F	AGTTCTCACGCAGGGAAGC	56°C	30
	HIPK1-R	GCAGTCTGTTCAACCAGCG		
HOXD3	HOXD3-F	TAATTGTGGTCACCTGGAGCC	60°C	38
	HOXD3-R	CTTCTGCATTGTGCACTCAGG		
LRP4	LRP4-F	ATGTTATCAGGCGAGCAGACC	58°C	35
	LRP4-R	CGCCTCAATGGTATTTTCGACC		
MSH2	MSH2-F	TGCAGTTTCATCACTGTCTGC	58°C	28
	MSH2-R	GCTCTGACTGCTGCAATATCC		
MSH4	MSH4-F	GGCCAGCTCAAGGTTTCATACT	58°C	35
	MSH4-R	CACTTGTGGTGTCCAGCTCTT		
MTUS1	MTUS1-F	ACTGGTGGACAACAACACAGC	60°C	28
	MTUS1-R	GGAAAGCTGCCTTGAGAT		
NABP1	NABP1-F	CGCGCCTGTCCCAATATGAAT	62°C	30
	NABP1-R	TGGTCACGCGTCCTATCTCC		
PDE4DIP	PDE4DIP-F	CTGTCAACAATGCAAACAACGC	60°C	32
	PDE4DIP-R	ACCAAGGATTACCTGTAAGC		
POLD4	POLD4-F	CTTCCAGTGCAGTCTCTGGC	64°C	34
	POLD4-R	GTTGAGCCTCTGACACCTCC		
RAD17	RAD17-F	AGTGGTGACTGGAATACACGC	60°C	28
	RAD17-R	AACTTGATCCTCCTCCTTGGC		
RAD51B	RAD51B-F	CAAGAGCTGTGTGACCGTCT	56°C	28
	RAD51B-R	TCCCATAAGCCGTTTGCATC		
SLC12A7	SLC12A7-F	CTGGAGAACCTGTGGAGTACG	60°C	30
	SLC12A7-R	CATGATACCGGTCACGGAAGG		
SLC39A12	SLC39A12-F	TCACCAAATGACAAGCACGGC	60°C	30
	SLC39A12-R	GCCTGTGAATCTTCTGGGCTT		
SLC7A7	SLC7A7-F	TATCCAGGACCAAGGAGGCA	58°C	30
	SLC7A7-R	ACGACGTTAAGCAGTGAG		
SPTBN1	SPTBN1-F	AGTGATGTCAACAACCGCTGG	60°C	28
	SPTBN1-R	TCACGCTCATCTGCCAGAGC		
STRN	STRN-F	CTGAAGCCTGGAATGTGGACC	60°C	28
	STRN-R	TGACCTATTGGGCCTCTTCACC		
QPolII	QPolII-F	CGCAATGAGCAGAACGGCGC	62°C	28
	QPolII-R	TCTGCATGGCACGGGGCAAG		

A amplificação de cada fragmento dos vários genes em estudo foi efetuada em paralelo com a amplificação de um fragmento do cDNA de RNA polimerase II (QPol II), como controlo interno, para normalização dos resultados, dado que é um gene de expressão ubiquitária e deste modo permite monitorizar e anular variações causadas por erros de pipetagem dos cDNAs utilizados.

3.6.3. Amplificação semi-quantitativa de fragmentos cDNA

As condições de amplificação de cada transcrito foram otimizadas experimentalmente para corresponderem à fase linear de amplificação do PCR, de forma a permitirem uma análise semi-quantitativa dos resultados. Para tal, em cada PCR, foram amplificadas diluições seriadas (1:1; 1:3; 1:6) de cDNA de DLD1 não transfetadas. Para cada amplificação realizou-se uma *mix* de reação, com um volume final de 25 µl, composta por Tampão B 1x (composição no Anexo VI), por 0,2 µM de cada um dos *primers* específicos e por 0,02 U/µl de *DreamTaq DNA Polymerase* (Thermo Fisher Scientific). Em seguida foi adicionado a cada uma das reações, com exceção das que desempenhavam função de controlo de PCR, 2 µL de cDNA da amostra respectiva. Depois de homogeneizadas, as amostras foram colocadas em termocicladores (Whatman Biometra®, www.biometra.de) e amplificadas segundo os programas descritos na Tabela 3.4.

Os programas basearam-se numa desnaturação inicial a 94°C durante 5 min, seguida de uma desnaturação a 94°C por 30 seg. As temperaturas de emparelhamento (do inglês *annealing*) foram específicas de cada reação de amplificação, de acordo com os *primers* utilizados (Tabela 3.3). O passo de extensão ocorreu durante 15 seg a 72°C para todas as reações dos genes em estudo e durante 30 seg a 72°C para a reação de QPol II. A extensão final foi feita a 72°C durante 5 min.

Tabela 3.4 – Composição da *mix* por amostra utilizada para cada PCR e definição dos programas de PCR. As temperaturas de *annealing* (Ta) e o número de ciclos utilizado na amplificação específica de cada um dos genes selecionado para validação experimental encontram-se na Tabela 3.3.

Transcrito	Mix por amostra	Programa	
Genes da via em estudo	- 0,2 µM <i>primer F</i> - 0,2 µM <i>primer R</i> - Tampão B (1x) - 0,02 U/µl <i>Dream Taq Polymerase</i>	94°C - 5 min 94°C - 30 seg Ta (°C) - 30 seg 72°C - 15 seg 72°C - 5 min	x ciclos
QPol II		94°C - 5 min 94°C - 30 seg 62°C-30 seg 72°C - 30 seg 72°C - 5 min	x 28 ciclos

3.6.4. Deteção de *Mycoplasma* por PCR

É essencial monitorizar as culturas celulares em laboratório para nos certificarmos de que estas se encontram livres de contaminações. Tal é feito através da sua observação recorrente ao microscópio, avaliando a sua taxa de crescimento, cor do meio, morfologia entre outras. Porém, nem todo o tipo de contaminações é visível ao microscópico, como é o caso da infeção pela bactéria *Mycoplasma*, um dos

maiores problemas da cultura de células *in vitro*, por crescer lentamente e não destruir as células hospedeiras, alterando, no entanto, o metabolismo da cultura celular de variadas formas. A contaminação por *Mycoplasma* é difícil de eliminar e pode espalhar-se a outras culturas que estejam em laboratório, pelo que é essencial testar regularmente a presença de *Mycoplasma* nas linhas celulares em cultura (Freshney 2000).

A detecção de *Mycoplasma* no laboratório de acolhimento foi feita, mensalmente, por PCR. Para cada teste realizado foram retirados 100 µL de meio de cultura do frasco de células, neste caso, DLD1 e colocaram-se a 95°C durante 10 min para desnaturar. Para um volume de reação, por amostra, de 25 µl que inclui Tampão B 1x (composição no Anexo VI), 0,2 µM de cada um dos *primers* específicos (Mico-F e Mico-R), 0,02 U/µl de *DreamTaq DNA Polymerase* (Thermo Fisher Scientific) e 3 µL da amostra desnaturada. O programa de PCR utilizado tem 45 ciclos e é baseado numa desnaturação inicial a 98°C durante 5 min, seguida de uma desnaturação a 94°C por 30 seg, um *annealing* de 56°C por 45 seg, uma extensão a 72°C por 30 seg e uma extensão final a 72°C durante 5 min.

Mico-F: 5' ACTCCTACGGGAGGCAGCAGTA 3'

Mico-R: 5' TGCACCATCTGTCACTCTGTAAACCTC 3'

3.6.5. Eletroforese de produtos de PCR

Posteriormente, os produtos de PCR foram separados por eletroforese, em tampão Tris-Borato-EDTA 1x (TBE – composição no Anexo VI), utilizando géis de agarose (Lonza, www.lonza.com) de 2,5% (p/v) de concentração, corados com brometo de etídio (0,4 µg/ml). Os géis foram sujeitos a uma diferença de potencial de 80 V durante 30-60 min e em seguida foram analisados e fotografados no aparelho *FireReaderV4* (Uvitec Cambridge, www.uvitec.co.uk).

3.6.6. Análise densitométrica

Para cada um dos genes em estudo foram realizados pelo menos três ensaios válidos de RT-PCR, que foram analisados e quantificados por densitometria da intensidade das bandas nos géis recorrendo ao software público *NIH ImageJ* (<https://imagej.nih.gov/ij/>).

3.6.7. qPCR

Outra variante do RT-PCR é a *Real-Time Reverse Transcription Polymerase Chain Reaction* – RT-qPCR, que é uma técnica estabelecida para a quantificação de mRNA em amostras biológicas (Huggett *et al.* 2005). Os benefícios do uso desta técnica incluem a maior sensibilidade, reprodutibilidade, precisão e melhor controlo da qualidade no processo de quantificação (Valones *et al.* 2009).

O PCR em tempo real monitoriza a amplificação do DNA enquanto esta ocorre, por meio de marcadores fluorogénicos que, ao interagirem com o DNA sintetizado *de novo* emitem fluorescência proporcional à quantidade presente em cada novo ciclo do PCR (Seifi *et al.* 2012). Este procedimento requer um termociclador com um sistema ótico para capturar a fluorescência e um computador com um software

capaz de recolher e processar os dados obtidos. A análise das curvas dos incrementos de fluorescência, permitem determinar entre que ciclos a reação de PCR se encontrava na fase exponencial, i.e., a fase em que o número de moléculas de DNA alvo duplica a cada ciclo, sendo proporcional a $2^{n^{\circ} \text{ ciclos}}$ da quantidade inicial de DNA na amostra. Definindo-se um nível ou patamar de fluorescência (T) em que todas as amostras estejam nesta fase, podemos, por comparação com séries de diluição de uma amostra com concentração inicial conhecida (quantificação absoluta) ou por comparação com uma amostra de referência única (quantificação relativa), extrapolar, respetivamente, a quantidade inicial em cada amostra ou em quantas vezes esta varia relativamente à amostra de referência, através da comparação entre o número de ciclos necessários, em cada amostra, para se atingir esse patamar (CT) (Arya *et al.* 2005).

Neste estudo, o composto fluorescente usado foi o *SYBR Green*, um fluoróforo com afinidade para se ligar ao DNA em dupla cadeia (dsDNA), sendo 100 vezes melhor que o brometo de etídio (Valones *et al.* 2009) e o aparelho de *real-time*, o ABI Prism 7500 *Sequence Detection System* (Applied Biosystems, Thermo Fisher Scientific).

Utilizou-se a quantificação relativa (método $\Delta\Delta CT$), em que a amostra de referência para cada um dos 6 ensaios foram as células transfetadas com GFP-ctrl. Para cada amostra preparou-se, em gelo, uma *mix* dos seguintes reagentes para um volume total de 25 μ l: 12,5 μ l SYBR Green PCR Master Mix (2x) (Applied Biosystems, Thermo Fisher Scientific), 2,5 μ l de *primer F* e *primer R*, 5 μ l de cDNA diluído 1:10 (cerca de 300 ng) e 2,5 μ l de água bidestilada para perfazer o volume. O programa de qPCR consistiu num passo de desnaturação inicial a 95°C durante 10 min, seguido de 40 ciclos de amplificação compostos de uma desnaturação durante 15 seg a 95°C e uma extensão de 45 seg a 64°C. À semelhança do RT-PCR foi também amplificado, em paralelo, o cDNA de QPol II para servir de controlo interno e normalizar os resultados à quantidade de RNA total em cada amostra.

3.7. Transformação de bactérias, culturas líquidas e extração de DNA plasmídico

Transformação é o processo pelo qual um organismo procarionte adquire DNA exógeno. Existem dois tipos de transformação bacteriana: transformação natural e artificial. A transformação natural descreve a absorção e incorporação de DNA em células do ambiente natural. Ao passo que a transformação artificial se refere a um vasto número de métodos para induzir a absorção de DNA exógeno (Sambrook and Russell 1999). Dos métodos utilizados na transformação artificial podem resultar bactérias quimio-competentes ou eletro-competentes. A competência de uma célula define-se pela capacidade da mesma receber DNA extra-cromossomal ou plasmídico. A necessidade de se trabalhar com células competentes advém do facto de o DNA ser uma molécula altamente hidrofílica não conseguindo penetrar na membrana celular bacteriana. Para que o DNA consiga atravessar a membrana celular bacteriana, as bactérias são ressuspensas numa solução com elevada concentração de cálcio que produz pequenos buracos na membrana das células bacterianas, por onde o DNA é forçado a entrar por choque térmico. Outra forma muito conhecida de transformação artificial é realizada por eletroporação, na qual as células bacterianas são sujeitas a um pulso elétrico de alta intensidade que aumenta significativamente a condutividade elétrica da membrana plasmática e por conseguinte, aumenta a sua permeabilidade (Das and Dash 2015). O método químico é o mais utilizado por ser menos agressivo, mais rentável e o mais adequado a bactérias *gram* negativas por terem uma parede celular menos substancial (Aune and Achmann 2010).

Em protocolos de clonagem, a transformação artificial é usada para introduzir o DNA recombinante na bactéria hospedeira. O crescimento das bactérias transformadas, depende da incorporação do plasmídeo pois este contém geralmente no seu *backbone* um gene de resistência a um antibiótico que é incluído no meio seletivo em que as bactérias irão crescer.

Neste estudo, as transformações realizadas foram utilizadas para amplificar DNA plasmídico purificado dos construtos de interesse (pEGFP-ctrl e pEGFP-Tiam1) que serão transfetados. Adicionou-se 1 µl (~10 ng) de DNA plasmídico a 10 µl (~10⁷) células quimicamente competentes de *Escherichia coli* da estirpe DH5α, seguindo-se um período de incubação de 15 a 30 min em gelo. A mistura foi, depois, sujeita a um choque térmico realizado por 40 seg a 42°C, regressando de imediato ao gelo. Após um período de 2 min adicionaram-se, à chama, 50 µl (5x o volume de bactérias) de LB líquido pré-aquecido à temperatura ambiente e, de seguida, incubou-se a mistura durante 1 h a 37°C, com agitação constante (~600 rpm). Posteriormente, plaquearam-se 20 µl da reação de transformação em placas, pré-aquecidas, de LB-Agar com canamicina (Kan, 30 µg/ml), uma vez que os plasmídeos utilizados têm um gene de resistência para este antibiótico. As placas foram colocadas a 37°C, invertidas, e no dia seguinte, uma colónia de cada placa (correspondente a cada um dos plasmídeos de interesse) foi inoculada em LB líquido com Kan e incubada a 37°C, O/N, com agitação constante (~220 rpm).

No dia seguinte procedeu-se à extração e purificação do DNA plasmídico das culturas líquidas utilizando o *innuPREP Plasmid Mini kit* (AnalytikJena, www.analytik-jena.de), tendo em conta as instruções do fabricante.

Este procedimento baseia-se num método modificado de lise alcalina que separa seletivamente o DNA cromossomal e proteínas celulares do DNA plasmídico, que é depois purificado por adsorção a uma matriz de sílica e recuperado quando em presença de água ou tampão TE.

O processo termina com a quantificação de DNA plasmídico obtido para cada amostra e com a medida do seu grau de pureza. Para tal foi medida, por espectrofotometria, a absorvância a 260 nm (Concentração = A_{260nm} x 50 x fator de diluição), bem como o rácio A_{260nm}/A_{280nm}, indicativo do grau de pureza da preparação (o valor ótimo é de 1.8).

3.8. Ensaio do Cometa

O ensaio do cometa, também designado de eletroforese em gel de uma única célula (do inglês, *single-cell gel electrophoresis*), tornou-se num dos métodos mais comuns para avaliar danos no DNA, com aplicações em testes de genotoxicidade, biomonitorização humana e epidemiologia molecular, ecogenotoxicologia bem como, na investigação dos mecanismos moleculares de geração de danos no DNA e sua reparação. Esta metodologia, encontrando-se ainda em desenvolvimento, apresenta características vantajosas como a sua simplicidade, sensibilidade, versatilidade, velocidade e economia. É também uma das técnicas usadas na área de investigação do cancro para a avaliação da genotoxicidade (A. R. Collins 2004).

Este procedimento foi assim designado, por a imagem resultante se assemelhar a um “cometa” com cabeça e cauda distintos, onde a cabeça é composta pelo DNA intacto, enquanto que a cauda consiste nos fragmentos de DNA resultantes de danos ou quebras simples ou duplas na cadeia de DNA (Sanadi *et al.* 2004). O “cometa” pode ser analisado digitalmente, extrapolando-se a extensão dos danos no DNA pela quantificação da intensidade do sinal (pixéis) na “cabeça” e “cauda”, a partir de imagens digitais

captadas por um microscópio com um software desenvolvido para este propósito (Olive and Banáth 2006).

Tecnicamente, este método envolve o tratamento das células com um agente genotóxico ou outro que leve ao dano das moléculas de DNA, sendo as células posteriormente embebidas em agarose, numa lâmina de microscópio, sujeitas a lise com uma solução hipertônica contendo um detergente não-iônico que remove todas as membranas e proteínas citoplasmáticas, preservando, no entanto, a estrutura da cromatina nos núcleos. Estes são então sujeitos a condições alcalinas ou neutras, o que torna a estrutura da cromatina mais relaxada, expondo o DNA e permitindo que ao ser aplicada uma corrente elétrica os fragmentos de DNA danificados migrem mais que o DNA intacto. Após coloração com um marcador fluorescente específico de DNA (por exemplo, o brometo de etídio), o gel é então observado ao microscópio de fluorescência e através de um software apropriado calculada a percentagem de fluorescência existente na cabeça e na cauda do cometa, bem como o seu comprimento (A. R. Collins 2004; Nandhakumar *et al.* 2011). A percentagem de DNA fragmentado na cauda do cometa é uma medida direta dos danos no DNA.

Para avaliar a influência do estado de ativação da via Rac1/PAK1 na indução de danos no DNA e na capacidade de os reparar foi realizado um ensaio de cometa em células DLD1 em três condições diferentes de ativação desta via: condições basais; ativação da via, através do tratamento das células com HGF (do inglês, *Hepatocyte Growth Factor*), um fator de crescimento responsável pela ativação de Rac1; e inibição da via, pelo tratamento das células com EHT. Cada uma das condições anteriores foi exposta a um agente genotóxico, neste caso, o sulfonato de etil-metano (EMS) ou ao veículo no qual este agente foi dissolvido. Adicionalmente, o desenho experimental incluiu o tempo de recuperação 0 h e 4 h após a exposição, ou não, ao agente genotóxico. Todos os tratamentos foram feitos em duplicado para amostra. A Figura 3.1 apresenta um esquema resumindo o procedimento experimental atrás descrito.

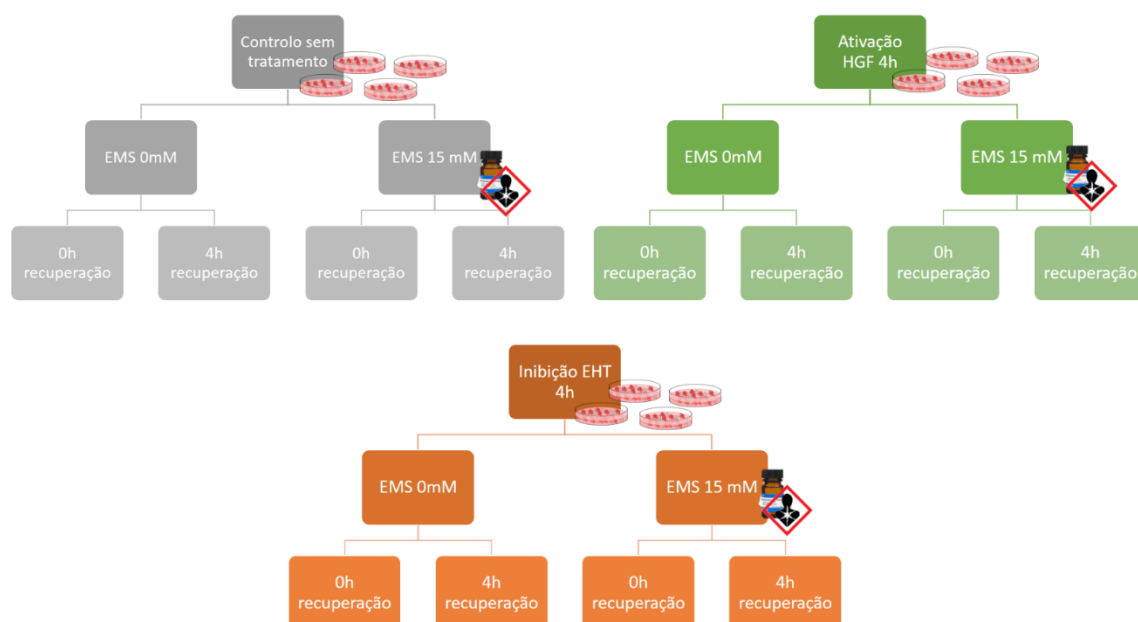


Figura 3.1 – Esquema representativo do procedimento experimental do ensaio do cometa.

Assim, prepararam-se células DLD1 com uma confluência de 40 a 50% (aproximadamente $0,25 \times 10^5$ células/poço) numa placa 24-well (Thermo Fisher Scientific). No dia seguinte mudou-se o meio para DMEM com apenas 1% (v/v) de FBS (DMEM 1%), um procedimento chamado de “starvation”. Este passo é essencial ser realizado antes das células serem tratadas com um fator de crescimento, dado que o soro (FBS) é também rico em fatores de crescimento e é importante anular o efeito desses antes de observarmos o resultado de outro tratamento. Passadas 24 h, com as células a uma confluência de cerca de 80%, procedeu-se ao tratamento com HGF (50 ng/ml), EHT (20 μ M), ou apenas novo DMEM 1%, nos poços respetivos, durante 4 h. Durante este tempo as células foram incubadas a 37°C, com 5% de CO₂.

Posteriormente, removeu-se o meio de todos os poços e lavou-se com PBS 1x (Gibco®, Thermo Fisher Scientific). Aos poços que não foram sujeitos a tratamento genotóxico foi adicionado DMEM 1%, enquanto aos outros foi adicionado 15 mM EMS diluído em DMEM 1% (500 μ l/poço), durante 1 h, mantendo mais uma vez as células a 37°C, com 5% de CO₂. Em seguida tirou-se o meio e lavaram-se todos os poços com PBS. Aos poços que seriam avaliados após 4 h de recuperação da exposição ao EMS foi adicionado DMEM 1%. Quanto aos poços com 0 h de tempo de recuperação foi iniciado o processo de tripsinização (utilizando 300 μ l de tripsina) e ficando a incubar durante sensivelmente 5 min na estufa a 37°C. Quando as células já se encontravam soltas, ressuspenderam-se em 600 μ l de DMEM 1% e transferiu-se a suspensão celular para tubos de 1,5 ml. Procedeu-se à sua centrifugação a 1200 rpm, a 4°C, durante 10 min. Após a centrifugação, descartou-se o sobrenadante e os *pellets* foram ressuspensos no volume residual remanescente.

Previamente, prepararam-se e identificaram-se lâminas revestidas com agarose (*Electrophoresis grade*, Sigma) e dissolveu-se 1% de *low melting point* agarose (composição no Anexo VI) a 70°C que, posteriormente, se colocou a 38°C e onde permaneceu até ser utilizada.

Às ressuspensões dos *pellets* anteriores foram adicionados 80 μ L de 1% *low melting point* agarose (composição no Anexo VI) a 38°C. Esta mistura foi rapidamente homogeneizada e distribuída em duas gotas numa lâmina revestida com agarose (Fig. 3.2), colocando-se de imediato uma lamela por cima de cada gota e em seguida, posicionando a lâmina em cima de uma caixa metálica incubada no frio. Este procedimento dá origem à formação de mini-géis contendo os núcleos das células obtidos nas várias condições do ensaio. Terminada esta fase, as lâminas são protegidas da luz para evitar qualquer tipo de efeito genotóxico ambiental e deixadas a 4°C entre 10 a 15 min.

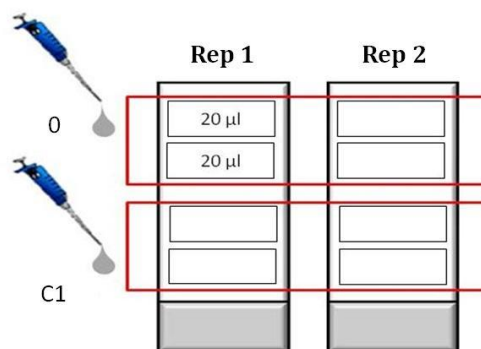


Figura 3.2 – Imagem representativa das lâminas com os mini-géis.

Por fim, depois de se retirarem as lamelas, as lâminas são embebidas em solução de lise (composição no Anexo VI) e incubadas entre 1 h a 24 h, a 4°C.

O procedimento anterior foi repetido para as células com tempo de recuperação de 4 h, até que todas as lâminas se encontrassem em solução de lise.

Finalizado o processo de lise, incubaram-se as lâminas em tampão de eletroforese alcalino (composição no Anexo VI), previamente mantido a 4°C, já na tina de eletroforese, durante 30 min. Em seguida, realizou-se a eletroforese com uma voltagem constante de 28 V, durante 25 min, tentando manter a amperagem perto dos 300 mA. Após a eletroforese, colocaram-se as lâminas em tampão de neutralização (composição no Anexo VI) frio, durante 10 min, seguida de outra incubação de 10 min em água bidestilada fria. Por fim, deixaram-se as lâminas a secar O/N à temperatura ambiente.

No dia seguinte, iniciou-se o processo de coloração das lâminas para posterior análise. Este processo incluiu a adição de 20 µL de brometo de etídio (0,125 µg/µL) a cada mini-gel, cobrindo-se cada um deles com uma lamela e incubando em câmara húmida, protegida da luz, durante 30-45 min a 4°C.

As lâminas foram observadas ao microscópio de fluorescência *Zeiss AxioPlan2*, acoplado a um sistema digital de aquisição de imagens. O *software OpenComet* (www.opencomet.org) foi usado para medir os cometas de 100 células entre os replicados de cada condição experimental.

