

Trabalho de conclusão do curso

Cliques consensuais em redes de co-expressão gênica em estudos de transcriptoma relacionados ao diabetes do tipo 2.

Aluno: Marcus Fabiano de Almeida Mendes
Orientador: Diego Bonatto
Co-Orientador: Ronnie Alves

Cliques consensuais em redes de co-expressão gênica em estudos de transcriptoma relacionados ao diabetes do tipo 2.

Marcus Fabiano A. Mendes¹, Diego Bonatto², Ronnie Alves³

¹ Graduando de ciências biológicas da Universidade Federal do Rio Grande do Sul

² Professor do departamento de biotecnologia, da Universidade Federal do Rio Grande do Sul

³ Pós-doutorando do departamento de informática, da Universidade Federal do Rio Grande do Sul

Diabetes tipo 2 (DT2) é a forma mais comum de diabetes. Milhões de pessoas são diagnosticadas com este tipo de diabetes anualmente, e esta taxa está em ascensão. Neste projeto, foi desenvolvido um modelo de meta-análise utilizando dados de transcriptomas de quatro estudos, previamente selecionados, para inferir potenciais módulos (cliques) consensuais em redes de co-expressão gênica relacionados ao DT2. Uma elucidação dos genes regulatórios desta doença e suas ligações irão prover um melhor entendimento da topologia destas redes gênicas e suas funções na progressão e tratamento desta doença. Neste trabalho foram utilizados fundamentos como mineração de grafos, e métricas para identificação de agrupamentos e associações para detectar os módulos gênicos mais significativos ao DT2. Para cada módulo detectado foi feita uma análise de enriquecimento funcional direcionando anotações relevantes para caracterização dos módulos de forma local (intra-estudo) e global (inter-estudo). Espera-se que os cliques consensuais possam delimitar potenciais marcadores gênicos para futuras observações em estudos clínicos. Os estudos de transcriptoma foram obtidos via os bancos de dados de expressão gênica *Gene Expression Omnibus* (GEO) e *Array Express*, e para modelagem computacional foi utilizada a linguagem R. Com esta técnica, foram encontrados nos módulos diversos genes relevantes para a DT2, comprovando assim a eficiência da metodologia.

INTRODUÇÃO

Nas últimas duas décadas temos assistido uma “avalanche” de dados gerados por experimentos bioquímicos. Consequência de avanços tecnológicos, e de estudos teóricos em biologia molecular que permitiram saltos significativos na compreensão do genoma de diversos organismos.

Dada a complexidade dos experimentos tanto ao nível biológico quanto para a tecnologia envolvida, é necessária a participação de diversas áreas do conhecimento, como ciência da computação, estatística e biologia molecular, para gerar, armazenar e analisar

grandes quantidades de dados. Emergiu então uma nova área de pesquisa chamada bioinformática.

A bioinformática surgiu para lidar com os dados resultantes de projetos relacionados a estudos *-ômicos*, como por exemplo genômica, proteômica, transcriptômica, entre outras [20]. Particularmente em transcriptômica, se busca um entendimento do perfil de expressão gênica de um indivíduo em situações e estados clínicos. De fato, desde o surgimento das tecnologias de microarranjos na década de 90 até o presente momento já foram publicados diversos trabalhos para a compreensão dos genes com funções importantes em diversas doenças, especialmente para diversos tipos de câncer. Além do câncer, a diabetes vem tomando um papel importante em pesquisas biomédicas dado o seu comportamento complexo e abrangência mundial.

A diabetes tipo 2 (DT2) é a forma mais comum de diabetes. É uma doença endêmica, que afeta mais de 170 milhões de pessoas no mundo, sendo que a tendência é que chegue a mais de 300 milhões de pessoas até 2030. É caracterizado por defeitos tanto na secreção como na ação da insulina, e está tipicamente associada com obesidade. [1] [3]

Sendo a diabetes uma doença multifatorial, ocorre a presença de muitas variáveis, tais como múltiplos genes, fatores metabólicos e ambientais [4]. Com isto, se torna difícil o desenvolvimentos de fármacos para esta doença, pois não se consegue apenas focar em apenas um gene, uma proteína, ou em um fator externo. A figura 1 nos mostra uma idéia de como funciona a tradução de um gene, desde o seu gene até sua função na célula, para se ter uma idéia da complexidade para o desenvolvimento de um fármaco [21].

A indústria farmacêutica investe milhões de dólares anuais na pesquisa de novos fármacos para a diabetes tipo 2. Várias técnicas utilizando a bioinformática, como a modelagem proteica, cristalografia de raios-X, análise de transcriptomas, entre outros, são muito utilizados para projeções de novos medicamentos [22].

