

RESSALVA

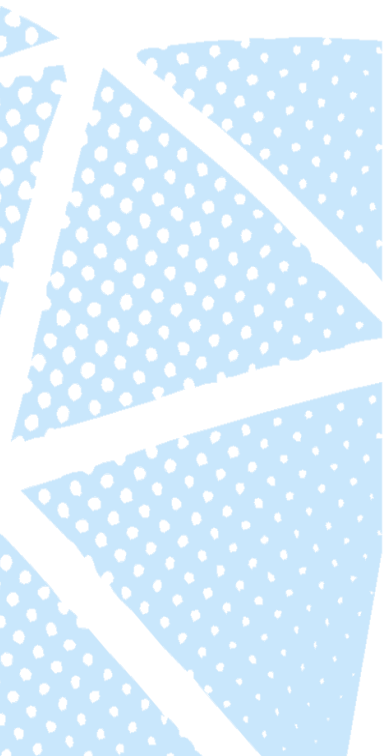
Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 30/10/2020.

Aprendizado de Máquina e Biologia de Sistemas aplicada ao estudo
da Síndrome de Microdeleção 22q11

CAMILA CRISTINA DE OLIVEIRA ALVES

BOTUCATU-SP

2019



Aprendizado de Máquina e Biologia de Sistemas aplicada ao estudo da Síndrome de Microdeleção 22q11

CAMILA CRISTINA DE OLIVEIRA ALVES

Orientadora: Profa. Dra. Lucilene Arilho Ribeiro Bicudo

Co-orientador: Prof. Dr. Guilherme Targino Valente

Dissertação apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Mestre no Programa de Pós-Graduação em Ciências Biológicas (Genética).

BOTUCATU-SP

2019

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Alves, Camila Cristina de Oliveira.

Aprendizado de máquina e biologia de sistemas aplicada
ao estudo da Síndrome de Microdeleção 22q11 / Camila
Cristina de Oliveira Alves. - Botucatu, 2019

Dissertação (mestrado) - Universidade Estadual Paulista
"Júlio de Mesquita Filho", Instituto de Biociências de
Botucatu

Orientador: Lucilene Arilho Ribeiro Bicudo

Coorientador: Guilherme Targino Valente

Capes: 20205007

1. Aprendizado de máquina. 2. Interação
proteína-proteína. 3. Cromossomos - Distúrbios. 4.
Cromossomos humanos par 22.

Palavras-chave: 22q11DS; Aprendizado de máquinas; Rede de
interação proteína-proteína; Síndrome DiGeorge.

*Aos meus pais Eliete e Neto, e a toda minha
família e amigos pelo auxílio e constante
incentivo.*

AGRADECIMENTOS

A minha orientadora Profa. Dra. Lucilene Arilho Ribeiro Bicudo pela oportunidade, pelo aprendizado profissional e pessoal e pela confiança depositada em mim.

Ao meu co-orientador Prof. Dr. Guilherme Targino Valente por me receber em seu laboratório e sua equipe. Obrigada pela colaboração, por todo ensinamento, incentivo, paciência e acolhimento. Sou muito grata por todo esforço e empenho durante esse tempo, tenho uma enorme admiração pela pessoa e profissional que você é.

Ao Me. Ivan Rodrigo Wolf por compartilhar seus conhecimentos e pelo auxílio desde a escrita do projeto, a colaboração em cada fase do trabalho até as sugestões e correções durante a elaboração da dissertação. Obrigada por me receber no laboratório, pela paciência e amizade.

Ao Dr. Bruno Faulin Gamba por todo o apoio e parceria. Muito obrigada pelos conselhos, orientações e toda ajuda na interpretação dos resultados e na elaboração dessa dissertação!

Aos meus amigos Lucas Farinazzo, Luiz Card, Lucas Lazari, Guilherme Luz, Eric Kawagoe e Giovanna Rock, que fizeram parte dessa jornada. Muito obrigada pelo acolhimento no laboratório, pelo apoio, orientações, risadas e por tantos momentos compartilhados. Guardarei boas lembranças.

A Camila Vaz Souza e Talita Aleixo que me acolheram em Botucatu da melhor forma e me deram todo o apoio durante o mestrado. Aprendi muito com vocês! Obrigada por todo o companheirismo.

A todos os outros amigos e colegas da Pós-graduação em Genética que contribuíram com esse trabalho tanto de forma direta quanto indireta.

Agradeço aos meus pais, Neto e Eliete por todo o apoio, conselhos, companheirismo, carinho, paciência e por estarem comigo em cada conquista. Vocês são meus exemplos de força e perseverança, amo vocês!

A CAPES pela bolsa de estudos concedida!