4. Resultados

4.1. Resumo da análise bioinformática dos dados de ChIP-seq

Como descrito na Introdução, o trabalho desenvolvido nesta dissertação teve por base um trabalho de colaboração entre o grupo de acolhimento e o grupo de bioinformática (GER) do centro de investigação BioISI da FCUL, liderado pela Professora Margarida Gama Carvalho. Desta colaboração, em que estiveram envolvidos a Doutora Andreia Fonseca e o Licenciado Tiago Oliveira, resultou um algoritmo desenvolvido especialmente para a análise dos dados de ChIP-seq obtidos pelo grupo de acolhimento, que é o foco de base da dissertação de mestrado do licenciado Tiago Oliveira e, como tal, a sua descrição pormenorizada encontra-se fora do âmbito desta tese. Contudo, para a contextualização do leitor, é necessário dizer que desta análise foi obtida uma lista de 2566 picos localizados num intervalo de 3000 pares de bases a montante e 1000 pares a jusante de *loci* anotados como correspondentes a locais de iniciação da transcrição génica (TSS do inglês, *Transcription Start Site*), doravante designados por *features* (Anexo I). Destes, 2419 surgiram aquando da ativação da via, correspondendo, portanto, a loci onde ocorreu a troca de BCL6 por STAT5, e 147 surgiram aquando da inibição da via, ou seja, correspondendo a loci onde ocorreu a troca de STAT5 por BCL6. Foram ainda encontrados 17 *features* que apresentavam picos próximos tanto na ativação como na inibição da via Rac1/PAK1. Após análise da localização de cada pico relativamente a cada *feature*, concluiu-se que estes representavam mais do que um *loci* responsivo à via na vizinhança destas *features*, provavelmente com respostas antagónicas à manipulação do seu estado de ativação (Anexo I e Fig. 4.1).

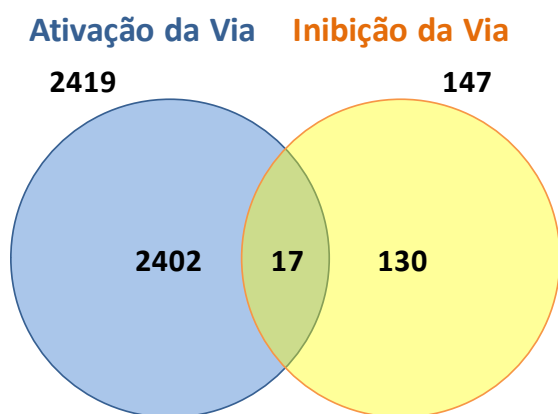


Figura 4.1 – Diagrama de Venn representativo das *features* identificadas por ChIP-seq na ativação (a azul) e na inativação da via Rac1/PAK1 (a amarelo) e em ambas as situações (a verde).

É de notar ainda que, tal como previsto, a amplitude dos picos encontrados, i.e., o número de *reads* (fragmentos de DNA sequenciado) por pico, é baixa, compreendendo-se entre 2 e 20 *reads* na ativação e entre 2 e 15 *reads*, na inibição da via (Tabela 4.1).

Tabela 4.1 – Resumo dos números máximos e mínimos de *reads* detetados por ChIP-seq mediante a ativação e inibição da via Rac1/PAK1.

Tratamento	Número de <i>reads</i> /pico	
	Máximo	Mínimo
Ativação	20	2
Inibição	15	2

É ainda de referir que foi também realizada uma análise bioinformática com vista a encontrar motivos putativos de *consensus* para a ligação dos fatores BCL6 e STAT5 em resposta à via Rac1/PAK1 (motivos específicos do *switch* destes dois fatores). Uma sequência de DNA *consensus* representa o resultado de alinhamentos múltiplos de sequências caracterizadas experimentalmente como passíveis de serem reconhecidas por um determinado fator. Estas são comparadas entre si e os motivos são calculados com base na frequência com que cada nucleótido aparece em determinada posição. Os motivos de sequência, são geralmente curtos, e consistem em padrões recorrentes no DNA, que se presume estarem associados a uma função biológica particular (D’Haeseleer 2006). Geralmente, estes indicam locais de ligação específica para proteínas como nucleases ou fatores de transcrição. Existem, atualmente, métodos computacionais de acesso público que permitem a identificação *in silico* quer de regiões regulatórias nos promotores de genes de interesse, quer a de novos motivos de *consensus* com base em sequências identificadas, por exemplo, através de ChIP-seq. Neste trabalho, foi utilizado o algoritmo “MEME Suite” (Bailey *et al.* 2006), disponível em <http://meme-suite.org/>, procedendo-se a uma análise das regiões de 250 pb a montante e a jusante do centro de cada um dos picos detetados (segundo o descrito por (Ma, Noble, and Bailey 2014)). Através desta análise foi encontrado o motivo representado na Figura 3.2 A, com $E = 5 \times 10^{-13}$ (valor representativo da possibilidade do motivo encontrado ser um falso positivo). Por comparação com os motivos canónicos dos fatores BCL6 e STAT5 (Fig. 4.2 B e C, respetivamente) observa-se uma homologia considerável para a sequência {G}R[AT]TYY[AT] (Fig 4.2, retângulo cinza).

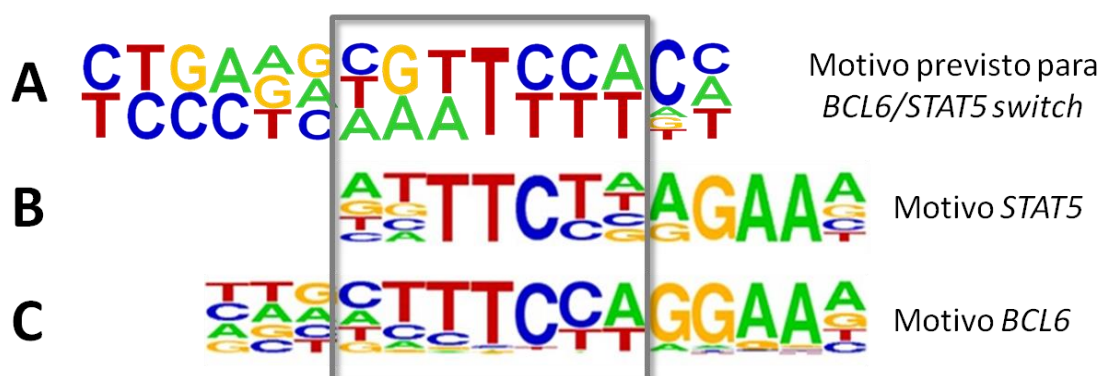


Figura 4.2 – (A) Representação gráfica do motivo de *consensus* previsto pelo algoritmo “MEME Suite” para a ocorrência do *switch* BCL6/STAT5 em resposta à via Rac1/PAK1. (B) Motivo *consensus* canónico para o fator STAT5. (C) Motivo *consensus* canónico para o fator BCL6. O retângulo cinza destaca a região de maior homologia entre os 3 motivos.

4.2. Seleção de genes para validação experimental

O passo seguinte foi seleccionar um conjunto de genes, definidos pelos *features* na vizinhança dos picos nesta lista, para validar experimentalmente se a manipulação da via Rac1/PAK1 induziria ou não variações nos seus níveis de expressão. Como os dados anteriores do grupo de acolhimento (Barros *et al.* 2012), bem como o número de picos obtidos nos indicou que, em estado basal, a via Rac1/PAK1 se apresentava pouco ativa nestas células, foi escolhido iniciar a validação experimental pela lista de 2402 *features* que respondiam em exclusivo à ativação da via, excluindo-se os 17 que variavam também na inibição da via, pois poderiam dificultar a interpretação dos dados resultantes da análise da expressão

dos genes associados. Foram ainda estabelecidos critérios de seleção adicionais, com vista a maximizar a hipótese de se observarem variações significativas ao nível da expressão dos genes analisados:

1. Excluíram-se *features* correspondentes a miRNAs, RNAs ribossomais, pseudogenes e genes “putativos” cuja estabilidade do mRNA e a sua tradução em proteína não estivesse confirmada e caracterizada;
2. Excluíram-se *features* na vizinhança de picos com dimensão inferior a 50 pb (dimensão média de uma *read*);
3. Excluíram-se *features* cujo pico correspondente estivesse a mais de 1500 pb a montante do TSS – região promotora clássica;
4. Excluíram-se *features* cujo número de *reads* médio (entre os replicados do ChIP seq) para o pico no controlo (condições basais) fosse igual ou superior ao número de *reads* médio para o pico correspondente nas condições de ativação da via.

Utilizando estes critérios reduziu-se a lista de *features* a 419 entradas (ver Anexo II). No sentido de facilitar o processo de seleção, estas 419 entradas foram ainda submetidas a uma análise de anotação e agrupamento funcional utilizando o algoritmo D.A.V.I.D. (*Database for Annotation, Visualization and Integrated Discovery*) disponível em <https://david.ncifcrf.gov/>. Verificou-se que, das 419 *features*, a maioria, 267, se agrupavam sob o classificador “*sequence variant*” (Fig. 4.3 e Anexo III), pelo que se optou por utilizar este grupo para selecionar o conjunto de genes a validar experimentalmente.

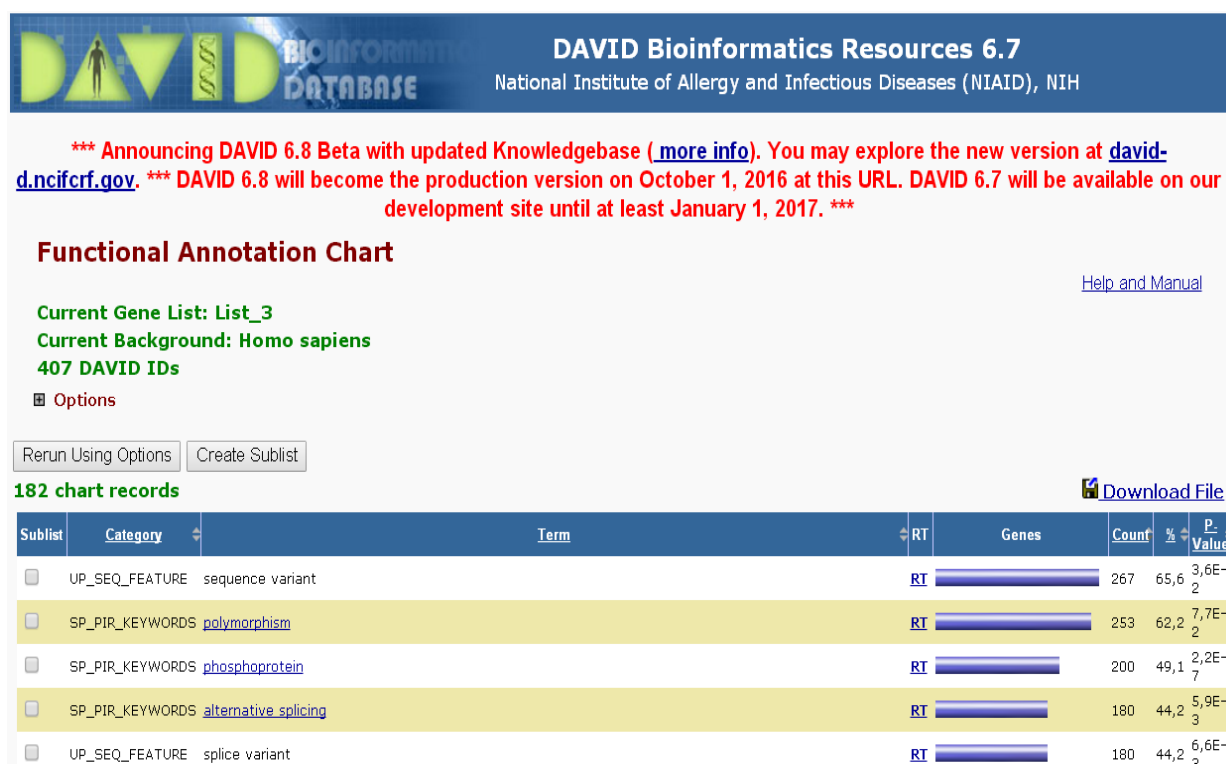


Figura 4.3 – Resultado da análise de anotação e agrupamento funcional, através do algoritmo D.A.V.I.D., das 419 features que passaram o primeiro conjunto de critérios de seleção.

Para tal, criou-se uma nova lista contendo os identificadores das *features* anotadas, juntamente com o número máximo de *reads* detetado para o pico correspondente aquando da ativação da via Rac1/PAK1 (Max_Act in Anexo IV). Na tentativa de verificar alguma correlação entre o número de *reads* e variações

na expressão dos genes definidos pelas respectivas *features*, definiram-se, após ordenação segundo o parâmetro Max_Act, 3 intervalos, representados no histograma da Figura 4.4.

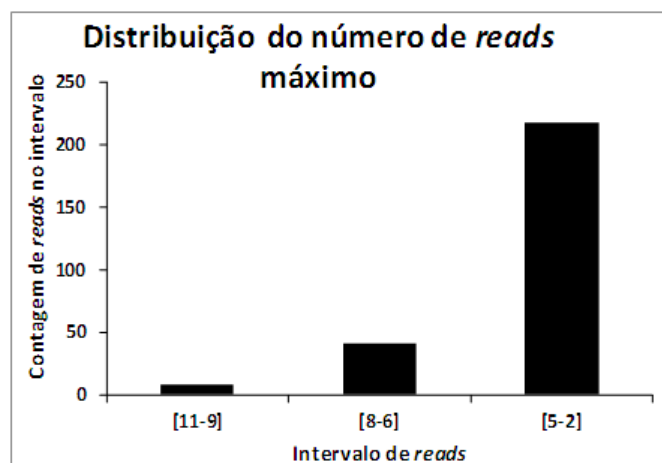


Figura 4.4 – Histograma representativo da distribuição do número máximo de *reads*/pico pelos 3 intervalos de contagens indicados.

Verificou-se que a maioria dos picos (218) apresentava um número máximo de *reads* baixo, entre 2 e 5, seguido por um conjunto de 41 picos com 6 a 8 *reads* e, por último, só 8 dos 267 picos anotados apresentava mais de 8 *reads*, com um máximo de 11 *reads* presentes em apenas dois deles (Fig. 4.4 e Anexo IV). Por forma a selecionar genes de interesse correspondentes a *features* contidos nestes 3 intervalos, calculou-se ainda um outro parâmetro, correspondente à diferença entre o número de *reads* presentes para cada pico em cada um dos replicados experimentais do ChIP-seq (ΔAct in Anexo IV). Selecionaram-se então os genes apresentados na Tabela 4.2, tendo em conta a sua representatividade em termos dos parâmetros, Max_Act, ΔAct e também a distribuição nos três intervalos indicados

Tabela 4.2 – Genes selecionados com os critérios estabelecidos, separados em 3 grupos mediante o seu número de *reads*: Alto (11 a 9 *reads*), Médio (8 a 6 *reads*) e Baixo (5 a 2 *reads*). Max-Act representa o número máximo de *reads* detetado para o pico correspondente na ativação da via; ΔAct representa a diferença entre o número de *reads* presentes para cada pico em cada um dos replicados experimentais.

	Símbolo_HGNC	Max_Act	ΔAct
Alto [11-9]	PDE4DIP	10	0
	MTUS1	9	7
Médio [8-6]	SLC7A7	8	4
	STRN	7	3
	HOXD3	6	2
	RAD17	6	3
	MSH2	6	4
	HIPK1	6	4
Baixo [5-2]	SLC12A7	5	1
	SPTBN1	5	3
	LRP4	4	1
	SLC39A12	4	1
	BMI1	3	1
	MSH4	2	0
	CBX7	2	0

4.3. Ativação e inibição da via Rac1/PAK1 em células DLD1

Neste estudo foi utilizada a linha celular de cancro colorretal, DLD1, que o laboratório de acolhimento já tinha utilizado nos ensaios de ChIP-seq, visto ter sido caracterizada como um modelo celular tumoral em que todos os componentes da via de sinalização se encontram adequadamente representados (Barros *et al.* 2012).

Para analisar se a expressão dos genes selecionados é ou não regulada pela via Rac1/PAK1 é, como já foi referido, necessário induzir a ativação e inibição desta via no nosso modelo celular. Para tal, seguindo uma metodologia já anteriormente validada pelo grupo de acolhimento (Barros, Jordan, and Matos 2009; Barros *et al.* 2012), transfetaram-se células DLD1 com as seguintes construções em vectores de expressão em células de mamífero (ver *Materiais e Métodos*): um vetor pEGFP controlo (GFP-ctrl), que expressa somente a proteína verde fluorescente do cnidário *Aequorea victoria*, mais conhecida por GFP (abreviatura do inglês, *green fluorescent protein*), para controlo do estado de ativação basal da via; ou, em duplicado, com o vetor pEGFP-Tiam1C1199 (doravante designado por GFP-Tiam1), uma construção onde a sequência codificante da proteína GFP foi fundida em fase com a sequência codificante de um mutante constitutivamente ativo de um GEF específico de Rac1 (Tiam1 (Michiels *et al.* 1997)), para estimular a ativação da via Rac1/PAK1. Após 16 h de expressão, um dos replicados transfectados com GFP-Tiam1, foi tratado durante 4 h com EHT 1846 (doravante designado como EHT), um inibidor seletivo da atividade de Rac1, que consegue bloquear a ativação endógena desta GTPase, mesmo na presença de Tiam1 (ver *Materiais e Métodos*). Em seguida, procedeu-se à lise das células para a obtenção quer de extratos proteicos, quer para o isolamento e purificação do RNA total (ver *Materiais e Métodos*). Os lisados proteicos de cada amostra foram resolvidos por SDS-PAGE, ao que se seguiu a sua análise por *Western blot* (WB) utilizando um anticorpo monoclonal contra o GFP, o que permite assim determinar os níveis de ambas as proteínas transfectadas (GFP-ctrl e GFP-Tiam1) (Fig. 4.5). Detetou-se também a proteína tubulina, utilizando um anticorpo monoclonal específico, para controlo da quantidade de proteína carregada em cada poço do gel de SDS-PAGE, uma vez que se trata de uma proteína *housekeeping* (ubiquamente expressa) que se encontra em elevada abundância em células eucarióticas, e cujos níveis deveriam ser equivalentes em todos os lisados (Gürtler *et al.* 2013).

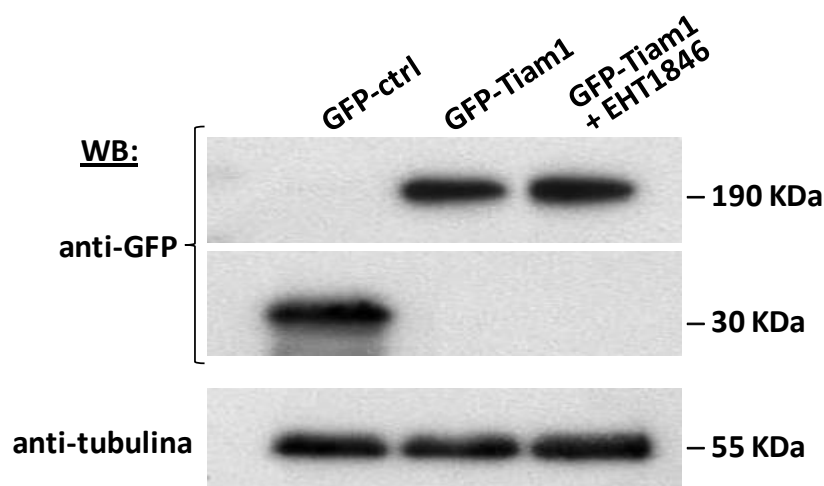


Figura 4.5 - Indução da ativação e inibição da via Rac1/PAK1 em células DLD1. GFP-ctrl corresponde às condições basais; GFP-Tiam1 corresponde às condições de ativação da via; GFP-Tiam1 + EHT 1846 corresponde às condições de inibição da via. A proteína tubulina foi utilizada para controlo da quantidade de proteína carregada em cada poço do gel de SDS-PAGE.

Confirmou-se, assim, que a transfecção foi eficiente em todas as condições e que, tendo como base os níveis de tubulina em cada lisado, os níveis de expressão de GFP-Tiam1 são semelhantes entre replicados, o que garante condições experimentais equivalentes para os ensaios subsequentes de avaliação da expressão génica.

É de referir que as transfecções foram realizadas em sextuplicado para cada condição (GFP-ctrl, GFP-Tiam1 e GFP-Tiam1 + EHT) tendo sido sempre validadas como descrito acima.

4.4. Avaliação morfológica de células DLD1 nas condições de ativação e inibição da via Rac1/PAK1

Como complemento da análise bioquímica avaliou-se, ainda, por imunofluorescência e microscopia confocal, a localização sub-celular, os níveis de expressão e o efeito na morfologia celular das proteínas transfetadas, nas condições retratadas acima. Para tal, células DLD1 transfetadas e tratadas nas condições expostas no ponto anterior foram fixadas, permeabilizadas e os microfilamentos de actina (constituintes do citoesqueleto, aqui usado como indicador da morfologia geral das células) marcados com faloidina (uma micotoxina produzida pelo cogumelo *Amanita phalloides* que se liga aos filamentos de actina estabilizando-os) complexada com o fluoróforo TRITC (isotiocianato de tetrametilrodamina), que emite na zona vermelha do espectro visível (595 nm). Os núcleos das células foram marcados com DAPI (4',6-diamino-2-fenilindol - um marcador fluorescente que se liga fortemente a regiões enriquecidas em adenina e timina no DNA) e os níveis de expressão exógena de GFP-ctrl e GFP-Tiam1 detetados pela fluorescência do GFP na zona verde do espectro visível (509 nm) (Fig. 4.6).

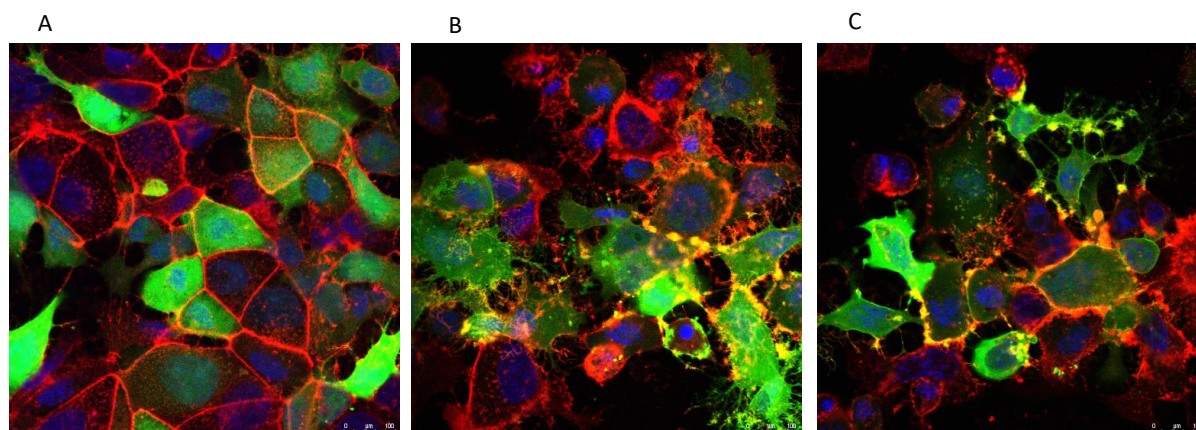


Figura 4.6 – Imagens de microscopia confocal representativas da imunofluorescência de células DLD1 em três condições distintas do estado de ativação da via Rac1/PAK1. (A) Células DLD1 transfetadas com GFP-ctrl, situação basal; (B) Células DLD1 transfetadas com GFP-Tiam1, situação de ativação; (C) Células DLD1 transfetadas com GFP-Tiam1 com tratamento de 4 h com EHT, situação de inibição. A azul encontram-se os núcleos (DAPI), a vermelho a actina (TRITC) e a verde o GFP.

Nas condições controlo (Figura 4.6A), as células DLD1 apresentam uma morfologia epitelial, em monocamada densamente disposta, com contactos célula-célula bem definidos e demarcados pelo anel cortical de actina junto à membrana das células. Como esperado, as células transfetadas com GFP-Tiam1 (Figura 4.6B), apresentam uma reorganização do citoesqueleto de actina, característica da sobre-ativação da GTPase Rac1 e da cinase PAK1, com a formação extensiva de *ruffles* em toda a periferia da membrana plasmática, e com uma clara desorganização dos contactos célula-célula, característica de fenótipos mesenquimais mais migratórios (Michiels *et al.* 1997; Hall 1998). Com a adição de EHT,

verifica-se que, em apenas 4 h, há uma redução clara do *ruffling* da membrana celular e uma recuperação parcial dos anéis corticais de actina, sugestiva do restabelecimento de contactos célula-célula, manifestas indicações da inibição da atividade da via Rac1/PAK1 nestas células (Fig. 4.6 C).

Em suma, esta avaliação morfológica confirma o que já havia sido demonstrado bioquimicamente pelo grupo de acolhimento relativamente à utilização de GFP-Tiam1 e EHT para induzir, respetivamente, a ativação e a inibição da via Rac1/PAK1 em células DLD1 (Barros *et al.* 2012).

4.5. Avaliação do papel da via Rac1/PAK1 na regulação da expressão dos genes selecionados

O passo seguinte consistiu em utilizar o RNA extraído dos 6 replicados para cada uma das três condições de ativação da via Rac1/PAK1 e determinar se seriam detetáveis variações da expressão dos genes da Tabela 4.2 entre elas. Para tal, otimizaram-se reações de RT-PCR (do Inglês, *reverse transcription-polymerase chain reaction*) semi-quantitativo (ver *Materiais e Métodos*) para cada um dos 15 genes selecionados, em paralelo com a amplificação de um fragmento do cDNA do RNA polimerase II (gene *housekeeping* ubiquamente expresso) como controlo interno (QPol II), para normalizar os resultados, eliminando potenciais variações causadas por erros de pipetagem dos diversos cDNAs utilizados (Fig. 4.7).