O transcriptoma é conjunto completo de transcritos produzidos pelo genoma em qualquer momento. Grande parte de nossas células contém o mesmo genoma, independentemente do tipo de célula, estágio de desenvolvimento, ou condições ambientais. Por outro lado, o transcriptoma varia consideravelmente nestas diferentes circunstancias em função dos diferentes padrões de expressão gênica. Transcriptômica, também chamado de estudo do transcriptoma, é portanto uma maneira global de visualizar

e tentar entender os padrões de expressão gênica. Com o avanço da análise GWA (*genome-wide association*) tem transformado o potencial de pesquisa em diabetes tipos 2, já que consegue-se estudar variantes que antes eram relevadas, pois eram de difícil observação, mas que em fenótipo do tipo DT2 é de grande relevância. [2]

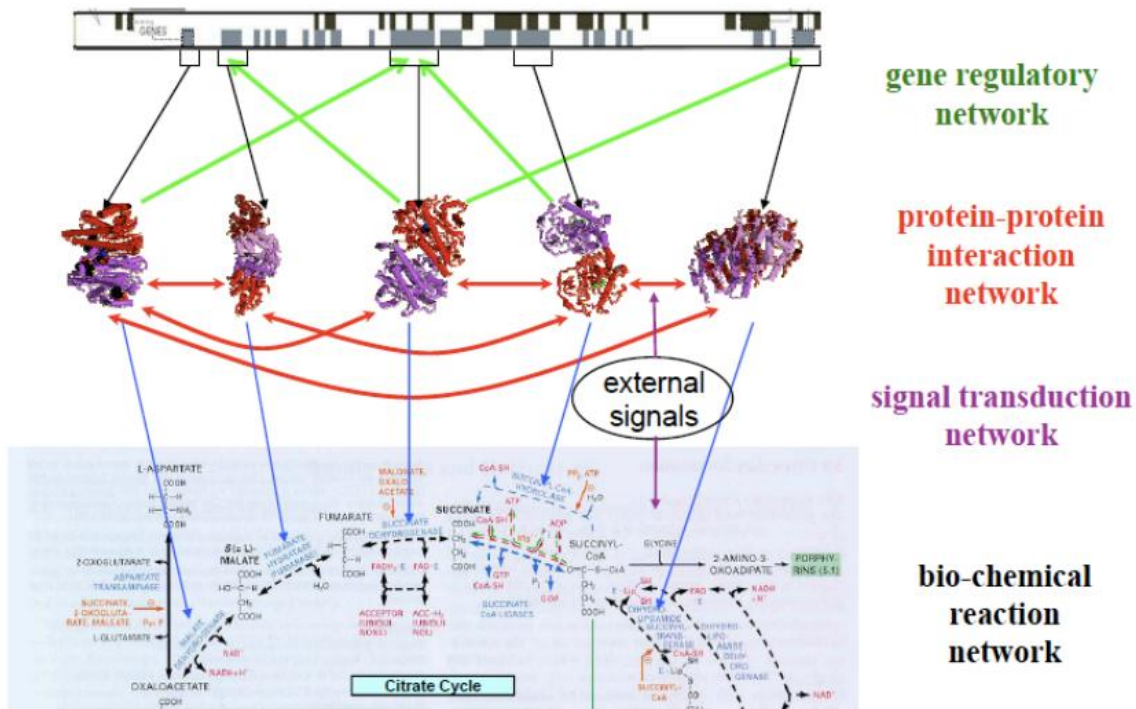


Figura 1: A complexidade das etapas, desde a transcrição dos genes, tradução em proteínas, interações entre si destas proteínas traduzidas e a participação em determinada rota metabólica dentro da célula.

No presente trabalho será proposta uma estratégia computacional para a detecção de potenciais módulos regulatórios em estudos relacionados ao DT2. Será realizado um estudo de meta análise com base em diversos estudos transcriptômicos, usando redes de co-expressão gênica [15] e métricas de biologia de sistemas para a identificação dos cliques (módulos) conservados em DT2. Espera-se que estes módulos possam delinear potenciais marcadores gênicos para uma melhor compreensão do DT2.

OBJETIVOS

O objetivo deste trabalho é encontrar um modelo capaz de detectar módulos de regulação conservados [14] em estudos transcriptômicos envolvendo indivíduos com

diabetes tipo 2, através da utilização de ferramentas computacionais, sendo neste caso utilizada a linguagem R [26] para o desenvolvimento do *pipeline*.

E, como consequência do desenvolvimento da estratégia computacional proposta, espera-se também que o *pipeline* possa ser utilizado para a detecção de módulos para outras doenças, os quais se utilizem dados transcriptômicos.

