Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística

Métodos de agrupamento aplicados a dados de microarranjo de DNA

Aluno: Caio Henrique de Sousa Lima
Orientadora: Samara Flamini Kiihl

Campinas
Setembro/2020

Conteúdo

Resumo

A tecnologia de microarranjo de DNA torna possível monitorar simultaneamente a expressão de milhares de genes para várias amostras. Extrair padrões genéticos a partir desses dados é um dos objetivos em genômica. No entanto, devido ao grande número de genes, interpretar resultados obtidos a partir desses experimentos é um desafio. Uma abordagem bastante utilizada tem sido empregar metodologias estatísticas de agrupamento de dados, para explorar e identificar estruturas e padrões interessantes nos dados. Neste trabalho, iremos estudar os algoritmos de agrupamento mais comuns e sua aplicação em conjunto de dados reais de microarranjo.

Palavras Chaves: Aprendizado Não Supervisionado, Agrupamento, K-means, Agrupamento Hierárquico, Microarranjo de DNA.

1 Introdução

Os avanços nas pesquisas trouxeram tanto conhecimento quanto novas tecnologias, que na área da biotecnologia resultaram em um enorme progresso. Pesquisas que antes só conseguiam analisar pequenas quantidades de genes por vez, agora geram um grande volume de dados pelo sequenciamento de genomas, que aliado com a quantidade de dados sobre expressão genética, tornou complexo a compreensão das finalidades dos genes nos organismos. Dessa forma, surge a tecnologia de Microarranjos, técnica que possibilitou a análise de grandes expressões genéticas, através de um experimento simples, rápido e eficaz.

A tecnologia de microarranjo consiste em chips de DNA, contendo amostras de RNA's que quando combinadas com reagentes químicos emitem uma luz fluorescente de acordo com a condição de interesse. Com as imagens geradas, é possível mensurar os dados biológicos (genes) cuja análise tem levado a várias descobertas. Enquanto o aprendizado de máquina é, segundo Mitchell, Thomas M. [1997], "a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência", sendo extremamente útil para a análise dos dados gerados pela tecnologia de microarranjo.

Deste modo, o presente trabalho possui o intuito de utilizar de técnicas de aprendizado de máquina não supervisionado, identificando possíveis grupos de observações de acordo com os genes dados. Para isto, será testado se é possível separar pacientes em subgrupos de forma que a diferença entre pacientes com choque séptico e sem choque séptico esteja bem delimitada.

2 Noções Biológicas

Todo organismo vivo armazena informações necessárias para coordenar o seu desenvolvimento e funcionamento através de instruções genéticas denominadas de genes, sendo um conjunto de genes um genoma. Essas informações se encontram no núcleo de células, onde os genomas estão codificados em sequência na forma de moléculas de ácidos nucléicos, divididos em dois nucleotídeos: ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA).

Um nucleotídeo é formado por três moléculas, as quais variam entre o DNA e o RNA, constituidos por bases nitrogenadas, radical fosfato (HPO4) proveniente do ácido fosfórico e pentose (açúcar, que no DNA é a desoxirribose e no RNA a ribose). O DNA é formado por uma fita, chamada de "dupla hélice", de pares de bases nitrogenadas e são quatro bases que a compõe: adenina (A), citosina (C), guanina (G) e timina (T). Para o RNA, formado por uma fita única, a timina é substituída pela uracila (U).

Além disso, o RNA possui uma variedade de estruturas secundárias, todas voltadas a produção de proteínas com funções específicas a serem efetuadas na célula. A produção de proteínas, fundamental para o funcionamento do organismo, ocorre através da transcrição do DNA, na qual as informações necessárias (expressões genéticas) são armazenadas no RNA mensageiro (mRNA), uma das estruturas citadas e, por conseguinte, é feito a tradução do mRNA, transformando-o em proteínas.

Comparado com o DNA, o mRNA é mais dinâmico e menos redundante [Gohlmann and Talloen, 2009]. Além do mRNA ser altamente complexo, é também suscetível a manipulações de acordo com as necessidades de célula, sendo constituído pela parte mais expressiva do DNA e com um foco maior nas atividades biológicas. Para analisar as informações das moléculas de mRNA é feito o processo de transcrição reversa, que as converte no DNA complementar correspondente.