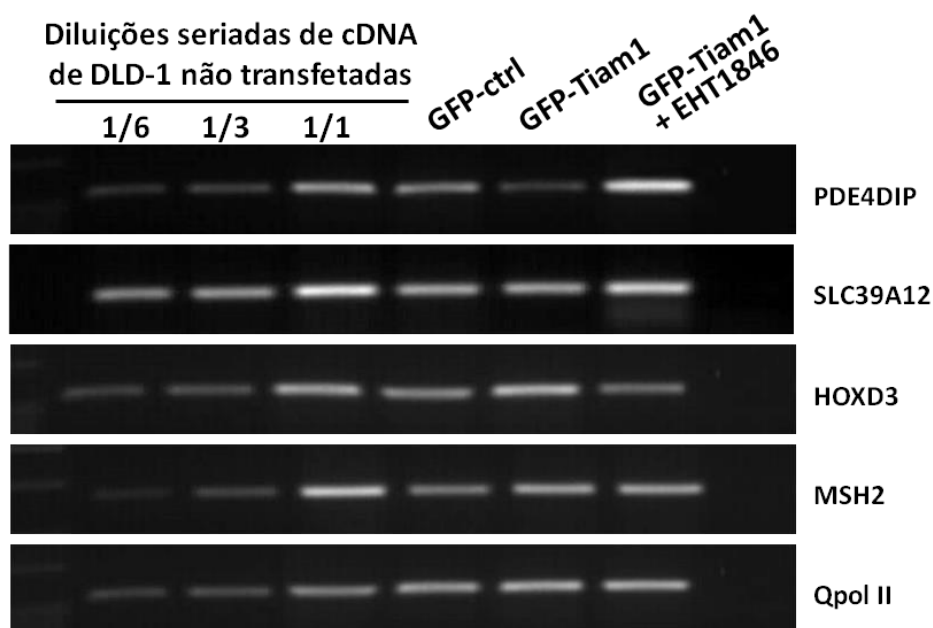


Figura 4.7 – Imagem representativa de RT-PCR semi-quantitativo para quatro dos genes da Tabela 4.2, juntamente com o controlo interno (QPol II).

Posto isto, procedeu-se à análise densitométrica das intensidades das bandas para cada condição, utilizando o *software* público *NIH ImageJ* (<https://imagej.nih.gov/ij/>). Após calibração dos resultados densitométricos utilizando as diluições seriadas de cDNA de DLD1 não transfectadas, foi possível extrair um valor semi-quantitativo (em unidades arbitrárias) das variações da expressão dos 15 genes selecionados, comparando as condições controlo (GFP-ctrl) com as condições de activação (GFP-Tiam) e inibição (GFP-Tiam + EHT) (Tabela 4.3).

Tabela 4.3 – Resultado da análise densitométrica semi-quantitativa das intensidades das bandas dos RT-PCRs para os 15 genes selecionados. Os valores indicados representam as médias e respectivos erros médios (SEM) calculados para os sextuplicados de cada condição experimental. $\Delta\text{Express}|\text{Tiam-EHT}|$ corresponde à amplitude, em valor absoluto, da variação na expressão dos genes analisados em resposta à manipulação da via Rac1/PAK1. Tiam corresponde à condição de ativação da via e EHT à condição de inativação da via.

Símbolo_HGNC	RT-PCR				
	Tiam	SEM Tiam	EHT	SEM EHT	$\Delta\text{Express} \text{Tiam-EHT} $
PDE4DIP	-1.6	0.1	1.6	0.2	3.2
MTUS1	1.1	0.1	1.1	0.1	0.0
SLC7A7	1.1	0.1	1.1	0.1	0.0
STRN	1.2	0.3	-1.0	0.0	2.3
HOXD3	1.2	0.3	-1.3	0.2	2.5
RAD17	1.1	0.0	-1.0	0.1	2.1
MSH2	1.0	0.1	1.1	0.2	0.1
HIPK1	1.5	0.1	1.2	0.2	0.3
SLC12A7	-1.3	0.1	-1.2	0.2	0.0
SPTBN1	-1.2	0.1	1.1	0.0	2.3
LRP4	1.1	0.4	2.3	0.4	1.2
SLC39A12	1.1	0.1	2.1	0.3	0.9
BMI1	-1.7	0.1	1.3	0.3	3.0
MSH4	1.4	0.2	-1.1	0.1	2.5
CBX7	-1.4	0.1	1.2	0.2	2.6

Como era antecipado, a amplitude das variações na expressão dos genes estudados, em resposta à manipulação do estado de ativação da via Rac1/PAK1, rondou as 2 vezes os valores basais, com valores entre os 2.3 e -1.7 vezes (Tabela 4.3). Cinco dos 15 genes (33%), não apresentaram variações de expressão claras (MTUS1, SLC7A7, MSH2, HIPK1, SLC12A7) e entre os restantes 10 genes, 6 variavam no sentido contrário ao esperado (PDE4DIP; SPTBN1; BMI1; CBX7; LRP4 e SLC39A12), i.e., a sua expressão diminui aquando da ativação da via (Tiam) e aumenta com a inibição da mesma (EHT).

No entanto, quando se calculou a amplitude da variação por manipulação da via (i.e., o módulo da diferença entre o nível de expressão na ativação e na inibição da via – $\Delta\text{Express}|\text{Tiam-EHT}|$) não se observou nenhuma associação óbvia entre esta e a distribuição dos genes pelos intervalos de alto, médio e baixo número de *reads* definidos anteriormente (Tabela 4.2 e 4.3).

4.5.1. Cálculo do coeficiente de correlação de *Pearson*

Com vista a analisar a existência de potenciais correlações entre os resultados da avaliação da expressão dos genes selecionados e os parâmetros com os quais estes poderiam estar relacionados e que condicionaram a sua seleção, criou-se a Tabela 4.4. Nesta comparam-se, lado a lado, os valores de $\Delta\text{Express}|\text{Tiam-EHT}|$ com: a distância do pico correspondente ao respetivo TSS (Dist_TSS); a largura do pico correspondente (Larg_pico); o número de *reads* máximo para o pico correspondente (*read_Max*); a diferença no número de *reads* entre replicados experimentais do ChIP-seq ($\Delta\text{read_Rep}$). Foi ainda escolhido um novo parâmetro, que se postulou poder também influenciar os resultados obtidos

– a diferença, para o pico correspondente, entre o número de *reads* no controlo e nas condições de ativação da via ($\Delta read_Act_Ctrl$).

Tabela 4.4 – Comparação da amplitude de variação na expressão dos genes candidatos ($\Delta Express|Tiam-EHT|$) com diversos parâmetros potencialmente relacionados: distância do pico correspondente ao respetivo TSS (*Dist_TSS*); largura do pico correspondente (*Larg_pico*); número de *reads* máximo para o pico correspondente (*read_Max*); diferença, entre replicados experimentais do ChIP-seq, no número de *reads* para o pico correspondente ($\Delta read_Rep$); diferença, para o pico correspondente, entre o número de *reads* no controlo e nas condições de ativação da via ($\Delta read_Act_Ctrl$).

Símbolo_HGNC	$\Delta Express Tiam-EHT $	<i>Dist_TSS</i>	<i>Larg_pico</i>	<i>read_Max</i>	$\Delta read_Rep$	$\Delta read_Act_Ctrl$
PDE4DIP	3.21	-273	51	10	0	6
MTUS1	0.02	-1239	107	9	7	1
SLC7A7	0.05	-842	311	8	4	3
STRN	2.26	-1489	85	7	3	1
HOXD3	2.52	-1112	112	6	2	2
RAD17	2.09	-546	182	6	3	3
MSH2	0.07	-19	94	6	4	3
HIPK1	0.32	-1278	159	6	4	2
SLC12A7	0.01	-1477	73	5	1	2
SPTBN1	2.31	-967	231	5	3	1
LRP4	1.16	-847	54	4	1	3
SLC39A12	0.94	-1105	312	4	1	2
BMI1	2.97	-1302	245	3	1	1
MSH4	2.46	-261	253	2	0	1
CBX7	2.64	-536	145	2	0	1

Mais uma vez, a análise direta da tabela não revelou associações óbvias. Optámos, por isso, por calcular o coeficiente de correlação de *Pearson* entre cada parâmetro e $\Delta Express|Tiam-EHT|$. O coeficiente de correlação de *Pearson* é uma medida estatística da robustez de uma relação linear entre duas variáveis aleatórias. Numa amostra é geralmente designada por r , e definida como: $-1 \leq r \leq 1$, onde valores positivos representam uma correlação linear positiva, valores negativos uma correlação linear negativa (ou inversa) e o valor zero indica a ausência de qualquer relação, sendo tanto mais forte a correlação linear quanto mais o valor se aproximar de 1 ou -1 (Evans, 1996). Os resultados deste cálculo encontram-se resumidos no gráfico da Figura 4.8, onde podemos constatar que somente a diferença no número de *reads* entre replicados experimentais do ChIP-seq ($\Delta read_Rep$) e o número de *reads* no controlo e nas condições de ativação da via ($\Delta read_Act_Ctrl$) parecem poder relacionar-se com as variações de expressão observadas. A diferença de *reads* entre replicados apresenta uma correlação inversa moderada ($r = -0.65$) com a amplitude de variação na expressão génica, i.e., quanto maior a diferença de *reads* entre os replicados experimentais menor a amplitude da variação na expressão génica, e vice-versa. Por outro lado, a diferença de *reads* entre o controlo e a ativação manifesta uma correlação direta, mas fraca ($r = 0.32$), com a amplitude de variação na expressão génica.

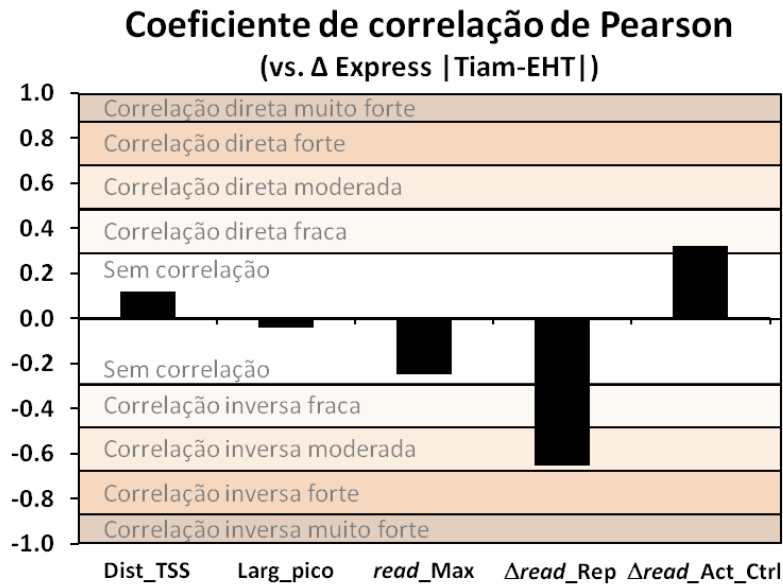


Figura 4.8 – Resultados dos cálculos do coeficiente de correlação de *Pearson* entre a amplitude de variação na expressão dos genes candidatos ($\Delta \text{Express|Tiam-EHT|}$) e os seguintes parâmetros: distância do pico correspondente ao respetivo TSS (*Dist_TSS*); largura do pico correspondente (*Larg_pico*); número de *reads* máximo para o pico correspondente (*read_Max*); diferença, entre replicados experimentais do ChIP-seq, no número de *reads* para o pico correspondente ($\Delta \text{read_Rep}$); diferença, para o pico correspondente, entre o número de *reads* no controlo e nas condições de ativação da via ($\Delta \text{read_Act_Ctrl}$).

4.5.2. Confirmação dos resultados por qPCR

Para fortalecer o significado da análise de dados efetuada, escolheram-se 4 genes que refletiam os padrões de variação observados (i.e., sentido canónico, sentido contrário ao canónico e sem variação), para validação por PCR quantitativo (qPCR) dos níveis de expressão nas 3 condições experimentais estudadas (Tabela 4.5).

Tabela 4.5 – Resultado da análise por qPCR da variação nos níveis de expressão dos genes indicados em resposta à ativação (qTiam) e inibição (qEHT) da via Rac1/PAK1. Os valores indicados representam as médias e respetivos erros médios (SEM) calculados para os sextuplicados de cada condição experimental. $q\Delta \text{Express|Tiam-EHT|}$ corresponde à amplitude, em valor absoluto, da variação na expressão dos genes indicados em resposta à manipulação da via Rac1/PAK1. qTiam corresponde à condição de ativação da via e qEHT à condição de inativação.

Símbolo_HGNC	qTiam	SEM qTiam	qEHT	SEM qEHT	$q\Delta \text{Express Tiam-EHT }$
PDE4DIP	-1.8	0.14	1.6	0.31	3.4
SLC39A12	1.1	0.12	2.1	0.35	0.9
HOXD3	1.4	0.25	-1.3	0.18	2.7
MSH2	1.0	0.15	1.1	0.14	0.1

Os resultados obtidos por qPCR vieram confirmar os obtidos por RT-PCR semi-quantitativo, não só ao nível do sentido da variação (canónico ou contrário), como ao nível do valor absoluto da magnitude da variação (comparar valores das Tabelas 4.3 e 4.5).

4.5.3. Identificação de motivos *consensus* para a resposta não-canônica do *switch* BCL6/STAT5

No sentido de se perceber se existiria algum motivo comum na vizinhança dos picos correspondentes aos 6 genes que apresentaram uma resposta transcricional contrária à esperada, recorreu-se novamente ao algoritmo “MEME Suite”. A análise das regiões de 250 pb a montante e a jusante do centro de cada um dos 6 picos detetou o motivo representado nas últimas linhas da Figura 4.9.

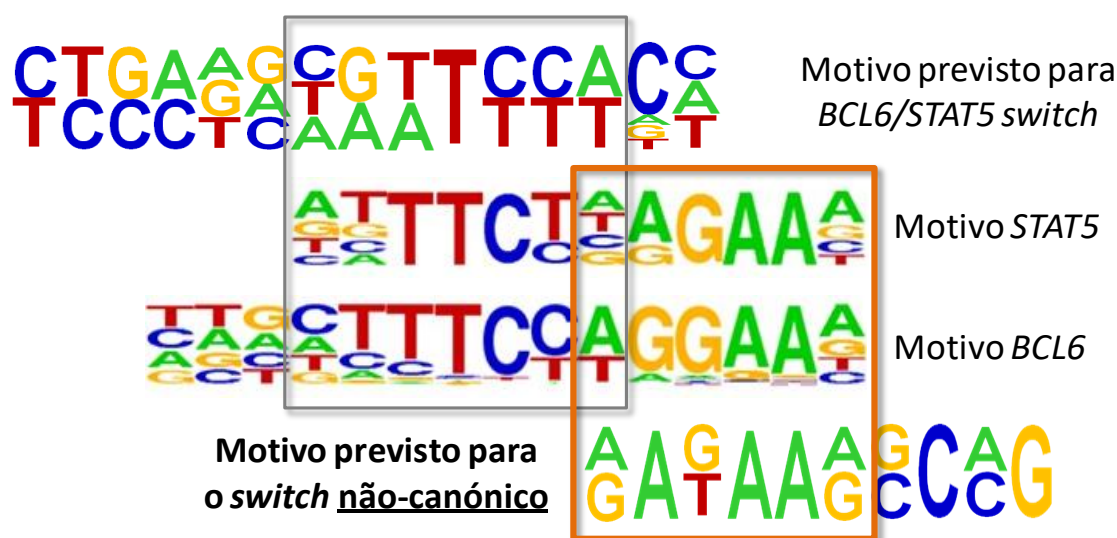


Figura 4.9 – Representação gráfica do motivo de *consensus* previsto pelo algoritmo “MEME Suite” para a ocorrência do *switch* BCL6/STAT5 na resposta não-canônica à via Rac1/PAK1. Estão indicados também o motivo *consensus* previsto para o *switch* de resposta canônica, e os motivos canônicos dos fatores STAT5 e BCL6. O retângulo cinza destaca a região de maior homologia entre os motivos STAT5, BCL6 e *switch* canônico e o laranja para o *switch* não-canônico.

Devido ao número reduzido de sequências disponíveis o valor *E* calculado pelo *software* para este motivo é de 7.2×10^2 . No entanto, para estas situações, o algoritmo dispõe de uma função *shuffle*, que altera aleatoriamente a posição dos nucleótidos em cada sequência e volta a procurar motivos de *consensus* (Bailey *et al.* 2006). Quanto maior a diferença entre os valores *E* dos motivos encontrados para sequências *shuffled* e originais maior a confiança do motivo encontrado para as últimas. Neste caso, todos os motivos *shuffled* tinham valores de *E* superiores a 2.8×10^4 , o que confere alguma confiança ao motivo encontrado nas sequências originais. É ainda interessante notar que a região de maior homologia entre os motivos STAT5, BCL6 e o motivo de *switch* não-canônico, corresponde exatamente à porção adjacente à de maior homologia com o motivo de *switch* canônico.

4.6. Seleção de genes candidatos por agrupamento funcional

Outro tipo de abordagem, comumente usada, para a seleção de genes candidatos entre os “*hits*” de estudos “ômicos”, é a análise de agrupamento funcional. Esta permite descortinar se a lista de “*hits*” se encontra especialmente enriquecida em genes envolvidos numa função biológica particular (Eisen *et al.* 1998). Neste sentido, recorreu-se novamente ao algoritmo público D.A.V.I.D., mas desta vez incluíram-se os 2419 *features* encontrados para a ativação da via Rac1/PAK1. Entre estes, 1665 encontravam-se anotados entre as várias bases de dados utilizadas pelo algoritmo. Da análise de enriquecimento,

destacou-se um grupo ou *cluster* de 50 genes anotados como estando envolvidos na resposta a estímulos por danos no DNA e sua reparação (Fig. 4.10 e Anexo V). Uma vez que esta é uma das formas mais comuns das células tumorais resistirem a agentes quimioterápicos genotóxicos (Kachalaki *et al.* 2016), decidiu-se investigar mais aprofundadamente os genes neste *cluster*.

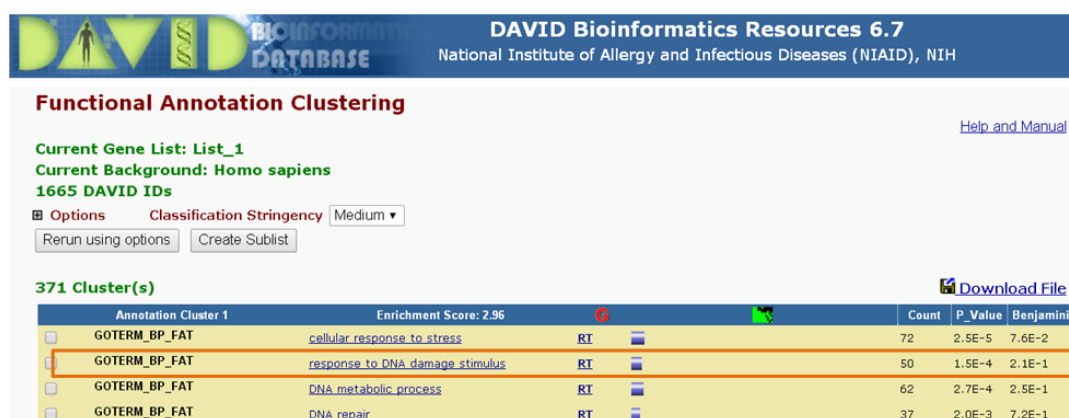


Figura 4.10 – Resultados obtidos da análise de agrupamento funcional via D.A.V.I.D. para as 2419 *features* identificadas em resposta à ativação da via Rac1/PAK1. O retângulo vermelho salienta um *cluster* de 50 genes anotados como estando envolvidos na resposta a estímulos por danos no DNA.

De forma a determinar se existia alguma via de sinalização particularmente representada entre estes 50 genes, recorreu-se à ferramenta bioinformática de análise de redes de proteínas (*protein networks*) GeneMANIA (<http://genemania.org/>). Esta permitiu correlacionar 11 dos 50 genes com vias de recombinação e reparação de DNA (Fig. 4.11).

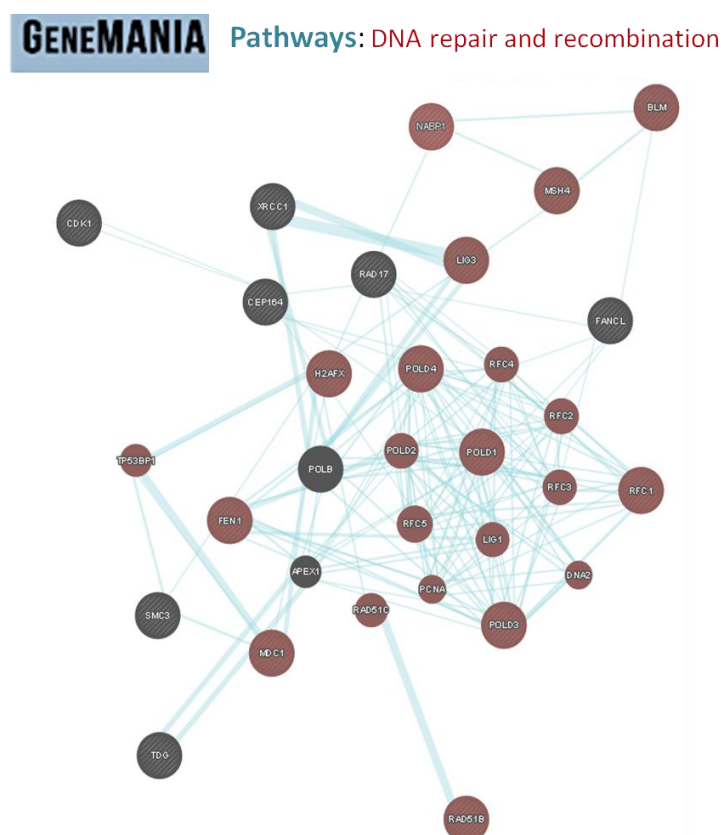


Figura 4.11 – Resultados obtidos na análise, via GeneMANIA, do cluster funcional de genes envolvidos na resposta a estímulos por danos no DNA. A rede de interações proteicas (*protein network*) foi extrapolada com base em associações anotadas como pertencentes à mesma via de sinalização celular.

Aos 11 picos correspondentes a estes *features*, aplicaram-se os critérios de seleção que na análise do primeiro conjunto de 15 genes selecionados revelaram maior associação com variações na expressão génica em resposta à modulação da via Rac1/PAK1 (menor diferença entre os replicados experimentais e maior diferença de *reads* entre ativação e controlo). Seguindo estes critérios, selecionaram-se 4 genes e procedeu-se ao estudo da influência da ativação e inibição da via Rac1/PAK1 nos seus níveis de expressão por RT-PCR semi-quantitativo, como descrito anteriormente (Tabela 4.6 e Fig. 4.12).

Tabela 4.6 – Resultado da análise, por RT-PCR semi-quantitativo, da influência da via Rac1/PAK1 na expressão de 4 dos genes do *cluster* funcional relacionado com vias de reparação e recombinação de DNA. Os valores indicados representam as médias e respetivos erros médios (SEM) calculados para os sextuplicados de cada condição experimental. Apresentam-se, ainda, o nome dos 4 genes e as referências das *features* correspondentes. Tiam corresponde à condição de ativação da via e EHT à condição de inativação da via.

<i>Feature</i>	Descrição	Símbolo_HGNC	Tiam	SEM Tiam	EHT	SEM EHT
ENSG00000182185	RAD51-like 1	RAD51B	1.4	0.1	1.0	0.0
ENSG00000188486	H2A histone family, member X	H2AFX	1.2	0.2	-1.6	0.1
ENSG00000173559	nucleic acid-binding protein 1	NABP1	1.3	0.1	-1.2	0.1
ENSG00000175482	polymerase (DNA-directed), delta 4	POLD4	1.5	0.1	-2.1	0.1

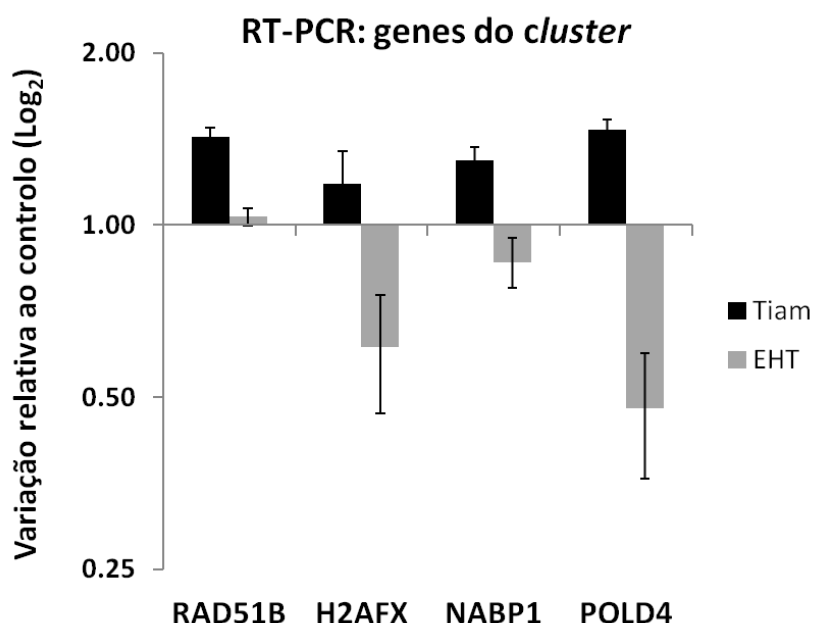


Figura 4.12 – Gráfico representando as variações, apresentadas na Tabela 4.6., na expressão de 4 dos genes do *cluster* funcional, em resposta à manipulação da via Rac1/PAK1.

Constatou-se que todos os genes deste grupo aparentavam responder à via Rac1/PAK e que esta resposta seguia o padrão esperado para o *switch* BCL6/STAT5, i.e., a sua expressão aumenta com a ativação da via e diminui com a sua inativação. De todos os genes o que apresentou uma maior variação na sua expressão foi o POLD4, tanto na ativação como na inibição da via. Os resultados obtidos foram, mais uma vez, confirmados por qPCR (Fig. 4.13).

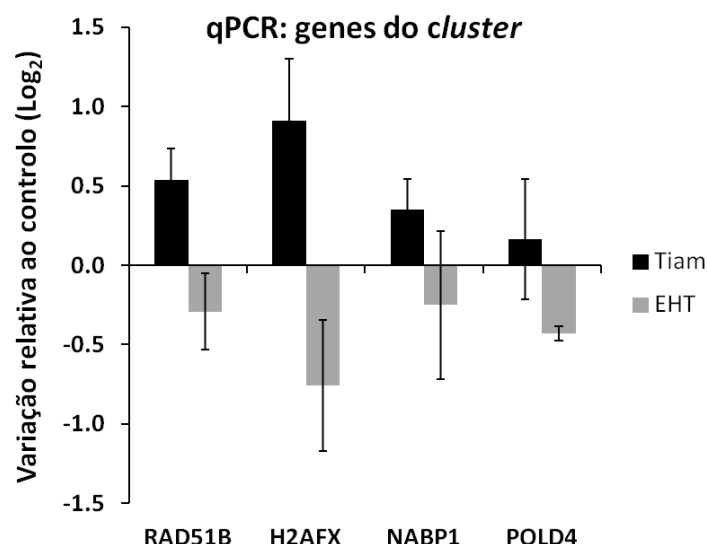


Figura 4.13 – Análise, por qPCR, das variações, em resposta à via Rac1/PAK1, na expressão de 4 dos genes selecionados a partir do *cluster* funcional.