MATERIAIS E MÉTODOS

Foram selecionados quatro estudos transcriptômicos que utilizam microarranjos processados na plataforma Affymetrix que estão armazenados no banco de dados público do *Gene Expression Omnibus* (GEO), também conhecido como GEO, o qual é um dos maiores bancos de dados para genoma funcional, sendo possível hospedar resultados de microarranjos e dados de sequenciamento gênico nas plataformas atuais de RNAseq. De fato, estes estudos foram selecionados dentro de um conjunto de 30 outros estudos transcriptômicos de grande impacto no estudo da diabetes tipo 2. Todos os estudos foram devidamente selecionados por uma colega pesquisadora, especialista em DT2, do *Institute of Developmental Biology and Câncer* (IBDC) em Nice na França:

- Dois destes estudos utilizam a espécie *Rattus Norvegicus*.
 - “Diabetes mellitus and the progression of post-infarction genetic regulatory expression.” [9],
 - “Diabetes biomarker disease progression study in rat liver” [10]
- Dois para o modelo *Mus Musculus*,
 - “Interferon- γ -dependent regulatory circuits in immune inflammation highlighted in diabetes” [11],
 - “Beta cells (MIN6) treated with amylin at different times and doses and growth at different concentrations of glucose” [12].

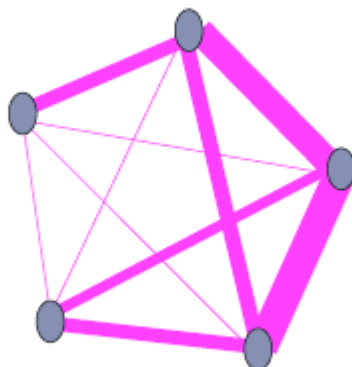
As duas espécies apresentadas acima são amplamente utilizadas em estudos envolvendo DT2, por serem de fácil manipulação e grande conhecimento do seu genoma funcional, tendo-se mapeado todos os seus genes e desenvolvido as anotações funcionais de grande parte do seu genoma.

Todos os quatro estudos utilizam a estratégia de microarranjos com objetivo de estudar a progressão do DT2 nos indivíduos selecionados em cada estudo. Com os

resultados dos microarranjos destes quatro estudos distintos, foram extraídos, separadamente, padrões de co-expressão gênica, com o qual se estruturou toda a base para a busca de cliques consensuais com os estudos de DT2 em questão.

O primeiro passo deste projeto, após a obtenção dos dados de microarranjos, foi um pré-processamento utilizando a função “RMA” do pacote “affy” presente no R, para a normalização e ajuste dos *cDNA chips*. Após a normalização, foi feita uma filtragem, onde apenas os genes “EXPRESSO” em 75% das condições foram selecionados para a construção da rede de co-expressão gênica de referência para cada estudo. Assim, um gene (sonda) é considerado “EXPRESSO” se estiver presente em pelo menos 75% de todas as amostras no estudo associado. Para esta seleção foi utilizada a função *mas5calls()* do pacote “affy” para realizar esse cálculo.

Rede Ponderada



- **Todos os genes estão conectados**
- **Cada ligação tem um valor diferente**

Figura 2: Uma ilustração da concepção de módulo. Cada ligação tem um valor que pode diferir em relações as outras, com isso facilita a implementação de um limiar.

A próxima etapa foi criar as conexões entre os genes co-expressados. Para isto, foi utilizado o conceito de *Rede Ponderada* [17], onde os genes estão todos conectados e cada conexão tem um valor diferente (potencializado), podendo-se criar *ranks* para estas conexões [23]. A rede é ponderada para permitir um ajuste de *fit* da rede de co-expressão, permitindo inferir um nível de conectividade que permita a rede ter um comportamento de livre escala (para a potenciação derivada do modelo de rede ponderada).

A implementação da *Rede Ponderada* ocorre da seguinte forma: através dos dados de expressão de microarranjos, se faz o cálculo da concordância da expressão gênica com a correlação de Pearson. Tradicionalmente é produzida uma matriz de correlação de Pearson, sendo dicotomizada para chegar a uma matriz de adjacência. Entretanto este processo é complexo, uma vez que existe um bias do qual seria o corte ideal para dicotomização da rede. Portanto a importância da inferência de *fit* segundo um modelo conhecido de crescimento em redes biológicas.

Por fim, a matriz de adjacência pode ser visualizada por um grafo, conforme podemos visualizar na figura 2, onde a ligação entre um gene e outro possuem um determinado valor, diferenciando assim a *Rede Ponderada* da *Não-Ponderada*. Todas as etapas podem ser visualizadas na Figura 3. Por fim, um *limiar* ótimo é calculado com base no conceito de livre escala, e este é aplicado como potenciação na rede de co-expressão.

Utilizando então o pacote WGCNA, foi escolhido um nível que representa bem uma conectividade média e topologia de livre escala da rede de referência para cada estudo. O *limiar p* aplicado a cada estudo foi GSE12389 ($p = 18$), GSE2253 ($p = 9$), GSE12639 ($p = 12$) e GSE13270 ($p = 12$). Após feito a *Rede Ponderada*, foram calculados os módulos de regulação, que são clusters de genes altamente conectados com um mesmo perfil de expressão.