RESUMO

ALVES, CCO. **Aprendizado de Máquina e Biologia de Sistemas aplicada ao estudo da Síndrome de Microdeleção 22q11**. 2019, 117 p. Dissertação de mestrado – Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

A Síndrome de Microdeleção 22q11 (SD22q11), causada por uma deleção de aproximadamente 3Mb na região 22q11, apresenta uma frequência média de 1 em 4000 a 9800 nascidos vivos sendo considerada a síndrome de microdeleção mais frequente e a segunda causa mais comum de atraso no desenvolvimento e de doença congênita grave, após a síndrome de Down. De acordo com o tamanho e a localização da deleção, diferentes genes podem ser afetados e o principal gene considerado como responsável pelos sinais clássicos da síndrome é o *TBX1*. A SD22q11 caracteriza-se por um espectro fenotípico bastante amplo, com efeitos pleiotrópicos que resultam no acometimento de praticamente todos os órgãos e/ou sistemas, altamente variáveis com mais de 180 sinais clínicos já descritos, tanto físicos como comportamentais. Nesse trabalho aplicamos ferramentas de bioinformática com o intuito de descobrir padrões clínicos e sistêmicos da deleção 22q11, classificando casos síndrômicos em típicos e atípicos e estudando o impacto da deleção em redes de interação proteína-proteína (PPI). Para avaliação dos sinais clínicos que pudessem diferenciar pacientes síndrômicos foi aplicado uma metodologia baseada em aprendizado de máquina para classificar os casos em típico e atípico de acordo com os sinais clínicos através do algoritmo J48 (um algoritmo de árvore de decisão). As árvores de decisão selecionadas foram altamente precisas. Sinais clínicos como fissura oral, insuficiência velofaríngea, atraso no desenvolvimento de fala e linguagem, incapacidade de aprendizagem específica, anormalidade comportamental e atraso de crescimento foram indicativos para classificação dos casos. Já a avaliação do impacto da deleção da região 22q11 foi realizada através de estudos envolvendo redes biológicas. Assim, os genes codificadores de proteínas envolvidos na deleção foram removidos da rede PPI humana para simular a deleção. Diferentes análises topológicas foram utilizadas para comparar a rede global (GN) com a rede paciente (PN). Além disso foi verificado as comunidades de ambas as redes e realizou-se uma análise de enriquecimento de ontologia. Os resultados mostraram que não há diferença significativa ao comparar GN e PN, porém observamos que há diferença entre as comunidades dessas redes. Além disso, foi possível analisar diferentes genes que estavam presentes em regiões enriquecidas com termos ontológicos semelhantes. Dessa forma, podemos concluir que estudos envolvendo Aprendizado de Máquina e Redes Biológicas podem apontar novas hipóteses no estudo da SD22q11 além de ter potencial para esclarecer diversos aspectos de diferentes patologias que não são prontamente acessíveis pela biologia molecular convencional ou abordagens genéticas.

Palavras-chaves: 22q11SD; Síndrome DiGeorge; Aprendizado de máquinas; Rede de interação proteína-proteína.

ABSTRACT

ALVES, CCO. **Machine Learning and Systems Biology applied to the study of the 22q11 Microdeletion Syndrome**. 2019, 117 p. Master's degree in Science – Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

The 22q11 Microdeletion Syndrome (22q11DS), caused by a deletion of approximately 3Mb in the 22q11 region, has an average frequency of 1 in 4000 to 9800 live births and is considered the most frequent microdeletion syndrome and the second most common cause of developmental delay and severe congenital disease after Down syndrome. According to the size and location of the deletion, different genes may be affected and the main gene considered to be responsible for the classic signs of the syndrome is *TBX1*. 22q11DS is characterized by a very broad phenotypic spectrum with pleiotropic effects that result in the involvement of variable organs and/or systems with more than 180 clinical signs already described, both physical and behavioral. In this work, we applied bioinformatics tools to detect clinical and systemic patterns of 22q11 deletion, classifying typical and atypical syndromic cases, and studying the impact of deletion on protein-protein interaction (PPI) networks. To evaluate clinical signs that could differentiate syndromic patients, a machine-learning based methodology was used to classify the cases into typical and atypical according to the clinical signs through the algorithm J48 (a decision tree algorithm). The selected decision trees were highly accurate. Clinical signs such as oral fissure, velopharyngeal insufficiency, speech and language development delay, specific learning disability, behavioral abnormality and growth delay were indicative for case classification. The evaluation of the impact of the 22q11 region deletion was performed through studies involving biological networks. To achieve this goal, the protein coding genes involved in the deletion were removed from the human PPI network to mimic the deletion. Different topological analyzes were used to compare the global network (GN) with the patient network (PN). In addition, the communities of both networks were verified and an ontology enrichment analysis was performed. The results showed that there is no significant difference when comparing GN and PN, but we observed that there is difference between the communities of these networks. In addition, it was possible to analyze different genes that were present in regions enriched with similar ontological terms. Thus, we can conclude that studies involving Machine Learning and Biological Networks may point out new hypotheses in the study of 22q11DS and have the potential to clarify several aspects of different pathologies that are not readily accessible by conventional molecular biology or genetic approaches.

Keywords: 22q11DS; DiGeorge syndrome; Machine learning; Protein-protein interaction network.