Embora os resultados variassem um pouco em termos absolutos, a análise quantitativa mostrou o mesmo padrão de variação para os 4 genes, onde a ativação da via promove o aumento da expressão génica e, em oposição, a inibição da via a diminui (Fig. 4.13).

4.7. Identificação de motivos de ligação ao DNA nos genes do *cluster*

Uma vez que qualquer um dos 4 genes do *cluster* estudados respondiam ao *switch* BCL6/STAT5 com variações canónicas nos seus níveis de expressão, investigou-se se as regiões delimitadas pelos picos de ChIP-seq correspondentes continham, ou não, sequências que correspondessem ao motivo de ligação de *consensus* definido bioinformaticamente no início deste trabalho, i.e., {G}R[AT]TYY[AT]. Para tal recorreu-se, novamente, ao *software* “MEME Suite” que devolveu um alinhamento das sequências dos 4 picos, realçando a cores a região de maior homologia encontrada (Fig. 4.14). Podemos verificar que esta região apresenta uma correspondência elevada com o motivo *consensus* previsto para o *switch* BCL6/STAT5 (Fig. 4.14).

Feature	Feat. strand	Gene Name	Site strand	Start	p-value	Homology sites
ENSG00000182185	+	RAD51B	+	128	8.37E-07	AAGAAGAATACGATCCACCTATTAGAAGGAA
ENSG00000188486	-	H2AFX	-	9	3.15E-07	TGCTTAAGCCATCCTCCCAACTCAGCCTT
ENSG00000173559	+	NABP1	+	209	1.67E-06	GCAAGTCTCTCGTTCCACCTAAGGAGAAAAA
ENSG00000175482	-	POLD4	-	79	3.15E-07	TGGCCTCAAGCGATCCTCCACCTTGGCATC

Motivo previsto para BCL6/STAT5 switch

CTGAAGCGTTCCACC
TCCCTCAAAATTTT

{G}R[AT]TYY[AT]

Figura 4.14 - Regiões de máxima homologia encontradas via “MEME Suite” nos picos correspondentes aos 4 genes estudados do *cluster* funcional. Indicados estão o nome dos 4 genes, as referências das *features* correspondentes e a sua orientação (*Feat. strand*). Apresentam-se, ainda, a posição de início das regiões de maior homologia (*start*), o alinhamento nessas regiões (a cores), as suas orientações (*Site strand*) e o *p-value* referente à possibilidade de estes ocorrerem aleatoriamente. O retângulo vermelho evidencia a região de maior homologia com o motivo *consensus* previsto para o *switch* BCL6/STAT5 canónico.

A relevância funcional desta sequência terá de ser, mais uma vez, confirmada experimentalmente recorrendo, por exemplo, a sistemas repórter luminescentes, como descrito anteriormente.

4.8. Envolvimento da via Rac1/PAK1 na suscetibilidade a agentes genotóxicos

No sentido de avaliar a relevância da modulação da via Rac1/PAK1 na ativação de genes envolvidos nos processos de reparação em resposta a danos no DNA, investigou-se de seguida se a estimulação desta conferiria às células DLD1 uma proteção adicional a agentes que induzissem danos no DNA. Para tal, utilizou-se a metodologia de ensaio do cometa (ver *Materiais e Métodos*) que permite avaliar e quantificar a extensão de danos no DNA causada, por exemplo, pela exposição das células a um agente genotóxico (Andrew R Collins, Ai-guo, and Duthie 1994) (Fig. 4.15).

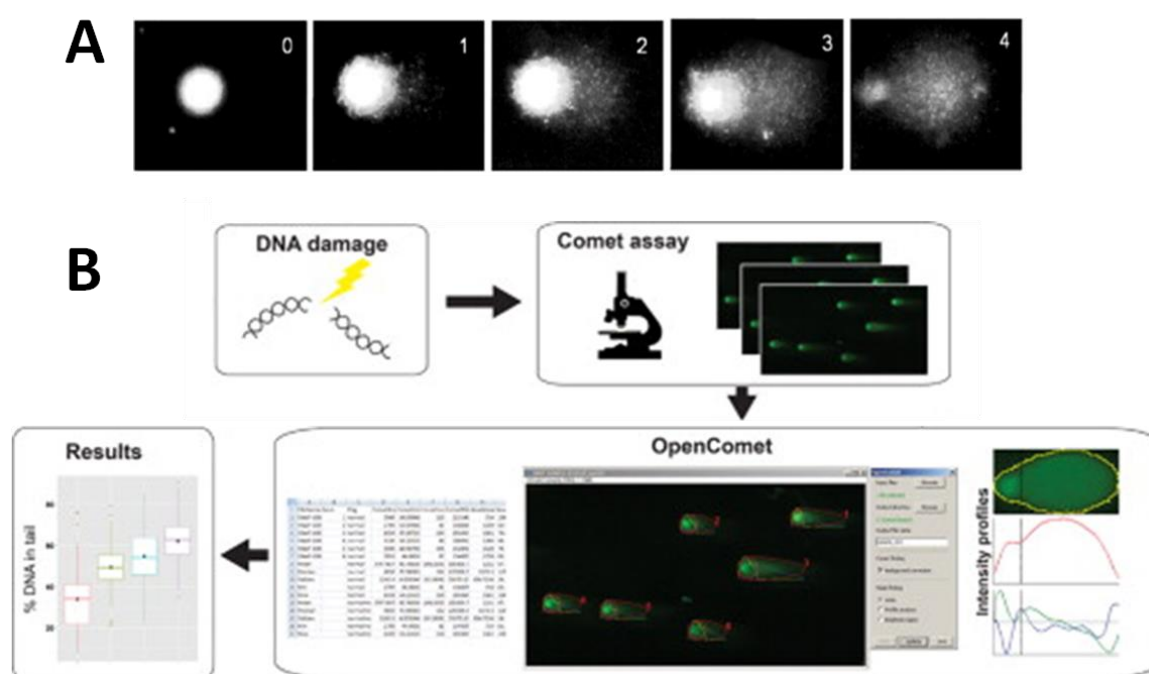


Figura 4.15 – Ensaio do cometa. (A) As imagens representam núcleos de células (linfócitos), corados com brometo de etídio, embebidos em matrizes de agarose e sujeitos a eletroforese. A nuvem à direita (vulgo, cauda do cometa) resulta da migração diferencial de fragmentos de DNA e é tanto maior quanto maior for a extensão dos danos genómicos. As fotografias apresentadas mostram as 5 classes (marcadas de 0 a 4) sugeridas por Collins *et al.* (1994) para classificação visual da extensão dos danos genómicos. (B) *Software* de acesso público, como o *OpenComet* (Gyori *et al.* 2014), permite, hoje em dia, fazer uma análise quantitativa mais precisa da extensão dos danos genómicos, com base em medições da área da cabeça e cauda do cometa, bem como da intensidade média dos pixels nestas regiões, a partir de imagens digitais recolhidas por microscopia de fluorescência (neste caso, via microscópio Zeiss – AxioPlan2).

Na nossa análise utilizámos o agente alquilante sulfonato de etil-metano (EMS), pelo seu efeito já documentado em células de cancro colorretal (Suggitt *et al.* 2003). Assim, células DLD1 foram expostas a 15 mM de EMS, ou ao mesmo volume de solvente (controlo), durante 1 h. A duplicados destas condições foram então aplicadas as condições habituais de manipulação da via Rac1/PAK, i.e., basais (controlo sem tratamento), ativação e inibição. Para a ativação da via Rac1/PAK1, e para evitar o desvirtuar dos resultados do cometa pelo sofrimento celular inerente à metodologia de transfeção,

utilizou-se um tratamento de 4 h com o fator de crescimento HGF (*hepatoocyte growth factor*) que o grupo de acolhimento já havia descrito como produzindo um efeito equivalente ao da transfeção com TiamC1199 na ativação de Rac1 (Moniz *et al.* 2013). Para a inativação da via utilizou-se o tratamento com EHT, novamente durante 4 h. Para cada uma das condições experimentais, foi lisado um duplicado a 0 h e 4 h de recuperação após a hora de exposição ao EMS (ver *Materiais e Métodos*). O DNA nuclear foi então marcado com brometo de etídio e as células processadas como descrito em *Materiais e Métodos*. As lâminas resultantes foram visualizadas num microscópio Zeiss AxioPlan2, acoplado a um sistema digital de aquisição de imagens, tendo-se utilizado o *software OpenComet* (www.opencomet.org) para medir os cometas de 100 células entre os replicados de cada condição experimental. Na Figura 4.16 apresentam-se os dados da quantificação da percentagem de DNA na cauda, normalizados relativamente às amostras que não foram expostas ao EMS bem como imagens representativas dos cometas observados em cada uma das condições experimentais.

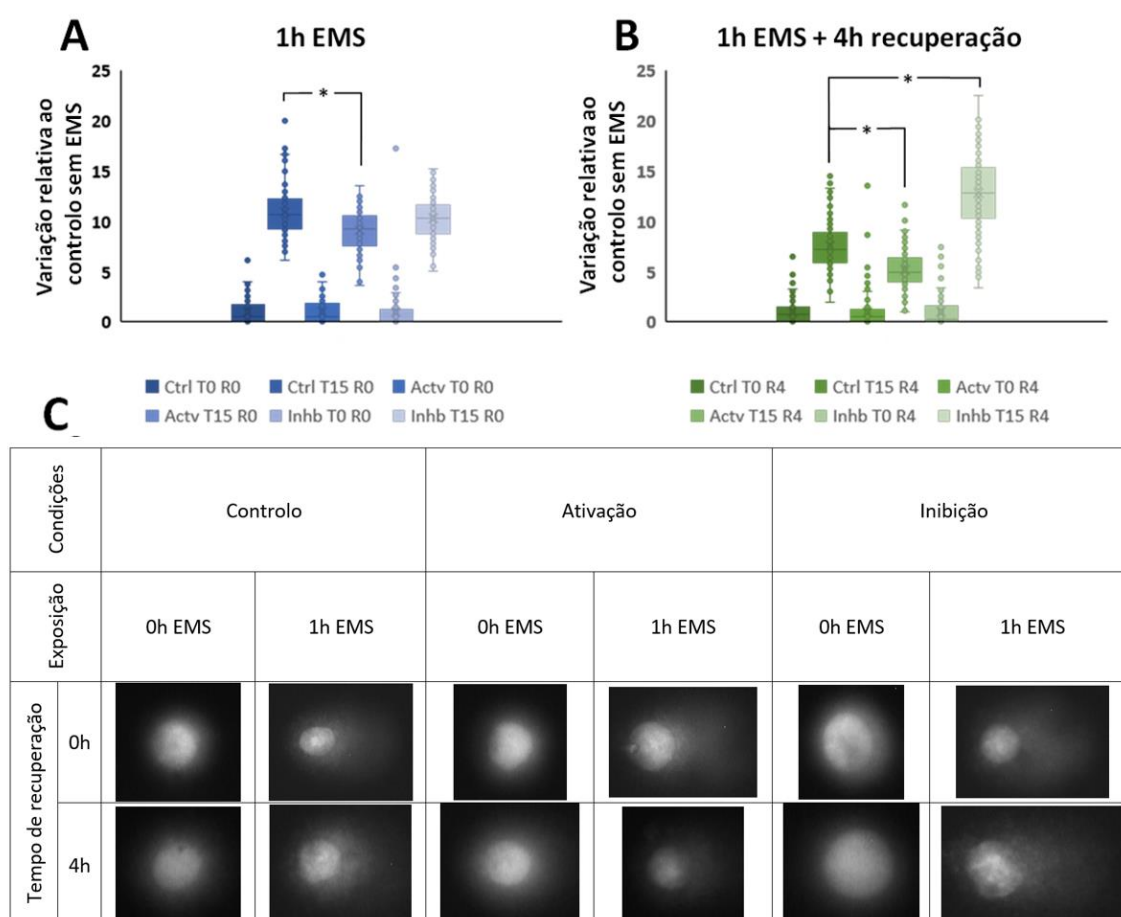


Figura 4.16 – Resultados obtidos do ensaio do cometa, após normalização dos valores às condições sem EMS, antes (A) e depois de (B) 4 h de recuperação após remoção do EMS. (C) Imagens representativas dos cometas observados em cada uma das condições experimentais. * $P < 0.001$. Ctrl, Actv e Inhb correspondem às condições controlo, de ativação e de inibição, respetivamente; T corresponde à concentração de EMS utilizada no ensaio e R corresponde ao tempo de recuperação.

Pode assim observar-se na Figura 4.16A que após 1 h de exposição ao EMS (T15, R0), a percentagem de DNA na cauda aumenta, quer no controlo quer na presença de EHT (Inhb), cerca de 11 vezes, ao passo que na presença de HGF (Actv) não ultrapassa as 9 vezes, indicativo da ocorrência de uma ligeira, mas significativa ($P < 0.001$) proteção aos danos no DNA. No entanto, os efeitos mais notórios manifestaram-se 4 h após a remoção do EMS, em que no caso das condições basais se observou uma

recuperação de ~30% na extensão dos danos (de 11 para cerca de 8 vezes o controlo sem EMS), na ativação da via de ~43% (de 8 para cerca de 5 vezes o controlo sem EMS), e na inibição um agravamento para ~124% dos danos ao tempo 0 h de recuperação (de 10.5 para cerca de 13 vezes o controlo sem EMS).

5. Discussão e Conclusão

Este trabalho procurava em primeiro lugar, validar uma nova estratégia de análise de dados de ChIP-seq, desenvolvida pelo laboratório de acolhimento, que visa encontrar genes alvo da via de sinalização Rac1/PAK1-BCL6/STAT5. Em seguida, pretendia-se seleccionar, entre os múltiplos “hits” encontrados, genes que, pela sua função e variação de expressão em resposta à manipulação desta via de sinalização, pudessem estar envolvidos nos efeitos associados a uma sobre-estimulação de Rac1 e PAK1 frequentemente observada em tumores mais invasivos e com pior prognóstico, que tipicamente apresentam resistência à quimioterapia convencional.

De facto, sabe-se que o papel de Rac1 na progressão do CRC encontra-se muitas vezes associado a vias dependentes de PAK1. Este eixo de sinalização Rac1/PAK1 encontra-se implicada na invasão de células tumorais, onde a ativação de Rac1 consegue promover a migração e invasão de células cancerígenas, sugerindo que a regulação mediada por Rac1 da via Rac1/PAK1 participa na progressão do cancro colorretal (Zhao *et al.* 2008). Alguns estudos têm vindo a demonstrar a importância da via Rac1/PAK1 para o crescimento e metástase em linhas celulares de cancro colorretal (Zhu *et al.* 2012), suscitando um interesse cada vez maior em perceber se esta via contribui de alguma forma para a resposta das células tumorais a agentes terapêuticos que induzem danos no DNA. De facto, evidências recentes sugerem que a sinalização a jusante de PAK1 contribui ativamente para este processo, nomeadamente através da activação de mecanismos celulares de proteção do genoma contra danos no DNA (R. Kumar and Li 2016). Sendo Rac1 uma GTPase com um papel fundamental na regulação da atividade da cinase PAK1, alguns autores sugerem que também Rac1 possa desempenhar um papel importante na regulação deste tipo de resposta (Zou *et al.* 2016). No entanto, os mecanismos moleculares através dos quais a via Rac1/PAK1 participa na resposta a danos no DNA ainda se encontram por caracterizar. Por esta razão, a caracterização dos alvos genómicos da via de sinalização Rac1/PAK1, através da sua ação no *switch* transcricional BCL6/STAT5 identificada anteriormente pelo grupo de acolhimento em tumores coloretais, é um passo importante na identificação de novos alvos terapêuticos que antagonizem os processos moleculares que conferem vantagem seletiva a tumores coloretais quimioresistentes.

Como em qualquer estudo piloto *in vitro*, que use linhas celulares tumorais estabelecidas, também neste estudo se encontraram algumas limitações técnicas inerentes ao modelo celular utilizado – as células de carcinoma colorretal DLD1. De facto, observámos que nas condições experimentais em que se procedeu à inibição da via Rac1/PAK1, apenas se obtiveram cerca de 6% dos picos de ChIP-seq comparativamente com o número de picos obtidos nas condições de ativação da via. Isto indica que, no seu estado basal, as células DLD1 apresentam um nível baixo de ativação endógena da via Rac1/PAK1, pelo que a sua inibição produz um efeito limitado no perfil de expressão génica regulada pelo *switch* BCL6/STAT5. Embora esta limitação nos tenha impedido de avaliar detalhadamente o impacto da inibição da sinalização via Rac1/PAK1 num contexto tumoral consistente com um fenótipo mais agressivo, possibilitou-nos, por outro lado, analisar de forma mais abrangente as consequências genómicas da aquisição, por parte das células tumorais do cólon, de uma hiperativação desta via de sinalização durante o seu desenvolvimento e progressão. De facto, a ativação desta via em células DLD1 conduziu ao aparecimento de picos representativos da troca de BCL6 por STAT5 na região promotora de 2402 genes (TSS *features*). No sentido de se validar experimentalmente o impacto da ativação da via Rac1/PAK1 nos níveis de expressão destes potenciais genes alvo, esta lista inicial foi sujeita a uma série de critérios de seleção que a reduziram a um conjunto final de 15 genes, cujos picos na sua região promotora apresentavam características representativas do conjunto geral de picos detetados. À semelhança do que Barros *et al.* (2012) verificaram para genes do ciclo celular, as variações de expressão

observadas para a maioria dos genes selecionados não ultrapassaram uma amplitude de cerca de 3 vezes os valores basais. Isto decorre, por certo, do facto da expressão de um determinado gene ser um evento bastante complexo e resultar da ação concertada de vários fatores e como, neste caso, apenas um desses fatores foi alterado, pode não ser o suficiente para se verificarem grandes variações na expressão dos genes alvo (Romero, Ruvinsky, and Gilad 2014; Pai, Pritchard, and Gilad 2015). Da análise da existência de potenciais correlações entre os resultados obtidos na avaliação da expressão génica e os parâmetros que levaram à seleção dos genes em causa, só se encontrou uma possível correlação inversa moderada com a diferença no número de *reads* entre replicados experimentais do ChIP-seq ($\Delta read_Rep$). O que sugere que o fator mais preponderante na análise crítica dos resultados, segundo este método, será considerar como robustos apenas os picos que apresentem uma amplitude semelhante, em termos do número de *reads*, entre replicados experimentais. Não obstante, observou-se ainda que do grupo de 15 genes selecionados, em 6 deles os níveis de expressão diminuía com a ativação da via Rac1/PAK1, apresentando o comportamento oposto na sua inibição. Como descrito anteriormente pelo grupo de acolhimento, esperava-se que a ativação desta via induzisse um incremento na expressão génica resultante da troca do repressor transcricional BCL6, pelo ativador STAT5 (Barros *et al.* 2012), pelo que este resultado era aparentemente contraditório. No entanto, após uma pesquisa bibliográfica sobre o assunto, constatou-se que esta observação pode simplesmente espelhar estudos anteriores em que se demonstrou que os elementos da família de fatores de transcrição STAT podem atuar como ativadores ou como repressores da expressão génica, consoante as modificações pós-traducionais que sofram que, por sua vez, condicionam o recrutamento de co-ativadores ou co-repressores transcricionais para os *loci* onde estão ligados (Desrivieres *et al.* 2006). De facto, à semelhança do que aqui se verificou, já foi demonstrado que no caso do co-repressor nuclear SMRT, também envolvido na regulação negativa da expressão dos genes alvo de STAT5, o uso de um inibidor químico da atividade histona desacetilase (HDAC) é suficiente para inverter o seu efeito na transcrição dependente de STAT5, sugerindo que a repressão dependente de SMRT deriva da atividade de HDAC (Litterst *et al.* 2005).

Utilizando, em paralelo, ferramentas bioinformáticas de agrupamento funcional com base em anotações de ontologia génica, chegou-se a um *cluster* de genes envolvidos na resposta a estímulos por danos no DNA e sua reparação, que se correlacionavam entre si com vias de recombinação e reparação de DNA e que respondiam aos critérios de seleção mais relevantes em termos de associação com variações na expressão génica em resposta à modulação da via Rac1/PAK1. Este *cluster* revelou o comportamento de expressão canónico aquando da manipulação da atividade da via Rac1/PAK1, i.e., a sua expressão aumentou com a ativação da via e vice-versa.

Dado que o primeiro grupo de genes selecionado e o segundo, correspondente ao *cluster*, apresentavam comportamentos distintos em termos de variação da expressão, procurou-se identificar a existência de algum motivo *consensus* entre estes genes e o motivo previsto para o *switch* BCL6/STAT5 canónico. Utilizando o algoritmo “MEME Suite” para análise de enriquecimento de sequências de ligação ao DNA entre os picos correspondentes aos diversos genes pôde constatar-se que, o primeiro grupo de genes selecionado apresenta um motivo previsto diferente do motivo canónico, que se denominou de motivo *consensus* não canónico. Curiosamente, este motivo partilha, juntamente com os motivos *consensus* para STAT5 e BCL6, uma sequência de homologia que corresponde à porção adjacente do motivo previsto para o *switch* BCL6/STAT5. Curiosamente, uma situação semelhante foi observada no contexto da transativação do gene supressor tumoral p53. Evidências da existência de mais motivos de ligação de p53 ao DNA, para além do motivo canónico já conhecido, surgiram na sequência da observação que a transativação deste supressor tumoral foi detetada em associação a um local não canónico, criado por um único polimorfismo de nucleótidos (SNPs), no promotor do gene FLT-1 (Jordan *et al.* 2008). Constatou-se, então, que o nível de expressão de p53 tem um grande impacto na sua capacidade de transativar a expressão de genes alvo, o seu aumento possibilita a este fator reconhecer sequências

distintas da sequência *consensus* canónica espalhadas no genoma humano (Jordan *et al.* 2008). É, portanto, relevante a confirmação experimental, em experiências futuras, de que estas sequências canónica e não-canónica de ligação do *switch* BCL6/STAT5, conseguem de facto produzir os comportamentos opostos observados na expressão dos dois grupos de genes estudados. Para tal, poder-se-á recorrer, por exemplo, a sistemas repórter, onde esta sequência, repetida em *tandem*, é introduzida num vetor, para expressão em células de mamífero, como *enhancer* de um promotor viral mínimo fundido ao gene da luciferase do pirilampo. Podemos, assim, monitorar os efeitos transcricionais da ativação e inibição da via Rac1/PAK1 sob estes motivos *consensus* medindo a atividade da luciferase em células transfectadas com estes repórteres. Esta estratégia ainda foi iniciada durante o presente mestrado, mas não foi possível concluí-la por limitações de tempo.

Para avaliar o impacto fisiológico da modulação da via Rac1/PAK1 na ativação dos genes envolvidos na resposta a danos no DNA genómico em células DLD1, utilizámos o agente genotóxico alquilante EMS para tratar estas células. Os resultados obtidos indicam que a via Rac1/PAK1 está de facto envolvida na resposta a danos no DNA e que a sua ativação parece conseguir conferir uma proteção parcial ao dano em si e acelera o processo de recuperação dos mesmos. Por outro lado, a inibição interfere significativamente com a sua reparação promovendo até o seu agravamento. Isto sugere que a ação da via é efetivamente ao nível da reparação dos danos causados por agentes genotóxicos e que a sua ativação constitutiva pode dar uma vantagem às células cancerígenas em resistirem a agentes quimioterápicos que atuem via indução de danos no DNA. Estes dados corroboram um estudo realizado anteriormente, em que os autores sugeriram a implicação de Rac1 numa rede que coordena as respostas celulares ao stress genotóxico proveniente de danos no DNA e especulam que estes efeitos eram também mediados pela cinase PAK1, não especificando, contudo, o modo como Rac1 ou PAK1 atuam perante este stress induzido (Fritz and Henninger 2015).

Estes resultados suportam a hipótese que esta via participa na regulação da resposta a danos no DNA, potencialmente complementando outras vias que têm ação neste tipo de resposta e que são também reguladas pela atividade de Rac1. Por exemplo, foi descrito recentemente em células de cancro da mama que, após tratamento com radiação, ocorre um aumento da expressão de Rac1 e por consequência, um aumento da sinalização das vias PI3K/AKT, MEK/ERK e NF-κB. Estas vias têm vindo a ser implicadas em mecanismos de reparação do DNA e sobrevivência das células, e estudos recentes demonstraram haver um envolvimento de Rac1 na sobrevivência das células de cancro da mama, sugerindo que a sua sinalização possa servir como um potencial alvo clínico para o estudo da radiosensibilização das células cancerígenas (Hein *et al.* 2016). Note-se ainda que, os genes RAD51B, H2AFX e NABP1 participam na reparação da quebra de dupla cadeia do DNA via recombinação homóloga (Ronen and Glickman 2001; Wood, Mitchell, and Lindahl 2005). Interessa aqui averiguar se existe algum conjunto de genes entre o *cluster* encontrado, cujos níveis de expressão possam ser utilizados para prever se o tumor responde a quimioterápicos que danifiquem o DNA. Tal requererá a análise de amostras de tumores recolhidas antes e após o tratamento com quimioterápicos, determinando-se o perfil transcricional através das variações ocorridas na expressão dos genes alvo em estudo. Esta abordagem tem sido comumente utilizada para averiguar o tipo de resposta celular a fármacos (Troester *et al.* 2004). A obtenção desta informação pode ser uma mais valia para a identificação de biomarcadores que permitam prever a resposta do tumor a agentes quimioterápicos genotóxicos, para auxiliar o clínico na escolha da terapia mais adequada ao tumor em questão.

Em suma, esta nova abordagem de ChIP-seq com a aplicação do filtro biológico, permitiu identificar uma resposta transcricional do *switch* BCL6/STAT5, modulada pela via Rac1/PAK1, com um papel relevante na resposta e reparação de danos no DNA, sendo fulcral, no futuro, aprofundar a função desta via e de outras vias complementares neste tipo de resposta, uma vez que a sua modulação farmacológica

poderá ter aplicação terapêutica. Será também importante averiguar a possibilidade de se encontrar uma assinatura génica entre os genes alvo desta via, que consiga prever a resposta do tumor a danos no DNA e particularmente, monitorizar a resposta a quimioterápicos.