Através destes padrões (módulos ou cliques), foram identificados agrupamentos, onde genes que tem uma forte correlação são agrupados em módulos [13]. Podemos ver na Figura 4 os diversos módulos detectados em cada estudo. Uma possibilidade interessante desta abordagem é o teste para inferir uma função para um módulo desconhecido tendo em conta outro módulo conhecido. Isto é feito a partir da análise de correlação do centróide (*eigenvector*) correspondente a cada módulo. Tal relação pode ser observada na Figura 4. Por um outro lado, a certeza de um função específica para o módulo só poderá ser alcançada através de uma análise de enriquecimento funcional para todos os módulos.

Depois da criação dos módulos, foram feitas mais algumas filtragens, tais como a ordenação de genes “hubs”, os quais conectam os módulos entre si, e também a criação de uma *pontuação*, isto é, a significância do módulo para aquela rede, onde com esta pontuação pode-se focar em determinados grupo de hub genes que são mais relevantes em

cada módulo. Todas as etapas deste processamento dos módulos foram calculados via programação em R, e através do pacote “WGCNA”.

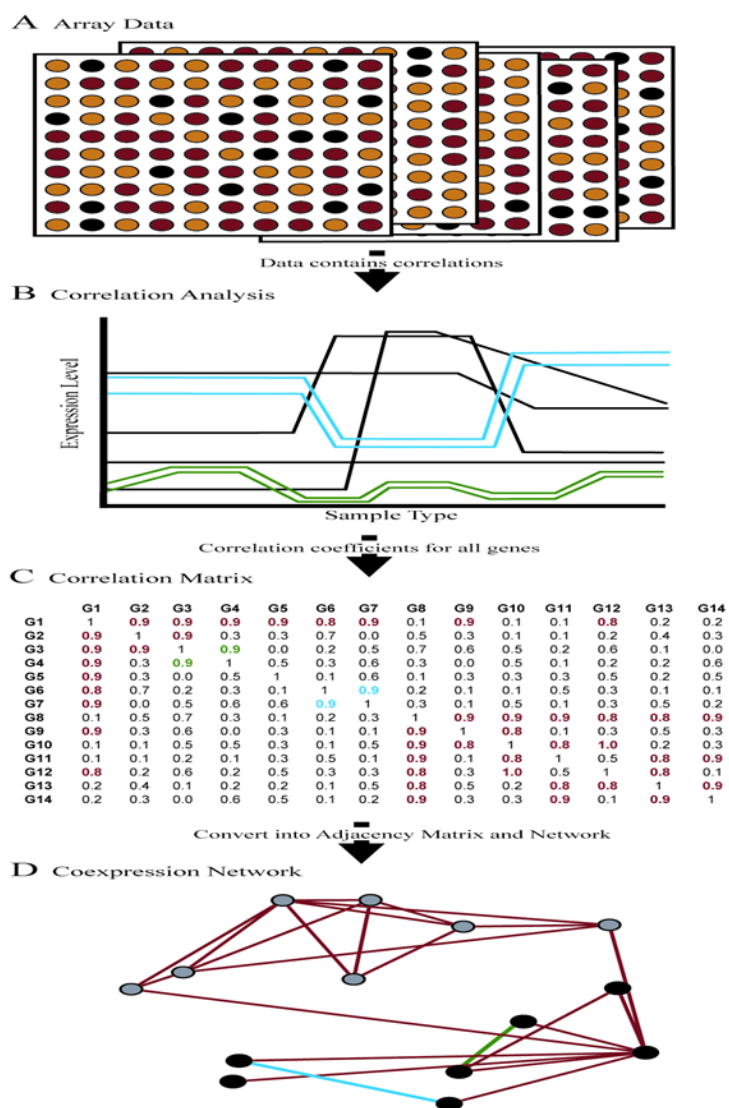


Figura 3: Passo a passo, da obtenção dos resultados do micro arranjo, extração dos perfis transcricionais, montagem da matriz de correlação e a rede de co-expressão genica.

Com os módulos criados e filtrados, o próximo objetivo foi obter o enriquecimento funcional destes módulos. O enriquecimento funcional é o processo em que se busca a anotação sobre a função daquele gene. O genoma nada mais é que um monte de bases nitrogenadas. Estas informações só fazem sentidos quando descobre em que proteína o determinado gene irar ser transcrito, e também qual é a função desta proteína.

