

UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico – CDTec
Curso de Graduação em Biotecnologia



Trabalho de Conclusão de Curso

**Análise de expressão diferencial de células epiteliais
alveolares tipo II de *Rattus norvegicus* utilizando a
técnica de microarranjo**

Jessica Rodrigues Praça

Pelotas, 2014

Jessica Rodrigues Praça

Análise de expressão diferencial de células epiteliais alveolares tipo II de *Rattus norvegicus* utilizando a técnica de microarranjo

Trabalho acadêmico apresentado ao Curso de Bacharelado em Biotecnologia da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Bacharel em Biotecnologia.

Orientador Acadêmico: Prof. Dr. Odir Antônio Dellagostin

Orientador de Estágio: Prof. PhD. Mark R. Segal

Pelotas, 2014

Dados de catalogação na fonte:
Ubirajara Buddin Cruz – CRB-10/901

Biblioteca de Ciência & Tecnologia - UFPel

P696a **Plaça, Jessica Rodrigues**

Análise de expressão diferencial de células epiteliais alveolares tipo II de *Rattus norvegicus* utilizando a técnica de microarranjo / Jessica Rodrigues Plaça. – 48f. : il. – Trabalho de conclusão de curso (Graduação em Biotecnologia). Universidade Federal de Pelotas. Centro de Desenvolvimento Tecnológico. Pelotas, 2014. – Orientador Odir Antônio Dellagostin.

1.Biotecnologia. 2.Célula epitelial alveolar tipo II. 3.Cultivo celular. 4.Técnica de microarranjo. 5.Expressão diferencial.

BANCA EXAMINADORA

Prof. Dr. Vinícius Farias Campos, Universidade Federal de Pelotas

Prof. Dra. Sibeles Borsuk, Universidade Federal de Pelotas

Prof. Dr. Odir Antônio Dellagostin, Universidade Federal de Pelotas

Dedico este trabalho de conclusão de curso a todas as pessoas que estiveram presentes durante a minha graduação, apoiando e incentivando meu trabalho. Especialmente à minha família.

Agradecimentos

À Deus, pelo dom da vida e por me abençoar e iluminar minha trajetória.

Aos meus pais, Agostinho e Lucia, por me tornarem a pessoa que sou hoje, me incentivando sempre e acreditando em mim, sem medir esforços para que eu conseguisse alcançar meus objetivos.

Aos meus irmãos Desirèe e Alex, pelo amor, alegria, apoio e companheirismo compartilhado em todos os momentos.

Ao meu namorado Lucas, por estar ao meu lado em todas as circunstâncias, me ajudando e apoiando com muito carinho na busca de novos objetivos. Me ajudou em grande parte de minhas vitórias.

Aos meus amigos que sempre me incentivaram e deram força para a realização das minhas metas, mesmo muitas vezes distantes.

Aos meus padrinhos Ivone e Cláudio, tias e tios que estiveram presentes, acompanhando os desafios e me trazendo muito amor e alegria para que eu conseguisse finalizar minhas tarefas.

Ao Professor Odir Dellagostin, por me proporcionar a oportunidade de ingressar na área científica, sempre fazendo o possível para que eu superasse novos obstáculos.

Ao Doutorando Marcus Eslabão, por me introduzir no mundo da bioinformática, orientando e me apoiando em todos os objetivos.

À Professora Luciana Bicca Dode que sempre esteve presente em todas as conquistas e desafios durante a graduação e que me ensinou, principalmente, os valores de identidade e pertencimento.

Ao Professor Mark Segal, por me aceitar em seu laboratório e me apoiar nesta brilhante oportunidade fundamental para minha formação.

Ao Professor Luciano Pinto, por acreditar em mim e dar suporte em todos os processos administrativos dentro da universidade.

Ao Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas por proporcionar uma completa estrutura de ensino e pesquisa.

Resumo

PLAÇA, Jessica Rodrigues. **Análise de expressão diferencial de células alveolares tipo II de *Rattus norvegicus* utilizando a técnica de microarranjo**. 2014. 48f. Trabalho de Conclusão de Curso — Curso de Bacharelado em Biotecnologia. Universidade Federal de Pelotas, Pelotas, 2014.

As células epiteliais alveolares tipo II (CEAT II) executam uma variedade de funções importantes no pulmão, incluindo a produção e regulação do metabolismo do surfactante pulmonar, transporte de íons e reparação alveolar em resposta à lesão. Trabalhar com linhagens dessas células permite a disponibilidade de um grande número celular, e a utilização de CEAT II de *Rattus norvegicus*, reflete em um melhor entendimento sobre o papel deste tipo de células em processos complexos, como na infecção. O conhecimento do transcriptoma tem sido cada vez mais aplicado para a compreensão das funções gênicas, sendo a técnica de microarranjo amplamente utilizada para esse tipo de análise. Sabe-se que quando CEAT II são cultivadas em placas de cultivo celular, perdem características associadas com o fenótipo maduro da célula e adquirem algumas características do fenótipo da célula epitelial alveolar tipo I. Dadas essas características é fundamental compreender o perfil de expressão gênica dessas células a fim de reconhecer mudanças que possam modificar a funcionalidade das CEAT II devido as condições de cultivo e assim, alterar resultados provenientes de pesquisas com esses tipos celulares. Desta forma, o objetivo do trabalho foi analisar *in silico* os dados de microarranjo de CEAT II de *Rattus norvegicus* frescas e cultivadas por 48 h, a fim de verificar diferenças na expressão gênica devido à condição de cultivo celular. Foram detectados cerca de 16.100 genes expressos nas amostras frescas e cultivadas das CEAT II. Não foram encontrados genes diferencialmente expressos, com todos os valores de p ajustado acima de 0,05, sendo o menor valor 0,211. Além disso, os valores do log *Folde-Change* ($\log_{2}FC > 1,5$) não foram estatisticamente significativos. Assim, os dados demonstram que as CEAT II de rato podem ser incubadas por 48 h, sem alteração nos padrões de expressão gênica nas condições de cultivo utilizadas neste estudo, podendo os experimentos com estas células ocorrer sem que o período e método de cultivo influenciem nos resultados.

Palavras chave: célula epitelial alveolar tipo II; cultivo celular; técnica de microarranjo; expressão diferencial.

Abstract

PLAÇA, Jessica Rodrigues. **Differential expression analysis of *Rattus norvegicus* alveolar type II cells using the microarray technology**. 2014. 48f. Finals period — Curso de Bacharelado em Biotecnologia. Universidade Federal de Pelotas, Pelotas, 2014.

The epithelial alveolar type II cells (EAC II) perform a variety of important functions in the lung, including the production and regulation of pulmonary surfactant metabolism, ion transport and alveolar repair in injury response. Work with these cell lines allows the availability of a large cell number, and the *Rattus norvegicus* EAC II reflects a better understanding of the role of these cells in complex processes such as infection. The knowledge of the transcriptome has been increasingly applied to the understanding of gene functions and the microarray technique is widely used for this type of analysis. It is known that EAC II loses the feature associated with cell mature phenotype and acquire certain properties of epithelial alveolar type I cell phenotype when EAC II are grown in cell culture plates. Given these characteristics, it is essential to understand the gene expression profile of these cells to recognize changes that could modify the functionality of the EAC II due to growing conditions and thus alter research results with these cell types. Therefore, the objective of the study was to analyze *in silico* microarray data, fresh and cultured *Rattus norvegicus* EAC II for 48 h, to detect differences in gene expression due to cell culture conditions. Approximately 16.100 expressed genes were detected in fresh and cultured samples of *Rattus norvegicus* EAC II. No differentially expressed genes were found, with all the adjusted *p-value* above 0.05, being the lowest value 0.211. Moreover the values of log Fold-Change ($\log_{2}FC > 1.5$) were not statistically significant. Thereby, these data demonstrate that rat EAC II can be incubated for 48h with no change in gene expression patterns at conditions used in this study, and research with EAC II may occur without the period and cultivation method influencing the results.

Key-words: epithelial alveolar type II cell; cell culture; microarray technology; differential expression.