6. Referências Bibliográficas

- Arya, Mani, Iqbal S Shergill, Magali Williamson, Lyndon Gommersall, Neehar Arya, and Hitendra RH Patel. 2005. "Basic Principles of Real-Time Quantitative PCR," 209–19.
- Aune, Trond Erik Vee, and Finn Lillelund Aachmann. 2010. "Methodologies to Increase the Transformation Efficiencies and the Range of Bacteria That Can Be Transformed," 1301–13. doi:10.1007/s00253-009-2349-1.
- Bailey, Timothy L., Nadya Williams, Chris Misleh, and Wilfred W. Li. 2006. "MEME: Discovering and Analyzing DNA and Protein Sequence Motifs." *Nucleic Acids Research* 34 (WEB. SERV. ISS.): 369–73. doi:10.1093/nar/gkl198.
- Barros, Patrícia, Peter Jordan, and Paulo Matos. 2009. "Rac1 Signaling Modulates BCL-6-Mediated Repression of Gene Transcription." *Molecular and Cellular Biology* 29 (15): 4156–66. doi:10.1128/MCB.01813-08.
- Barros, Patrícia, Eric W F Lam, Peter Jordan, and Paulo Matos. 2012. "Rac1 Signalling Modulates a STAT5/BCL-6 Transcriptional Switch on Cell-Cycle-Associated Target Gene Promoters." *Nucleic Acids Research* 40 (16): 7776–87. doi:10.1093/nar/gks571.
- Barski, Artem, and Keji Zhao. 2008. "Genomic Location Analysis by ChIP-Seq." *Journal of Cellular Biochemistry* 18 (December 2008): 11–18. doi:10.1002/jcb.22077.
- Benitah, Salvador A., Pilar F. Valerón, Linda Van Aelst, Christopher J. Marshall, and Juan Carlos Lacal. 2004. "Rho GTPases in Human Cancer: An Unresolved Link to Upstream and Downstream Transcriptional Regulation." *Biochimica et Biophysica Acta - Reviews on Cancer* 1705 (2): 121–32. doi:10.1016/j.bbcan.2004.10.002.
- Bosco, E E, J C Mulloy, and Y Zheng. 2009. "Rac1 GTPase : A ' Rac ' of All Trades" 66: 370–74. doi:10.1007/s00018-008-8552-x.
- Bostner, J, M Ahnstro, and O Sta. 2007. "Amplification of CCND1 and PAK1 as Predictors of Recurrence and Tamoxifen Resistance in Postmenopausal Breast Cancer," no. December 2006: 6997–7005. doi:10.1038/sj.onc.1210506.
- Brenner, Hermann, Matthias Kloor, and Christian Peter Pox. 2014. "Colorectal Cancer" 383.
- Chen, Guoting, Ning Han, Guofeng Li, Xin Li, Guang Li, Zengchun Li, and Qinchuan Li. 2016. "Time Course Analysis Based on Gene Expression Profile and Identification of Target Molecules for Colorectal Cancer." *Cancer Cell International* 16 (1). BioMed Central: 22. doi:10.1186/s12935-016-0296-3.
- Ciccio, Alberto, and Stephen J. Elledge. 2010. "The DNA Damage Response: Making It Safe to Play with Knives." *Molecular Cell* 40 (2). Elsevier Inc.: 179–204. doi:10.1016/j.molcel.2010.09.019.
- Collins, A. R. 2004. "The Comet Assay for DNA Damage and Repair: Principles, Applications, and Limitations." *Molecular Biotechnology* 26 (3): 249–61. doi:10.1385/MB:26:3:249.
- Collins, Andrew R, Ma Ai-guo, and Susan J Duthie. 1994. "The Kinetics of Repair of Oxidative DNA Damage (Strand Breaks and Oxidised Pyrimidines) in Human Cells" 336: 69–77.
- D'Haeseleer, Patrik. 2006. "What Are DNA Sequence Motifs?" *Nat Biotech* 24 (4): 423–25. doi:10.1038/nbt0406-423.
- Dalby, Brian, Sharon Cates, Adam Harris, Elise C. Ohki, Mary L. Tilkins, Paul J. Price, and Valentina C. Ciccarone. 2004. "Advanced Transfection with Lipofectamine 2000 Reagent: Primary Neurons, siRNA, and High-Throughput Applications." *Methods* 33 (2): 95–103.

doi:10.1016/j.ymeth.2003.11.023.

- Das, S., and H. R. Dash. 2015. "Cloning and Transformation 2." In *Microbial Biotechnology - A Laboratory Manual for Bacterial Systems*, 35–45. doi:10.1007/978-81-322-2095-4.
- De Rosa, Marina, Ugo Pace, Daniela Rega, Valeria Costabile, Francesca Duraturo, Paola Izzo, and Paolo Delrio. 2015. "Genetics, Diagnosis and Management of Colorectal Cancer (Review)." *Oncology Reports* 34 (3): 1087–96. doi:10.3892/or.2015.4108.
- Del Pulgar, Teresa Gómez, Eva Bandrés, Carolina Espina, Fátima Valdés-Mora, Rosa Pérez-Palacios, Fermín García-Amigot, Jesús García-Foncillas, and Juan Carlos Lacal. 2007. "Differential Expression of Rac1 Identifies Its Target Genes and Its Contribution to Progression of Colorectal Cancer." *International Journal of Biochemistry and Cell Biology* 39 (12): 2289–2302. doi:10.1016/j.biocel.2007.06.014.
- Del Pulgar, Teresa Gómez, Salvador A. Benitah, Pilar F. Valerón, Carolina Espina, and Juan Carlos Lacal. 2005. "Rho GTPase Expression in Tumourigenesis: Evidence for a Significant Link." *BioEssays* 27 (6): 602–13. doi:10.1002/bies.20238.
- Dent, Alexander L, Farha H Vasanwala, and Lisa M Toney. 2002. "Regulation of Gene Expression by the Proto-Oncogene BCL-6" 41: 1–9.
- Desrivieres, Sylvane, Christian Kunz, Itamar Barash, Vida Vafaizadeh, Corina Borghouts, and Bernd Groner. 2006. "The Biological Functions of the Versatile Transcription Factors STAT3 and STAT5 and New Strategies for Their Targeted Inhibition," 75–87. doi:10.1007/s10911-006-9014-4.
- Dokmanovic, Milos, Dianne S Hirsch, and Yi Shen. 2009. "Rac1 Contributes to Trastuzumab Resistance of Breast Cancer Cells : Rac1 as a Potential Therapeutic Target for the Treatment of Trastuzumab-Resistant Breast Cancer" 8 (June): 1557–69. doi:10.1158/1535-7163.MCT-09-0140.
- Eisen, Michael B., Paul T. Spellman, Patrick O. Brown, and David Botstein. 1998. "Cluster Analysis and Display of Genome-Wide Expression Patterns" 95 (22): 14863–14868.
- Ellenbroek, Saskia I J, and John G Collard. 2007. "Rho GTPases : Functions and Association with Cancer," 657–72. doi:10.1007/s10585-007-9119-1.
- Espina, Carolina, María Virtudes Céspedes, Miguel Angel García-Cabezas, María Teresa Gómez del Pulgar, Alicia Boluda, Lourdes García Oroz, Salvador A Benitah, et al. 2008. "A Critical Role for Rac1 in Tumor Progression of Human Colorectal Adenocarcinoma Cells." *The American Journal of Pathology* 172 (1): 156–66. doi:10.2353/ajpath.2008.070561.
- Evans, James D. 1996. "Straightforward Statistics for the Behavioral Sciences," 122.
- Freshney, R. Ian. 2000. "Culture of Animal Cells Set." *Medicina Clinica* 82: 608. doi:10.1002/9780471747598.
- Fritz, Gerhard, and Christian Henninger. 2015. "Rho GTPases: Novel Players in the Regulation of the DNA Damage Response?" *Biomolecules* 5 (4): 2417–34. doi:10.3390/biom5042417.
- Fritz, Gerhard, Ingo Just, and Bernd Kaina. 1999. "Rho GTPases Are over-Expressed in Human Tumors." *International Journal of Cancer* 81 (5): 682–87. doi:10.1002/(SICI)1097-0215(19990531)81:5<682::AID-IJC2>3.0.CO;2-B.
- Grady, W M. 2004. "Genomic Instability and Colon Cancer." *Cancer Metastasis Rev.* 23 (1–2): 11–27. doi:10.1023/A:1025861527711.
- Gürtler, Anne, Nancy Kunz, Maria Gomolka, Sabine Hornhardt, Anna A. Friedl, Kevin McDonald, Jonathan E. Kohn, and Anton Posch. 2013. "Stain-Free Technology as a Normalization Tool in Western Blot Analysis." *Analytical Biochemistry* 433 (2): 105–11. doi:10.1016/j.ab.2012.10.010.

- Gyori, Benjamin M, Gireedhar Venkatachalam, P S Thiagarajan, David Hsu, and Marie-veronique Clement. 2014. "Redox Biology OpenComet : An Automated Tool for Comet Assay Image Analysis." *Redox Biology* 2. Elsevier: 457–65. doi:10.1016/j.redox.2013.12.020.
- Hall, Alan. 1998. "Rho GTPases and the Actin Cytoskeleton." *Science* 509 (1998). doi:10.1126/science.279.5350.509.
- Heidebrecht, F., A. Heidebrecht, I. Schulz, S. E. Behrens, and A. Bader. 2009. "Improved Semiquantitative Western Blot Technique with Increased Quantification Range." *Journal of Immunological Methods* 345 (1–2). Elsevier B.V.: 40–48. doi:10.1016/j.jim.2009.03.018.
- Hein, A L, C M Post, Y M Sheinin, I Lakshmanan, A Natarajan, C A Enke, S K Batra, M M Ouellette, and Y Yan. 2016. "RAC1 GTPase Promotes the Survival of Breast Cancer Cells in Response to Hyper-Fractionated Radiation Treatment." *Nature Publishing Group*, no. March. Nature Publishing Group: 1–11. doi:10.1038/onc.2016.163.
- Helleday, T, E Petermann, C Lundin, B Hodgson, and R A Sharma. 2008. "DNA Repair Pathways as Targets for Cancer Therapy." *Nat Rev Cancer* 8 (3): 193–204. doi:nrc2342 [pii]\n10.1038/nrc2342.
- Helleday, Thomas, Saeed Eshtad, and Serena Nik-zainal. 2014. "Mechanisms Underlying Mutational Signatures in Human Cancers." *Nature Publishing Group*, no. July. Nature Publishing Group: 1–14. doi:10.1038/nrg3729.
- Huggett, J, K Dheda, S Bustin, and A Zumla. 2005. "Real-Time RT-PCR Normalisation; Strategies and Considerations." *Genes and Immunity* 6 (4): 279–84. doi:10.1038/sj.gene.6364190.
- Jackson, Stephen P, and Jiri Bartek. 2009. "The DNA-Damage Response in Human Biology and Disease." *Nature* 461 (7267): 1071–78. doi:10.1038/nature08467.The.
- Jaffe, Aron B, and Alan Hall. 2005. "Rho GTPases : Biochemistry and Biology." *Cell*, 247–72. doi:10.1146/annurev.cellbio.21.020604.150721.
- Jensen, Ellen C. 2012. "The Basics of Western Blotting." *Anatomical Record* 295 (3): 369–71. doi:10.1002/ar.22424.
- Jordan, Jennifer J, Daniel Menendez, Alberto Inga, Maher Nourredine, Douglas Bell, and A Michael. 2008. "Noncanonical DNA Motifs as Transactivation Targets by Wild Type and Mutant p53" 4 (6). doi:10.1371/journal.pgen.1000104.
- Kachalaki, Saeed, Mina Ebrahimi, Leila Mohamed Khosroshahi, Sina Mohammadinejad, and Behzad Baradaran. 2016. "Cancer Chemoresistance; Biochemical and Molecular Aspects: A Brief Overview." *PHASCI*. Elsevier B.V. doi:10.1016/j.ejps.2016.03.025.
- Kamata, H, H Yagisawa, S Takahashi, and H Hirata. 1994. "Amphiphilic Peptides Enhance the Efficiency of Liposome-Mediated DNA Transfection 327." *Nucleic Acids Res* 22 (3): 536–37.
- Kawashima, Toshiyuki, Chun Bao Ying, Yasushi Nomura, Yuseok Moon, Yukio Tonozuka, Yukinori Minoshima, Tomonori Hatori, et al. 2006. "Rac1 and a GTPase-Activating Protein, MgcRacGAP, Are Required for Nuclear Translocation of STAT Transcription Factors." *Journal of Cell Biology* 175 (6): 937–46. doi:10.1083/jcb.200604073.
- Kharchenko, Peter V, Michael Y Tolstorukov, and Peter J Park. 2008. "Design and Analysis of ChIP-Seq Experiments for DNA-Binding Proteins" 26 (12): 1351–59. doi:10.1038/nbt.1508.
- Kim, Tae Kyung, and James H. Eberwine. 2010. "Mammalian Cell Transfection: The Present and the Future." *Analytical and Bioanalytical Chemistry* 397 (8): 3173–78. doi:10.1007/s00216-010-3821-6.
- Kumar, R., and D. -Q. Li. 2016. "PAKs in Human Cancer Progression : From Inception to Cancer Therapeutic to Future Oncobiology" 130. doi:10.1016/bs.acr.2016.01.002.

- Kumar, Rakesh, Anupama E Gururaj, and Christopher J Barnes. 2006. "p21-Activated Kinases in Cancer" 6 (June): 459–71. doi:10.1038/nrc1892.
- Kurien, Biji T., and R. Hal Scofield. 2006. "Western Blotting." *Methods* 38 (4): 283–93. doi:10.1016/j.ymeth.2005.11.007.
- Liang, Kun, and Sündüz Kele. 2012. "Detecting Differential Binding of Transcription Factors with ChIP-Seq." *Bioinformatics* 28 (1): 121–22. doi:10.1093/bioinformatics/btr605.
- Litterst, Claudia M., Stefanie Kliem, Marco Lodrini, and Edith Pfitzner. 2005. "Coactivators in Gene Regulation by STAT5" 70: 359–86.
- Lord, Christopher J., and Alan Ashworth. 2012. "The DNA Damage Response and Cancer Therapy." *Nature* 481 (7381): 287–94. doi:10.1038/nature10760.
- Ma, Wenxiu, William S Noble, and Timothy L Bailey. 2014. "Motif-Based Analysis of Large Nucleotide Data Sets Using MEME- ChIP" 9 (6): 1428–50. doi:10.1038/nprot.2014.083.Motif-based.
- MacPhee, Daniel J. 2010. "Methodological Considerations for Improving Western Blot Analysis." *Journal of Pharmacological and Toxicological Methods* 61 (2). Elsevier Inc.: 171–77. doi:10.1016/j.vascn.2009.12.001.
- Malliri, Angeliki, Rob A Van Der Kammen, and Kristopher Clark. 2002. "Mice Deficient in the Rac Activator Tiam1 Are Resistant to Ras-Induced Skin Tumours" 417 (June): 867–71.
- Manneville, Sandrine Etienne-, and Alan Hall. 2002. "Rho GTPases in Cell Biology." *Nature* 420 (December): 629–35.
- Marone, Maria, Simona Mozzetti, Daniela De Ritis, Luca Pierelli, and Giovanni Scambia. 2001. "Semiquantitative RT-PCR Analysis to Assess the Expression Levels of Multiple Transcripts from the Same Sample." *Biological Procedures Online* 3 (1): 19–25. doi:10.1251/bpo20.
- Matos, Paulo, Jennifer Skaug, Sebastian Beck, Christian Gespach, Maria Guida Boavida, Stephen W Scherer, and Peter Jordan. 2000. "Small GTPase Rac1 : Structure , Localization , and Expression of the Human Gene" 751: 741–51. doi:10.1006/bbrc.2000.3743.
- Mernier, G., S. Majocchi, N. Mermod, and P. Renaud. 2012. "In Situ Evaluation of Single-Cell Lysis by Cytosol Extraction Observation through Fluorescence Decay and Dielectrophoretic Trapping Time." *Sensors and Actuators, B: Chemical* 166–167. Elsevier B.V.: 907–12. doi:10.1016/j.snb.2012.03.057.
- Michiels, Frits, Jord C Stam, Peter L Hordijk, Rob A Van Der Kammen, Lisette Ruuls-van Stalle, Constance A Feltkamp, and John G Collard. 1997. "Regulated Membrane Localization of Tiam1, Mediated by the NH 2 -Terminal Pleckstrin Homology Domain, Is Required for Rac-Dependent Membrane Ruffling and C-Jun NH 2 -Terminal Kinase Activation" 137 (2): 387–98.
- Moniz, Sónia, Marisa Sousa, Bruno José Moraes, Ana Isabel Mendes, Marta Palma, Celeste Barreto, José I. Fragata, Margarida D. Amaral, and Paulo Matos. 2013. "HGF Stimulation of Rac1 Signaling Enhances Pharmacological Correction of the Most Prevalent Cystic Fibrosis Mutant F508del- CFTR."
- Nandhakumar, S, S Parasuraman, MM Shanmugam, Kramachandra Rao, Parkash Chand, and Bvishnu Bhat. 2011. "Evaluation of DNA Damage Using Single-Cell Gel Electrophoresis (Comet Assay)." *Journal of Pharmacology and Pharmacotherapeutics* 2 (2): 107. doi:10.4103/0976-500X.81903.
- Nie, X, and R P Singh. 2001. "A Novel Usage of Random Primers for Multiplex RT-PCR Detection of Virus and Viroid in Aphids , Leaves , and Tubers." *Journal of Virological Methods* 91: 37–49.
- Olive, Peggy L, and Judit P Banáth. 2006. "The Comet Assay: A Method to Measure DNA Damage in

- Individual Cells.” *Nat. Protoc.* 1 (1): 23–29. doi:10.1038/nprot.2006.5.
- Pai, Athma A, Jonathan K Pritchard, and Yoav Gilad. 2015. “The Genetic and Mechanistic Basis for Variation in Gene Regulation” 11 (1). doi:10.1371/journal.pgen.1004857.
- Park, Peter J. 2009. “Of a Maturing Technology.” *Nature Reviews Genetics* 10 (10). Nature Publishing Group: 669–80. doi:10.1038/nrg2641.
- Rio, Donald C. 2014. “Reverse Transcription-Polymerase Chain Reaction.” *Cold Spring Harbor Protocols* 2014 (11): 1207–16. doi:10.1101/pdb.prot080887.
- Robinson, J Paul, Jennifer Sturgis Bs, and George L Kumar. 2013. “Immunofluorescence.” *Immunohistochemical Staining Methods Education Guide*, no. 3: 160.
- Rojas, Ana Maria, Gloria Fuentes, Antonio Rausell, Alfonso Valencia, Ana Maria Rojas, Gloria Fuentes, Antonio Rausell, and Alfonso Valencia. 2012. “The Ras Protein Superfamily : Evolutionary Tree and Role of Conserved Amino Acids.” *Cell Biology* 196 (2): 189–201. doi:10.1083/jcb.201103008.
- Romero, Irene Gallego, Ilya Ruvinsky, and Yoav Gilad. 2014. “Comparative Studies of Gene Expression and the Evolution of Gene Regulation.” *Nat Rev Genet.* 13 (7): 505–16. doi:10.1038/nrg3229.Comparative.
- Ronen, Amiram, and Barry W Glickman. 2001. “Human DNA Repair Genes.” *Environmental and Molecular Mutagenesis* 283: 241–83.
- Rossman, Kent L., Channing J. Der, and John Sodek. 2005. “GEF Means Fo: Turning on Rho GTPases with Guanine Nucleotide-Exchange Factors.” *Nature Reviews Molecular Cell Biology* 6 (2): 167–80. doi:10.1038/nrm1587.
- Sahai, Erik, and Christopher J. Marshall. 2002. “Rho–Gtpases and Cancer.” *Nature Reviews Cancer* 2 (2): 133–42. doi:10.1038/nrc725.
- Sambrook, Joseph, and David W. Russell. 1999. “Molecular Cloning - Technical Guide.” doi:10.1016/B978-0-12-418687-3.00007-0.
- Sanadi, Rizwan M., Nitin Khuller, P. Basavaraj, and Rahul Hegde. 2004. “The Comet Assay for DNA Damage and Repair.” *Molecular Biotechnology* 26 (18): 249–61. doi:10.1385/MB:26:3:249.
- Seifi, Morteza, Asghar Ghasemi, Siamak Heidarzadeh, Mahmood Khosravi, Atefeh Namipashaki, Vahid Mehri Soofiany, Ali Alizadeh Khosroshahi, and Nasim Danaei. 2012. “Overview of Real-Time PCR Principles.” *Polymerase Chain Reaction*, 405–42. doi:10.5772/39220.
- Suggitt, M, J Fearnley, D J Swaine, M Volpato, R M Phillips, M C Bibby, P M Loadman, and D Anderson. 2003. “Comet Assay and Flow Cytometry Analysis of DNA Repair in Normal and Cancer Cells Treated With Known Mutagens With Different Mechanisms of Action.” *Teratogenesis, Carcinogenesis, and Mutagenesis Supplement* 29: 13–29. doi:10.1002/tcm.10077.
- Swank, Michael W., Vikas Kumar, Jing Zhao, and Gang Yi Wu. 2006. “A Novel Method of Loading Samples onto Mini-Gels for SDS-PAGE: Increased Sensitivity and Western Blots Using Sub-Microgram Quantities of Protein.” *Journal of Neuroscience Methods* 158 (2): 224–33. doi:10.1016/j.jneumeth.2006.06.002.
- Tran, Mai Quynh Thanh, Yvonne Nygren, Christina Lundin, Peter Naredi, and Erik Björn. 2010. “Evaluation of Cell Lysis Methods for Platinum Metallomic Studies of Human Malignant Cells.” *Analytical Biochemistry* 396 (1). Elsevier Inc.: 76–82. doi:10.1016/j.ab.2009.08.044.
- Troester, Melissa A, Katherine A Hoadley, Therese Sørli, Brittney-shea Herbert, Per Eystein Lønning, Jerry W Shay, William K Kaufmann, and Charles M Perou. 2004. “Cell-Type-Specific Responses to Chemotherapeutics in Breast Cancer,” 4218–26.

- Valones, Marcela Agne Alves, Rafael Lima Guimarães, Lucas André Cavalcanti Brandão, Paulo Roberto Eleutério De Souza, Alessandra De Albuquerque Tavares Carvalho, and Sergio Crovela. 2009. "Principles and Applications of Polymerase Chain Reaction in Medical Diagnostic Fields: A Review." *Brazilian Journal of Microbiology* 40 (1): 1–11. doi:10.1590/S1517-83822009000100001.
- Valouev, Anton, David S Johnson, Andreas Sundquist, Catherine Medina, Elizabeth Anton, Serafim Batzoglou, Richard M Myers, and Arend Sidow. 2008. "Genome-Wide Analysis of Transcription Factor Binding Sites Based on ChIP-Seq Data." *Nature* 5 (9): 829–34. doi:10.1038/NMETH.1246.
- Watson, Alastair J M. 2006. "An Overview of Apoptosis and the Prevention of Colorectal Cancer." *Critical Reviews in Oncology/Hematology* 57 (2): 107–21. doi:10.1016/j.critrevonc.2005.06.005.
- Wood, Richard D, Michael Mitchell, and Tomas Lindahl. 2005. "Human DNA Repair Genes , 2005." *Mutation Research* 577: 275–83. doi:10.1016/j.mrfmmm.2005.03.007.
- Zhao, Shi-Yi, Yan Sun, Zhuo-Sheng Lai, Qing-Zhen Nan, Kang Li, and Zhen-Shu Zhang. 2008. "Inhibition of Migration and Invasion of Colorectal Cancer Cells via Deletion of Rac1 with RNA Interference." *Molecular Cell Biochemistry*, 179–84. doi:10.1007/s11010-008-9955-6.
- Zhu, G, Y Wang, B Huang, J Liang, Y Ding, A Xu, and W Wu. 2012. "A Rac1 / PAK1 Cascade Controls B -Catenin Activation in Colon Cancer Cells." *Nature* 1: 1001–12. doi:10.1038/onc.2011.294.
- Zou, Ting, Xiaoyuan Mao, Jiye Yin, Xi Li, Juan Chen, Tao Zhu, Honghao Zhou, Zhaoqian Liu, Xiangya Hospital, and Zhao-qian Liu. 2016. "Emerging Roles of RAC1 in Treating Lung Cancer Patients." *Clinical Genetics*.