SUMÁRIO

INTRODUÇÃO E REVISÃO DA LITERATURA	11
1. Estrutura do genoma	11
2. Síndrome de microdeleção 22q11	12
2.1. Nomenclatura	13
2.2. Frequência	14
2.3. Etiologia	15
2.4. Características clínicas	18
2.5. Diagnóstico	21
2.6. Tratamento	23
3. Aprendizado de máquinas	24
3.1. Dados de entrada ou Input	25
3.2. Árvores de decisão	26
3.3. Algoritmo de classificação J48	31
3.4. Medidas de desempenho	32
4. Rede de interação proteína-proteína	34
4.1. Biologia de sistemas	34
4.2. Teoria dos grafos	35
4.3. Propriedades gerais das redes	37
4.4. Modelos de redes biológicas	39
OBJETIVOS	42
1. Objetivo geral	42
2. Objetivos específicos	42
REFERÊNCIAS	44
CAPÍTULO 1	53
1. Introduction	54
2. Methods	55
2.1. Data collection	55
2.2. Data preparation	55
2.3. Decision-tree Modeling	56
3. Results	57
3.1 Data collation and preparation	57

3.2 Decision tree model	59
4. Discussion	61
5. Conclusion	65
6. Reference	65
7. Supplementary material	68
7.1. Supplementary table 1	68
7.2. Supplementary table 2	69
CAPÍTULO 2	74
1. Introduction	75
2. Methods	77
2.1. Data sources	77
2.2. Establish patient and global network	77
2.3. Network metrics	77
2.4. Neighbouring genes in the context of communities	78
2.5. Gene Ontology Enrichment Analysis	78
3. Results	78
3.1 Analysis of the PPI networks of each established group	78
3.2. Community context analysis of neighbouring proteins	79
3.3. Gene Ontology Enrichment Analysis	83
4. Discussion	87
5. Conclusion	92
6. References	93
7. Supplementary material	101
7.1 Supplementary material 1	101
7.2 Supplementary material 2	102
7.3. Supplementary material 3	103

***INTRODUÇÃO E
REVISÃO DA LITERATURA***

INTRODUÇÃO E REVISÃO DA LITERATURA

1. ESTRUTURA DO GENOMA

Distúrbios genômicos são doenças resultantes da perda ou ganho de material cromossômico. Os rearranjos cromossômicos em humanos são diversos, frequentes e geralmente resultam em anormalidades fenotípicas, defeitos de nascimento e letalidade embrionária (Shaffer and Lupski, 2000). As desordens genômicas mais comuns e delineadas são divididas em duas categorias principais: as que resultam da perda do número de cópias (deleções) e do ganho de número de cópias (duplicações) (Picchi, 1997). Além disso, alguns rearranjos cromossômicos ocorrem em segmentos menores, assim, deleções cromossômicas muito pequenas que não são detectadas pela microscopia, utilizando métodos citogenéticos tradicionais, são denominadas microdeleções (Shaffer and Lupski, 2000).

Ao longo do genoma, várias cópias de Regiões de repetição de pequeno número de cópias (LCRs - *Low Copy Repeats*) podem ser encontrados (Cardoso *et al.*, 2016). Os LCRs são sequências homologas, com comprimento maior ou igual a 1Kb, que foram gerados através de eventos de duplicação (Cardoso *et al.*, 2016; Harel and Lupski, 2018). Os LCRs com sequências de alta homologia podem promover uma recombinação homóloga não alélica (NAHR) (Shaffer and Lupski, 2000; Shaikh, Kurahashi and Emanuel, 2001; Burnside, 2015).

Dois tipos de NAHR podem ocorrer entre os LCRs: eventos Intercromossômicos entre LCRs parálogos ou eventos Intracromossômicos (Figura 1). Assim, LCRs levam a ocorrência de NAHR que resultam em variações no número de cópias (CNVs). O tamanho das CNVs é de mais de 50pb e pode resultar em 1,2% de diferença em relação ao genoma humano de referência (Zarrei *et al.*, 2015; Nowakowska, 2017). Além disso, um contínuo espectro de fenótipos é resultante de recorrentes CNVs (Zarrei *et al.*, 2015; Harel and Lupski, 2018).