Lista de Figuras

Figura 1	Ilustração sistemática das células especializadas do tecido epitelial alveolar.....	16
Figura 2	Representação do protocolo para microarranjo de uma cor e de duas cores.....	24
Figura 3	Gráfico Volcano.....	35

Lista de Tabelas

Tabela 1	Atividade dos 12 genes que apresentaram maior significância estatística nos níveis de expressão gênica.....	34
Tabela 2	Genes que apresentaram maior significância estatística nos níveis de expressão gênica.....	35

Lista de Abreviaturas e Siglas

A	Adenina
ANOVA	Análise de variância
C	Citosina
CEAT I	Célula Epitelial Alveolar Tipo I
CEAT II	Célula Epitelial Alveolar Tipo II
CEM	Células Endoteliais Microcapilares
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
cDNA	Ácido desoxirribonucleico complementar
Cy3	Cianina-3
Cy5	Cianina-5
DPOC	Doença Pulmonar Obstrutiva Crônica
DNA	Ácido desoxirribonucleico
FC	<i>Fold-Change</i>
FDR	<i>False Discovery Rate</i>
FWER	<i>Family Wise Error Rate</i>
G	Guanina
GEO	<i>Gene Expression Omnibus</i>
IL -1	Interleucina 1
IL-6	Interleucina 6
IL-8	Interleucina 8
log	Logarítimo na base 10
SAM	<i>Significance Analysis of Microarrays</i>
mRNA	Ácido ribonucleico mensageiro
pb	Pares de base
RNA	Ácido ribonucleico
SP	Proteína surfactante
SP-A	Proteína surfactante A
SP-B	Proteína surfactante B
SP-C	Proteína surfactante C
SP-D	Proteína surfactante D

T Timina
TGF- β Fator de crescimento transformante
TNF- α Fator de necrose tumoral
U Uracila
UCSF *University of California, San Francisco*

Sumário

1 Introdução Geral	14
2 Objetivos	15
2.1 Objetivo Geral	15
2.2 Objetivos Específicos	15
3 Revisão Bibliográfica.....	16
3.1 Epitélio Alveolar Respiratório	16
3.2 Célula Epitelial Alveolar Tipo II.....	18
3.3 Cultivo <i>in vitro</i> de Células Epiteliais Alveolares Tipo II	19
3.4 O transcriptoma.....	20
3.5 Estudos do transcriptoma de Células Epiteliais Alveolares Tipo II	21
3.6 A técnica de microarranjo.....	22
3.6.1 Tipos de microarranjo.....	24
3.6.2 Normalização	27
3.7 Análise dos dados de microarranjo	27
4 Materiais e Métodos.....	33
4.1 Cultivo celular.....	33
4.2 Extração do RNA e preparo da Técnica de Microarranjo	33
4.3 Análise do Microarranjo.....	33
4.4 Análises estatísticas.....	34
4.5 Anotação dos genes	34
5 Resultados e discussão	35
6 Conclusão	37
Referências bibliográficas	38

1 Introdução Geral

A partir do aceite recebido pelo Professor Mark Segal, do departamento de Epidemiologia e Bioestatística na Universidade da Califórnia – São Francisco (*University of California, San Francisco - UCSF*), e conseguinte contemplação com bolsa modalidade Graduação Sanduíche pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), tive a oportunidade de realizar o meu estágio de final de curso no Centro de Bioinformática e Estatística Molecular da UCSF. Durante este período foi possível aprender sobre assuntos nunca explorados durante a graduação, possibilitando o contato com pesquisadores renomados da área de Informática Biomédica. Além disso, a vivência dentro do laboratório e a participação em reuniões sistemáticas para discussão de artigos e palestras de pesquisadores da área de outras universidades americanas foram fundamentais para aprimorar os meus conhecimentos em bioinformática.

O trabalho desenvolvido em uma das atividades realizadas neste período gerou base para a realização desta monografia. Ele consiste na análise *in silico* de dados de expressão gênica, disponibilizados pela realização da técnica de microarranjo, de Células Epiteliais Alveolares Tipo II de *Rattus norvegicus* frescas e cultivadas por 48 h, a fim de verificar diferenças de expressão gênica devido as condições de cultivo. Para tanto foi necessário um extenso estudo de métodos estatísticos para este tipo de análises bem como o entendimento da metodologia empregada.

2 Objetivos

2.1 Objetivo Geral

Analisar *in silico* dados de expressão gênica, disponibilizados pela realização da técnica de microarranjo, de Células Epiteliais Alveolares Tipo II de *Rattus norvegicus* frescas e cultivadas por 48 h, a fim de verificar diferenças na expressão gênica devido a condição de cultivo celular.

2.2 Objetivos Específicos

- Identificar o perfil de expressão das sondas (genes) disponíveis na plataforma de microarranjo das Células Epiteliais Alveolares Tipo II de rato para cada condição.
- Verificar se há diferenças estatisticamente significativas de expressão gênica nas Células Epiteliais Alveolares Tipo II de ratos entre as condições frescas e cultivadas por 48 h, utilizando o pacote do Bioconductor Limma.

3 Revisão Bibliográfica

3.1 Epitélio Alveolar Respiratório

A estrutura ramificada das vias respiratórias permite a eficiente transferência bidirecional de ar entre o ambiente externo e os alvéolos, sendo que o tecido epitelial constitui essa interface entre o meio interno do organismo e o ambiente externo (GEREKE et al, 2007; RACKLEY; STRIPP, 2012). Assim o epitélio alveolar do aparelho respiratório compreende cerca de 99% da área da superfície interna do pulmão (GONZALEZ et al., 2009) e é o ponto inicial de contato com micro-organismos patogênicos, alérgenos e poluentes ambientais. Por sua vez, sua principal função consiste em uma barreira física (TAM et al, 2011).

A troca gasosa durante a respiração celular ocorre através da barreira de ar-sangue formada pela justaposição das Células Epiteliais Alveolares Tipo I (CEAT I) e Células Endoteliais Microcapilares (CEM), sendo a CEAT I correspondente a aproximadamente 95% da superfície alveolar e possuindo formato escamoso (DOBBS et al., 2010). Outro tipo de célula epitelial alveolar especializada é a Célula Epitelial Alveolar Tipo II (CEAT II). Ambos os tipos de células desempenham importante papel na produção de fluidos e transporte de íons (FEHRENBACH, 2001). Entre esses dois tipos de células especializadas encontra-se uma ou duas finas folhas citoplasmática que se estendem a partir do núcleo da CEAT I para cobrir a lâmina basal que separa o epitélio do intersticial (GONZALEZ et al., 2005). O epitélio pulmonar cumpre, então, múltiplas funções essenciais para a homeostase do tecido (GEREKE et al., 2007).

Estudos recentes indicam claramente que as células epiteliais alveolares são metabolicamente muito ativas, com a capacidade de modular uma variedade de processos inflamatórios através da ação de uma série de acontecimentos mediados por receptores. Após a ativação, elas também possuem a capacidade de produzir um número de mediadores ou reguladores pró-inflamatórios, incluindo os produtos como o ácido araquidônico, o óxido nítrico, endotelina-1, fator de crescimento transformante (TGF- β), fator de necrose tumoral (TNF- α), e citocinas, tais como interleucina IL -1, IL-6 e IL-8 (KNIGHT; HOLGATE, 2003).

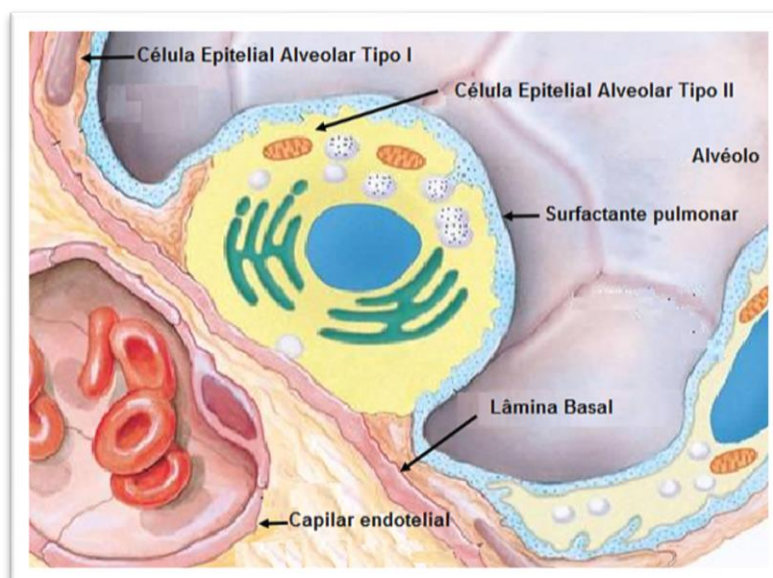


Figura 1 - Ilustração sistemática das células especializadas do tecido epitelial alveolar.

Fonte: Human Physiology site.

Segundo RACKLEY, STRIPP (2012), diferentes necessidades regionais para demais funções, tais como a defesa do hospedeiro ou a troca gasosa, são acompanhadas por diferenças no tipo e abundância de células epiteliais alveolares especializadas em cada local. As propriedades e destino das células epiteliais alveolares são em grande parte ditadas durante o desenvolvimento, mas o seu fenótipo e função são dinâmicos e podem ser influenciados pela mudança das micro ou macro condições ambientais. A capacidade do epitélio para se adaptar a mudanças das condições ambientais é uma propriedade importante que lhe permite tolerar o estresse crônico. No entanto, as respostas exageradas a estímulos crônicos ou respostas aberrantes aos estímulos pequenos podem resultar em remodelagem patológica, levando a alterações na barreira e em propriedades do epitélio, diferenciação na composição do fluido de revestimento extracelular e na taxa de transporte mucociliar, todos os quais têm a potencial de comprometer a mecânica pulmonar e troca gasosa.

3.2 Célula Epitelial Alveolar Tipo II

As CEAT II representam uma subpopulação altamente especializada do epitélio alveolar respiratório. Elas consistem em cerca de 15% das células do pulmão distal e ocupam 5% da superfície alveolar (CHEN et al., 2004). As CEAT II executam uma variedade de funções importantes dentro do pulmão, incluindo a regulação do metabolismo do surfactante pulmonar, transporte de íons e reparação alveolar em resposta à lesão (LESUR; ARSALANE, 1996).