Um processo bioquímico também importante é o da hibridização, que ocorre quando duas sequências complementares de nucleotídeos se combinam, ou seja, em que duas cadeias pareiam suas bases complementares (A com T e C com G), mediante a formação de pontes de hidrogênio.

3 A tecnologia de Microarranjo

Microarranjos de DNA, popularmente conhecidos como chips de DNA, são superfícies sólidas compostas por um vidro especial (ou lâminas), onde os segmentos de DNA ficam distribuídos ordenadamente, com uma posição única para cada componente do arranjo. Cada uma dessas posições é chamada de *spot* (ponto) e seus componentes contêm pequenas sequências de DNA conhecido.

Para a detecção dos arranjos são utilizados *probes* (sondas) formados por: oligonucleotídeos, pequenas moléculas de DNA que hibridizam apenas com um dos mRNA; e por sequências de DNA complementar (cDNA), produzidas a partir de um mRNA que se deseja observar e que representam apenas um gene do genoma cada. Deste modo, cada uma destas sondas está propensa a se ligar às sequências de nucleotídeos correspondentes, devido ao processo de hibridização.

A hibridização ocorre com mRNAs de uma solução biológica que foi exposta a uma situação de interesse, mRNAs esses que foram marcados previamente com uma substância fluorescente. Durante o processo de marcação fluorescente é feito a transcrição reversa dos mRNA em seus respectivos cDNA e em seguida, após a hibridização, as lâminas são expostas a raios lasers que incitam as substâncias fluorescentes a emitirem uma luz com intensidade que varia de acordo com a quantidade da expressão do gene. Essa intensidade da luz serve de medição, mesmo que indireta, da expressão de cada gene do genoma estudado.

4 Aprendizado de Máquina Não Supervisionado

Aprendizado de máquina é a área da IA responsável por desenvolver técnicas computacionais sobre o aprendizado e construir sistemas capazes de adquirir conhecimento de forma autônoma [Rezende, 2003].

Com o Aprendizado de Máquina (AM), os computadores são programados para aprenderem com a experiência passada, sendo assim capazes de obter conhecimento de forma autônoma. Seguindo uma linha oposta aos métodos clássicos de programação, que resolve problemas matematicamente ou algoritmicamente, a aprendizagem utiliza um conhecimento prévio, com os dados e os resultados provindos deles, para encontrar um algoritmo que possa ser utilizado na previsão de resultados futuros. O AM se divide em dois métodos: o aprendizado supervisionado e o não supervisionado.

O método não supervisionado diz respeito a exploração ou descrição de um conjunto de dados, os quais não possuem um atributo de saída/resposta, assim o objetivo passa a ser de agrupar ou encontrar regras de associação dos dados. Este método tende a ser mais desafiador, uma vez que as avaliações dos algoritmos são mais subjetivas, comparadas àquelas de métodos supervisionados. O presente trabalho visa utilizar métodos de aprendizado de máquina não supervisionado.

4.1 Análise de Componentes Principais

A Análise de Componentes Principais (ACP) é uma técnica de análise multivariada muito útil para o AM, em que se analisa as inter-relações de um grande número de variáveis para então condensar as informações em um conjunto menor de variáveis (componentes), perdendo o mínimo de informação e assim aumentando a eficiência de soluções e a facilidade de interpretação.

Segundo Mingoti [2007], o objetivo principal é explicar a estrutura da variância e da covariância de um vetor aleatório, composto de p-variáveis aleatórias, por meio da construção de combinações lineares das variáveis originais. Algebricamente, os componentes principais são combinações das variáveis originais $X_1, X_2, \ldots, X_n$ e, portanto, podem nos dizer as mesmas informações estando em menor dimensão.

4.2 Análise de Agrupamentos

Análise de Agrupamentos (Clustering) é um conjunto de métodos para encontrar subgrupos (clusters) em um banco de dados, agrupando pela similaridade de seus elementos, de forma que haja uma homogeneidade dentro dos grupos e uma heterogeneidade entre eles. Há diversos métodos para aplicar clusters em um conjunto de dados. Neste trabalho será abordado o K-means e a Clusterização Hierárquica, sendo K-means um método utilizado com um número previamente estipulado de clusters, enquanto para a Clusterização Hierárquica não se sabe a princípio quantos clusters são desejados.