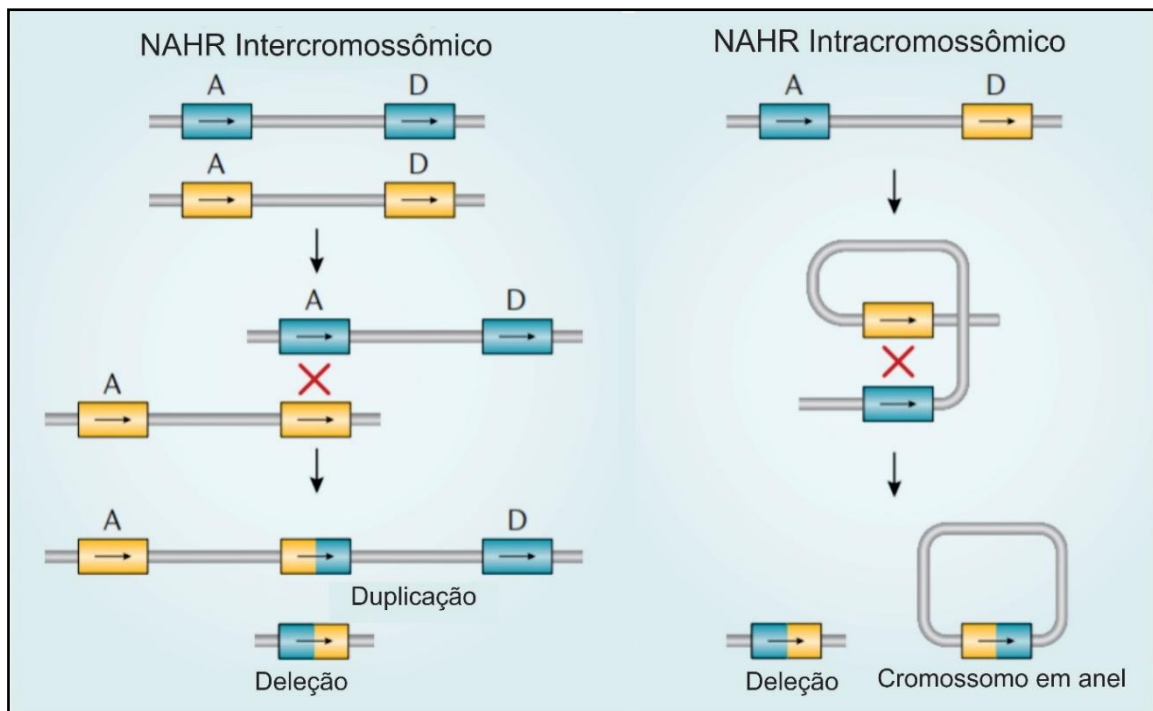


Figura 1. Diagrama dos dois tipos diferentes NAHR (Recombinação Homóloga Não Alélica) que podem ocorrer entre LCRs (Repetições de Pequeno Número de Cópias). No lado esquerdo, observa-se um rearranjo intercromossômico entre dois LCRs, indicados como A e D respectivamente. Esse processo resulta em uma duplicação ou deleção de genes nos gametas resultantes (O "X" mostra o cruzamento dos dois cromossomos). Já no lado direito, está esquematizado uma recombinação intracromossômica que ocorre devido cruzamento (indicado por "X") dentro de um alelo, resultando em um deleção ou um cromossomo em anel (não viável). Imagem adaptada de McDonald-McGinn *et al.*, 2015.

2. SÍNDROME DE MICRODELEÇÃO 22q11

Distúrbios genômicos, resultantes de recorrentes CNVs, foram descritos nos cromossomos 2, 7, 15, 16, 17 e 22 (Lupski, 1998; Shaffer and Lupski, 2000). Dentre estes, destaca-se a região q11 do cromossomo 22, uma área rica em genes que apresenta um conjunto de regiões de LCRs as quais predispõe à deleção ou duplicação dessa região (Guo *et al.*, 2011). Dessa forma, a deleção (que geralmente possui 3Mb) na região 1 banda 1 do braço longo (q) do cromossomo 22 é considerada a etiologia da Síndrome de Microdeleção 22q11 (SD22q11) (Figura 2).

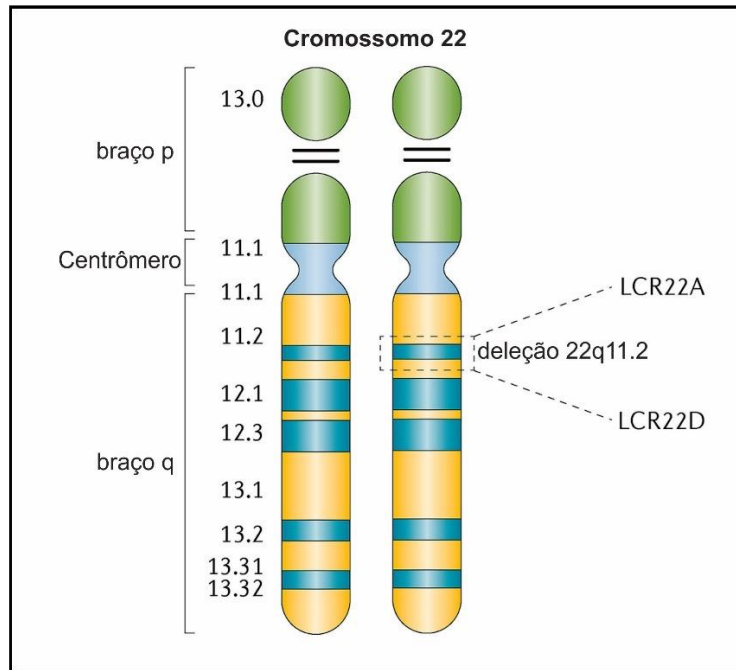


Figura 2. Representação do cromossomo 22 mostrando os braços curtos (p), braços longos (q) e o centrômero. A deleção na região 22q11 ocorre no braço longo de um dos dois cromossomos (representado pelas linhas tracejadas) devido a recombinação homóloga não alélica (NAHR) entre as regiões de repetição de pequeno número de cópias (LCRs) A e D (Modificado de McDonald-McGinn *et al.*, 2015).