Os critérios utilizados para identificação das CEAT II são seu formato cuboidal e a presença de corpos lamelares, microvilosidades apicais e proteínas específicas de junção. As CEAT II também mantêm a integridade do epitélio alveolar por proliferação e diferenciação em CEAT I em resposta à lesão e, podem regular o fluido alveolar por uma variedade de mecanismos (ANDREEVA et al., 2007).

Além disso, as CEAT II sintetizam e secretam o surfactante pulmonar que forma uma monocamada de lipídeos na interface de ar e de líquido reduzindo ao longo do epitélio alveolar a tensão superficial. O surfactante pode também desempenhar um papel na proteção contra oxidantes e contra infecções (ANDREEVA et al., 2007). Ele consiste de cerca de 80% de glicerofosfolípidos (fosfatidilcolina, fosfatidilglicerol, fosfatidiletanolamina, fosfatidilserina, fosfatidilinositol e esfingomiélna), 10% de colesterol e 10% de proteínas surfactantes tipo SP (SP-A, SP-B, SP-C e SP-D) (JOHANSSON et al., 1994). A composição de ácidos graxos por fosfolípidios pode estar sujeita a regulação, por exemplo, por condições hiperóxicas (DOMBROWSKY et al., 2006), ou diferir entre as células fetais, neonatais e adultas (POSTLE et al., 2006). Teoricamente todos os fosfolípidios surfactantes se originam da exocitose de corpos lamelares, evidenciado pela semelhança da composição de fosfolípidios de corpos lamelares isoladas de CEAT II e do surfactante pulmonar total da lavagem broncoalveolar (KIKKAWA Y, 1983).

As CEAT II expressam um certo número de moléculas necessárias para a transdução, bem como a geração de sinais envolvidos em vias célula-célula como também nas interações célula-matriz. As interações célula-célula podem ser diretas através do contato por proteínas de junção, ou indiretas através de sinais secretados

e difusíveis (FEHRENBACH, 2001). Interações de CEAT II com leucócitos têm sido também objeto de intensa investigação e não há evidências que suportam seu papel na função acessória de ativação dos linfócitos T. Além disso, a expressão de quimiocinas pela CEAT II é induzida pelo reconhecimento específico do antígeno CD8 e células T, desempenhando um importante papel na perpetuação da intersticial experimental da pneumonia, por exemplo (ZHAO et al., 2001). Consequentemente, a CEAT II foi descrita como unidade integrativa do alvéolo (MASON; WILLIAMS, 1977).

3.3 Cultivo *in vitro* de Células Epiteliais Alveolares Tipo II

Dada a importância das funções das CEAT II no pulmão, é necessário isolar populações enriquecidas com viabilidade suficiente e pureza para estudos funcionais (CHEN et al., 2004). Desta forma trabalhar com linhagens dessas células oferece uma série de benefícios, tais como a disponibilidade de um grande número de células para análise extensiva. Acredita-se que a utilização de CEAT II de animais, como os ratos, permite um melhor entendimento sobre o papel deste tipo de células em processos complexos como infecção ou inflamação autoimune (GEREKE et al., 2012). As CEAT II podem então ser isoladas diretamente de animais, inclusive para estudo de doenças respiratórias em ratos modelos, o que significa que têm sido objeto de todos os fatores extrínsecos adicionais desempenhando um papel no cenário analisado.

O extenso estudo para compreensão da ação de meios de cultivo a fim de estabelecer um protocolo eficiente para o experimento em questão é fundamental para estabelecimento de protocolos. A exemplo disso, CORTI et al. (1996) cultivaram pela primeira vez CEAT II de *Rattus norvegicus* em lâminas com a câmara revestida de fibronectina e após 5 dias, mostrando que as células atingiram uma morfologia poligonal achatada e perderam seus corpos lamelares, apesar da transição para o fenótipo de cultura das CEAT I ter ocorrido. No entanto a transição foi apenas deferida pela expressão do fenótipo, pois marcadores específicos não foram avaliados, mostrando estudos iniciais.

Em estudo subsequente, as CEAT II de rato foram cultivadas em placas de plástico revestidas com uma matriz Matrigel de colágeno da cauda de *Rattus*

norvegicus e mantidos em meio suplementado com fator de crescimento de queratinócitos, mostrando que ativamente secretaram e sintetizaram fosfolipídios surfactantes por 7 dias, de acordo as características CEAT II sob estas condições de cultura (SHI et al., 2007).

3.4 O transcriptoma

Todo organismo vivo armazena informação biológica na forma de moléculas de ácidos nucléicos, formadas por nucleotídeos. Cada nucleotídeo por sua vez, consiste de uma molécula de açúcar (desoxirribose ou ribose), um grupo fosfato e, uma base nitrogenada. Os ácidos nucléicos são, portanto, classificados como Desoxirribonucleico (DNA) ou Ribonucleico (RNA).

A sequência de milhões de bases emparelhadas que formam o DNA é subdividida em fragmentos de diversos tamanhos denominados genes. Neles estão contidas as informações que especificam a estrutura das proteínas, macromoléculas que de fato realizam as principais ações nos organismos. No DNA são encontrados quatro tipos de bases nitrogenadas: adenina (A), citosina (C), guanina (G) e timina (T). Ele é representado por uma fita dupla (emparelhamentos de nucleotídeos) complementar e antiparalela. No esquema de emparelhamento, o nucleotídeo A sempre se liga ao nucleotídeo T e o nucleotídeo C sempre se liga ao G através de pontes de hidrogênio.

O RNA também engloba quatro tipos de bases trocando-se timina pela uracila (U). Apresenta-se como uma cadeia única de tamanho menor que o DNA e com grande diversidade de estruturas secundárias relacionadas às suas funções desempenhadas na célula. Quando duas sequências complementares de ácido nucléicos se combinam, esse processo bioquímico é chamado de hibridização.

Por expressão gênica, entende-se o processo em que um gene é utilizado na construção de uma proteína ou para controlar a expressão de outros genes. A síntese das proteínas pode ser resumida em dois passos: O DNA é transcrito ao ácido ribonucleico mensageiro (mRNA) em um processo chamado transcrição; O mRNA é

transformado em proteína ou em parte dela (aminoácido), em um processo chamado tradução.

A transcrição é a reprodução de uma fita de DNA em uma sequência de RNA complementar. O modelo onde o DNA serve como molde para a síntese de moléculas de RNA, que por sua vez, coordenam a síntese de proteínas, é conhecido como o Dogma Central da Biologia Molecular. Outro mecanismo importante é a replicação dos genes que permite a hereditariedade das informações carregadas no DNA. O conjunto de toda a informação codificada no DNA de um organismo é chamado de genoma.

O conjunto de genes expressos ou transcritos é por vezes referido como o perfil de expressão, ou transcriptoma da célula (TSENG et al., 2012). Especificamente o transcriptoma corresponde ao conjunto de RNAs mensageiros, RNAs ribossômicos, RNAs transportadores, pequenos RNAs como os microRNAs e RNAs longos.

O conjunto dos transcritos é dinâmico, respondendo de forma contínua às mudanças nas condições fisiológicas e ambientais em uma célula, tecido ou organismo, e a sua análise proporciona a primeira leitura funcional entre o genoma e o fenótipo expresso (MCHALE et al., 2013). Transcriptômica, a análise do transcriptoma, tem sido um dos pilares dos estudos em genômica funcional e tem sido cada vez mais aplicados em para a compreensão das funções gênicas. Assim é considerado um dos métodos centrais para investigação do papel da proteína codificada e sua atividade biológica (NAPOLI et al., 2003).

3.5 Estudos do transcriptoma de Células Epiteliais Alveolares Tipo II

Embora vários genes e produtos terem sido demonstrados como diferencialmente expressos entre as células epiteliais alveolares especializadas (KALINA et al., 1992; VANDERBILT et al., 2003), o fenótipo molecular detalhado das CEAT II ainda não foi descrito. Como as células CEAT II cultivadas em placas de plástico perdem características associadas com o fenótipo celular maduro e adquirem algumas características do fenótipo celular CEAT I (DOBBS et al., 1985, 1988; CAMPBELL et al., 1999), tem sido proposto que as células cultivadas são transdiferenciadas em CEAT I. Desta forma seria de extrema importância a

compreensão do perfil de expressão gênica dessas células para compreender a biologia celular e reconhecer mudanças que possam modificar a funcionalidade das CEAT II devido as condições de cultivo e assim, alterar resultados provenientes de pesquisas com esses tipos celulares.

Um estudo realizado por FUJINO et al (2012), por exemplo, teve como objetivo identificar o padrão específico da expressão gênica em CEAT II isoladas a partir de pacientes com a Doença Pulmonar Obstrutiva Crônica (DPOC), demonstrando que genes estimulados pelo interferon envolvidos no processamento do antígeno via apresentação e os genes envolvidos na progressão do ciclo celular foram diferencialmente expressos nas CEAT II dos pacientes com DPOC e estes genes poderiam ser responsáveis pelo fenótipo da DPOC. Além disso, outras mudanças nos padrões de expressão multigênicos das CEAT II podem fornecer pistas sobre os mecanismos de regulação e funções bioativas mais amplas, auxiliando na identificação de indicadores moleculares.

Outros estudos, como o desenvolvido por O'REILLY et al. (2012) utilizou a técnica de microarranjo para a análise de expressão do genoma total de um subconjunto geneticamente marcado de CEAT I a fim de identificar novos mecanismos a respostas imunitárias inatas de infecção viral respiratória. Eles encontraram a expressão de eosinófilos-RNase associados 1 e sua perda contribui na lesão epitelial associada ao oxigênio e fibrose após a infecção.