4.2.1 K-means

O algoritmo K-means é uma técnica de clusterização na qual é necessário ser informado previamente a quantidade de k clusters que será formada, que, segundo e Vilma França e Nizam Omar [2003], tem como objetivo criar k clusters iniciais para então realizar realocações com interações baseadas em similaridade, de tal forma que melhore as posições dos centróides de cada grupo. Utilizar as distâncias médias entre pontos é um bom método de mostrar o grau de similaridade entre os pontos, sendo a distância Euclidiana a mais comumente usada, dada por:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

De forma resumida, o algoritmo atribui P pontos a K clusters e é calculado a distância Euclidiana entre os pontos, definindo a média dos vetores de pontos de cada grupo (centróides). Em seguida, os pontos são atribuídos aos seus grupos correspondentes de acordo com o centróide mais próximo dos mesmos. Esse processo de redistribuição de pontos entre os grupos de acordo com os centróides mais próximos continua até que todos estejam em seus respectivos grupos de forma correta, ou seja, que estejam mais próximos do centróide de seus grupos ao invés de outros.

4.2.2 Clusterização Hierárquica

Analogamente a outros métodos de agrupamento de clusters, a Clusterização Hierárquica agrupa os elementos de acordo com suas similaridades, porém essa abordagem não exige uma quantidade de clusters a ser formada e fornece mais de um tipo de partição dos dados, descritos de forma hierárquica. Um conjunto de dados pode conter uma série de clusters, esses, por sua vez, podem conter sub-clusters e assim por diante, sendo desta forma a composição hierárquica dos dados. Uma maneira frequentemente utilizada para a representação da hierarquia dos clusters é o dendograma, caracterizado por sua estrutura em forma de árvore, semelhante a uma árvore genealógica, que interliga um grupo de pais ao grupo de seus filhos.

Antes de inserir os dados no algoritmo, é recomendado que os dados sejam normalizados, para que as variáveis sejam comparáveis em uma mesma escala. Para decidir a qual subgrupos cada observação pertence, é calculado medidas de similaridade entre as observações, comumente sendo usada a distância Euclidiana. O método de aglomeração pode variar, em que cada um agrupa objetos em clusters baseado-se em diferentes formas de similaridades. O métodos geralmente recomendados são conhecidos como: *ward.D* e *ward.D2*, que buscam minimizar a variância total de cada cluster; e *complete*, em que a distância das observações entre clusters seja máxima.

5 Filtragem de Genes mais Expressivos

Tipicamente, o número de genes expressados em um microarranjo está na casa de dezenas de milhares e, diante disso, existem diversas propostas para a seleção de genes que sejam diferencialmente expressos entre as condições, considerando a média e a variância geral ou o coeficiente de variação, ou até mesmo a covariância e correlação entre os genes. Porém, em muitos contextos com dados de alta dimensão, é testado estatisticamente variável por variável se o comportamento se diferencia entre as condições estudados, como é estabelecido por Bourgon et al. [2010]. Cada variável (gene) é associada a uma hipótese nula de que seu comportamento não se difere entre as condições, que é rejeitada caso os dados indiquem que, pelo p-valor, que a hipótese é inconsistente. Assim, é possível comparar gene por gene entre as classes e selecionar aqueles que se mostrarem significativos.

Devido ao imenso número de hipóteses testadas, muitos falsos positivos, ou erros do Tipo I podem aparecer por acaso. Existem procedimentos para contornar o problema de múltiplos testes, em que o p-valor é ajustado e diminui a taxa de falsos positivos, como a taxa de descoberta falsa (FDR). Em contra partida, controlar a existência de falsos positivos diminui o poder de detectar verdadeiros positivos. Visto isso, é recomendado realizar uma filtragem prévia antes de aplicar múltiplos testes, com o intuito de reduzir o impacto que os ajustes têm sobre o poder dos testes ([Bourgon et al., 2010]). Assim, há duas estapas a serem seguidas: *(i)* identificar e remover conjuntos de genes aparentemente não informativos, que no presente trabalho será considerado o critério do coeficiente de variação (cv), que fornece a variação dos dados em relação a média em que quanto menor o cv, mais homogêneo são os dados; *(ii)* testar as hipóteses dos genes restantes serem iguais ou diferentes entre as classes.