6. References

- Albert, R., Jeong, H. and Barabási, A.-L. (2000) 'Error and attack tolerance of complex networks. *Nature*, 406: 378–482, 2000', *Nature*, 406(6794), pp. 378–382. doi: 10.1038/35019019.
- Ashburner, M. *et al.* (2000) 'Gene Ontology: tool for the unification of biology', *Nature Genetics*, 25(1), pp. 25–29. doi: 10.1038/75556.
- Ashley, E. A. *et al.* (2006) 'Network analysis of human in-stent restenosis', *Circulation*, 114(24), pp. 2644–2654. doi: 10.1161/CIRCULATIONAHA.106.637025.
- Baldini, A., Fulcoli, F. G. and Illingworth, E. (2017) *Tbx1: Transcriptional and Developmental Functions*. 1st edn, *Current Topics in Developmental Biology*. 1st edn. Elsevier Inc. doi: 10.1016/bs.ctdb.2016.08.002.
- Barabási, A. L. and Oltvai, Z. N. (2004a) 'Network biology: Understanding the cell's functional organization', *Nature Reviews Genetics*, 5(2), pp. 101–113. doi: 10.1038/nrg1272.
- Barabási, A. L. and Oltvai, Z. N. (2004b) 'Network biology: Understanding the cell's functional organization', *Nature Reviews Genetics*, 5(2), pp. 101–113. doi: 10.1038/nrg1272.
- Barabasi and Albert (1999) 'Emergence of scaling in random networks', *Science (New York, N.Y.)*, 286(5439), pp. 509–12. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10521342>.
- Baryshnikova, A. (2016a) 'Spatial Analysis of Functional Enrichment (SAFE) in Large

Biological Networks', *bioRxiv*, p. 094904. doi: 10.1101/094904.

Baryshnikova, A. (2016b) 'Systematic Functional Annotation and Visualization of Biological Networks', *Cell Systems*. The Author(s), 2(6), pp. 412–421. doi: 10.1016/j.cels.2016.04.014.

Bassett, A. S. *et al.* (2011) 'Practical guidelines for managing patients with 22q11.2 deletion syndrome', *Journal of Pediatrics*. Mosby, Inc., 159(2), p. 332–339.e1. doi: 10.1016/j.jpeds.2011.02.039.

Bates, G. P. *et al.* (2015) 'Huntington disease', *Nat Rev Dis Primers*, 1(April), pp. 1–21. doi: 10.1038/nrdp.2015.5.

Bengoa-Alonso, A. *et al.* (2016) 'Delineation of a recognizable phenotype for the recurrent LCR22-C to D/E atypical 22q11.2 deletion', *American Journal of Medical Genetics, Part A*, 170(6), pp. 1485–1494. doi: 10.1002/ajmg.a.37614.

Bertini, V. *et al.* (2017) 'Deletion Extents Are Not the Cause of Clinical Variability in 22q11.2 Deletion Syndrome : Does the Interaction between DGCR8 and miRNA-CNVs Play a Major Role ?', 8(May). doi: 10.3389/fgene.2017.00047.

Blom, U. M. *et al.* (2011) 'Prioritizing candidate disease genes by network-based boosting of genome-wide association data', *Genome Research*, 21(7), pp. 1109–1121. doi: 10.1101/gr.118992.110.

Bossi, G. *et al.* (2016) 'Failure to thrive as presentation in a patient with 22q11.2 microdeletion', *Italian Journal of Pediatrics*. Italian Journal of Pediatrics, 42(1), pp. 1–4. doi: 10.1186/s13052-016-0224-0.

Breckpot, J. *et al.* (2012) 'Congenital heart defects in a novel recurrent 22q11.2 deletion harboring the genes CRKL and MAPK1', *American Journal of Medical Genetics, Part A*, 158 A(3), pp. 574–580. doi: 10.1002/ajmg.a.35217.

Burnside, R. D. (2015) '22q11.21 deletion syndromes: A review of proximal, central, and distal deletions and their associated features', *Cytogenetic and Genome Research*, 146(2), pp. 89–99. doi: 10.1159/000438708.

Chan, S. Y. and Loscalzo, J. (2012) 'The Emerging Paradigm of Network Medicine in the Study of Human Disease', *Circulation Research*, 111(3), pp. 359–374. doi: 10.1161/CIRCRESAHA.111.258541.

Chang, Y. H. *et al.* (2019) 'TRMT2A is a novel cell cycle regulator that suppresses cell proliferation', *Biochemical and Biophysical Research Communications*. Elsevier Ltd, 508(2), pp. 410–415. doi: 10.1016/j.bbrc.2018.11.104.

Chen, J. *et al.* (2016) 'Identification of a Novel ENU-Induced Mutation in Mouse Tbx1 Linked to Human DiGeorge Syndrome', *Neural Plasticity*. Hindawi Publishing Corporation, 2016. doi: 10.1155/2016/5836143.

Clauset, A., Newman, M. E. J. and Moore, C. (2004) 'Finding community structure in very large networks', *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related*

Interdisciplinary Topics, 70(6), p. 6. doi: 10.1103/PhysRevE.70.066111.

Consortium, T. U. (2019) 'UniProt: a worldwide hub of protein knowledge', *Nucleic Acids Research*. Oxford University Press, 47(D1), pp. D506–D515. doi: 10.1093/nar/gky1049.

Csárdi, G. and Nepusz, T. (2006) 'The igraph software package for complex network research', *InterJournal Complex Systems*, 1695, pp. 1–9. doi: 10.3724/SP.J.1087.2009.02191.

Diez, D. *et al.* (2010) 'The use of network analyses for elucidating mechanisms in cardiovascular disease', *Molecular BioSystems*, 6(2), pp. 289–304. doi: 10.1039/b912078e.