3.6 A técnica de microarranjo

Com o desenvolvimento de novas tecnologias científicas, a pesquisa de ciências biológicas vêm se inserindo na era da biologia sistêmica exercendo um efeito multiplicador sobre o desenvolvimento de vários campos da biotecnologia (LI, 2013). Apesar do desenvolvimento contínuo da tecnologia de sequenciamento de DNA e RNA, esta técnica permanece ainda financeiramente inviável para alguns laboratórios que almejam a sequenciar genomas inteiros, inclusive pela arduosidade em se analisar esses dados. Além disso, para comparações globais entre os genomas dentro das espécies, em que espera-se um número relativamente pequeno de diferenças na

sequência de todo o genoma, determinar a sequência inteira é desnecessária. Em tais casos, é suficiente avaliar a extensão e localização da variação da sequência de maneira análoga à genômica comparativa que compara o número de mudanças entre genomas estreitamente relacionados (GRESHAM et al, 2006). A exemplo disso temos a técnica de microarranjo, primeiramente descrita por SCHENA et al. (1995).

Esta metodologia se tornou desde então uma tecnologia amplamente utilizada para análise de transcriptomas conhecidos devido ao seu baixo custo, facilidade de uso e análise, e estrutura otimizada de controle de qualidade (BRAZMA, 2003). Apesar disso, este método baseado na hibridação do DNA sofre alguns problemas que conduzem a um sinal de saída com ruídos e questões de hibridização cruzada levando a resultados falsos positivos. Além disso, os microarranjos são semi-quantitativos, pois existem uma faixa dinâmica limitada devido à saturação do sinal e falta de sensibilidade para detectar baixas transcritos abundantes (SHI et al., 2006). No entanto, grandes avanços em métodos estatísticos para análise de microarranjo e os dados de sequenciamento de alto rendimento tem sido demonstrados (YAFFE, 2008), e ainda outras vantagens são vistas como sua baixa complexidade computacional.

A maioria dos estudos com microarranjo almejam identificar padrões de expressão gênica em um sistema particular, por exemplo, comparando tecido doente com tecido sadio, analisando respostas de expressão de genes em um estímulo de estresse, ou comparando os padrões de expressão de um tecido em particular ao longo do tempo. Conceitualmente, essas diferenças entre as várias condições devem refletir em mudanças, em particular, das vias celulares (QUACKENBUSH, 2003).

Os conceitos básicos de experimentos de microarranjos utilizam as propriedades físicas e químicas das moléculas de DNA. Uma molécula de DNA é constituída por duas cadeias complementares. Cada fita do DNA pode se hibridizar a sua fita complementar, mas não nos casos das seqüências de nucleotídeos serem muito diferentes de si próprias. Uma vez que as seqüências de milhares de genes diferentes são conhecidas e armazenadas em bases de dados públicas, elas podem ser utilizadas como sondas, que serão fixadas a um suporte de vidro (matriz). Assim

os *spots* (sondas) de DNA são aderidos em um dos lados da matriz, sendo cada sonda um DNA clonado sequenciado com função conhecida ou desconhecidas da célula em estudo (PICARD et al., 2004).

Em paralelo, são extraídos RNAs a partir de amostras biológicas e convertidos em DNA complementar (cDNA). O cDNA é amplificado e marcados com corantes fluorescentes [cianina-3 (Cy3) e cianina-5 (Cy5)]. Esta mistura de transcritos, ou alvos, é hibridizada nas sondas, ou seja, cDNAs se vinculam ao seu modelo complementar. Uma vez que as sondas são localizadas unicamente em um dos lados da matriz, a quantificação dos sinais de fluorescência por pixel irá definir uma medida da abundância de milhares de transcritos em uma célula de uma determinada condição (DUGGAN et al., 1999).

As imagens são geradas por dispositivos de varredura (*scanner*) especiais que utilizam *lasers* microscópicos. Cada imagem é uma representação do microarranjo varrido com várias matrizes em duas dimensões. Como resultado, as imagens digitalizadas apresentam a reação de fluorescência de todos os spots contidos na lâmina varrida pelo laser (CAMPOS, 2008).

3.6.1 Tipos de microarranjo

Uma das primeiras decisões encontradas ao planejar um experimento de microarranjo é o seu tipo abordagem decorrente das classificações de acordo com os critérios de: número de amostras que podem ser simultaneamente perfiladas sob uma matriz, comprimento das sondas e o método de produção.

O número de amostras que podem ser perfilados em uma única matriz pode ser uma ou duas, sendo chamadas então de microarranjos de uma cor ou de duas cores. O procedimento de microarranjos de uma cor envolve a hibridização de uma única amostra de cada matriz depois de ter sido marcada com um fluoróforo único (tais como a ficoeritrina, Cy3 e Cy5), enquanto que no processo de duas cores, duas amostras (por exemplo, experimental e controle) são marcados com diferentes fluoróforos e hibridizadas juntas em um único microarranjo, conforme Figura 2 (SHI et al., 2006).

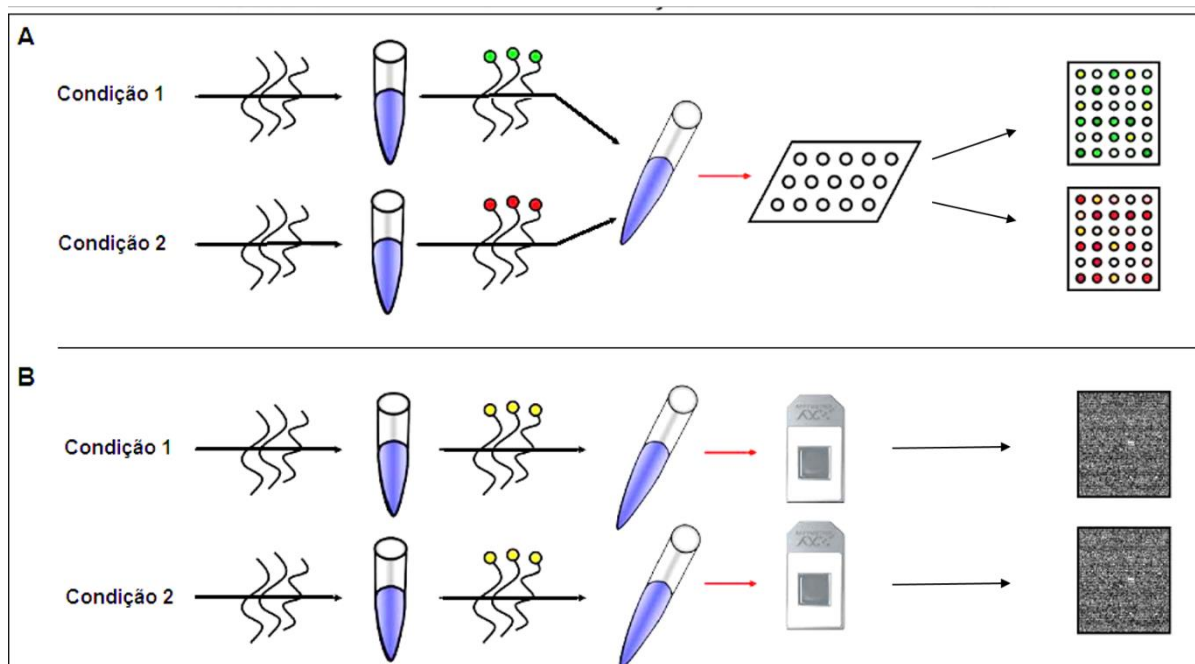


Figura 2 – Representação do protocolo para microarranjo de uma cor e de duas cores.

O painel superior (A) ilustra a tecnologia de duas cores enquanto o painel inferior (B), ilustra a tecnologia de cor única. No painel A, os mRNA de cada amostra são rotulados com dois corantes diferentes, misturados e, em seguida, hibridizados na mesma matriz. Após a lavagem, a matriz é digitalizada em dois comprimentos de onda diferentes para produzir duas imagens, sendo uma para cada condição. No painel B cada amostra é rotulada com o mesmo corante fluorescente, mas hibridizadas de forma independente em diferentes matrizes.

Fonte: TARCA et al (2006).

As principais vantagens de plataformas de cor única são a simplicidade de desenho experimental e flexibilidade. A hibridação de uma amostra única por microarranjo facilita comparações entre microarranjos e entre os grupos de amostras. A inconsistência de dados através de ensaios devido a múltiplas fontes de variabilidade, incluindo fabricação e processamento de microarranjos, pode ser reduzida com microarranjos de uma só cor, realizando suficientes ensaios técnicos e replicas biológicas (TARCA et al., 2006).

Já plataforma do microarranjo de duas cores foi inicialmente desenvolvida para reduzir o erro associado com a variabilidade na produção da técnica, sendo que a disponibilidade de microarranjos comerciais de alta qualidade conseguiu cumprir este

objetivo com êxito. Quando ocorre a hibridização de duas amostras no mesmo microarranjo permite-se uma comparação direta, minimizando a variabilidade advinda do processamento de vários microarranjos por ensaio. Esta variabilidade reduzida teoricamente resulta no aumento da sensibilidade e precisão na determinação dos níveis de expressão diferencial entre os pares de amostras (PATTERSON et al., 2006).