7. Anexos

Anexo I

Picos Exclusivos da Activação da Via Rac1/PAK1 - Total:						2419
Cromossomo	Início	Fim	Largura	TSS anotado	Distância ao TSS	Simbolo, HGNC
1	27961957	27962274	318	ENSG00000009038	-169	FCBR
1	160702552	160702753	202	ENSG00000007908	-668	SELE
2	23859681	23859920	240	ENSG00000007968	-1969	E2F2
2	213409111	213417014	264	ENSG00000001072	-2940	SPRNT
2	211472424	211472505	82	ENSG00000001072	-427	SPRNT
2	24068997	24069275	279	ENSG00000001007	-649	TECB3
2	24115409	24115505	97	ENSG00000001009	-2052	LYPLA2
2	24116821	24116900	80	ENSG00000001009	-640	LYPLA2
2	87013342	87013624	283	ENSG00000001602	982	CLCA4
2	36021672	36021924	253	ENSG00000002019	-1403	NCN9
2	94373979	94374135	157	ENSG00000002309	987	GCLM
2	115322840	115322911	83	ENSG00000005723	457	SRK1
2	36609087	36609048	52	ENSG000000054118	879	TIRAP3
2	235494458	235494639	182	ENSG000000054267	-2924	ARID4B
2	85516351	85516327	177	ENSG00000005732	-2149	MYCLN3
2	76262307	76262559	253	ENSG000000057468	-261	MSH1
2	24105680	24105855	176	ENSG000000057757	784	PTIBD1
2	29562199	29562468	268	ENSG000000060656	-830	PTPRU1
2	173862208	173862262	55	ENSG000000074321	-1873	KILB20
2	210403792	210404034	243	ENSG000000082497	-2353	SERTAD4
2	63832462	63832770	309	ENSG000000080035	-800	ALG6
2	120202921	120202945	23	ENSG000000092621	901	PDEHD1
2	147010614	147010879	266	ENSG000000116128	-2569	BCL9
2	8086884	8087103	220	ENSG000000116285	-516	EBRFT1
2	29173559	29173687	129	ENSG000000116329	-1096	OPRD1
2	111993524	111993699	176	ENSG000000116455	-1526	WDR77
2	111994098	111994108	11	ENSG000000116455	-2100	WDR77
2	111994268	111994381	117	ENSG000000116455	-2207	WDR77
2	33430727	33430972	246	ENSG000000116514	-778	RNF90B
2	33646713	33647037	325	ENSG000000116525	958	TRIM62
2	156471238	156471401	163	ENSG000000116604	-84	MEP2D
2	70668081	70668029	52	ENSG000000116754	-5653	SRF1
2	154250801	154250925	124	ENSG000000116774	-1262	OLFM3L
2	72001651	72001715	65	ENSG000000116971	-2559	CRYZ
2	236229044	236229276	232	ENSG000000117353	-145	NID1
2	40369168	40369234	157	ENSG000000116990	-742	MYCL
2	241692544	241692894	351	ENSG000000117009	-2891	KMO
2	206765904	206766384	480	ENSG000000118416	-1415	CR2
2	44441921	44441988	68	ENSG000000117411	-895	BAGAL2T
2	46016679	46016907	229	ENSG000000117448	-465	AKR1A1
2	171108572	171108140	269	ENSG000000117507	-1008	FMO4P
2	171108607	171109951	125	ENSG000000117507	-145	VAMP4
2	26759749	26759781	33	ENSG000000117682	-975	DHSD5
2	28156227	28156322	86	ENSG000000117751	-1053	PPV1R3
2	25179591	25181168	167	ENSG000000117751	-1053	PPV1R3
2	45479333	45479526	194	ENSG000000126107	-2306	HECTD3
2	27932769	27932933	165	ENSG000000126705	-1827	AHDC1
2	19536249	19536281	33	ENSG000000117533	-121	UBR4
2	10859003	10859034	302	ENSG000000130940	-2296	CASZ1
2	160120418	160120580	163	ENSG000000132081	-935	ATP1A4
2	10923599	10923593	25	ENSG000000134186	-1177	PIPF3B3
2	109235651	10923561	41	ENSG000000134186	-1295	PIPF3B3
2	113009248	113009331	84	ENSG000000134245	-84	WNTB2
2	17725370	17725314	56	ENSG000000134376	-207	CRB1
2	38061349	38061356	7	ENSG000000134376	-1837	GNL2
2	38062164	38062166	3	ENSG000000134697	-628	GNL2
2	52860407	52860480	134	ENSG000000134748	-2190	PEPBBP8A
2	23432266	23432129	137	ENSG000000134748	-1407	PCN3L2
2	22979136	22979459	324	ENSG000000135763	-2846	UBR2
2	17241472	17241507	36	ENSG000000135845	-1542	PRCC
2	209459754	20945960	147	ENSG000000136634	45	IL10
2	21574070	215740387	128	ENSG000000136636	-658	KCTD3
2	94314868	94313903	226	ENSG000000137936	-2162	BCAR3
2	78445483	78445550	68	ENSG000000137960	-258	GIPC2
2	94706584	94706583	1	ENSG000000137960	-258	RECAP4P2
2	8878384	8878384	301	ENSG000000142599	-685	REEE
2	32042971	32042926	10	ENSG000000142910	830	TINAC1L
2	109235911	109235292	59	ENSG000000142910	-1066	FNDC7
2	165667368	16566727	260	ENSG000000143149	732	ALDRB1
2	167902817	167902861	45	ENSG000000143164	-205	DCAF6
2	16779977	16778026	250	ENSG000000143185	-1550	TMCO1
2	167109544	167109684	140	ENSG000000143196	-125	TPST3
2	168699713	168699227	55	ENSG000000143196	-671	DPT
2	16147439	161474066	212	ENSG000000143226	-826	FCGR2A
2	156702713	156702417	256	ENSG000000143294	-228	PRC1
2	156712201	156712448	246	ENSG000000143314	-821	MRPL24
2	18257417	182574263	87	ENSG000000143263	-634	RCG16
2	15182411	151824273	163	ENSG000000143376	-1178	SNX27
2	15117042	151170569	228	ENSG000000143398	-84	PIPSK1A
2	150601401	150601420	20	ENSG000000143420	697	ENSA
2	212208583	212208601	19	ENSG000000143476	-2337	DTL
2	22191489	221914501	597	ENSG000000143501	-977	DCAF10
2	153949370	153949561	192	ENSG000000143543	794	JTB
2	154167730	154167475	106	ENSG000000143549	-2761	TPM3
2	153939354	153939476	113	ENSG000000143570	824	SLC39A1
2	153939354	153939476	113	ENSG000000143578	-467	CREBB1A
2	153881418	153881564	147	ENSG000000143622	-223	RIT1
2	94232225	94232429	205	ENSG000000143622	-1146	FAM80A
2	10022322	100223428	107	ENSG000000158699	-973	FRRS1
2	54875062	54875093	32	ENSG000000157216	-2970	SSBP1
2	12805612	12805813	177	ENSG000000157300	-522	Clorf158
2	52869745	52869946	121	ENSG000000157978	-327	LILR4
2	158259953	158259970	18	ENSG000000158481	29	CDIC
2	201368995	201369031	37	ENSG000000159166	-259	LAD1
2	201374070	201374081	12	ENSG000000159466	-1734	LAD1
2	201479970	201480337	368	ENSG000000159176	-1386	CSRP1
2	153125813	153125997	185	ENSG000000159516	-2468	SPRR2G
2	46150621	46150701	81	ENSG000000159516	-2468	TMEM469
2	46150770	46150848	115	ENSG000000159596	-2117	TMEM69
2	46151430	46151524	95	ENSG000000159596	-1457	TMEM69
2	155277224	155277391	168	ENSG000000160753	-1318	PDPS
2	15602281	156052491	111	ENSG000000160789	-218	LAMN
2	156571256	156571444	189	ENSG000000160818	46	GPATCH4
2	6614980	6615263	284	ENSG000000162408	-385	NOL9
2	8782109	8782161	53	ENSG000000162426	-1738	SLC45A3
2	62191274	62191495	222	ENSG000000162604	-1778	TMED1
2	78445483	78445550	68	ENSG000000162613	-689	FUBP1
2	244213702	244213960	258	ENSG000000162616	623	DNAH8
2	200380910	20038107	198	ENSG000000162702	-1726	ZNF281
2	247494732	247494797	65	ENSG000000162714	415	ZNF496
2	159925757	159926107	321	ENSG000000162723	-1743	SLAMP1
2	161956888	161956798	111	ENSG000000162745	-2044	OLFM1B2
2	209958112	209959991	180	ENSG000000162757	-1908	Clorf74
2	224625259	224625355	97	ENSG000000162923	-524	WDR26
2	224625010	224625315	214	ENSG000000162923	-1367	WDR26
2	224626354	224626413	60	ENSG000000162923	-1619	WDR26
2	151170324	151170569	228	ENSG000000163159	-2545	VP52
2	153017209	153017247	39	ENSG000000163216	-2802	SPRR2D
2	114470667	114470825	159	ENSG000000163439	-1278	HPK1
2	202976916	202977164	249	ENSG000000163444	-401	TMEM183A
2	155145329	155145593	265	ENSG000000163462	-1018	TRIM46
2	205293740	205293936	197	ENSG000000163455	-2482	NUAK2
2	37939688	37939747	60	ENSG000000163874	-466	ZC3H12A
2	43310312	43310452	141	ENSG000000164011	-1969	ZNF991
2	43311847	43312088	242	ENSG000000164011	-234	ZNF991
2	110198778	110199005	228	ENSG000000168765	74	GSTM4
2	55503268	55503347	80	ENSG000000169174	-1953	PKCNR1
2	24911058	24911280	213	ENSG000000171161	-1422	ZNF712
2	67268842	67268873	32	ENSG000000172410	-1903	NSL5
2	116653092	116653141	50	ENSG000000173212	-1285	MAR21L3
2	16456587	1645772	115	ENSG000000173212	-1285	FBPFB1
2	6614980	6615263	284	ENSG000000173662	-202	TAS1R1
2	24451514	24451563	30	ENSG000000173728	-324	Clorf100
2	84974496	84974496	1	ENSG000000174021	-2248	GNCS
2	145724528	145724775	247	ENSG000000174021	-2248	GNCS
2	26437623	26437822	200	ENSG000000175087	-42	BKD1L
2	1164949	1165169	221	ENSG000000176022	-2681	BXGAL1E
2	204010400	204010808	398	ENSG000000176022	-2681	LINC001804
2	182929241	182929459	309	ENSG000000176855	299	KRT18P28
2	11795987	11796005	18	ENSG000000176764	-155	ACTRAP
2	153962294	153962618	325	ENSG000000176764	-155	RSF2
2	44676619	44676861	243	ENSG000000178028	-2509	DMAPI
2	145075985	145076097	113	ENSG000000178014	210	PDE4DIP
2	145076468	145076516	49	ENSG000000178014	-273	PDE4DIP
2	145079801	145079345	456	ENSG000000178014	-1888	PDE4DIP
2	145078509	145078764	218	ENSG000000178014	-2314	PDE4DIP
2	43922478	43922551	74	ENSG000000179022	-2844	HY1
2	244213702	244213960	258	ENSG000000179456	-487	ZBTB18
2	158697223	158697532	310	ENSG000000180437	-2321	OR6A6
2	146648731	146648732	147	ENSG000000180867	-1447	PDA3P1
2	26197656	26197670	215	ENSG000000182749	88	PAQR7
2	152061806	152062306	501			

1	42926172	42926337	166	ENSG00000186409	-2830	CCDC30
1	15853309	158533310	252	ENSG00000186440	-315	OROP1
1	16345687	16345772	86	ENSG00000186510	-356	CLCNKA
1	22137230	22137422	193	ENSG00000187942	-1529	LDLRAD2
1	20504216	20504225	20	ENSG00000187980	-2529	PLA2G2C
1	203462074	203462238	165	ENSG00000188770	-1198	OPTC
1	1625577	1625847	271	ENSG00000189339	-1410	SLC35E2B
1	44497247	44497471	225	ENSG00000196517	-113	SLC6A9
1	153779797	153779842	46	ENSG00000197084	-690	LCE1C
1	46218086	46218258	173	ENSG00000197429	-1764	IPP
1	151969716	151969865	150	ENSG00000197747	-2850	S100A10
1	109620520	109620664	145	ENSG00000197780	-1996	TAFI3
1	153511397	153511522	126	ENSG00000197956	-2677	S100A6
1	38272370	38272399	30	ENSG00000197982	-282	C1orf122
1	38272469	38272586	218	ENSG00000197982	-183	C1orf122
1	43885288	43885576	289	ENSG00000198198	-1141	SZT2
1	158459793	158459884	92	ENSG00000198703	-1217	OR10R3P
1	109558434	109558521	88	ENSG00000198734	-2608	F5
1	52490846	52490903	218	ENSG00000198841	-358	KTLL2
1	184018384	184018555	172	ENSG00000198860	-2428	TSEN15
1	162040148	162040299	152	ENSG00000198929	581	NO51AP
1	156253779	156253861	83	ENSG00000198952	-1163	SMG5
1	147733865	147733995	131	ENSG00000198979	-1887	RNU1-120P
1	28835374	28835450	77	ENSG00000200087	302	SNORA73B
1	163427950	163427951	2	ENSG00000200227	445	RNASSP62
1	173941302	173941564	263	ENSG00000200755	-2728	RNASSP68
1	38862751	38862831	81	ENSG00000200796	419	RNU6-753P
1	52438464	52438674	211	ENSG00000200839	-619	RNASSP48
1	45342104	45342223	120	ENSG00000200913	99	SNORD46
1	45242104	45242223	120	ENSG00000202031	-1412	SNORD38A
1	199983937	199984244	308	ENSG00000203239	-14	RNU6-609P
1	247652387	247652457	171	ENSG00000203664	-616	OR2W5
1	150534463	150534471	9	ENSG00000203804	128	ADAMTSL4-AS1
1	27151432	27151544	113	ENSG00000204160	-1770	ZEBH18
1	15497919	15497978	60	ENSG00000204464	-106	C1orf195
1	168254148	168254254	107	ENSG00000206880	-1429	RNU6-1310P
1	224491700	224491882	183	ENSG00000206887	-1383	RNU6-1008P
1	26289537	26289704	168	ENSG00000207237	-1152	RNU6-110P
1	45342104	45342223	120	ENSG00000207421	-1999	SNORD15B
1	3478325	3478487	163	ENSG00000207776	-971	MR551A
1	172108981	172109007	25	ENSG00000207949	-936	MR214
1	172116667	172116900	234	ENSG00000208024	-2883	MR19A2
1	197746501	197746595	95	ENSG00000213047	-1675	DENN1B
1	238652822	238652963	142	ENSG00000215807	-1791	KKTRP65
1	54666492	54666719	228	ENSG00000215883	-746	CTVSRL
1	23291592	23291647	56	ENSG00000215906	205	LACTBL1
1	154161730	154167475	106	ENSG00000215938	-1151	MR190B
1	16942901	16942943	43	ENSG00000215981	-2818	NBP1
1	167966698	167966877	180	ENSG0000021545	-1201	MR125B2
1	159147993	159148174	182	ENSG00000222552	-280	RNASSP60
1	144944207	144944269	63	ENSG00000222788	-466	RNU2-38P
1	144944778	144945094	316	ENSG00000222788	106	RNU2-38P
1	236017590	236017625	36	ENSG00000222831	-1230	MR1537
1	28149158	28149417	260	ENSG00000223062	-1892	RNU6-1245P
1	151923322	151923493	171	ENSG00000223810	-67	KRT2P28
1	1071385	1071648	264	ENSG00000223823	-1013	LINC01342
1	174727472	174727492	21	ENSG00000223828	-1215	BANFP4
1	211834952	211835176	225	ENSG00000223763	-333	FOPSP8
1	46916367	46916374	8	ENSG00000224863	-991	LINC01398
1	46916682	46916877	196	ENSG00000224863	-1306	LINC01398
1	43367155	43367338	184	ENSG00000225099	-1749	ATP9V1E1P1
1	100674900	100677094	195	ENSG00000225169	-1950	BRIP1
1	46113889	46113909	21	ENSG00000225447	-2020	RPS15AP10
1	146675777	146675944	168	ENSG00000226015	736	CTBP1
1	202764051	202764353	303	ENSG00000226148	892	SLC25A39P1
1	158484570	158484690	121	ENSG00000227445	-6	OR10R1P
1	155367996	155368130	135	ENSG00000227773	-1227	ASHL1-IT1
1	155368259	155368432	174	ENSG00000227773	-1490	ASHL1-IT1
1	199340185	199340204	20	ENSG00000228190	-806	RPL23AP16
1	226855321	226855398	78	ENSG00000228458	-1278	ITPKB-AS1
1	113464689	113464885	197	ENSG00000229020	-1284	AKR7A2P1
1	172116667	172116900	234	ENSG00000238636	-2733	DNMA3C
1	173114189	173114389	201	ENSG00000238649	-2700	GOT2P2
1	174822290	174822343	144	ENSG00000231378	-2894	NDU4AF4P4
1	203254605	203254913	309	ENSG00000231507	-1676	LINC01353
1	193271680	193271830	151	ENSG00000232077	-2196	LINC01031
1	207398646	207398686	41	ENSG00000232621	-498	C4BPAP2
1	198467218	198467359	142	ENSG00000232624	98	PER1-IP1
1	95117093	95117142	50	ENSG00000232857	-98	KATNB1L1P2
1	21330545	21330803	259	ENSG00000233072	-118	RPS15AP6
1	3999392	3999582	191	ENSG00000233304	-1282	LINC01346
1	92590011	92590210	200	ENSG00000233401	-448	PRKAR1AP
1	171831871	171832034	163	ENSG00000233540	-1456	DNM3-IT1
1	37939888	37939747	60	ENSG00000233621	324	LINC01337
1	10116326	10116366	41	ENSG00000233623	-2404	PGAM1P11
1	41062509	41062574	66	ENSG00000233910	-2484	GTF3P22
1	217958702	217958912	211	ENSG00000234070	-240	SPATA17-AS1
1	94391009	94391313	305	ENSG00000234099	-1508	MTND4P11
1	94391735	94391771	37	ENSG00000234099	-2234	MTND4P11
1	6862957	68629847	191	ENSG00000234383	-884	CTBP2P8
1	59247880	59248065	186	ENSG00000234807	-2953	LINC01135
1	113160627	113160907	281	ENSG00000235299	-780	MRPL5P1
1	169958626	169958895	270	ENSG00000235378	-583	MRPS10P1
1	234162485	234162517	33	ENSG00000235601	488	RAC1P7
1	53968783	53968870	88	ENSG00000236253	-777	SLC25A3P1
1	203845140	203845458	209	ENSG00000236430	-1825	KRT2P29
1	24105680	24105855	173	ENSG00000236810	-903	CTCB3-AS1
1	7427379	74274541	163	ENSG00000237402	-1625	CANTAT1-IT1
1	68240314	68240500	187	ENSG00000238778	-1978	RNU7-80P
1	116146722	116146770	49	ENSG00000239984	-2371	RN7SL20P
1	161706105	161706332	228	ENSG00000234147	506	RN7SL46P8
1	89292766	89292868	103	ENSG00000241975	-2020	TCB1P19
1	1344882	1344998	117	ENSG00000242485	-2189	MRP20
1	220748239	220748445	207	ENSG00000243872	-2862	RN7SL46P4
1	202662676	202662719	444	ENSG00000243931	864	ALS2
2	37311936	37311953	18	ENSG00000000869	-451	HEATR5B
2	65281469	65281640	172	ENSG00000011523	-2032	CEP68
2	63813077	63813400	324	ENSG00000011641	-2667	MDM1
2	32859913	32851043	131	ENSG00000013699	-2187	TTCT2
2	37385716	37385891	176	ENSG00000053532	-1508	EIF2AK2
2	47222356	47222356	101	ENSG00000057935	646	MTA3
2	70325116	70325161	46	ENSG00000059728	295	MXD1
2	231281291	231281494	204	ENSG00000060766	633	SP100
2	47141806	47141875	70	ENSG00000060874	-1491	TTCTA
2	47142803	47142722	220	ENSG00000060874	-794	TTCTA
2	47143109	47143368	260	ENSG00000060874	-188	TTCTA
2	54011863	54012099	447	ENSG00000060912	-2319	ERLEC1
2	102312699	102312808	110	ENSG00000071054	-614	MAP4K4
2	131112580	131112563	304	ENSG00000072135	-1321	PTPN18
2	122409138	122409348	211	ENSG00000074054	-1975	CLASP1
2	709970737	70997229	193	ENSG00000075340	-1680	ADJ2
2	98613306	98613555	250	ENSG00000075568	-952	TMEM1131
2	234262290	234262377	88	ENSG00000077044	-864	DGKD
2	202565567	202565679	113	ENSG00000082126	-2150	MPP4
2	47630900	47630183	94	ENSG00000095902	-19	MSH2
2	47630428	47630647	220	ENSG00000095902	319	MSH2
2	47631028	47631051	24	ENSG00000095902	919	MSH2
2	74154447	74154683	237	ENSG00000114956	-893	DCCOK
2	113543762	113544065	304	ENSG00000115008	-1595	IL1A
2	162162960	162162993	34	ENSG00000115233	-1590	PSMD14
2	54083319	54083498	180	ENSG00000115239	-1022	ASB3
2	74679332	74679421	90	ENSG00000115274	-2819	INOB0B
2	54682456	54682686	231	ENSG00000115306	-967	SPTBN1
2	55649922	55650194	273	ENSG00000115355	-2805	CTCDBA
2	58470590	58470850	192	ENSG00000115392	-2152	PANCL
2	216948809	216948971	163	ENSG00000115425	-1131	PECR
2	233897650	233897815	166	ENSG00000115488	267	NEU2
2	62900636	62900726	91	ENSG00000115504	-351	EBHP1
2	198299042	198299359	318	ENSG00000115524	773	SF3B1
2	198379874	198379954	81	ENSG00000115540	-422	MOB4
2	22085888	22085957	70	ENSG00000115657	-176	ABCB6
2	171784159	171784388	230	ENSG00000115806	-816	GORASP2
2	37195104	37195188	85	ENSG00000115808	-1489	STRN
2	37195494	37195553	62	ENSG00000115808	-1879	STRN
2	37461081	37461751	69	ENSG00000115816	-2827	CEBPZ
2	37552618	37552928	311	ENSG00000115825	-467	PRKD3
2	37553510	37553707	198	ENSG00000115825	-1539	PRKD3
2	173861473	173867338	266	ENSG00000115840	-2707	SLC25A12
2	65212859	65212976	118	ENSG00000115902	-2721	SLC1A4
2	43823332	43823577	246	ENSG00000115970	-147	THADA
2	69901902	69902039	138	ENSG00000115977	-421	AKK1
2	70482862	70482890				