Du, Y. *et al.* (2017) 'Increased cerebral expressions of MMPs, CLDN5, OCLN, ZO1 and AQP5 are associated with brain edema following fatal heat stroke', *Scientific Reports*, 7(1), pp. 1–10. doi: 10.1038/s41598-017-01923-w.

Dugoff, L., Mennuti, M. T. and McDonald-McGinn, D. M. (2017) 'The benefits and limitations of cell-free DNA screening for 22q11.2 deletion syndrome', *Prenatal Diagnosis*, 37(1), pp. 53–60. doi: 10.1002/pd.4864.

Edelmann, L. *et al.* (2001) 'Two functional copies of the DGCR6 gene are present on human chromosome 22q11 due to a duplication of an ancestral locus', *Genome Research*, 11(2), pp. 208–217. doi: 10.1101/gr.GR-1431R.

Edelmann, L., Pandita, R. K. and Morrow, B. E. (1999) 'Low-Copy Repeats Mediate the Common 3-Mb Deletion in Patients with Velo-cardio-facial Syndrome', *The American Journal of Human Genetics*, 64(4), pp. 1076–1086. doi: 10.1086/302343.

Friedel, C. C. and Zimmer, R. (2007) 'Influence of degree correlations on network structure and stability in protein-protein interaction networks', *BMC Bioinformatics*, 8, pp. 1–10. doi: 10.1186/1471-2105-8-297.

Fulcoli, F. G. *et al.* (2016) 'Rebalancing gene haploinsufficiency in vivo by targeting chromatin', *Nature Communications*. Nature Publishing Group, 7, pp. 1–11. doi: 10.1038/ncomms11688.

Gao, S., Li, X. and Amendt, B. A. (2013) 'Understanding the role of Tbx1 as a candidate gene for 22q11.2 deletion syndrome', *Current Allergy and Asthma Reports*, 13(6), pp. 613–621. doi: 10.1007/s11882-013-0384-6.

Gao, W. *et al.* (2015) 'DGCR6 at the proximal part of the DiGeorge critical region is involved in conotruncal heart defects', *Human Genome Variation*. Nature Publishing Group, 2(1), pp. 1–7. doi: 10.1038/hgv.2015.4.

Gene, T. and Consortium, O. (2019) 'The Gene Ontology Resource: 20 years and still GOing strong', *Nucleic acids research*. Oxford University Press, 47(D1), pp. D330–D338. doi: 10.1093/nar/gky1055.

Giuliani, S. *et al.* (2016) 'Coagulation Gene Expression Profiling in Infants With Necrotizing Enterocolitis', *Journal of Pediatric Gastroenterology and Nutrition*, 63(6), pp. e169–e175. doi: 10.1097/MPG.0000000000001215.

- Guris, D. L. *et al.* (2001) 'Mice lacking the homologue of the human 22q11.2 gene CRLK phenocopy neurocristopathies of DiGeorge syndrome', *Nature Genetics*, 27(3), pp. 293–298. doi: 10.1038/85855.
- Habel, A. *et al.* (2012) 'Syndrome-specific growth charts for 22q11.2 deletion syndrome in Caucasian children', *American Journal of Medical Genetics, Part A*, 158 A(11), pp. 2665–2671. doi: 10.1002/ajmg.a.35426.
- Hall, M. *et al.* (2009) 'The WEKA data mining software', *SIGKDD Explorations Newsletter*, 11(1), p. 10. doi: 10.1145/1656274.1656278.
- Han, J. D. J. (2008) 'Understanding biological functions through molecular networks', *Cell Research*, 18(2), pp. 224–237. doi: 10.1038/cr.2008.16.
- Hawkinson, J. E. *et al.* (2017) 'Potent Pyrimidine and Pyrrolopyrimidine Inhibitors of Testis-Specific Serine/Threonine Kinase 2 (TSSK2)', *ChemMedChem*, 12(22), pp. 1857–1865. doi: 10.1002/cmdc.201700503.
- Hicks, D. G. *et al.* (2010) 'The expression of TRMT2A, a novel cell cycle regulated protein, identifies a subset of breast cancer patients with HER2 over-expression that are at an increased risk of recurrence', *BMC Cancer*, 10. doi: 10.1186/1471-2407-10-108.
- Hooper, S. R. *et al.* (2012) 'Dysregulation of DGCR6 and DGCR6L: psychopathological outcomes in chromosome 22q11.2 deletion syndrome', *Translational Psychiatry*, 2(4), pp. e105–e105. doi: 10.1038/tp.2012.31.
- Jensen, M. K. *et al.* (2011) 'Protein Interaction-Based Genome-Wide Analysis of Incident Coronary Heart Disease', *Circulation: Cardiovascular Genetics*, 4(5), pp. 549–556. doi: 10.1161/CIRCGENETICS.111.960393.
- Jerome, L. A. and Papaioannou, V. E. (2001) 'DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1', *Nature Genetics*, 27(3), pp. 286–291. doi: 10.1038/85845.
- Ju, Z. R. *et al.* (2016) 'HIRA gene is lower expressed in the myocardium of patients with tetralogy of Fallot', *Chinese Medical Journal*, 129(20), pp. 2403–2408. doi: 10.4103/0366-6999.191745.
- Kiryluk, K. *et al.* (2017) 'Genetic Drivers of Kidney Defects in the DiGeorge Syndrome', *New England Journal of Medicine*, 376(8), pp. 742–754. doi: 10.1056/nejmoa1609009.
- Kolaczyk, E. D. and Csárdi, G. (2014) *Statistical Analysis of Network Data with R, International Statistical Review*. New York, NY: Springer New York (Use R!). doi: 10.1007/978-1-4939-0983-4.
- Li, C. *et al.* (2018) 'An analysis of plasma reveals proteins in the acute phase response pathway to be candidate diagnostic biomarkers for depression', *Psychiatry Research*, 272(November 2016), pp. 404–410. doi: 10.1016/j.psychres.2018.11.069.
- Lindsay, E. A. (2001) 'Chromosomal microdeletions: Dissecting DEL22Q11 syndrome', *Nature Reviews Genetics*, 2(11), pp. 858–868. doi: 10.1038/35098574.