6 Aplicação em Dados Reais

6.1 Sepse e Choque Séptico

A Sepse é uma condição em que uma infecção chega a corrente sanguínea e causa inflamações em outras partes do corpo. É considerada como uma resposta desregulada do sistema inflamatório e imunológico a uma invasão microbiana, tendo uma taxa de mortalidade de 15% a 25%, chegando a produzir lesões a órgãos e produzindo febre (ou hipotermia), taquicardia (aceleramento dos batimentos cardíacos), taquipneia (aumento da frequência respiratória) e mudanças de leucócitos no sangue [Hotchkiss et al., 2016]. O Choque Séptico é a evolução do quadro de Sepse, com sintomas de hiperlactatemia e hipotensão, surgindo a necessidade de introduzir agentes anti-hipotensivos no paciente, e desta forma, a mortalidade passa a ser de 30% a 50%. Existem outros tipos de choques que se manifestam de maneiras diferentes, mas que, segundo [Standl et al., 2018], levam ao mesmo estágio final de falência múltipla de órgãos como resultado do desequilíbrio entre a demanda e fornecimento de oxigênio.

Pacientes submetidos a procedimentos cirúrgicos ficam mais expostos a infecções, que podem atingir um quadro de Sepse ou de Choque Séptico. Porém, atualmente não há um padrão a ser seguido para diagnosticar a Sepse, e, desta forma, há um desafio em diferenciar um choque séptico e um choque não séptico após uma cirurgia, já que os pacientes de ambas as condições apresentam sintomas similares. Neste sentido, é necessário um diagnóstico rápido e preciso de choque séptico, para permitir um tratamento imediato desta condição.

Tendo em mente os desafios estabelecidos, o objetivo deste trabalho é avaliar expressões genéticas de pacientes pós-cirúrgicos com choque séptico e com choque não séptico, realizando técnicas de agrupamento com o fim de obter subgrupos que indiquem uma boa divisão.

6.2 Materiais e Métodos

O conjunto de dados utilizado foi fornecido pelo Martínez-Paz P. [2020], em domínio público na plataforma *Gene Expression Omnibus* (GSE) do National Center for Biotechnology Information (NCBI), como número de série GSE131761. Os dados consistem em uma amostra de 129 pacientes, dos quais 81 foram diagnosticado com Choque Séptico pós-cirúrgico, 33 pacientes com Choque Não Séptico pós-cirúrgico e 15 pacientes controle. Ao todo, a expressão genética coletada corresponde a 34127 genes. Os dados utilizados já estavam pré-processados.

Para a execução dos métodos de agrupamento foi utilizado o software *RStudio*. Os dados foram obtidos através das bibliotecas *GEOquery* e *Biobase*, disponibilizadas pelo *Biocunductor*. Os algoritmos dos métodos de agrupamento foram feitos com a biblioteca *stats*, e para a visualização dos resultados foi usado o pacote *kableExtra* para tabelas, e o pacote *factoExtra* para gráficos. Por fim, foi necessário o uso do pacote *genefilter* para a filtragem dos genes.

6.3 Resultados

6.3.1 Análise de Agrupamentos

Antes de aplicar o método do K-means, os dados foram normalizados para que o algoritmo não dependa da unidade de uma variável. Como o objetivo é encontrar os 3 grupos citados, foi indicado que o algoritmo de k-means dividisse de tal forma. O resultado do método pode ser visto na Tabela 1. Os pacientes controle foram alocados em um único subgrupo, mas os pacientes com Choque estão misturados nos subgrupos restantes, impossibilitando uma divisão adequada. Isso pode ser visto melhor com a Figura 1, que através de PCA, construiu duas componentes que possibilitam a visualização dos dados. É visto que há um confundimento entres os subgrupos 1 e 2, exemplificando as misturas entre pacientes com algum tipo de Choque.