Foram utilizados os bancos de dados do *Gene ontology* [16], onde este possui as anotações de milhares de produtos gênicos. Para isto, o primeiro passo foi utilizar o pacote “*biomaRt*”, o qual busca, utilizando o nome da sonda na plataforma referência do estudo *Affymetrix*, os genes associados, e posteriormente estes genes passaram por um enriquecimento funcional do módulo relacionado com o pacote “*GOstats*”.

Depois de resgatados as anotações para todos os módulos dos quatro estudos, foram utilizadas regras de associação para buscar anotações conservadas entre os estudos. Essa função busca associações (co-ocorrências de padrões frequentes) num determinado conjunto de dados, sendo normalmente definidos em 2 variáveis, o “*Transaction ID*” (TID) e o “*Itemset*” (Items) [24].

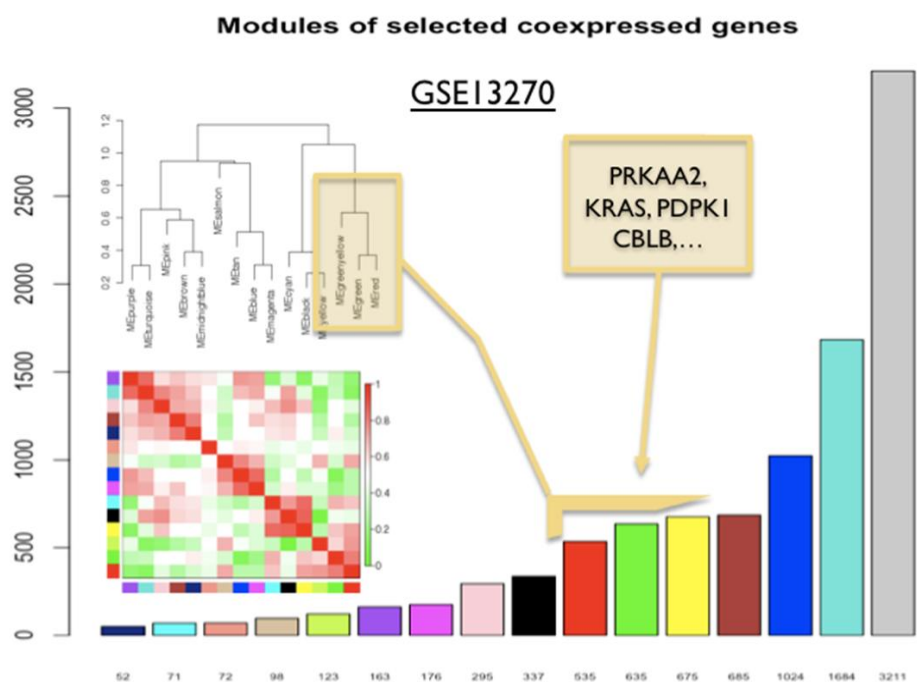


Figura 4: Exemplo mostrando os 16 módulos gênicos obtidos do estudo GSE13270, após a utilização dos pacotes RMA e WGCNA.

Para uma adequação ao projeto o TID pode ser visto como o módulo, e os Items os termos GO recuperados. Uma vez que existem três plataformas Affymetrix, uma etapa custosa seria a adequação e mapeamento de todos genes para um identificador único. A estratégia de utilizar a anotação evita o custo de integração, porém como veremos mais a

frente, é necessária uma engenharia reversa pra avaliar os genes mais significativos na função compartilhada (clique ou módulo consensual) por mais de um estudo.

Para a análise dos cliques consensuais, foram selecionadas as anotações mais significantes entres os módulos do mesmo estudo. Utilizando a estratégia de padrões frequentes, com um suporte de 0,15, foram selecionados em média de 3 a 5 anotações relevantes no mesmo estudo, isto é, anotações aparecem em vários módulos daquele mesmo estudo. Para uma diminuição na quantidade de dados gerados, as anotações foram classificadas conforme o termo GO de raiz, tendo 3 tipos: “*Biological Process*”, “*Molecular Function*” ou “*Cellular Component*”.

Após esta etapa, foram comparados as anotações mais frequentes entre os estudos. Nesta etapa, se utilizou como *limiar* o suporte de 0,5 e uma confiança de 0,95. As anotações mais significantes são guardadas e com elas realiza-se um estudo de engenharia reversa, onde buscamos a qual gene está anotação está ligada e com isso procuramos o impacto deste gene na literatura relacionada ao DT2.

Para a engenharia reversa, foi desenvolvido um programa com o objetivo de buscar os genes relacionado a aquela anotação relevante, em seguida foram mapeados para o gene similar em humanos e cruzados com uma lista atualizada de genes relacionados à diabetes tipo 2, encontrados na *Phenopedia*[25]. Assim, conseguimos uma lista dos genes relacionados a DT2 encontrados nos cliques consensuais, e o número de citações sobre aquele gene na literatura.