- Liu, Y. *et al.* (2018) 'Infertility in a man with oligoasthenozoospermia associated with mosaic chromosome 22q11 deletion', *Molecular Genetics and Genomic Medicine*, 6(6), pp. 1249–1254. doi: 10.1002/mgg3.487.
- Lui, L. T. *et al.* (2017) 'Characterization of the Molecular Mechanisms Underlying the Chronic Phase of Stroke in a Cynomolgus Monkey Model of Induced Cerebral Ischemia', *Journal of Proteome Research*, 16(3), pp. 1150–1166. doi: 10.1021/acs.jproteome.6b00651.
- Ma, S. C. *et al.* (2017) 'Claudin-5 regulates blood-brain barrier permeability by modifying brain microvascular endothelial cell proliferation, migration, and adhesion to prevent lung cancer metastasis', *CNS Neuroscience and Therapeutics*, 23(12), pp. 947–960. doi: 10.1111/cns.12764.
- Ma, S. C. *et al.* (2018) 'CLDN5 affects lncRNAs acting as ceRNA dynamics contributing to regulating blood-brain barrier permeability in tumor brain metastasis', *Oncology Reports*, 39(3), pp. 1441–1453. doi: 10.3892/or.2018.6208.
- Macfarlan, T. *et al.* (2005) 'Human THAP7 is a chromatin-associated, histone tail-binding protein that represses transcription via recruitment of HDAC3 and nuclear hormone receptor corepressor', *Journal of Biological Chemistry*, 280(8), pp. 7346–7358. doi: 10.1074/jbc.M411675200.
- McDonald-McGinn, D. M. *et al.* (2015) '22Q11.2 Deletion Syndrome', *Nature Reviews Disease Primers*, 1(November). doi: 10.1038/nrdp.2015.71.
- Methylation, D. N. a *et al.* (2014) 'Biomarker Insights Gene as a Biomarker for Head and Neck Cancers', *Biomarker insights*, 9, pp. 53–60. doi: 10.4137/BMI.S16199.Received.
- Michaelovsky, E. *et al.* (2019) 'Risk gene-set and pathways in 22q11.2 deletion-related schizophrenia: a genealogical molecular approach', *Translational Psychiatry*. Springer US, 9(1). doi: 10.1038/s41398-018-0354-9.
- Mikelsaar, R., Lissitsina, J. and Bartsch, O. (2011) 'Small supernumerary marker chromosome (sSMC) derived from chromosome 22 in an infertile man with hypogonadotropic hypogonadism', *Journal of Applied Genetics*, 52(3), pp. 331–334. doi: 10.1007/s13353-011-0041-5.
- Moon, A. M. *et al.* (2006) 'Crkl Deficiency Disrupts Fgf8 Signaling in a Mouse Model of 22q11 Deletion Syndromes', *Developmental Cell*, 10(1), pp. 71–80. doi: 10.1016/j.devcel.2005.12.003.
- Morrow, B. E. *et al.* (2018) 'Molecular genetics of 22q11.2 deletion syndrome', *American Journal of Medical Genetics Part A*, 176(10), pp. 2070–2081. doi: 10.1002/ajmg.a.40504.
- Morrow, B. E. *et al.* (2018) 'Molecular genetics of 22q11.2 deletion syndrome', *American Journal of Medical Genetics, Part A*, 176(10), pp. 2070–2081. doi: 10.1002/ajmg.a.40504.
- Motta, M. *et al.* (2019) 'Dominant Noonan syndrome-causing LZTR1 mutations specifically affect the Kelch domain substrate-recognition surface and enhance RAS-MAPK signaling', *Human Molecular Genetics*, 28(6), pp. 1007–1022. doi: 10.1093/hmg/ddy412.

Nahorski, M. S. *et al.* (2015) 'A novel disorder reveals clathrin heavy chain-22 is essential for human pain and touch development', *Brain*, 138(8), pp. 2147–2160. doi: 10.1093/brain/awv149.

Ophoff, R. A. *et al.* (2016) 'Peripheral blood gene expression profiles linked to monoamine metabolite levels in cerebrospinal fluid', *Translational Psychiatry*. Nature Publishing Group, 6(12), pp. e983–e983. doi: 10.1038/tp.2016.245.