Tabela 1: Resultados por K-means

	Cluster		
Diagnóstico	1	2	3
Choque Não Séptico	12	19	2
Choque Séptico	22	58	1
Controle	0	0	15

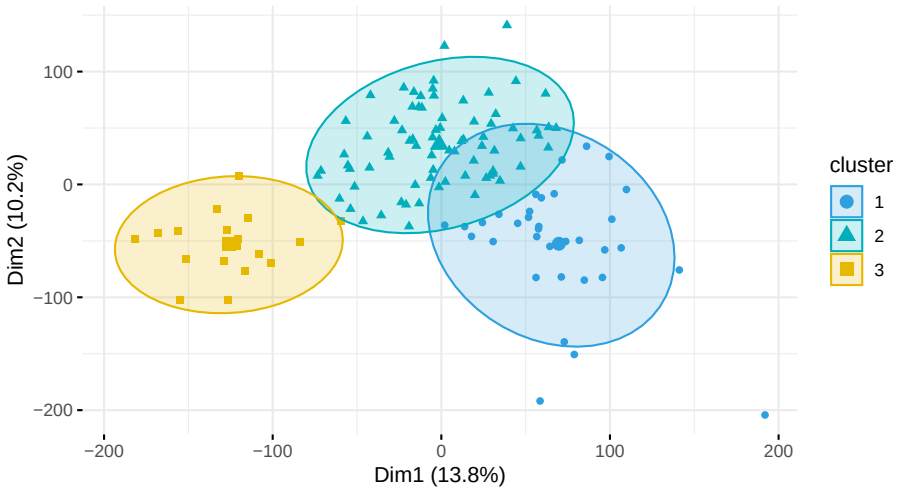

Figura 1: Cluster K-means

Para a Clusterização Hierárquica os dados também foram normalizados e então

obter a distância Euclidiana. Diferentes métodos de aglomeração foram considerados, mas o *ward.D* mostrou atingir um melhor resultado, que pode ser visto na Figura 2. A Tabela 2 mostra que o mesmo problema ocorreu de confundimento entre pacientes com algum Choque, apesar de indicar um cluster exclusivamente composto por pacientes controle.

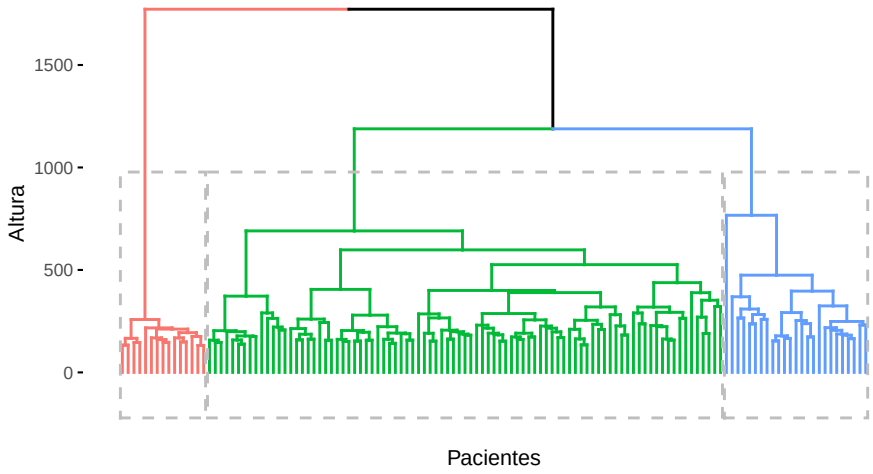

Figura 2: Dendograma com 3 Clusters

Tabela 2: Resultados por Cluster Hierárquico

	Cluster		
Diagnóstico	1	2	3
Choque Não Séptico	22	11	0
Choque Séptico	67	14	0
Controle	0	0	15

6.3.2 Agrupamento com as Componentes Principais

Foi visto que simplesmente aplicando algoritmos de agrupamentos não resultou em divisões satisfatórias. Levando em consideração que 34 mil genes é uma quantidade muito elevada, é interessante reduzir o número de variáveis através da Análise de Componentes Principais. Assim, será utilizado uma quantidade baixa de variáveis, mas que ainda explicam muito bem a variância da amostra. Os resultados do PCA podem ser vistos na Figura 3, em que as primeiras 20 componentes correspondem a cerca de 65% da variabilidade.