Todos esses procedimentos foram desenvolvidos via programação com o R, e através da utilização de pacotes especializado para a montagem do *pipeline*. Algumas informações sobre os módulos/cliques gerados podem ser buscados on-line no site do projeto: <http://www.inf.ufrgs.br/~ralves/mobio.html>

RESULTADOS

Na parte de pré-processamento e seleção de genes mais expressados, foi obtido uma redução significativa, conforme mostrado na tabela 1. Assim, amenizando o custo computacional para a criação das redes de co-expressão gênica.

Na parte da criação da *Rede Ponderada*, as sub-redes (cliques ou módulos) mais significativos foram agrupadas e, assim, vários módulos foram detectados nos estudos, sendo mostrado na Figura 2 um exemplo dos módulos, onde as arestas são os genes e estes estão conectados a outras arestas, de forma onde cada ligação tem um valor e estes valores podem ser diferentes entre si. A seguir, foram feitos mais algumas filtrações (ex., hub genes), para tentar reduzir ainda mais a quantidade de genes a serem analisados, como mostrado na Figura 5. A explicação sobre as reduções irar ser dada na seção de discussão.

Na análise de anotações, foram separados genes que conseguiram e que não conseguiram os resgates das anotações. Dos que foram resgatados, foi feita a análise de cliques consensuais, e através do algoritmo de *apriori*, foram obtidas as anotações mais frequentes intra-módulos e inter-módulos, respectivamente. Essas anotações foram utilizadas para buscar os genes relacionados, através de engenharia reversa, e depois uma avaliação dos genes que apresentam um histórico de estudo por estarem relacionadas a DT2. A Figura 6 mostra um fluxograma, mostrando os passos que foram seguidos e as funções utilizadas durante todo o processo para criação do *pipeline*.

Tabela 1: Quantidade de genes antes e depois do pré-processamento utilizando a função *RMA*

Study	Samples	Num.Of.Genes		Affymetrix Chip
		Before	After	
GSE12389	8	45101	5316	Mouse430_2
GSE2253	20	22690	11686	Mouse430A
GSE12639	12	31099	14998	Rat230_2
GSE13270	101	31099	13935	Rat230_2

Conclusão

Essa abordagem de análise de resultados em microarranjos é uma técnica relativamente nova, e que seu uso está cada em grande ascensão, já que cada vez mais está sendo utilizado o microarranjos.

- ▶ GSE12389[β 18]: start(5316) > end(1556) x 12 modules
- ▶ GSE2253[β 9]: start(11686) > end(5570) x 17 modules
- ▶ GSE12639[β 12]: start(14998) > end(5660) x 34 modules
- ▶ GSE13270[β 12]: start(13935) > end(9836) x 16 modules

Figura 5: Exemplo quantificando a redução de genes após as filtrações.

Com esta metodologia computacional que desenvolvemos, foi possível encontrar genes que são bem descritos na literatura e estão altamente relacionados com a DT2, assim como obesidade. Os genes relacionados abaixo são todos bem descritos como possíveis biomarcadores e estão em um mesmo módulo, isto é, estão interligados entre si.

Um exemplo da eficiência da técnica foi o resgate do gene NR3C1, o qual deu como significativo entre os estudos. Este gene já é bem conhecido na literatura, com vários artigos o descrevendo, onde o produto dele é um fator de transcrição e que está muito relacionado a diabetes tipo 2.