Ainda, as técnicas de microarranjo podem ser classificadas de acordo com o comprimento das sondas presentes nas matrizes, podendo ser de cDNA ou de oligonucleotídeos. As matrizes com sondas de cDNA, molécula sintética mais estável, possuem centenas ou milhares de pares de base (pb) que se estendem em ensaios necessariamente de duas amostras biológicas diferencialmente marcadas por corantes fluorescentes, e competitivamente hibridizadas em uma matriz para proporcionar uma medida relativa da quantidade de transcritos. Já as plataformas de oligonucleotídeos utilizam sondas curtas (geralmente 50 pb) constituídas por segmentos não repetidos e que hibridizam apenas com um dos mRNA a ser utilizado. Um exemplo de matriz de cor única e oligonucleotídeo é o Affymetrix GeneChip (CAMPOS, 2008).

Além disso, os diferentes métodos de fabricação incluem a deposição de sequências previamente sintetizados e *in situ*. Normalmente, os arranjos de cDNA são fabricados usando a deposição prévia com pinos, jatos de tinta e outros dispensadores para realizar a deposição das sondas (ROSE, 2000). As matrizes de oligonucleotídeos são fabricados com tecnologias *in situ*. As tecnologias *in-situ*, realizam a síntese das sequências de DNA diretamente na matriz de microarranjo, através dos nucleotídeos em suspensão. Para produzir as sondas incluem-se as metodologias de fotolitografia (por exemplo, Affymetrix, Santa Clara, CA), impressão a jato de tinta (por exemplo, Agilent, Palo Alto, CA) e síntese eletroquímica (por exemplo, CombiMatrix, Mukilteo, WA) (TARCA et al., 2006; GRESHAM et al., 2008).

3.6.2 Normalização

A normalização é um passo de pré-processamento que visa corrigir diferenças sistemáticas entre genes ou matrizes, sendo que as técnicas específicas dependem da tecnologia de matriz utilizada. Por exemplo, em uma matriz de cDNA de duas cores, as intensidades da amostra marcada com o corante verde (Cy3) podem aparecer consistentemente mais elevadas do que as da amostra marcada com o corante vermelho (Cy5). Devido a isso, meramente considerando as razões entre as intensidades de vermelhas e verdes, as relações entre o quantidades de mRNA na amostra não refletem com precisão. Este desequilíbrio entre os dois canais é conhecido como "viés do corante" (QUACKENBUSH, 2001).

Em plataformas Affymetrix, as intensidades das sondas em uma determinada matriz podem ser consistentemente mais elevadas ou mais baixas. Tais diferenças são coletivamente referidas como "viés de matriz". Assim, comparando as intensidades da mesma sonda sobre as diferentes matrizes podem introduzir erros graves se um procedimento de normalização não é realizado pela primeira vez. Têm sido propostos vários métodos para resolver este problema (LI; WONG, 2001; IRIZARRY et al., 2003).

Abordagens comuns para normalizar estes problemas de intensidades de hibridação envolvem uma escala linear, suavização local e ponderada da dispersão e normalização por quantil (BILBAN et al., 2002). Para as tecnologias de matriz que incluem várias sondas por transcrito, estas intensidades por sonda hibridizada são resumidas a uma estimativa de expressão por transcrição, muitas vezes levando em conta o grau de hibridação do controle negativo e as estimativas da eficiência de diferentes hibridizações de cada sonda (CAMPBELL et al., 2011).

3.7 Análise dos dados de microarranjo

O principal objetivo da análise dos estudos de microarranjos é identificar características genéticas individuais associadas a um fenótipo, principalmente através da expressão diferencial. O método utilizado para alcançar esta meta depende das condições celulares que estão sendo examinadas, inclusive o número de réplicas (PICARD et al., 2004).

A réplica em um experimento de microarranjo é essencial para obter a variação na expressão do gene pelos cálculos estatísticos. Existem dois tipos de réplicas: biológica e técnica. A réplica biológica refere-se a análise de múltiplas amostras biológicas independentes (por exemplo, um tipo de tecido obtido a partir de diferentes pacientes com a mesma doença, ou amostras individuais de uma determinada linhagem de células sob o mesmo tratamento), enquanto a réplica técnica refere-se à repetição do experimento de microarranjo usando as mesmas amostras de RNA extraídas. A réplica biológica é particularmente importante para detectar os perfis de expressão das amostras, porque pode haver variabilidade de expressão entre o mesmo tipo de tecido. Qualquer amostra pode não ser representante de todo o grupo condicional. A réplica técnica fornece uma medição precisa do gene expresso para uma determinada amostra e elimina muitas variações técnicas introduzidas durante o experimento. Sugere-se que cada experimento deve ser realizado em triplicata técnica para aumentar confiabilidade dos dados (SLONIM; YANAI, 2009).

Apenas a obtenção de uma medição de expressão exata de uma amostra pela réplica técnica não vai solucionar o problema da variação biológica. Mas para tanto, técnicas estatísticas para identificação de perfis de expressão gênica podem ser utilizadas. Geralmente estes métodos se enquadram em duas grandes categorias: supervisionadas e não supervisionadas. Métodos supervisionados contam com o conhecimento prévio sobre a condição biológica da qual as amostras se originaram. Um exemplo seria utilizar as amostras no experimento de tecidos doentes *versus* tecidos saudáveis. Os testes estatísticos então usam essas informações para identificar genes que são significativamente diferencialmente expressos entre os tecidos de amostras ou encontrar padrões de expressão gênica que se correlacionam com características celulares (CAMPBELL et al., 2011).

Quase todas as abordagens estatísticas para análise das variáveis contínuas podem ser aplicadas neste tipo de análise de dados, e baseiam-se na comparação da magnitude da diferença da expressão dos genes associados com a condição celular e a magnitude da variabilidade na expressão do gene, inclusive de amostras semelhantes. O teste pode ser paramétrico ou não paramétrico e é realizado por meio

do cálculo sobre os valores de expressão do gene de interesse medido nos dois grupos em estudo (condição 1 e condição 2), como, por exemplo, o Teste t de *Student's*, o teste de soma das ordens de Wilcoxon e Análise de Variância (ANOVA) (TSENG et al., 2012). Estes métodos geralmente dão um valor para cada gene representando a probabilidade de observar a quantidade de expressão diferencial associada ao fenótipo de um determinado gene dada a hipótese nula de que a expressão gênica não varia com o fenótipo (FAN; REN, 2006).

Os métodos estatísticos utilizados para identificar genes diferencialmente expressos em, dois grupos são baseados *Fold-Change* (FC). O FC foi um dos primeiros métodos estatísticos a serem usados na análise de microarranjo e basicamente descreve o valor de quanto uma amostra em determinada condição altera sua expressão gênica quando comparada com outra amostra em outra condição. Todavia ele sozinho não consegue quantificar nenhuma probabilidade estatística e por isso não é utilizado nas análises de microarranjo.

A fim de avaliar o significado estatístico do FC, alguns pesquisadores têm utilizado o Teste t . Teste t compara a diferença nos níveis de expressão entre os dois grupos, levando em consideração a variabilidade dos dados (diferença significa entre grupos dividido pelo desvio padrão). No entanto, o desvio padrão pode ser muito pequeno simplesmente ao acaso. Quando o denominador aproxima-se de zero, o valor- p se torna grande e, por conseguinte, o gene parece ser altamente significativo quando, na realidade não é (TARCA et al., 2006).

O Teste b também possui a mesma função do Teste t , pois é o log da razão das probabilidades entre a proporção de genes diferencialmente expressos *versus* não diferencialmente expressos. Todavia ele permite variações específicas e combina informações através de muitos genes, portanto é mais estável do que o Teste t (CUI; CHURCHILL, 2003). Métodos empíricos Bayesianos que minimizam estimativas de variância individuais em direção a um valor comum também têm sido propostos para melhorar o comportamento do Teste t em muitas configurações por gene (SMYTH, 2004; VERDUCCI et al., 2006).

Já o teste ANOVA pode ser considerado uma generalização do Teste t , quando se compara dois grupos por apenas uma variável independente. De maneira geral é atribuído quando mais de duas amostras biológicas são estudadas. Um outro método, descrito por TUSHER et al. (2001) denominado de *Significance Analysis of Microarrays* (SAM), também tem sido bastante adotado. Nele, é atribuído um valor a cada gene, com base na mudança relativa de sua expressão e o desvio padrão dessa mudança. Genes com valores maiores do que um limiar são considerados significativos.

A transformação logarítmica desses testes estatísticos é muito utilizada. Sem a transformação logarítmica, a distribuição dos dados pode ser muito assimétrica, enquanto com a transformação em logaritmo (ou outras transformações semelhantes, reconhecendo que a transformação em log não pode lidar com o valor zero), mesmo que a distribuição não seja ainda uma perfeita distribuição normal, é muito mais parecida com normal. Há outras vantagens na transformação em log, como a variância ser mais estabilizada sem tender ao aumento com a média (LI, 2012).