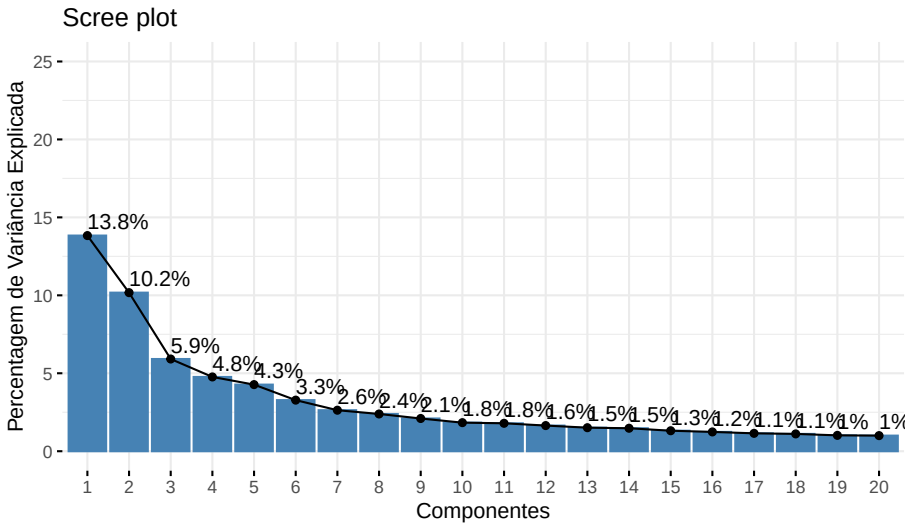

Figura 3: Autovalores das Componentes da PCA

É recomendado analisar o Scree plot para a escolha das componentes a serem utilizadas, observando onde a linha, que representa os autovalores para cada componente, para de decrescer precipitadamente e se nivela ([Brown, 2016]). Como apresentado pela Figura 4, seria adequado o uso de até 7 componentes.

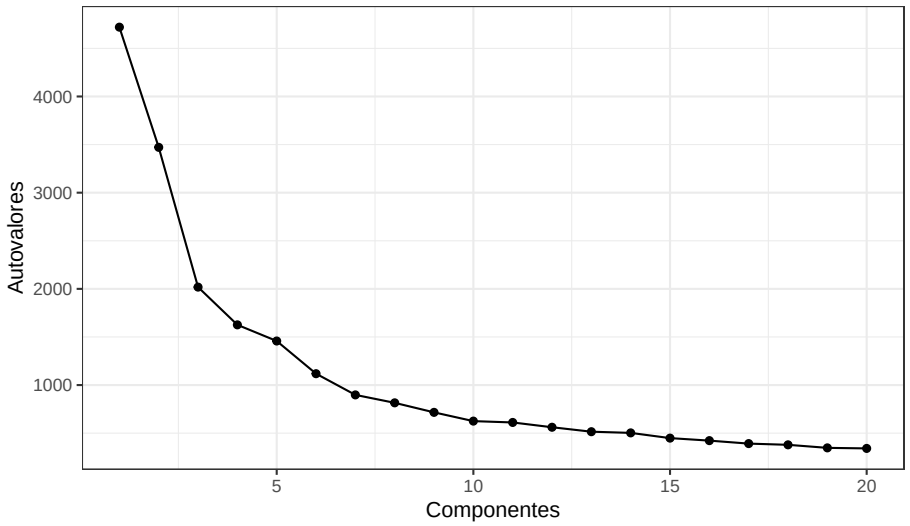

Figura 4: Scree Plot dos Autovetores das Componentes

Considerando o uso das 7 primeiras componentes, que correspondem a cerca de 45% da variância explicada, foram realizados novamente os métodos de agrupamentos com componentes esolhidas, obtendo os resultados presentes na Tabela 3. As conclusões foram muito similares com o uso de números diferentes de componentes. Desta forma, foi necessário realizar outras abordagens para obter os resultados desejados.