Anexo II

Largura_pico	Feature	Distancia à feature	Descrição	Símbolo_HGNC	Max_reads_Act	Med_reads_Act	Med_reads_Ctrl
318	ENSG00000000938	-169	Gardner-Rasheed feline sarco	FGR	4	3	1
444	ENSG00000000393	-364	amyotrophic lateral sclerosis 2	ALS2	6	5	2
85	ENSG000000004766	-1402	coiled-coil domain containing	VP550	4	3	1
260	ENSG000000005135	-621	ligase III, DNA, ATP-depende	LG3	4	3	1
136	ENSG000000005194	-1303	cytokine induced apoptosis-ini	CIAPN1	3	3	1
244	ENSG000000006530	-1021	acylglycerol kinase [Source:H	AGK	2	2	1
57	ENSG000000006831	-618	adiponectin receptor 2 [Source	ADIPOR2	5	4	3
192	ENSG000000007866	-394	TEA domain family member 2	TEAD3	4	3	1
335	ENSG000000010318	-803	PHD finger protein 7 [Source:	PHF7	6	5	1
279	ENSG000000011007	-649	transcription elongation factor	TCEB3	5	4	1
290	ENSG000000014138	-34	polymerase (DNA directed), a	POLA2	3	3	1
253	ENSG000000020129	-1403	neurochondrin [Source:HGNC	NCDN	4	4	1
194	ENSG000000024048	-418	ubiquitin protein ligase E3 cor	UBR2	2	2	1
310	ENSG000000029363	-846	BCL2-associated transcription	BCLAF1	4	3	1
242	ENSG000000030582	-201	granulin [Source:HGNC Syml	GRN	3	3	1
154	ENSG000000039068	-271	cadherin 1, type 1, E-cadherin	CDH1	2	2	1
243	ENSG000000049130	-154	KIT ligand [Source:HGNC Sy	KITLG	2	2	1
195	ENSG000000049167	-776	excision repair cross-comple	ERCC8	2	2	1
228	ENSG000000050130	-926	JNK1/MAPK8-associated me	JKAMP	5	4	1
245	ENSG000000050733	-286	TNFAIP3 interacting protein 2	TNIP3	2	2	1
97	ENSG000000052841	-421	tetratricopeptide repeat domai	TTC17	2	2	1
119	ENSG000000055957	-219	inter-alpha (globulin) inhibitor	ITIH1	2	2	1
253	ENSG000000057468	-261	mutS homolog 4 (E. coli) [Sou	MSH4	2	2	1
70	ENSG000000058091	-865	cyclin-dependent kinase 14 [S	CDK14	5	5	1
232	ENSG000000058729	-151	RIO kinase 2 (yeast) [Source:	RIOK2	2	2	1
288	ENSG000000060656	-830	protein tyrosine phosphatase, i	PTPRU	4	3	2
253	ENSG000000064666	-213	calponin 2 [Source:HGNC Sy	CNN2	5	4	1
291	ENSG000000065357	-309	diacylglycerol kinase, alpha 8c	DGKA	3	3	1
136	ENSG000000065970	-1432	forkhead box J2 [Source:HGN	FOXJ2	3	3	2
236	ENSG000000066739	-997	ATG2 autophagy related 2 ho	ATG2B	2	2	1
220	ENSG000000068724	-794	tetratricopeptide repeat domai	TTC7A	5	4	1
73	ENSG000000068971	-1042	protein phosphatase 2, regulat	PPP2R5B	2	2	1
252	ENSG000000069849	-986	ATPase, Na+/K+ transporting	ATP1B3	4	3	1
116	ENSG000000070808	-942	calcium/calmodulin-dependen	CAMK2A	4	3	1
92	ENSG000000070961	-1500	ATPase, Ca++ transporting, p	ATP2B1	4	4	1
110	ENSG000000071054	-614	mitogen-activated protein kina	MAP4K4	3	3	1
304	ENSG000000072135	-1321	protein tyrosine phosphatase, i	PTPN18	4	4	1
152	ENSG000000072310	-269	sterol regulatory element bindi	SREBF1	5	4	2
92	ENSG000000073050	-1065	X-ray repair complementing d	XRCC1	6	4	1
174	ENSG000000073803	-1100	mitogen-activated protein kina	MAP3K13	3	3	1
204	ENSG000000075223	-126	sema domain, immunoglobuli	SEMA3C	5	4	1
50	ENSG000000075303	-1278	solute carrier family 25, mem	SLC25A40	4	4	1
239	ENSG000000075388	-53	fibroblast growth factor 4 [Sou	HGF4	3	3	1
250	ENSG000000075568	-952	transmembrane protein 131 [S	TMEM131	2	2	1
88	ENSG000000077044	-864	diacylglycerol kinase, delta 13	DGKD	8	5	1
131	ENSG000000078487	-978	zinc finger, CW type with PW	ZCWPW1	4	3	1
299	ENSG000000082781	-692	integrin, beta 5 [Source:HGNC	ITGB5	5	4	2
279	ENSG000000082996	-639	ring finger protein 13 [Source	RNF13	5	4	2
318	ENSG000000085733	-1059	cortactin [Source:HGNC Sym	CTTN	4	4	1
309	ENSG000000088035	-800	asparagine-linked glycosylatio	ALG6	7	5	1
172	ENSG000000088247	-1445	KH-type splicing regulatory pr	KHSRP	3	3	2
62	ENSG000000088782	-1134	defensin, beta 127 [Source:HG	DEFB127	4	4	2
200	ENSG000000088833	-832	NSFL1 (p97) cofactor (p47) [E	NSFL1C	2	2	1
224	ENSG000000089094	-314	lysine [K]-specific demethylas	KDM2B	4	4	1
218	ENSG000000089558	-524	potassium voltage-gated chan	KCNH4	3	3	1
125	ENSG000000091181	-1086	interleukin 5 receptor, alpha [S	IL5RA	2	2	1
200	ENSG000000092010	-854	proteasome (prosome, macrop	PSME1	5	4	3
433	ENSG000000092098	-1361	ring finger protein 31 [Source:	RNF31	9	7	2
195	ENSG000000093010	-921	catechol-O-methyltransferase	COMT	2	2	1
270	ENSG000000094916	-992	chromobox homolog 5 (HP1) i	CBX5	5	4	1
94	ENSG000000095902	-19	mutS homolog 2, colon cance	MSH2	6	4	1
60	ENSG000000095794	-917	cAMP responsive element mo	CREM	6	6	1
165	ENSG000000100206	-1106	DMC1 dosage suppressor of r	DMC1	3	3	1
253	ENSG000000100227	-691	polymerase (DNA-directed), d	POLDIP3	7	5	3
55	ENSG000000100258	-1407	lipase maturation factor 2 [Sou	LMF2	4	4	1
224	ENSG000000100302	-1302	RASD family, member 2 [Sou	RASD2	3	3	2
145	ENSG000000100307	-536	chromobox homolog 7 [Source	CBX7	2	2	1
233	ENSG000000100325	-1043	activating signal cointegrator 1	ASC22	5	4	1
195	ENSG000000100523	-1376	DDHD domain containing 1 [I	DDHD1	6	4	1
248	ENSG000000100554	-1055	ATPase, H+ transporting, lys	ATP6V1D	7	5	1
264	ENSG000000100612	-744	dehydrogenase/reductase (SD	DHRS7	7	5	2
110	ENSG000000100897	-341	DDB1 and CUL4 associated f	DCAF11	8	6	2
265	ENSG000000100906	-380	nuclear factor of kappa light p	NFKBIA	3	3	1
145	ENSG000000100968	-475	nuclear factor of activated T-c	NFATC4	9	6	2
321	ENSG000000101152	-1166	DnaJ (Hsp40) homolog, subfa	DNAJC5	5	4	1
164	ENSG000000101204	-1005	cholinergic receptor, nicotinic,	CHRNA4	3	3	1
126	ENSG000000101311	-273	fermitin family homolog 1 (Di	FERMT1	3	3	1
51	ENSG000000101751	-840	polymerase (DNA directed) io	POLI	2	2	1
54	ENSG000000101940	-1304	WD repeat domain 13 [Source	WDR13	2	2	1
182	ENSG000000102699	-1178	poly (ADP-ribose) polymerase	PARP4	8	5	1
190	ENSG000000102893	-1419	phosphorylase kinase, beta [Se	PHKB	2	2	1
320	ENSG000000102984	-167	zinc finger protein 821 [Sourc	ZNF821	3	3	1
296	ENSG000000103507	-149	branched chain ketoacid dehy	BCKDK	5	4	2
279	ENSG000000104064	-991	GA binding protein transcripti	GABPB1	8	5	1
295	ENSG000000104979	-182	Leydig cell tumor 10 kDa prot	CL9aorf53	5	4	1
219	ENSG000000105443	-937	cytohesin 2 [Source:HGNC Sy	CYTH2	2	2	1
92	ENSG000000105755	-1052	ethylmalonic encephalopathy	ETHE1	5	4	2
179	ENSG000000105852	-1291	paraoxonase 3 [Source:HGNC	PON3	5	4	1
306	ENSG000000105866	-563	Sp4 transcription factor [Sour	SP4	6	4	2
122	ENSG000000105953	-1317	oxoglutarate (alpha-ketoglutar	OGDH	2	2	1
275	ENSG000000106665	-1340	CAP-GLY domain containing	CLIP2	4	3	2
225	ENSG000000108064	-1028	transcription factor A, mitoch	TFAM	4	3	2
331	ENSG000000108557	-603	retinoic acid induced 1 [Sourc	RAI1	3	3	2
239	ENSG000000108588	-756	coiled-coil domain containing	CCDC47	2	2	1
193	ENSG000000108784	-161	N-acetylglucosaminidase, alph	NAGLU	4	4	1
267	ENSG000000109670	-102	F-box and WD repeat domain	FBXW7	4	3	1
194	ENSG000000110063	-319	decapping enzyme, scavenger	DCPS	4	3	1
184	ENSG000000110344	-625	ubiquitination factor E4A (UF	UBE4A	4	3	1
317	ENSG000000110906	-586	potassium channel tetramerisa	KCTD10	5	4	1
301	ENSG000000111261	-1072	MANSC domain containing 1	MANSC1	6	4	1
250	ENSG000000111452	-419	G protein-coupled receptor 13	ADGRL1	4	4	2
154	ENSG000000111540	-1156	RAB5B, member RAS oncog	RAB5B	3	3	1
253	ENSG000000111711	-604	golgi transport 1 homolog B (C	GOLT1B	5	4	1
301	ENSG000000111912	-1332	nuclear receptor coactivator 7	NCOA7	7	5	3
125	ENSG000000112877	-744	centrosomal protein 72kDa [S	CEP72	5	4	1
270	ENSG000000112893	-663	mannosidase, alpha, class 2A,	MAN2A1	4	3	2
123	ENSG000000113312	-26	tetratricopeptide repeat domai	TTC1	3	3	1
73	ENSG000000113504	-1477	solute carrier family 12 (potas	SLC12A7	5	5	3
254	ENSG000000114209	-1427	programmed cell death 10 [Sou	PDCD10	2	2	1
261	ENSG000000114423	-701	Cas-Br-M (murine) ecotropic	CBLB	3	3	2
235	ENSG000000114850	-1365	signal sequence receptor, gam	SSR3	3	3	2
180	ENSG000000115239	-1022	ankyrin repeat and SOCS box	ASB3	5	4	1
231	ENSG000000115306	-967	spectrin, beta, non-erythrocyti	SPTBN1	5	4	3
163	ENSG000000115425	-1131	peroxisomal trans-2-enoyl-Co	PECR	2	2	1
85	ENSG000000115808	-1489	striatin, calmodulin binding pr	STRN	7	6	5

311	ENSG00000115825	-667	protein kinase D3 [Source:HG	PRKD3	8	5	1
246	ENSG00000115970	-147	thyroid adenoma associated [S	THADA	11	7	2
138	ENSG00000115977	-421	AP2 associated kinase 1 [Sour	AAK1	6	4	1
220	ENSG00000116285	-516	ERBB receptor feedback inhib	ERRFI1	6	4	1
129	ENSG00000116329	-1096	opioid receptor, delta 1 [Sourc	OPRD1	2	2	1
246	ENSG00000116514	-378	ring finger protein 19B [Sourc	RNF19B	4	3	1
198	ENSG00000116604	-584	myocyte enhancer factor 2D [MEF2D	3	3	1
234	ENSG00000116962	-562	nidogen 1 [Source:HGNC Sym	NID1	2	2	1
157	ENSG00000116990	-1243	v-myc myelocytomatosis viral	MYCL	4	3	1
269	ENSG00000117507	-1008	flavin containing monooxygen	FMO6P	5	4	2
86	ENSG00000117751	-1053	protein phosphatase 1, regulat	PPP1R8	2	2	1
254	ENSG00000117906	-949	reticulocalbin 2, EF-hand calci	RCN2	4	3	2
57	ENSG00000117983	-939	mucin 5B, oligomeric mucus/	MUC5B	6	6	1
250	ENSG00000118407	-229	filamin A interacting protein 1	FILIP1	3	3	1
253	ENSG00000119203	-823	cleavage and polyadenylation	CPSF3	4	3	1
270	ENSG00000119718	-1202	eukaryotic translation initiatio	EIF2B2	3	3	2
180	ENSG00000119737	-1193	G protein-coupled receptor 75	GPR75	5	4	1
118	ENSG00000119844	-1178	atfiphilin [Source:HGNC Sym	AFTPH	7	5	2
195	ENSG00000119862	-1317	Galectin-related protein [Sour	LGALS1	10	6	1
78	ENSG00000120729	-419	myotilin [Source:HGNC Sym]	MYOT	2	2	1
286	ENSG00000120738	-1391	early growth response 1 [Sour	EGR1	7	5	2
239	ENSG00000120885	-795	clusterin [Source:HGNC Sym	CLU	2	2	1
220	ENSG00000122718	-909	olfactory receptor, family 2, se	OR2S2	2	2	1
299	ENSG00000124155	-457	phosphatidylinositol glycan an	PIGT	4	3	2
137	ENSG00000124357	-394	N-acetylglucosamine kinase [S	NAGK	12	8	1
255	ENSG00000124613	-1341	zinc finger protein 391 [Sourc	ZNF391	5	4	1
154	ENSG00000125089	-775	SH3 domain and tetratricopep	SH3TC1	2	2	1
121	ENSG00000125503	-1153	protein phosphatase 1, regulat	PPP1R12C	3	3	1
134	ENSG00000126216	-435	tubulin, gamma complex asso	TUBGCP3	2	2	1
122	ENSG00000126522	-587	argininosuccinate lyase [Sourc	ASL	3	3	2
109	ENSG00000126821	-894	sphingosine-1-phosphate phos	SGPP1	4	4	1
184	ENSG00000126856	-758	PR domain containing 7 [Sou	PRDM7	3	3	1
185	ENSG00000127080	-156	inositol 1,3,4,5,6-pentakispho	IPPK	3	3	2
178	ENSG00000128000	-559	zinc finger protein 780B [Sou	ZNF780B	3	3	2
112	ENSG00000128652	-1112	homeobox D3 [Source:HGNC	HOXD3	6	5	3
261	ENSG00000128654	-1291	metaxin 2 [Source:HGNC Sym	MTX2	2	2	1
255	ENSG00000128881	-800	tau tubulin kinase 2 [Source:H	TUBK2	2	2	1
260	ENSG00000128918	-612	aldehyde dehydrogenase 1 fan	ALDH1A2	5	4	1
107	ENSG00000129422	-1239	microtubule associated tumor	MTUS1	9	6	5
116	ENSG00000129991	-735	troponin I type 3 (cardiac) [So	TNNI3	2	2	1
265	ENSG00000130254	-227	scaffold attachment factor B2	SAFB2	4	3	1
230	ENSG00000130511	-1090	single stranded DNA binding	SSBP4	4	3	1
227	ENSG00000130702	-640	laminin, alpha 5 [Source:HGNC	LAMA5	2	2	1
202	ENSG00000130706	-1241	adhesion regulating molecule	ADRM1	2	2	1
123	ENSG00000132004	-654	F-box and WD repeat domain	FBXW9	6	5	2
163	ENSG00000132681	-935	ATPase, Na+/K+ transporting	ATP1A4	2	2	1
252	ENSG00000132874	-997	solute carrier family 14 (urea t	SLC14A2	4	3	1
73	ENSG00000132911	-24	neuromedin U receptor 2 [Sou	NMUR2	2	2	1
327	ENSG00000133740	-816	E2F transcription factor 5, p1	E2F5	4	4	1
313	ENSG00000133835	-585	hydroxysteroid (17-beta) dehy	HSD17B4	5	4	1
187	ENSG00000134352	-1194	interleukin 6 signal transducer	IL6ST	10	6	2
54	ENSG00000134569	-847	low density lipoprotein recept	LRP4	4	4	1
149	ENSG00000134755	-388	desmocollin 2 [Source:HGNC	DSC2	4	4	1
253	ENSG00000134757	-795	desmoglein 3 (pemphigus vul	DSG3	3	3	2
100	ENSG00000134987	-281	WD repeat domain 36 [Sourc	WDR36	2	2	1
179	ENSG00000135334	-85	akirin 2 [Source:HGNC Sym	AKIRIN2	5	5	1
271	ENSG00000135414	-1481	growth differentiation factor 1	GDF11	2	2	1
178	ENSG00000136026	-1437	cytoskeleton-associated protein	CKAP4	5	4	2
209	ENSG00000136247	-1179	zinc finger, DHHC-type conta	ZDHHC4	5	4	2
83	ENSG00000136280	-1311	cerebral cavernous malformati	CCM2	2	2	1
310	ENSG00000136636	-658	potassium channel tetramerisa	KCTD3	4	3	1
62	ENSG00000136938	-1126	acidic (leucine-rich) nuclear p	ANP32B	6	5	2
189	ENSG00000137033	-330	interleukin 33 [Source:HGNC	IL33	2	2	1
231	ENSG00000137055	-681	phospholipase A2-activating p	PLAA	2	2	1
194	ENSG00000137100	-903	dynactin 3 (p22) [Source:HG	DCTN3	2	2	1
206	ENSG00000137473	-1008	tetratricopeptide repeat domai	TTC29	2	2	1
219	ENSG00000137648	-1283	transmembrane protease, serin	TMPPRSS4	3	3	1
79	ENSG00000138032	-1090	protein phosphatase 1B (form	PPM1B	3	3	1
136	ENSG00000138068	-542	sulfotransferase family, cytos	SULT6B1	2	2	1
74	ENSG00000138071	-1324	ARP2 actin-related protein 2 l	ACTR2	5	4	1
270	ENSG00000138363	-1303	5-aminimidazole-4-carboxam	ATIC	4	3	2
123	ENSG00000138379	-951	myostatin [Source:HGNC Sym	MSTN	5	5	1
83	ENSG00000138670	-1452	RasGEF domain family, mem	RASGEF1B	2	2	1
279	ENSG00000139372	-778	thymine-DNA glycosylase [So	TDG	4	3	2
116	ENSG00000139515	-1185	pancreatic and duodenal home	PDX1	3	3	1
58	ENSG00000139547	-535	retinol dehydrogenase 16 (all-	RDH16	2	2	1
56	ENSG00000140009	-694	estrogen receptor 2 (ER beta)	ESR2	2	2	1
149	ENSG00000140153	-961	WD repeat domain 20 [Sourc	WDR20	4	4	2
308	ENSG00000140406	-1068	mesoderm development candi	MESDC1	4	4	1

A lista completa de dados encontra-se disponível por pedido a phmatos@fc.ul.pt

Anexo III

ID	Gene Name	Species
ENSG00000000938	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog	Homo sapiens
ENSG00000003393	amyotrophic lateral sclerosis 2 (juvenile)	Homo sapiens
ENSG000000005156	ligase III, DNA, ATP-dependent	Homo sapiens
ENSG000000005194	cytokine induced apoptosis inhibitor 1; cytokine induced apoptosis inhibitor 1 pseudogene	Homo sapiens
ENSG000000006530	acylglycerol kinase	Homo sapiens
ENSG000000006831	adiponectin receptor 2	Homo sapiens
ENSG000000007866	TEA domain family member 3	Homo sapiens
ENSG00000010318	PHD finger protein 7	Homo sapiens
ENSG00000011007	transcription elongation factor B (SIII), polypeptide 3 (110kDa, elongin A)	Homo sapiens
ENSG00000014138	polymerase (DNA directed), alpha 2 (70kD subunit)	Homo sapiens
ENSG00000020129	neurochondrin	Homo sapiens
ENSG00000024048	ubiquitin protein ligase E3 component n-recognin 2	Homo sapiens
ENSG00000029363	similar to Bcl-2-associated transcription factor 1 (Btf); BCL2-associated transcription factor 1	Homo sapiens
ENSG00000030582	granulin	Homo sapiens
ENSG00000039068	cadherin 1, type 1, E-cadherin (epithelial)	Homo sapiens
ENSG00000049130	KIT ligand	Homo sapiens
ENSG00000049167	excision repair cross-complementing rodent repair deficiency, complementation group 8	Homo sapiens
ENSG00000050730	TNFAIP3 interacting protein 3	Homo sapiens
ENSG00000052841	tetratricopeptide repeat domain 17	Homo sapiens
ENSG00000055957	inter-alpha (globulin) inhibitor H1	Homo sapiens
ENSG00000057468	mutS homolog 4 (E. coli)	Homo sapiens
ENSG00000058091	PFTAIRE protein kinase 1	Homo sapiens
ENSG00000058729	RIO kinase 2 (yeast)	Homo sapiens
ENSG00000060656	protein tyrosine phosphatase, receptor type, U	Homo sapiens
ENSG00000065357	diacylglycerol kinase, alpha 80kDa	Homo sapiens
ENSG00000065970	forkhead box J2	Homo sapiens
ENSG00000066739	ATG2 autophagy related 2 homolog B (S. cerevisiae)	Homo sapiens
ENSG00000068724	tetratricopeptide repeat domain 7A	Homo sapiens
ENSG00000070961	ATPase, Ca++ transporting, plasma membrane 1	Homo sapiens
ENSG00000071054	mitogen-activated protein kinase kinase kinase 4	Homo sapiens
ENSG00000072135	protein tyrosine phosphatase, non-receptor type 18 (brain-derived)	Homo sapiens
ENSG00000072310	sterol regulatory element binding transcription factor 1	Homo sapiens
ENSG00000073050	X-ray repair complementing defective repair in Chinese hamster cells 1	Homo sapiens
ENSG00000073803	mitogen-activated protein kinase kinase kinase 13	Homo sapiens
ENSG00000075223	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	Homo sapiens
ENSG00000075303	solute carrier family 25, member 40	Homo sapiens
ENSG00000078487	zinc finger, CW type with PWWP domain 1	Homo sapiens
ENSG00000082781	integrin, beta 5	Homo sapiens
ENSG00000088035	asparagine-linked glycosylation 6, alpha-1,3-glucosyltransferase homolog (S. cerevisiae)	Homo sapiens
ENSG00000088782	defensin, beta 127	Homo sapiens
ENSG00000088833	NSFL1 (p97) cofactor (p47)	Homo sapiens
ENSG00000089558	potassium voltage-gated channel, subfamily H (eag-related), member 4	Homo sapiens
ENSG00000091181	interleukin 5 receptor, alpha	Homo sapiens
ENSG00000092010	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	Homo sapiens
ENSG00000092098	ring finger protein 31	Homo sapiens
ENSG00000093010	catechol-O-methyltransferase	Homo sapiens
ENSG00000095002	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)	Homo sapiens
ENSG00000095794	cAMP responsive element modulator	Homo sapiens
ENSG00000100206	DMC1 dosage suppressor of mck1 homolog, meiosis-specific homologous recombination (yeast)	Homo sapiens
ENSG00000100258	lipase maturation factor 2	Homo sapiens
ENSG00000100307	chromobox homolog 7	Homo sapiens
ENSG00000100325	activating signal cointegrator 1 complex subunit 2	Homo sapiens
ENSG00000100612	dehydrogenase/reductase (SDR family) member 7	Homo sapiens
ENSG00000100897	WD repeat domain 23	Homo sapiens
ENSG00000100906	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	Homo sapiens
ENSG00000100968	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 4	Homo sapiens
ENSG00000101204	cholinergic receptor, nicotinic, alpha 4	Homo sapiens
ENSG00000101311	fermitin family homolog 1 (Drosophila)	Homo sapiens
ENSG00000101751	polymerase (DNA directed) iota	Homo sapiens
ENSG00000101940	WD repeat domain 13	Homo sapiens
ENSG00000102699	poly (ADP-ribose) polymerase family, member 4	Homo sapiens
ENSG00000102893	phosphorylase kinase, beta	Homo sapiens
ENSG00000104064	GA binding protein transcription factor, beta subunit 1	Homo sapiens
ENSG00000104979	chromosome 19 open reading frame 53	Homo sapiens
ENSG00000105755	ethylmalonic encephalopathy 1	Homo sapiens
ENSG00000105852	paraoxonase 3	Homo sapiens
ENSG00000105866	Sp4 transcription factor	Homo sapiens
ENSG00000105953	oxoglutarate (alpha-ketoglutarate) dehydrogenase (lipoamide)	Homo sapiens
ENSG00000106665	CAP-GLY domain containing linker protein 2	Homo sapiens
ENSG00000108064	transcription factor A, mitochondrial	Homo sapiens
ENSG00000108557	retinoic acid induced 1	Homo sapiens
ENSG00000108784	N-acetylglucosaminidase, alpha-	Homo sapiens
ENSG00000109670	F-box and WD repeat domain containing 7	Homo sapiens
ENSG00000110063	decapping enzyme, scavenger	Homo sapiens
ENSG00000111261	MANSC domain containing 1	Homo sapiens
ENSG00000111452	G protein-coupled receptor 133	Homo sapiens
ENSG00000111912	nuclear receptor coactivator 7	Homo sapiens
ENSG00000112877	centrosomal protein 72kDa	Homo sapiens
ENSG00000113504	solute carrier family 12 (potassium/chloride transporters), member 7	Homo sapiens
ENSG00000114209	programmed cell death 10	Homo sapiens
ENSG00000114423	Cas-Br-M (murine) ecotropic retroviral transforming sequence b	Homo sapiens
ENSG00000115306	spectrin, beta, non-erythrocytic 1	Homo sapiens

ENSG00000115425	peroxisomal trans-2-enoyl-CoA reductase	Homo sapiens
ENSG00000115808	striatin, calmodulin binding protein	Homo sapiens
ENSG00000115825	protein kinase D3	Homo sapiens
ENSG00000115970	thyroid adenoma associated	Homo sapiens
ENSG00000115977	AP2 associated kinase 1	Homo sapiens
ENSG00000116285	ERBB receptor feedback inhibitor 1	Homo sapiens
ENSG00000116329	opioid receptor, delta 1	Homo sapiens
ENSG00000116604	myocyte enhancer factor 2D	Homo sapiens
ENSG00000116962	nidogen 1	Homo sapiens
ENSG00000116990	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)	Homo sapiens
ENSG00000117983	mucin 5B, oligomeric mucus/gel-forming	Homo sapiens
ENSG00000118407	filamin A interacting protein 1	Homo sapiens
ENSG00000119203	cleavage and polyadenylation specific factor 3, 73kDa	Homo sapiens
ENSG00000119718	eukaryotic translation initiation factor 2B, subunit 2 beta, 39kDa	Homo sapiens
ENSG00000119737	G protein-coupled receptor 75	Homo sapiens
ENSG00000119844	atipphilin	Homo sapiens
ENSG00000120729	myotilin	Homo sapiens
ENSG00000120738	early growth response 1	Homo sapiens
ENSG00000120885	clusterin	Homo sapiens
ENSG00000122718	olfactory receptor, family 2, subfamily S, member 2	Homo sapiens
ENSG00000124155	phosphatidylinositol glycan anchor biosynthesis, class T	Homo sapiens
ENSG00000124613	zinc finger protein 391	Homo sapiens
ENSG00000125089	SH3 domain and tetratricopeptide repeats 1	Homo sapiens
ENSG00000125503	protein phosphatase 1, regulatory (inhibitor) subunit 12C	Homo sapiens
ENSG00000126216	tubulin, gamma complex associated protein 3	Homo sapiens
ENSG00000126856	PR domain containing 7	Homo sapiens
ENSG00000127080	inositol 1,3,4,5,6-pentakisphosphate 2-kinase	Homo sapiens
ENSG00000128652	homeobox D3	Homo sapiens
ENSG00000128881	tau tubulin kinase 2	Homo sapiens
ENSG00000128918	aldehyde dehydrogenase 1 family, member A2	Homo sapiens
ENSG00000129422	mitochondrial tumor suppressor 1	Homo sapiens
ENSG00000129991	troponin I type 3 (cardiac)	Homo sapiens
ENSG00000130702	laminin, alpha 5	Homo sapiens
ENSG00000132004	F-box and WD repeat domain containing 9	Homo sapiens
ENSG00000132681	ATPase, Na+/K+ transporting, alpha 4 polypeptide	Homo sapiens
ENSG00000132874	solute carrier family 14 (urea transporter), member 2	Homo sapiens
ENSG00000132911	neuromedin U receptor 2	Homo sapiens
ENSG00000133740	E2F transcription factor 5, p130-binding	Homo sapiens
ENSG00000133835	hydroxysteroid (17-beta) dehydrogenase 4	Homo sapiens
ENSG00000134352	interleukin 6 signal transducer (gp130, oncostatin M receptor)	Homo sapiens
ENSG00000134569	low density lipoprotein receptor-related protein 4	Homo sapiens
ENSG00000134755	desmocollin 2	Homo sapiens
ENSG00000134757	desmoglein 3 (pemphigus vulgaris antigen)	Homo sapiens
ENSG00000134987	WD repeat domain 36	Homo sapiens
ENSG00000136026	cytoskeleton-associated protein 4	Homo sapiens
ENSG00000136247	zinc finger, DHHC-type containing 4	Homo sapiens
ENSG00000136280	cerebral cavernous malformation 2	Homo sapiens
ENSG00000136636	potassium channel tetramerisation domain containing 3	Homo sapiens
ENSG00000137033	interleukin 33	Homo sapiens
ENSG00000137473	tetratricopeptide repeat domain 29	Homo sapiens
ENSG00000137648	transmembrane protease, serine 4	Homo sapiens
ENSG00000138068	sulfotransferase family, cytosolic, 6B, member 1	Homo sapiens
ENSG00000138363	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	Homo sapiens
ENSG00000138379	myostatin	Homo sapiens
ENSG00000139372	similar to G/T mismatch-specific thymine DNA glycosylase; thymine-DNA glycosylase	Homo sapiens
ENSG00000139515	pancreatic and duodenal homeobox 1	Homo sapiens
ENSG00000140153	WD repeat domain 20	Homo sapiens
ENSG00000140527	WD repeat domain 93	Homo sapiens
ENSG00000140740	ubiquinol-cytochrome c reductase core protein II	Homo sapiens
ENSG00000142173	collagen, type VI, alpha 2	Homo sapiens
ENSG00000143190	POU class 2 homeobox 1	Homo sapiens
ENSG00000143196	dermatopontin	Homo sapiens
ENSG00000143226	Fc fragment of IgG, low affinity IIa, receptor (CD32)	Homo sapiens
ENSG00000143294	papillary renal cell carcinoma (translocation-associated)	Homo sapiens
ENSG00000143333	regulator of G-protein signaling 16	Homo sapiens
ENSG00000143870	protein disulfide isomerase family A, member 6	Homo sapiens
ENSG00000144224	UBX domain protein 4	Homo sapiens
ENSG00000144331	zinc finger protein 385B	Homo sapiens
ENSG00000146373	ring finger protein 217	Homo sapiens
ENSG00000146904	EPH receptor A1	Homo sapiens
ENSG00000147437	gonadotropin-releasing hormone 1 (luteinizing-releasing hormone)	Homo sapiens
ENSG00000147454	solute carrier family 25, member 37	Homo sapiens
ENSG00000147614	ATPase, H+ transporting, lysosomal 38kDa, V0 subunit d2	Homo sapiens
ENSG00000148187	mitochondrial ribosome recycling factor	Homo sapiens
ENSG00000148444	COMM domain containing 3	Homo sapiens
ENSG00000148482	solute carrier family 39 (zinc transporter), member 12	Homo sapiens
ENSG00000148688	ribonuclease P/MRP 30kDa subunit	Homo sapiens
ENSG00000149212	sestrin 3	Homo sapiens
ENSG00000149257	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	Homo sapiens
ENSG00000149582	transmembrane protein 25	Homo sapiens
ENSG00000150627	WD repeat domain 17	Homo sapiens