Rejeitar uma hipótese nula verdadeira é chamado de Erro Tipo I ou um falso-positivo enquanto não rejeitar uma hipótese nula falsa é chamado de Erro Tipo II ou uma falso negativo. A não rejeição da hipótese nula não implica que a hipótese nula é verdadeira, sendo este um equívoco comum. Significa, simplesmente, que não há evidência suficiente com base nos dados observados para concluir que a hipótese nula deve ser rejeitada em favor da alternativa (BARRY, 2011).

Todos os genes com um valor- p inferior a um determinado limiar são considerados significativos se assim expressos diferencialmente em resposta à condição de estudo, sendo necessário fazer a distinção entre as possíveis causas das diferenças na expressão dos genes. Uma dificuldade neste tipo de análise é a chamada Comparação Múltipla ou Teste de Hipóteses Múltiplas, pois quando inferências estatísticas são feitas para milhares de genes em única análise, um número considerável de dados de expressão passa do limiar do valor- p (por exemplo, $p < 0,05$) apenas por acaso, mesmo quando não existe nenhuma expressão diferencial verdadeira no que diz respeito ao fenótipo (CAMPBELL et al., 2011). Um exemplo de

Teste de Hipóteses Múltiplas é o Teste de Bonferroni que controla o aumento da probabilidade global do Erro do Tipo I e a diminuição do intervalo de confiança global (SUÁREZ et al., 2009).

O método de controle mais comumente usado para testes de Hipóteses Múltiplas é o *Family Wise Error Rate (FWER)* que significa a probabilidade de produzir um ou mais falsos positivos de todas as hipóteses testadas (SUÁREZ et al., 2009). No entanto, apesar desta técnica ser geralmente escolhida entre as demais aplicações não é uma boa escolha para estudos de microarranjos, porque a probabilidade de identificar corretamente genes diferencialmente expressos é muito baixa. Métodos alternativos, como a Taxa de Detecção Falsa (*False Discovery Rate - FDR*), propostos por BENJAMINI e HOCHBERG (1995) podem ajustar o valor- p , a fim de estimar a fração de genes que passam do limiar do valor- p e são susceptíveis a serem falsos positivos e geralmente devem ser aplicados antes das estatísticas inferenciais serem feitas (REINER et al., 2003).

Outros métodos supervisionados utilizam o domínio para previsão de classificação. Estes métodos usam um conjunto de amostras casos e controles de treinamento para identificar e combinar os níveis de expressão de múltiplos genes em um único composto que pode prever o estado da condição das amostras onde é desconhecido. Métodos que foram aplicados para dados de microarranjos combinando esses recursos em preditores compostos incluem votação ponderada (*weighted voting*), k -vizinhos mais próximos (*k-nearest neighbours*), florestas aleatórias (*random forests*), redes neurais artificiais (*artificial neural networks*) e máquinas de vetor de suporte (*support vector machines*) (GOLUB et al., 1999; FUREY et al., 2000; LI et al., 2001; KHAN et al., 2001; ZHANG et al., 2003).

Há muitos pacotes comerciais para análises de microarranjo, no entanto ferramentas comerciais podem ser caras e ter flexibilidade limitada. Uma série de ferramentas *on-line*, como *softwares* de código aberto estão disponíveis para a análise de dados de microarranjo. Uma estratégia comum é criar um *pipeline* de análise dos dados personalizado usando pacotes estatísticos de *softwares* de análise, tais como Matlab ou R. Ambos permitem grande flexibilidade, análise personalizada, e acesso a

muitos pacotes especializados projetados para analisar dados de expressão de gene. A plataforma R, por exemplo, também permite o uso do *BioConductor* que fornece uma série de ferramentas para a análise e compreensão de dados genômicos de alto rendimento (SLONIM; YANAI, 2009).

4 Materiais e Métodos

4.1 Cultivo celular

As CEAT II foram isoladas de *Rattus norvegicus* livres de patógenos (Simonsen Laboratories, Gilroy, CA). Para as células que seriam cultivadas, houve o semeio em triplicata em placas revestidas com colágeno e cultivadas durante 48h em DMEM-H21, F-12 de Ham (1:1) e 10% de soro fetal bovino. Para as células frescas também foram utilizadas triplicadas para realização do estudo.

4.2 Extração do RNA e preparo da Técnica de Microarranjo

O RNA total foi extraído de todas as amostras a partir das células frescas e das células cultivadas, utilizando o protocolo *RNA STAT* (Tel -Test, Friendswood, TX, EUA) segundo as instruções do fabricante. Todas as amostras de foram avaliadas pelo Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, EUA). Para a realização da análise de microarranjo o RNA foi convertido em cDNA marcado com biotina, utilizando reagentes e protocolos *Affymetrix* (www.affymetrix.com). O cDNA foi hibridizado com chips de microarranjo *Affymetrix GeneChip® Rat Gene 1.0 ST Array*. Um total de 3 chips foram analisados com células Alveolares Tipo II cultivadas por 48h e células frescas. As hibridações foram realizadas com cDNA preparado a partir de células frescas e células cultivadas por 48h. A Hibridização, lavagem, coloração e digitalização foram realizadas utilizando os procedimentos descritos no manual técnico *Affymetrix GeneChip Expression Analysis*.

4.3 Análise do Microarranjo

O programa *Affymetrix Microarray Suite 5.0* foi utilizado para quantificar o conteúdo de mRNA de genes expressos. Este *software* usa o teste de Wilcoxon para avaliar se uma transcrição é detectável no arranjo (presente, marginal, ausente) e para calcular a probabilidade de uma significativa mudança entre as matrizes (ou seja, células frescas e células cultivadas), atribuindo os valores-*p* e uma chamada de mudança (aumento, diminuição, sem alteração) para cada sonda. Quando mais do que um conjunto de sonda estava presente para o mesmo gene, os dados foram

combinados para fornecer um valor médio. Os valores padrões fornecidos pela *Affymetrix* foram aplicados a todos os parâmetros de análise.

As medidas da intensidade da fluorescência em cada canal de cor da imagem podem ser distorcidas durante os passos de preparação de dados. Logo, para garantir a confiabilidade nas comparações entre os resultados obtidos de diferentes lâminas de microarranjos é necessário o uso de técnicas de normalização dos dados. Para isso sondas de 70 genes controle em cada chip foram usadas para normalizar a intensidade de fluorescência entre os chips e as matrizes foram escaladas a uma intensidade média de 1.500 unidades de fluorescência e analisadas de forma independente.

4.4 Análises estatísticas

A análise estatística dos dados foi realizada via pacote Limma, uma biblioteca do ambiente *R Bioconductor* que utiliza determinações com base no FDR dos números de genes, sendo esta probabilidade encontrada por métodos Bayesianos Empíricos para fornecer resultados estáveis, mesmo quando o número de matrizes é pequena. As funções de normalização e de análise de dados são para microarranjos com sinal de duas cores. A linguagem de programação R também foi utilizada para manejo dos dados e construção de gráficos.

4.5 Anotação dos genes

Os genes foram anotados com informações de dois bancos de dados: ENSEMBL (disponível *online* em <http://www.ensembl.org/index.html>) e o GenBank (disponível *online* em <http://www.ncbi.nlm.nih.gov/genbank/>). A função molecular dos genes foi determinada a partir da base de dados Swiss-Prot (disponível *online* em <http://us.expasy.org/sprot/sp-docu.html>), e a distribuição relativa dos genes entre as categorias foi comparada com dados de todas as entradas de ratos em esta base de dados. Vários termos foram geralmente atribuídos com a função molecular do gene.

5 Resultados e discussão

Foram detectados 16.145 genes expressos entre as amostras frescas e cultivadas das CEAT II nas sondas.

O resultado da análise de expressão diferencial, realizado pelo pacote *Limma* gerou a lista com os valores estatísticos obtidos respectivos ao \log *Fold-change*, Média de Expressão entre as condições, valor-*t*, valor-*p* e valor-*b*, para cada sonda (genes) presente. Na tabela 1, pode-se ver a lista dos 12 genes que, segundo o valor-*p*, são os mais diferencialmente expressos ($p < 0,05$). Percebe-se que estes genes estão relacionados principalmente a estruturação da cromatina, como por exemplo os genes HIST1H2BL e HIST1H1A, controle do ciclo celular, como os genes ENC1, HAUS3 e NUF2 e, sinalização celular como os genes RGS18 e MOXD1.

Tabela 1 - Atividade dos 12 genes que apresentaram maior significância estatística nos níveis de expressão gênica.

Gene	Atividade
Moxd1	Catalítica
LOC100360880	Transcricional
Kif18b	Motora dos microtúbulos
Rgs18	Transdutor de sinal
Nuf2	Ligação de proteínas
Fam71a	Desconhecida
Lin9	Controle de expressão
Haus3	Formação de microtúbulos
LOC691979	DNA polimerase direcionada
Hist1h1a	Ligação ao DNA
Hist1h2bl	Ligação ao DNA
Enc1	Liagação a actina

No entanto, quando analisa-se os resultados gerados na Tabela 2, percebe-se que o valor-*p* quando normalizado e ajustado de acordo com *FDR* não apresenta genes diferencialmente expressos, com todos os valor-*p* ajustado acima de 0,05, com o menor valor de aproximadamente 0,211. Além disso os valores *do log Fold-Change* ($\log FC > 1,5$), valor-*t*, e valor-*b* não foram estatisticamente significativos.

Tabela 2 - Genes que apresentaram maior significância estatística nos níveis de expressão gênica.