Tabela 3: Agrupamento com as Componentes Principais

	K-means			Hierárquico		
Diagnóstico	1	2	3	1	2	3
Choque Não Séptico	6	26	1	18	15	0
Choque Séptico	25	56	0	52	29	0
Controle	15	0	0	0	0	15

6.3.3 Filtrando Genes

É comum na literatura realizar um filtro de genes que seriam mais relevantes, que poderiam contribuir para a divisão desejada. Primeiramente, foi calculado o Coeficiente de Variação, que pode ser visto na Figura 5. Como não existe uma regra que estabeleça a partir de qual medida os genes são filtrados, foram testadas diversas possibilidades cortes a partir do quantil do conjunto de dados. Conjuntamente, para cada filtro testado, foram realizados testes de hipóteses para os genes restantes e que tiveram seus respectivos p-valores ajustado por FDR. Além disso, para cada filtro, foi testado diferentes níveis de significância para a escolha dos genes pelos testes de hipóteses.

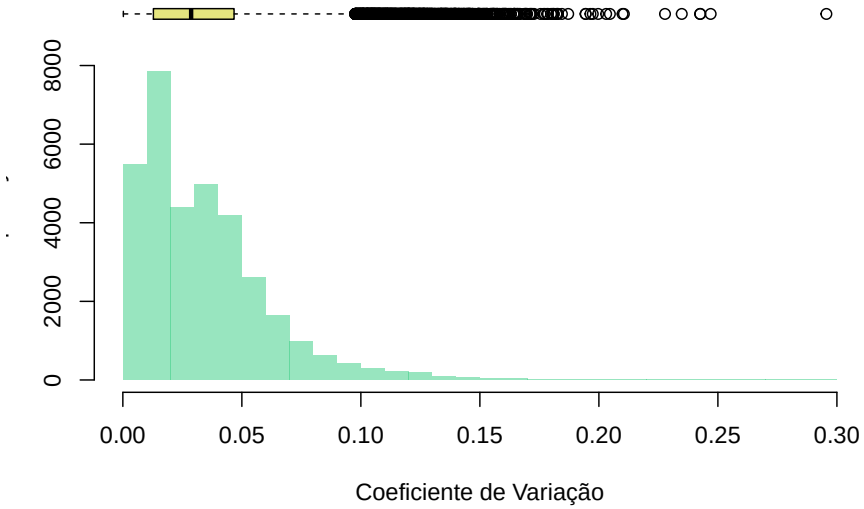

Figura 5: Histograma dos Coeficientes de Variação dos Genes

Considerando uma linha de corte de 24% para o quantil e a um nível de significância de 0.01, chegou-se em 154 genes finais, pois assim obteve-se o melhor resultado, usando Cluster Hierárquico, que pode ser visualizado na Figura 6. Ao invés de obter 3 clusters, verificou-se que ao separar um dos subgrupos o resultado ficou mais interessante. A Tabela 4 mostra que os subgrupos 1 e 2 são compostos majoritariamente por pacientes com Choque Séptico, e, analogamente, o subgrupo 3 é composto por paciente com Choque Não Séptico em sua maioria, restando o último subgrupo com apenas pacientes controle.

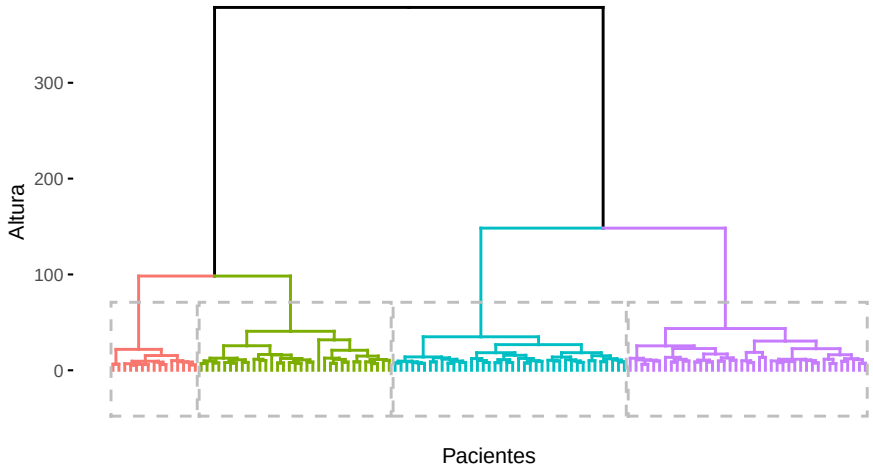

Figura 6: Dendograma com 4 Clusters após a filtragem de genes

Tabela 4: Resultados por Cluster Hierárquico com Genes Filtrados

	Cluster			
Diagnóstico	1	2	3	4
Choque Não Séptico	0	7	26	0
Choque Séptico	41	33	7	0
Controle	0	0	0	15

Os resultados obtidos acima são completamente reprodutíveis e a análise pode ser encontrada na plataforma *GitHub*, pelo link: https://github.com/CaioHSLima/Proj-IniciacaoCientifica-CNPq.