[A lista completa de dados encontra-se disponível por pedido a phmatos@fc.ul.pt](mailto:phmatos@fc.ul.pt)

Anexo IV

Feature	Descrição	Símbolo_HGNC	Dist_TSS	Larg_pico	Max_Act	DAct
ENSG00000115970	thyroid adenoma associated [S	THADA	-147	246	11	8
ENSG00000173114	leucine rich repeat neuronal 3	LRRN3	-981	127	11	7
ENSG00000134352	interleukin 6 signal transducer	IL6ST	-1194	187	10	8
ENSG00000178104	phosphodiesterase 4D interact	PDE4DIP	-273	51	10	0
ENSG00000092098	ring finger protein 31 [Source:	RNF31	-1361	433	9	5
ENSG00000100968	nuclear factor of activated T-c	NFATC4	-475	145	9	7
ENSG00000129422	microtubule associated tumor	MTUS1	-1239	107	9	7
ENSG00000167920	transmembrane protein 99 [So	TMEM99	-964	86	9	7
ENSG00000100897	DDB1 and CUL4 associated fi	DCAF11	-341	110	8	4
ENSG00000102699	poly (ADP-ribose) polymerase	PARP4	-1178	182	8	6
ENSG00000104064	GA binding protein transcripti	GABPB1	-991	279	8	6
ENSG00000115825	protein kinase D3 [Source:HG	PRKD3	-667	311	8	6
ENSG00000140740	ubiquinol-cytochrome c reduct	UQCRC2	-674	158	8	5
ENSG00000155465	solute carrier family 7 (cationi	SLC7A7	-842	311	8	4
ENSG00000158417	eukaryotic translation initiatio	EIF5B	-887	246	8	6
ENSG000000088035	asparagine-linked glycosylation	ALG6	-800	309	7	5
ENSG00000100612	dehydrogenase/reductase (SDI	DHRS7	-744	264	7	5
ENSG00000111912	nuclear receptor coactivator 7	NCOA7	-1332	301	7	4
ENSG00000115808	striatin, calmodulin binding pr	STRN	-1489	85	7	3
ENSG00000119844	afthiphilin [Source:HGNC Sym	AFTPH	-1178	118	7	5
ENSG00000120738	early growth response 1 [Sour	EGR1	-1391	286	7	5
ENSG00000146373	ring finger protein 217 [Source	RNF217	-576	280	7	5
ENSG00000173559	oligonucleotide/oligosaccharid	NABP1	-1018	255	7	5
ENSG00000198833	ubiquitin-conjugating enzyme	UBE2J1	-53	303	7	5
ENSG00000003393	amyotrophic lateral sclerosis 2	ALS2	-364	444	6	2
ENSG0000010318	PHD finger protein 7 [Source:	PHF7	-803	335	6	2
ENSG00000073050	X-ray repair complementing d	XRCC1	-1065	92	6	4
ENSG00000095002	mutS homolog 2, colon cancer	MSH2	-19	94	6	4
ENSG00000095794	cAMP responsive element mo	CREM	-917	60	6	1
ENSG00000105866	Sp4 transcription factor [Sour	SP4	-563	306	6	4
ENSG00000111261	MANSC domain containing 1	MANSC1	-1072	301	6	4
ENSG00000115977	AP2 associated kinase 1 [Sour	AAK1	-421	138	6	4
ENSG00000116285	ERBB receptor feedback inhib	ERRFI1	-516	220	6	4
ENSG00000117983	mucin 5B, oligomeric mucus/g	MUC5B	-939	57	6	1
ENSG00000128652	homeobox D3 [Source:HGNC	HOXD3	-1112	112	6	2
ENSG00000132004	F-box and WD repeat domain	FBXW9	-654	123	6	3
ENSG00000143294	papillary renal cell carcinoma (PRCC	-228	243	6	4
ENSG00000152942	RAD17 homolog (S. pombe) [RAD17	-546	182	6	3
ENSG00000162408	nucleolar protein 9 [Source:HO	NOL9	-385	284	6	4
ENSG00000163349	homeodomain interacting prot	HIPK1	-1278	159	6	4
ENSG00000165655	zinc finger protein 503 [Source	ZNF503	-99	323	6	4
ENSG00000167346	matrix metalloproteinase 26 [St	MMP26	-1187	117	6	1
ENSG00000170956	carcinoembryonic antigen-rela	CEACAM3	-96	168	6	4
ENSG00000173662	taste receptor, type 1, member	TAS1R1	-262	284	6	4
ENSG00000176253	olfactory receptor, family 4, su	OR4K13	-1324	188	6	4
ENSG00000180113	tudor domain containing 6 [So	TDRD6	-1202	236	6	4
ENSG00000198728	LIM domain binding 1 [Source	LDB1	-1201	253	6	4
ENSG00000198944	ankyrin repeat domain 43 [Sot	SOWAHA	-307	196	6	4
ENSG00000006831	adiponectin receptor 2 [Source	ADIPOR2	-618	57	5	3
ENSG000000011007	transcription elongation factor	TCEB3	-649	279	5	3
ENSG00000058091	cyclin-dependent kinase 14 [Sc	CDK14	-865	70	5	1
ENSG00000068724	tetratricopeptide repeat domai	TTC7A	-794	220	5	2
ENSG00000072310	sterol regulatory element bindi	SREBF1	-269	152	5	3
ENSG00000075223	sema domain, immunoglobulin	SEMA3C	-126	204	5	3
ENSG00000082781	integrin, beta 5 [Source:HGNC	ITGB5	-692	299	5	2
ENSG00000092010	proteasome (prosome, macrop	PSME1	-854	200	5	2
ENSG00000100325	activating signal cointegrator 1	ASCC2	-1043	233	5	2
ENSG00000104979	Leydig cell tumor 10 kDa prot	C19orf53	-182	295	5	3
ENSG00000105755	ethylmalonic encephalopathy 1	ETHE1	-1052	92	5	3
ENSG00000105852	paraoxonase 3 [Source:HGNC	PON3	-1291	179	5	3
ENSG00000112877	centrosomal protein 72kDa [S	CEP72	-744	125	5	3
ENSG00000113504	solute carrier family 12 (potas	SLC12A7	-1477	73	5	1
ENSG00000115306	spectrin, beta, non-erythrocyti	SPTBN1	-967	231	5	3
ENSG00000119737	G protein-coupled receptor 75	GPR75	-1193	180	5	3
ENSG00000124613	zinc finger protein 391 [Source	ZNF391	-1341	255	5	3
ENSG00000128918	aldehyde dehydrogenase 1 fam	ALDH1A2	-612	260	5	3
ENSG00000133835	hydroxysteroid (17-beta) dehy	HSD17B4	-585	313	5	2
ENSG00000136026	cytoskeleton-associated protei	CKAP4	-1437	178	5	3
ENSG00000136247	zinc finger, DHHC-type conta	ZDHHC4	-1179	209	5	2
ENSG00000138379	myostatin [Source:HGNC Syn	MTN	-951	123	5	0
ENSG00000143190	POU class 2 homeobox 1 [Sot	POU2F1	-577	325	5	3
ENSG00000143333	regulator of G-protein signalin	RGS16	-634	87	5	3
ENSG00000147454	solute carrier family 25, memb	SLC25A37	-569	109	5	3
ENSG00000147614	ATPase, H+ transporting, lyso	ATP6V0D2	-1474	151	5	3
ENSG00000149212	sestrin 3 [Source:HGNC Sym	SESN3	-651	298	5	3
ENSG00000156802	ATPase family, AAA domain c	ATAD2	-1050	102	5	2
ENSG00000163217	bone morphogenetic protein 10	BMP10	-1488	104	5	3
ENSG00000165338	HECT domain containing 2 [S	HECTD2	-1180	210	5	3
ENSG00000167588	glycerol-3-phosphate dehydrog	GPD1	-378	275	5	2
ENSG00000167858	tektin 1 [Source:HGNC Symb	TEKT1	-841	104	5	3
ENSG00000174276	zinc finger, HIT type 2 [Sour	ZNFHIT2	-258	195	5	3
ENSG00000176834	V-set and immunoglobulin do	VSIG10	-463	144	5	3

ENSG00000177669	membrane bound O-acyltransf	MBOAT4	-425	261	5	3
ENSG00000185252	zinc finger protein 74 [Source:	ZNF74	-1033	318	5	3
ENSG00000187242	keratin 12 [Source:HGNC Syr	KRT12	-978	257	5	1
ENSG00000000938	Gardner-Rasheed feline sarcor	FGR	-169	318	4	2
ENSG000000005156	ligase III, DNA, ATP-depende	LIG3	-621	260	4	2
ENSG00000007866	TEA domain family member 3	TEAD3	-394	192	4	2
ENSG00000020129	neurochondrin [Source:HGNC	NCDN	-1403	253	4	0
ENSG00000029363	BCL2-associated transcription	BCLAF1	-846	310	4	2
ENSG00000060656	protein tyrosine phosphatase, 1	PTPRU	-830	288	4	2
ENSG00000070961	ATPase, Ca++ transporting, p	ATP2B1	-1500	92	4	0
ENSG00000072135	protein tyrosine phosphatase, 1	PTPN18	-1321	304	4	1
ENSG00000075303	solute carrier family 25, memb	SLC25A40	-1278	50	4	1
ENSG00000078487	zinc finger, CW type with PW	ZCWPW1	-978	131	4	2
ENSG00000088782	defensin, beta 127 [Source:HC	DEFB127	-1134	62	4	0
ENSG00000100258	lipase maturation factor 2 [Sou	LMF2	-1407	55	4	1
ENSG00000106665	CAP-GLY domain containing	CLIP2	-1340	275	4	2
ENSG00000108064	transcription factor A, mitoch	TFAM	-1028	225	4	2
ENSG00000108784	N-acetylglucosaminidase, alph	NAGLU	-161	193	4	1
ENSG00000109670	F-box and WD repeat domain	FBXW7	-102	267	4	2
ENSG00000110063	decapping enzyme, scavenger	DCPS	-319	194	4	2
ENSG00000111452	G protein-coupled receptor 13	ADGRD1	-419	250	4	1
ENSG00000116990	v-myc myelocytomatosis viral	MYCL	-1243	157	4	2
ENSG00000119203	cleavage and polyadenylation	CPSF3	-823	253	4	2
ENSG00000124155	phosphatidylinositol glycan an	PIGT	-457	299	4	2
ENSG00000132874	solute carrier family 14 (urea t	SLC14A2	-997	252	4	2
ENSG00000133740	E2F transcription factor 5, p1	E2F5	-816	327	4	1
ENSG00000134569	low density lipoprotein recept	LRP4	-847	54	4	1
ENSG00000134755	desmocollin 2 [Source:HGNC	DSC2	-388	149	4	1
ENSG00000136636	potassium channel tetramerisa	KCTD3	-658	310	4	2
ENSG00000138363	5-aminoimidazole-4-carboxam	ATIC	-1303	270	4	2
ENSG00000139372	thymine-DNA glycosylase [So	TDG	-778	279	4	2
ENSG00000140153	WD repeat domain 20 [Source	WDR20	-961	149	4	1
ENSG00000144224	UBX domain protein 4 [Sourc	UBXN4	-310	209	4	2
ENSG00000146904	EPH receptor A1 [Source:HG	EPHA1	-556	273	4	2
ENSG00000148482	solute carrier family 39 (zinc ti	SLC39A12	-1105	312	4	1
ENSG00000148688	ribonuclease P/MRP 30kDa su	RPP30	-808	307	4	2
ENSG00000149257	serpin peptidase inhibitor, clad	SERPINH1	-1423	181	4	2
ENSG00000152672	C-type lectin domain family 4,	CLEC4F	-1031	182	4	1
ENSG00000157985	ArfGAP with GTPase domain,	AGAP1	-1087	135	4	2
ENSG00000159176	cysteine and glycine-rich prote	CSRP1	-1386	368	4	1
ENSG00000163638	ADAM metalloproteinase with	ADAMTS9	-942	257	4	1
ENSG00000163956	low density lipoprotein recept	LRPAP1	-1109	211	4	2
ENSG00000164023	sphingomyelin synthase 2 [Sou	SGMS2	-934	118	4	2
ENSG00000167325	ribonucleotide reductase M1 [RRM1	-1396	287	4	1
ENSG00000168398	bradykinin receptor B2 [Sourc	BDKRB2	-747	227	4	2
ENSG00000168589	dynein, light chain, roadblock	DYNLRB2	-912	168	4	2
ENSG00000169508	G protein-coupled receptor 18	GPR183	-968	187	4	2
ENSG00000170006	transmembrane protein 154 [S	TMEM154	-1139	231	4	1
ENSG00000172769	olfactory receptor, family 5, su	OR5B3	-1338	133	4	2
ENSG00000172819	retinoic acid receptor, gamma	RARG	-282	110	4	2
ENSG00000178802	mannose phosphate isomerase	MPI	-503	239	4	2
ENSG00000185245	glycoprotein Ib (platelet), alph	GP1BA	-815	244	4	2
ENSG00000197706	olfactory receptor, family 6, su	OR6C74	-243	59	4	1
ENSG00000198523	phospholamban [Source:HGNC	PLN	-954	55	4	2
ENSG00000198841	KTI12 homolog, chromatin as	KTI12	-358	218	4	2
ENSG00000204344	Serine/threonine-protein kinas	STK19	-828	246	4	1
ENSG00000204536	coiled-coil alpha-helical rod pr	CCHCR1	-556	155	4	2
ENSG000000005194	cytokine induced apoptosis inf	CIAPIN1	-1303	136	3	0
ENSG000000014138	polymerase (DNA directed), a	POLA2	-34	290	3	1
ENSG00000030582	granulin [Source:HGNC Synt	GRN	-201	242	3	1
ENSG00000052841	tetratricopeptide repeat domai	TTC17	-421	97	3	1
ENSG00000065357	diacylglycerol kinase, alpha 80	DGKA	-309	291	3	1
ENSG00000065970	forkhead box J2 [Source:HGNC	FOXJ2	-1432	136	3	1
ENSG00000071054	mitogen-activated protein kina	MAP4K4	-614	110	3	1
ENSG00000073803	mitogen-activated protein kina	MAP3K13	-1100	174	3	1
ENSG00000089558	potassium voltage-gated chan	KCNH4	-524	218	3	1
ENSG00000100206	DMC1 dosage suppressor of n	DMC1	-1106	165	3	0
ENSG00000100906	nuclear factor of kappa light p	NFKBIA	-380	265	3	1
ENSG00000101204	cholinergic receptor, nicotinic,	CHRNA4	-1005	164	3	0
ENSG00000101311	fermitin family homolog 1 (Dn	FERMT1	-273	126	3	1
ENSG00000108557	retinoic acid induced 1 [Source	RAI1	-603	331	3	1
ENSG00000114423	Cas-Br-M (murine) ecotropic	CBLB	-701	261	3	0
ENSG00000116604	myocyte enhancer factor 2D [S	MEF2D	-584	198	3	1
ENSG00000118407	filamin A interacting protein 1	FILIP1	-229	250	3	1
ENSG00000119718	eukaryotic translation initiat	EIF2B2	-1202	270	3	0
ENSG00000125503	protein phosphatase 1, regulat	PPP1R12C	-1153	121	3	0
ENSG00000126856	PR domain containing 7 [Sour	PRDM7	-758	184	3	1
ENSG00000127080	inositol 1,3,4,5,6-pentakisphos	IPPK	-156	185	3	1
ENSG00000134757	desmoglein 3 (pemphigus vulg	DSG3	-795	253	3	1
ENSG00000137648	transmembrane protease, serin	TMPRSS4	-1283	219	3	1
ENSG00000139515	pancreatic and duodenal home	PDX1	-1185	116	3	0

A lista completa de dados encontra-se disponível por pedido a phmatos@fc.ul.pt

Anexo V

ID	Gene Name	Species
ENSG00000069399	B-cell CLL/lymphoma 3	Homo sapiens
ENSG00000197299	Bloom syndrome, RecQ helicase-like	Homo sapiens
ENSG00000115392	Fanconi anemia, complementation group L	Homo sapiens
ENSG00000188486	H2A histone family, member X	Homo sapiens
ENSG00000151657	KIN, antigenic determinant of recA protein homolog (mouse)	Homo sapiens
ENSG00000159593	NEDD8 activating enzyme E1 subunit 1	Homo sapiens
ENSG00000170430	O-6-methylguanine-DNA methyltransferase	Homo sapiens
ENSG00000152942	RAD17 homolog (S. pombe)	Homo sapiens
ENSG00000119318	RAD23 homolog B (S. cerevisiae)	Homo sapiens
ENSG00000182185	RAD51-like 1 (S. cerevisiae)	Homo sapiens
ENSG00000197275	RAD54 homolog B (S. cerevisiae)	Homo sapiens
ENSG00000067066	SP100 nuclear antigen	Homo sapiens
ENSG00000073050	X-ray repair complementing defective repair in Chinese hamster cells 1	Homo sapiens
ENSG00000134982	adenomatous polyposis coli	Homo sapiens
ENSG00000100601	alkB, alkylation repair homolog 1 (E. coli)	Homo sapiens
ENSG00000176194	cell death-inducing DFFA-like effector a	Homo sapiens
ENSG00000136305	cell death-inducing DFFA-like effector b	Homo sapiens
ENSG00000170312	cell division cycle 2, G1 to S and G2 to M	Homo sapiens
ENSG00000110274	centrosomal protein 164kDa	Homo sapiens
ENSG00000101204	cholinergic receptor, nicotinic, alpha 4	Homo sapiens
ENSG00000010072	chromosome 1 open reading frame 124	Homo sapiens
ENSG00000148153	chromosome 9 open reading frame 80	Homo sapiens
ENSG00000143476	denticless homolog (Drosophila)	Homo sapiens
ENSG00000171320	establishment of cohesion 1 homolog 2 (S. cerevisiae)	Homo sapiens
ENSG00000049167	excision repair cross-complementing rodent repair deficiency, complementation group 8	Homo sapiens
ENSG00000168496	flap structure-specific endonuclease 1	Homo sapiens
ENSG00000163349	homeodomain interacting protein kinase 1	Homo sapiens
ENSG000000005156	ligase III, DNA, ATP-dependent	Homo sapiens
ENSG00000101898	malignant T cell amplified sequence 1	Homo sapiens
ENSG00000137337	mediator of DNA-damage checkpoint 1	Homo sapiens
ENSG00000125901	mitochondrial ribosomal protein S26	Homo sapiens
ENSG00000061794	mitochondrial ribosomal protein S35	Homo sapiens
ENSG00000095002	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)	Homo sapiens
ENSG00000057468	mutS homolog 4 (E. coli)	Homo sapiens
ENSG00000109674	nei endonuclease VIII-like 3 (E. coli)	Homo sapiens
ENSG00000169189	non-SMC element 1 homolog (S. cerevisiae)	Homo sapiens
ENSG00000156831	non-SMC element 2, MMS21 homolog (S. cerevisiae)	Homo sapiens
ENSG00000173559	oligonucleotide/oligosaccharide-binding fold containing 2A	Homo sapiens
ENSG00000164611	pituitary tumor-transforming 1; pituitary tumor-transforming 2	Homo sapiens
ENSG00000102699	poly (ADP-ribose) polymerase family, member 4	Homo sapiens
ENSG00000101751	polymerase (DNA directed) iota	Homo sapiens
ENSG00000062822	polymerase (DNA directed), delta 1, catalytic subunit 125kDa	Homo sapiens
ENSG00000077514	polymerase (DNA-directed), delta 3, accessory subunit	Homo sapiens
ENSG00000175482	polymerase (DNA-directed), delta 4	Homo sapiens
ENSG00000148985	post-GPI attachment to proteins 2	Homo sapiens
ENSG00000035928	replication factor C (activator 1) 1, 145kDa	Homo sapiens
ENSG00000139372	similar to G/T mismatch-specific thymine DNA glycosylase; thymine-DNA glycosylase	Homo sapiens
ENSG00000108055	structural maintenance of chromosomes 3	Homo sapiens
ENSG00000042088	tyrosyl-DNA phosphodiesterase 1	Homo sapiens
ENSG00000119048	ubiquitin-conjugating enzyme E2B (RAD6 homolog)	Homo sapiens

Anexo VI

Composição das soluções utilizadas na lise de células

- **Laemmli Sample Buffer 5x**
 - 250 mM Tris-HCl, pH 6,8
 - 25% Glicerol (v/v)
 - 10% SDS (w/v)
 - 325 mM DTT (w/v)
 - 1,25 mg Azul de Bromofenol (w/v)
 - ddH₂O
- **Tampão de lise NP-40**
 - 50 mM Tris-HCl pH 7,5
 - 100 mM NaCl
 - 2 mM MgCl₂
 - 10% Glicerol (v/v)
 - 1% Nonidet P-40 (NP-40) (v/v)
 - ddH₂O
 - 1x Cocktail de Inibidores de Proteases (Sigma®)

Composição dos géis de poliacrilamida

- **Running – 10%**
 - 2,45 ml H₂O
 - 1,25 ml *Lower Buffer*
 - 1,25 ml Solução Bis-Acrilamida 40%, 1:37,5
 - 50 µl SDS 10%
 - 50 µl Persulfato de amónio (APS) 10%
 - 2,5 µl TEMED
- **Stacking – 4%**
 - 1,5 ml H₂O
 - 0,25 ml *Upper Buffer*
 - 0,20 ml Solução Bis-Acrilamida 40%, 1:37,5
 - 20 µl SDS 10%
 - 40 µl APS 10%
 - 2 µl TEMED
- **Lower Buffer** (1L) (1,5M Tris-HCl, pH 8,8)
 - 181,71 g Tris

- ~ 30 ml HCl 37%

▪ **Upper Buffer (1L) (0,5M Tris-HCl, pH 6,8)**

- 60,6 g Tris
- ~ 40 ml HCl 37%

Composição dos tampões utilizados em SDS-PAGE, *Western blot* e ECL

▪ **Tampão SDS-PAGE 10x (1 L) (630 mM Tris, 4,8 M Glicina, 87 mM SDS)**

- 30,4 g Tris
- 144 g Glicina
- 10 g SDS

▪ **Tampão SDS-PAGE 1x (1L)**

- 100 ml Tampão SDS-PAGE 10X
- 900 ml H₂O

▪ **Tampão de transferência 25x (1 L) (248 mM Tris, 1,92 M Glicina)**

- 145 g Tris
- 72,5 g Glicina
- 9,3 g SDS

▪ **Tampão de transferência 1x (1L)**

- 200 ml Metanol
- 40 ml Tampão de transferência 25x
- 760 ml H₂O

▪ **Azul de Coomassie (1L)**

- 2,5 g Coomassie Brilliant Blue G
- 450 ml Metanol
- 100 ml Ácido acético
- 450 ml H₂O

▪ **Solução Destain (1L)**

- 450 ml Metanol
- 100 ml Ácido Acético
- 450 ml H₂O

▪ **TBS 10x, pH 7,6 (1 L) (1,25 M Tris, 3,75 M NaCl)**

- 60,4 g Tris
- 87,6 g NaCl
- ~ 36 ml HCl 37%

▪ **TBS 1x + 0,05% Triton X-100 – TBST (1L)**

- 100 ml TBS 10x

- 500 ul Triton X-100
- 900 ml H₂O
- **ECL (para uma membrana)**
 - **Solução 1:**
 - 980 µL Tampão ECL
 - 5 µL Ácido Cumárico
 - 15 µL Luminol
 - **Solução 2:**
 - 1ml Tampão ECL
 - 1 µL H₂O₂
- Tampão ECL
 - 3 ml Tris-HCl pH 8,8
 - 42 ml H₂O

Composição dos tampões utilizados em PCR e eletroforese

- **Tampão B 10x**
 - 500 mM KCl
 - 100 mM Tris-HCl pH 9
 - 15 mM MgCl₂
 - 0,1 % Gelatina (w/v)
- **Tampão B 1x**
 - 5 ml Tampão B (10x)
 - 45 ml ddH₂O
 - 100 µL de cada dNTP (100 mM)
- **TBE 10x, pH 8**
 - 890 mM Tris
 - 890 mM Ácido bórico
 - 20 mM EDTA
- **TBE 1x (1L)**
 - 100 ml TBE 10x
 - 900 ml H₂O

Composição das soluções utilizadas no Ensaio do Cometa

- **1% Low Melting point agarose em PBS**
 - 0,5 g
 - 50 ml PBS
- **Solução de Lise, pH 10 (1L) (2,5 M NaCl, 100 mM Na₂EDTA.2H₂O, 10 mM Tris-HCl, 10 M NaOH)**
 - 146,10 g NaCl
 - 37,20 g Na₂EDTA.2H₂O
 - 1,20 g Tris-HCl
 - ~ 18 ml NaOH

- **Tampão de lise (100 ml) (preparado imediatamente antes de usar)**
 - 89 ml Solução de lise
 - 10 ml DMSO
 - 1 ml Triton X-100

- **Tampão de Neutralização, pH 7,5 (1L) (0,4 M Tris, 4 M HCl)**
 - 48,50 g Tris
 - 95 ml HCl

- **Tampão de Eletroforese, pH 13 (1L) (300 mM NaOH, 1mM Na₂EDTA.2H₂O)**
 - 12 g NaOH
 - 0,37 g Na₂EDTA.2H₂O