Gene	logFC	MExp	T	P	P ajustado	B
Moxd1	1.457251	6.921113333	10.04373587	1.30864000058229e-05	0.211266842	2.220510966
LOC100360880	-0.854525	4.887492667	-7.553909664	9.29178329116006e-05	0.305841858	1.159452005
Kif18b	0.71628	4.292725	7.058337036		0.000146024	0.305841858
Rgs18	-1.017468	7.5279175	-6.8083087		0.000185158	0.305841858
Nuf2	-1.093259	3.204073667	-6.625082451		0.00022128	0.305841858
Fam71a	-0.672415	6.372040333	-6.593108813		0.000228356	0.305841858
Lin9	-0.917715	5.91635	-6.555768647		0.00023694	0.305841858
Haus3	-1.671272	10.1030805	-6.438893261		0.000266218	0.305841858
LOC691979	-0.456952	6.543021333	-6.326999843		0.000298066	0.305841858
Hist1h1a	-0.591561	5.943685	-6.128439373		0.000365558	0.305841858
Hist1h2bl	-0.609967	6.408788833	-6.111232283		0.000372164	0.305841858
Enc1	-0.731833	8.0880975	-5.956836126		0.00043774	0.305841858

Para melhor visualização do resultado pode-se utilizar a Figura 3, pois mostra duas regiões de interesse, os pontos que se encontram na extremidade superior da área e os pontos que estão nas extremidades esquerda ou direita da área do gráfico. No entanto, no gráfico apresentado é possível perceber que valores de *Fold-Change* são muito baixos e não há significância estatística entre as condições.

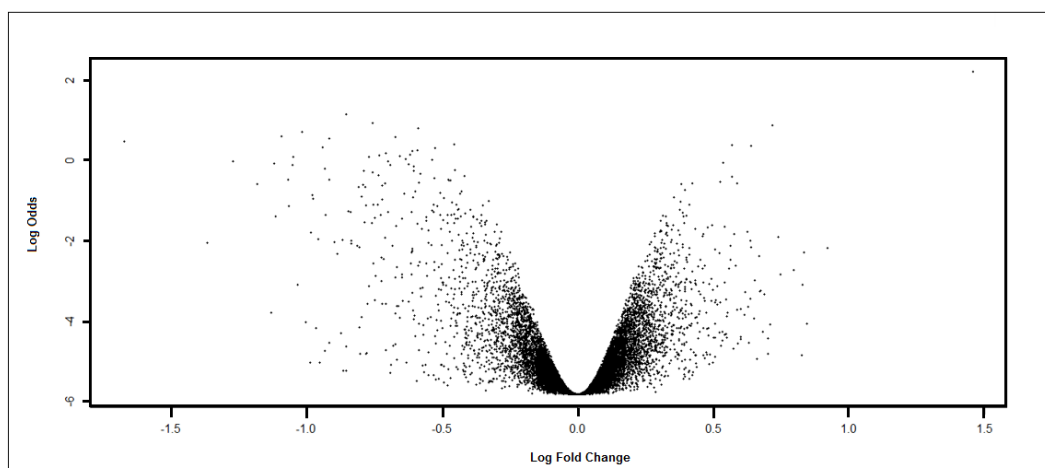


Figura 3 -Gráfico Volcano. No eixo y encontra-se o logaritmo negativo do valor da Taxa de Probabilidade da hipótese (*Log Odds*) e no, eixo x, encontra-se o logaritmo do *Fold-Change* entre as duas condições (células Alveolares Tipo II frescas e cultivadas).

6 Conclusão

Estes dados encontrados demonstram que as Células Epiteliais Alveolares Tipo II de rato podem ser incubadas por 48 h, sem alteração nos padrões de expressão gênica das aproximadamente 16.000 sondas analisadas e nas condições de cultivo utilizadas neste estudo, podendo as pesquisas com estas células ocorrer sem que o período e método de cultivo influenciem nos resultados. Percebe-se também que a escolha das análises estatísticas corretas são fundamentais para um resultado de expressão diferencial confiável, pois apesar de alguns métodos estatísticos mostrarem que os dados de expressão poderiam ser diferencialmente expressos quando ajustados demonstram que este resultado seria um falso positivo.

Apesar de não terem sido encontrados genes diferencialmente expressos entre as duas condições, esta metodologia é de extrema importância para realização de pesquisas que necessitem do cultivo celular. No futuro outros trabalhos possam ser realizados a fim de identificar variações de expressão gênica em um tempo de cultivo maior, ou até mesmo com outros meios de cultivo ou tipos celulares.

Referências bibliográficas

ANDREEVA, A. V; KUTUZOV, M. A.; VOYNO-YASENETSKAYA, T. A. Regulation of surfactant secretion in alveolar type II cells. **Am J Physiol Lung Cell Mol Physiol** **293**:, v. 60612, n. Mc 868, 2007.

BARRY, W. T. Statistical Considerations for Analysis of Microarray Experiments
Statistical Considerations for Analysis of Microarray Experiments. **Duke University**, 2011.

BENJAMINI, Y; HOCHBERG, Y. Controlling the False Discovery Rate: a Practical and Powerful approach to Multiple Testing. **J R Stat Soc Ser B**, v. 57, p. 289–300, 1995.

BILBAN, M.; BUEHLER, L. K.; HEAD, S.; DESOYE, G.; QUARANTA, V. Normalizing DNA Microarray Data. **Curr. Issues Mol. Biol.**, p. 57–64, 2002.

BRAZMA, A. ArrayExpress--a public repository for microarray gene expression data at the EBI. **Nucleic Acids Research**, v. 31, n. 1, p. 68–71, 2003.

CAMPBELL, J. D.; SPIRA, A.; LENBURG, M. E. Applying gene expression microarrays to pulmonary disease. **Respirology (Carlton, Vic.)**, v. 16, n. 3, p. 407–18, 2011.

CAMPBELL, L.; HOLLINS, A. J.; AL-EID, A.; et al. Caveolin-1 Expression and Caveolae Biogenesis during Cell Transdifferentiation in Lung Alveolar Epithelial Primary Cultures. **Biochemical and Biophysical Research Communications**, v. 262, n. 3, p. 744–751, 1999.

CAMPOS, R. S. UM ESTUDO SOBRE PROCESSAMENTO E ANÁLISE DE IMAGENS DE MICROARRANJOS DE DNA, 2008.

CHEN, J.; CHEN, Z.; NARASARAJU, T.; JIN, N.; LIU, L. Isolation of highly pure alveolar epithelial type I and type II cells from rat lungs. **Laboratory investigation; a journal of technical methods and pathology**, v. 84, n. 6, p. 727–35, 2004.

CORTI, M.; BRODY, A. R.; HARRISON, J. H. Isolation and primary culture of murine alveolar type II cells. **American journal of respiratory cell and molecular biology**, v. 14, n. 4, p. 309–15, 1996. American Public Health Association.

CUI, X.; CHURCHILL, G. A. Statistical tests for differential expression in cDNA microarray experiments. **Genome biology**, v. 4, n. 4, p. 210, 2003.

DOBBS, L. G.; JOHNSON, M. D.; VANDERBILT, J.; GONZALEZ, R. The Great Big Alveolar TI Cell : Evolving Concepts and Paradigms. **Cellular Physiology and Biochemistry**, p. 55–62, 2010.

DOBBS, L. G.; WILLIAMS, M. C.; BRANDT, A. E. Changes in biochemical characteristics and pattern of lectin binding of alveolar type II cells with time in culture. **Biochimica et biophysica acta**, v. 846, n. 1, p. 155–66, 1985.

DOBBS, L. G.; WILLIAMS, M. C.; GONZALEZ, R. Monoclonal antibodies specific to apical surfaces of rat alveolar type I cells bind to surfaces of cultured, but not freshly isolated, type II cells. **Biochimica et Biophysica Acta (BBA) - Molecular Cell Research**, v. 970, n. 2, p. 146–156, 1988.

DOMBROWSKY, H., TSCHERNIG, T., VIETEN, G., RAU, G. A., OHLER, F., ACEVEDO, C., BEHRENS, C., POETS, C. F., VON DER HARDT, H. AND BERNHARD, W. Molecular and functional changes of pulmonary surfactant in response to hyperoxia. **Pediatr. Pulmonol.**, v. 41, p. 1025–1039, 2006.

DUGGAN, D. J.; BITTNER, M.; CHEN, Y.; MELTZER, P.; TRENT, J. M. Expression profiling using cDNA microarrays. **Nature genetics**, v. 21, n. 1 Suppl, p. 10–4, 1999.

FAN, J.; REN, Y. Statistical analysis of DNA microarray data in cancer research. **Clinical cancer research : an official journal of the American Association for Cancer Research**, v. 12, n. 15, p. 4469–73, 2006.

FEHRENBACH, H. Alveolar epithelial type II cell: defender of the alveolus revisited. **Respiratory research**, v. 2, n. 1, p. 33–46, 2001.

FUJINO, N.; OTA, C.; TAKAHASHI, T.; et al. Gene expression profiles of alveolar type II cells of chronic obstructive pulmonary disease: a case-control study. **BMJ open**, v. 2, n. 6, p. 1–9, 2012.