7 Discussão e Considerações Finais

O objetivo do presente trabalho foi compreender e aplicar métodos de aprendizado não supervisionado, voltado para o agrupamento de dados de microarranjo de DNA, com fim de encontrar subgrupos de interesse. Para isto, foram escolhidos dados com expressões genéticas de pessoas que tiveram Choque Séptico ou outro tipo de Choque, e assim, usou-se o k-means e o cluster hierárquico como métodos de agrupamentos na tentativa de separar os dois grupos pelas expressões genéticas.

Foi abordado o que é o microarranjo de DNA e sua importância, bem como alguns conceitos biológicos. Em conjunto com os métodos de aprendizado de máquina não supervisionado, pode-se agrupar pessoas em diferentes condições de acordo com suas expressões genética, o que poderia auxiliar em um diagnóstico mais preciso e rápido.

Ao aplicar ambos os métodos nas expressões genéticas, ocorreu um confundimento entre os grupos de pessoas com Choque Séptico e Choque Não Séptico, ou seja, nos clusters formados os grupos ficaram misturados, com exceção dos pacientes controle. Em seguida, reduziu-se o número de variáveis através do PCA, que indicou componentes principais que mais explicam a variabilidade dos dados. Porém, agrupar utilizando as componentes principais obteve resultados similares que anteriormente, com confundimento entre grupos.

A presença de tantos genes podem ter contribuído para a má divisão, visto que há muitos com baixa variância e acabam não se diferenciando entre os grupos de interesse. Tendo isso em mente, foi proposta uma filtragem dos genes, com o intuito de utilizar aqueles que mais se diferenciavam e contribuíssem em agrupamentos mais adequados. Primeiramente filtrou-se os genes que possuíam um baixo coeficiente de variação, que estavam abaixo da linha de corte de 24% do quantil do conjunto de dados. Em seguida, realizou-se teste de hipóteses gene por gene, para identificar os que são diferentes entre as categorias. Por ser uma grande quantidade de testes, ajustou-se os p-valores obtidos com FDR para reduzir a taxa de falsos positivos. A um nível de significância de 0.01, chegou-se em 154 genes finais. O método de cluster hierárquico se mostrou melhor, que agrupou os dados em 4 grupos: dois compostos majoritariamente por pessoas com Choque Séptico, um por Choque não Séptico em sua maioria, e o último apenas por pessoas controle.

Referências

Bourgon, R., R. Gentleman, and W. Huber
2010. Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences*, 107(21):9546–9551.

Brown, J. D.
2016. *Statistics Corner: Questions and Answers About Language Testing Statistics*. Createspace Independent Pub.

e Vilma França e Nizam Omar, E. P.
2003. A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 1(1):495–504.

Gohlmann, H. and W. Talloen
2009. *Gene Expression Studies Using Affymetrix Microarrays*. Chapman and Hall/CRC.

Hotchkiss, R. S., L. L. Moldawer, S. M. Opal, K. Reinhart, I. R. Turnbull, and J.-L. Vincent
2016. Sepsis and septic shock. *Nature Reviews Disease Primers*, 2(1).

Martínez-Paz P., Tamayo E., G.-M. E.
2020. Gene expression patterns in septic and non-septic shock postsurgical patients. Data accessible at NCBI GEO database, access by GSE131761.

Mingoti, S.
2007. *Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada*. Editora UFMG.

Mitchell, Thomas M.
1997. *Machine Learning*, 1 edition. USA: McGraw-Hill, Inc.

Rezende, S. O.
2003. *Sistemas Inteligentes: Fundamentos e Aplicações*. Barueri, SP: Editora Manole Ltda.

Standl, T., T. Annecke, I. Cascorbi, A. R. Heller, A. Sabashnikov, and W. Teske
2018. The nomenclature, definition and distinction of types of shock. *Deutsches Aerzteblatt Online*.