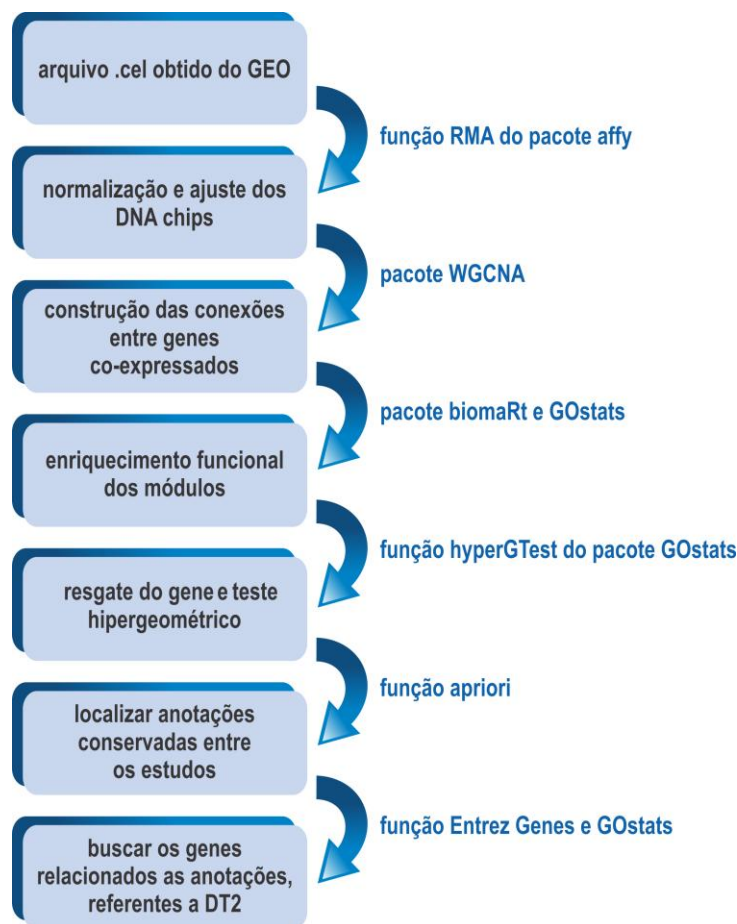


Figura 6: Todas as etapas do desenvolvimento do *pipeline*, indicando quais funções e pacotes do R foram utilizadas durante o processo de detecção dos cliques consensuais.

O gene “Adiponectin Receptor 1”, também conhecido como ADIPOR1, tem como função atuar na AMP cinase para melhorar a oxidação de ácidos graxos [5], e está relacionado com a melhora na sensibilidade a insulina, sendo assim, diabéticos possuem baixa atividades deste gene, comparados a pessoas normais [6].

O gene CDC123 está associado com a produção de insulina, os quais variantes deste gene têm uma baixa produção e uma resposta inata a este hormônio [7]. Um ponto interessante é que esta variante atinge mais asiáticos que outras etnias [8].

Um próximo passo que pode ser dado, utilizando essa mesma ideia de meta análise de estudos, seria buscar módulos consensuais entre a diabetes tipo 2 e a obesidade, já que ambos são doenças que tem uma correlação bem conhecida, como por exemplo a proteína SERPINE 1, o qual está relacionada tanto com a rota metabólica da diabetes tipo 2 como no desenvolvimento da obesidade [3], com a utilização da metodologia proposta, podem ser e confirmadas e descobertos novos marcadores gênico, auxiliando no desenvolvimentos de novos fármacos.

A metodologia é válida na obtenção de resultados de genes que podem estar conservados entre diversas espécies, com isso podendo obter marcadores para outras doenças como a diabetes tipo 2, Alzheimer [18], esquizofrenia [19] entre outras, além da vantagem de poder utilizar estudos em animais modelos, como o *Mus musculus* e, utilizando o nosso *pipeline*, conseguir genes homólogos para a nossa espécie.

Uma dificuldade do pipeline é a grande quantidade de dados gerados ao longo do processo, e a forma não compacta na detecção de cliques/módulos consensuais. Apenas 4 estudos foram selecionados para este projeto piloto. Vários ajustes foram necessários para obter a melhor seleção dos dados mais interessantes para análise. Utilizando uma maior quantidade de estudos, pode se fazer necessário uma maior quantidade de filtros, para tentar diminuir a quantidade dos dados. Portanto, duas observações são críticas para um estudo de meta-análise em larga escala (+10 estudos):

- Identificação de cliques/módulos, uma forma mais automática e compacta para identificar os módulos em redes de co-expressão gênica.
- Detecção de cliques consensuais, uma forma otimizada de cálculo de limiar para automação da seleção das anotações compartilhadas mais frequentes e de maior impacto funcional (biológico).

Referencias Bibliograficas

- 1 - Stumvoll, M. et al. (2005). **Type 2 diabetes: principles of pathogenesis and therapy.** *Lancet* 2005; 365:1333–1346
- 2 - Sladek, R. et al. (2007). **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature* 2007; 445:881–885
- 3 - Punit Kaur, Michael D. Reis, Glen R. Couchman, Samuel N. Forjuoh, John F. Greene Jr, and Alexander Asea. **SERPINE 1 Links Obesity and Diabetes: A Pilot Study.** *J Proteomics Bioinform.* 2010; 3(6): 191–199
- 4 - Narayan KM, Boyle JP, Thompson TJ, Sorensen SW, Williamson DF. **Lifetime risk for diabetes mellitus in the United States.** *JAMA.* 2003; 290:1884–1890.
- 5 - Yamauchi T, Kamon J, Waki H, Terauchi Y, Kubota N, Hara K, Mori Y, Ide T, Murakami K, Tsuboyama-Kasaoka N, Ezaki O, Akanuma Y, Gavrilova O, Vinson C, Reitman ML, Kagechika H, Shudo K, Yoda M, Nakano Y, Tobe K, Nagai R, Kimura S, Tomita M, Froguel P, Kadowaki T. **The fat-derived hormone adiponectin reverses insulin resistance associated with both lipotrophy and obesity.** *Nat Med* 2001; 7:941–946
- 6 - Tomas E, Tsao TS, Saha AK, Murrey HE, Zhang CC, Itani SI, Lodish HF, Ruderman NB. **Enhanced muscle fat oxidation and glucose transport by ACRP30 globular domain: acetyl-CoA carboxylase inhibition and AMPactivated protein kinase activation.** *Proc Natl Acad Sci U S A.* 2002; 99:16309–16313.
- 7 - Grarup N, Andersen G, Krarup NT, Albrechtsen A, Schmitz O, Jørgensen T, Borch-Johnsen K, Hansen T, Pedersen O. **Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes.** *Diabetes* 2008; 57:2534–2540
- 8 - Sanghera DK, Been L, Ortega L, Wander GS, Mehra NK, Aston CE, Mulvihill JJ, Ralhan S. **Testing the association of novel meta-analysis-derived diabetes risk genes with type II diabetes and related metabolic traits in Asian Indian Sikhs.** *J Hum Genet* 2009; 54:162–168
- 9 - Song GY, Wu YJ, Yang YJ, Li JJ et al. **The accelerated post-infarction progression of cardiac remodelling is associated with genetic changes in an untreated streptozotocin-induced diabetic rat model.** *Eur J Heart Fail,* 2009; 11(10):911-21.