FUREY, T. S.; CRISTIANINI, N.; DUFFY, N.; DAVID, W. Support vector machine classification and validation of cancer tissue samples using microarray expression data. **Bioinformatics**, v. 16, n. 10, p. 906–914, 2000.

GEREKE, M.; GRÖBE, L.; PRETTIN, S.; et al. Phenotypic alterations in type II alveolar epithelial cells in CD4+ T cell mediated lung inflammation. **Respiratory research**, v. 8, p. 47, 2007.

GOLUB, T. R.; SLONIM, D. K.; TAMAYO, P.; et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. **Science (New York, N.Y.)**, v. 286, n. 5439, p. 531–7, 1999.

GONZALEZ, R. F.; ALLEN, L.; DOBBS, L. G. Rat alveolar type I cells proliferate, express OCT-4, and exhibit phenotypic plasticity in vitro. **American journal of**

physiology. Lung cellular and molecular physiology, v. 297, n. 6, p. L1045–55, 2009.

GONZALEZ, R.; YANG, Y. H.; GRIFFIN, C.; et al. Freshly isolated rat alveolar type I cells , type II cells , and cultured type II cells have distinct molecular phenotypes. **Am J Physiol Lung Cell Mol Physiol**, v. 94118, p. 179–189, 2005.

GRESHAM, D.; DUNHAM, M. J.; BOTSTEIN, D. Comparing whole genomes using DNA microarrays. **Nature reviews. Genetics**, v. 9, n. 4, p. 291–302, 2008.

GRESHAM, D.; RUDERFER, D. M.; PRATT, S. C.; et al. Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. **Science (New York, N.Y.)**, v. 311, n. 5769, p. 1932–6, 2006.

IRIZARRY, R. A.; HOBBS, B.; COLLIN, F.; et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. **Biostatistics (Oxford, England)**, v. 4, n. 2, p. 249–64, 2003.

JOHANSSON, J.; CURSTEDT, T.; ROBERTSON, B. The proteins of the surfactant system. **European Respiratory Journal**, v. 7, n. 2, p. 372–391, 1994.

KALINA, M.; MASON, R. J.; SHANNON, J. M. Surfactant protein C is expressed in alveolar type II cells but not in Clara cells of rat lung. **American journal of**

respiratory cell and molecular biology, v. 6, n. 6, p. 594–600, 1992. American Lung Association.

KHAN, J.; WEI, J. S.; RINGNÉR, M.; et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. **Nature medicine**, v. 7, n. 6, p. 673–9, 2001.

KIKKAWA Y, S. F. Cellular and biochemical aspects of pulmonary surfactant in health and disease. **Lab Invest.**, v. 48, n. 3, p. 122–39, 1983.

KNIGHT, D. A.; HOLGATE, S. T. INVITED REVIEW SERIES : CELLS OF THE LUNG The airway epithelium : Structural and functional properties in health and disease. **Respirology (2003)**, p. 432–446, 2003.

LI, C.; WONG, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 1, p. 31–6, 2001.

LI, H. Systems genetics in “-omics” era: current and future development. **Theory in biosciences = Theorie in den Biowissenschaften**, v. 132, n. 1, p. 1–16, 2013.

LI, L.; WEINBERG, C. R.; DARDEN, T. A. gene expression data : study of sensitivity to choice of parameters of the GA / KNN method. **Bioinformatics**, v. 17, n. 12, p. 1131–1142, 2001.

LI, W. Volcano plots in analyzing differential expressions with mRNA microarrays. **Journal of bioinformatics and computational biology**, v. 10, n. 6, p. 1231003, 2012.

MARCUS GEREKE, ANDREA AUTENGRUBER, LOTHAR GRÖBE, ANDREAS JERON, DUNJA BRUDER, S. S.-K. Flow Cytometric Isolation of Primary Murine Type II Alveolar Epithelial Cells for Functional and Molecular Studies. **oVE Immunology and Infection**, 2012.

MASON RJ, W. M. Type II alveolar cell. Defender of the alveolus. **Am Rev Respir Dis.**, v. 115, p. 81–92, 1977.

MCHALE, C. M.; ZHANG, L.; THOMAS, R.; SMITH, M. T. Review Article Analysis of the Transcriptome in Molecular Epidemiology Studies. , v. 517, n. August, p. 500–517, 2013.

NAPOLI, C.; LERMAN, L. O.; SICA, V.; et al. Microarray analysis: a novel research tool for cardiovascular scientists and physicians. **Heart (British Cardiac Society)**, v. 89, n. 6, p. 597–604, 2003.

O. LESUR , K. ARSALANE, D. L. Lung alveolar epithelial cell migration in vitro: modulators and regulation processes. **American Journal of Physiology - Lung Cellular and Molecular Physiology**, v. 270, p. L311–L319, 1996.

O'REILLY, M. A; YEE, M.; BUCZYNSKI, B. W.; et al. Neonatal oxygen increases sensitivity to influenza A virus infection in adult mice by suppressing epithelial expression of Ear1. **The American journal of pathology**, v. 181, n. 2, p. 441–51, 2012. Elsevier Inc.

PATTERSON, T. A; LOBENHOFER, E. K.; FULMER-SMENTEK, S. B.; et al. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. **Nature biotechnology**, v. 24, n. 9, p. 1140–50, 2006.

PICARD, F.; ROBIN, S.; PARIS-GRIGNON, I. N. A.; BERNARD, C.; PARIS, F.-. Introduction to Statistical Methods for Microarray Data Analysis, 2004.

POSTLE, A. D.; GONZALES, L. W.; BERNHARD, W.; et al. Lipidomics of cellular and secreted phospholipids from differentiated human fetal type II alveolar epithelial cells. **Journal of lipid research**, v. 47, n. 6, p. 1322–31, 2006.

QUACKENBUSH, J. Computational analysis of microarray data. **Nature reviews. Genetics**, v. 2, n. 6, p. 418–27, 2001.

QUACKENBUSH, J. Genomics. Microarrays--guilt by association. **Science (New York, N.Y.)**, v. 302, n. 5643, p. 240–1, 2003.

RACKLEY, C. R.; STRIPP, B. R. Review series Building and maintaining the epithelium of the lung. **The Journal of Clinical Investigation**, v. 122, n. 8, 2012.

REINER, A.; YEKUTIELI, D.; BENJAMINI, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. **Bioinformatics**, v. 19, n. 3, p. 368–375, 2003.

SCHENA, M.; SHALON, D.; DAVIS, R. W.; BROWN, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science (New York, N.Y.)**, v. 270, n. 5235, p. 467–70, 1995.

SHI, L.; REID, L. H.; JONES, W. D.; et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. **Nature biotechnology**, v. 24, n. 9, p. 1151–61, 2006.

SHI, X.; ZHANG, H.; XIONG, S. Primary culture of alveolar epithelial type II cells and its bionomic study. **Journal of Huazhong University of Science and Technology. Medical sciences**, v. 27, n. 6, p. 653–6, 2007.

SLONIM, D. K.; YANAI, I. Getting started in gene expression microarray analysis. **PLoS computational biology**, v. 5, n. 10, p. e1000543, 2009.

SMYTH, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. **Statistical applications in genetics and molecular biology**, v. 3, n. 1, p. Article3, 2004.

SUÁREZ, E.; BURGUETE, A.; MCLACHLAN, G.; BIOARRAY, C. review article Microarray Data Analysis for Differential Expression: a Tutorial. **PRHSJ**, v. 28, n. 2, 2009.

TAM A, WADSWORTH S, DORSCHIED D, MAN SF, S. D. The airway epithelium: more than just a structural barrier. **Ther Adv Respir Dis**, v. 5, n. 4, p. 8, 2011.

TARCA, A. L.; ROMERO, R.; DRAGHICI, S. Analysis of microarray experiments of gene expression profiling. **American journal of obstetrics and gynecology**, v. 195, n. 2, p. 373–88, 2006.

TSENG, G. C.; GHOSH, D.; FEINGOLD, E. Comprehensive literature review and statistical considerations for microarray meta-analysis. **Nucleic acids research**, v. 40, n. 9, p. 3785–99, 2012.

TUSHER, V. G.; TIBSHIRANI, R.; CHU, G. Significance analysis of microarrays applied to the ionizing radiation response. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 9, p. 5116–21, 2001.

VANDERBILT, J. N.; MAGER, E. M.; ALLEN, L.; et al. CXC chemokines and their receptors are expressed in type II cells and upregulated following lung injury. **American journal of respiratory cell and molecular biology**, v. 29, n. 6, p. 661–8, 2003. American Thoracic Society.

VERDUCCI, J. S.; MELFI, V. F.; LIN, S.; et al. Microarray analysis of gene expression : considerations in data mining and statistical treatment. **Physiol Genomics 25:**, p. 355–363, 2006.

YAFFE, M. B. Beyond microarrays. **Science signaling**, v. 1, n. 51, p. eg11, 2008.

ZHANG, H.; YU, C.-Y.; SINGER, B. Cell and tumor classification using gene expression data: construction of forests. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 7, p. 4168–72, 2003.

ZHAO, M. Q.; AMIR, M. K.; RICE, W. R.; ENELOW, R. I. Type II pneumocyte-CD8+ T-cell interactions. Relationship between target cell cytotoxicity and activation. **American journal of respiratory cell and molecular biology**, v. 25, n. 3, p. 362–9, 2001.