- 10 - Richard R Almon, Debra C DuBois, William Lai, Bai Xue, Jing Nie, William J Jusko. **Gene expression analysis of hepatic roles in cause and development of diabetes in Goto-Kakizaki rats.** *The Journal of endocrinology.* 2009; 200(3): 331-46
- 11 - Boris Calderon, Anish Suri, Xiaou O. Pan, Jason C. Mills and Emil R. Unanue. **IFN- γ -Dependent Regulatory Circuits in Immune Inflammation Highlighted in Diabetes.** *The Journal of Immunology,* 2008; 181:6964 -6974
- 12 - Casas S, Gomis R, Gribble FM, Altirriba J et al. **Impairment of the ubiquitin-proteasome pathway is a downstream endoplasmic reticulum stress response induced by extracellular human islet amyloid polypeptide and contributes to pancreatic beta-cell apoptosis.** *Diabetes* 2007; 56(9):2284-94.
- 13 - Carter, S. L., Brechbuhler, C. M., Griffin, M. and Bond, A. T. **Gene co-expression network topology provides a framework for molecular characterization of cellular state.** *Bioinformatics* 2004; 20:2242-2250.
- 14 - Bergmann, S., Ihmels, J. and Barkai, N. **Similarities and Differences in Genome-Wide Expression Data of Six Organisms.** *PLoS Biol.* 2004; 2:E9.
- 15 - Langfelder P, Luo R, Oldham MC, Horvath S (2011) **Is My Network Module Preserved and Reproducible?.** *PLoS Comput Biol* 2011; 7(1): e1001057.
- 16- The Gene Ontology Consortium. **Gene Ontology: tool for the unification of biology.** *Nature Genet* 2000; 25:25-29.
- 17 – Tesson BM, Breitling R, Jansen RC, ***DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules.*** *BMC Bioinformatics.* 2010; 11:497
- 18 - Eric M. Blalock, James W. Geddes, Kuey Chu Chen, Nada M. Porter, William R. Markesbery, and Philip W. Landfield. **Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses.** *PNAS* 2004; 101(7):2173-2178
- 19 - Vawter MP, Crook JM, Hyde TM, Kleinman JE, Weinberger DR, Becker KG, Freed WJ. **Microarray analysis of gene expression in the prefrontal cortex in schizophrenia: a preliminary study.** *Schizophr Res.* 2002 ; 58(1):11-20.
- 20 - Junker B., Schreiber F., 2008. **Analysis of Biological Networks.** Wiley-Interscience; 1 edition, 368p.
- 21 - Volkhard H., 2008.**Principles of Computational Cell Biology: From Protein Complexes to Cellular Networks.** Wiley-VCH; 1 edition, 289p.

22 - Pevsner J., 2009. **Bioinformatics and Functional Genomics**. Wiley-Blackwell; 2 edition, 992p.

23 - Horvath S., Dong J., 2008. **Geometric Interpretation of Gene Coexpression Network Analysis**. Plos computational biology, 27p.

24 - ALVES, R. ; Rodriguez-Baena, D. S. ; Aguilar-Ruiz, J. S. **Gene Association Analysis: A Survey of Frequent Pattern Mining from Gene Expression Data**. BRIEFINGS IN BIOINFORMATICS 2010; 210-224.

25 - W. Yu, M. Clyne, M. J. Khoury and M. Gwinn. **Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations**. Bioinformatics 2010; 26(1):145-146.

26 - The R Foundation for Statistical Computing c/o Department of Statistics and Mathematics. **A Guidance Document for the Use of R in Regulated Clinical Trial Environments**. R: Regulatory Compliance and Validation Issues 2008; 1-23.