Ozcan, A. and Sahin, Y. (2017) 'DiGeorge Syndrome Associated with Azoospermia: First case in the literature', *Türk Üroloji Dergisi/Turkish Journal of Urology*, 43(3), pp. 390–392. doi: 10.5152/tud.2017.08555.

Panamonta, V. *et al.* (2016) 'Birth Prevalence of Chromosome 22q11.2 Deletion Syndrome: A Systematic Review of Population-Based Studies.', *Journal of the Medical Association of Thailand = Chotmaihet thangphaet*, 99 Suppl 5(18), pp. S187-93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29906080>.

Pržulj, N., Wigle, D. A. and Jurisica, I. (2004) 'Functional topology in a network of protein interactions', *Bioinformatics*, 20(3), pp. 340–348. doi: 10.1093/bioinformatics/btg415.

Racedo, S. E. *et al.* (2015) 'Mouse and human CRKL is dosage sensitive for cardiac outflow tract formation', *American Journal of Human Genetics*. The American Society of Human Genetics, 96(2), pp. 235–244. doi: 10.1016/j.ajhg.2014.12.025.

Raman, K. (2010) 'Construction and analysis of protein–protein interaction networks', *Automated Experimentation*, 2(1), p. 2. doi: 10.1186/1759-4499-2-2.

Rolland T. Tasan M., C. B. P. S. Z. Q. *et al.* (2014) 'A proteome-scale map of the human interactome network', *Cell*, 159(5), pp. 1213–1226. doi: 10.1016/j.cell.2014.10.050.A.

Rosa, R. F. M. *et al.* (2009) 'Síndrome de deleção 22q11.2: compreendendo o CATCH22', *Revista Paulista de Pediatria*, 27(2), pp. 211–220. doi: 10.1590/S0103-05822009000200015.

Rual, J. F. *et al.* (2005) 'Towards a proteome-scale map of the human protein-protein interaction network', *Nature*, 437(7062), pp. 1173–1178. doi: 10.1038/nature04209.

Ryan, A. K. *et al.* (1997) 'Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study.', *Journal of Medical Genetics*, 34(10), pp. 798–804. doi: 10.1136/jmg.34.10.798.

Santoro, M. L. *et al.* (2019) 'Downregulation of genes outside the deleted region in individuals with 22q11.2 deletion syndrome', *Human Genetics*. Springer Berlin Heidelberg, 138(1), pp. 93–103. doi: 10.1007/s00439-018-01967-6.

Scambler, P. J. (2000) 'The 22q11 deletion syndromes', *Human Molecular Genetics*, 9(16), pp. 2421–2426. doi: 10.1093/hmg/9.16.2421.

Scambler, P. J. (2010) '22q11 Deletion syndrome: A role for TBX1 in pharyngeal and cardiovascular development', *Pediatric Cardiology*, 31(3), pp. 378–390. doi: 10.1007/s00246-009-9613-0.

- Schmith, J. *et al.* (2005) ‘Damage, connectivity and essentiality in protein-protein interaction networks’, *Physica A: Statistical Mechanics and its Applications*, 349(3–4), pp. 675–684. doi: 10.1016/j.physa.2004.10.038.
- Sengupta, U. *et al.* (2009) ‘Expression-based network biology identifies alteration in key regulatory pathways of type 2 diabetes and associated risk/complications’, *PLoS ONE*, 4(12). doi: 10.1371/journal.pone.0008100.
- Shetty, J. *et al.* (2016) ‘Recombinant production of enzymatically active male contraceptive drug target hTSSK2 - Localization of the TSKS domain phosphorylated by TSSK2’, *Protein Expression and Purification*, 121(3), pp. 88–96. doi: 10.1016/j.pep.2016.01.009.
- Sisak, F. *et al.* (2018) ‘Protein expression in the liver and blood serum in chickens in response to Salmonella Enteritidis infection’, *Veterinary Immunology and Immunopathology*. Elsevier B.V., 205, pp. 10–16. doi: 10.1016/j.vetimm.2018.10.006.
- Soubeyran, P. *et al.* (2017) ‘Regulation of NUB1 Activity through Non-Proteolytic Mdm2-Mediated Ubiquitination’, *Plos One*, 12(1), p. e0169988. doi: 10.1371/journal.pone.0169988.
- Stark, C. (2005) ‘BioGRID: a general repository for interaction datasets’, *Nucleic Acids Research*, 34(90001), pp. D535–D539. doi: 10.1093/nar/gkj109.
- Stoller, J. Z. *et al.* (2011) ‘Ash2l interacts with Tbx1 and is required during early embryogenesis’, *Exp Biol Med*, 235(5), pp. 569–576. doi: 10.1258/ebm.2010.009318.
- Sullivan, K. E. (2019) ‘Chromosome 22q11.2 deletion syndrome and DiGeorge syndrome’, *Immunological Reviews*, 287(1), pp. 186–201. doi: 10.1111/imr.12701.
- Sullivan, K. E. (2019) ‘Chromosome 22q11.2 deletion syndrome and DiGeorge syndrome’, *Immunological Reviews*, 287(1), pp. 186–201. doi: 10.1111/imr.12701.
- Syring, I. *et al.* (2018) ‘The knockdown of the Mediator complex subunit MED15 restrains urothelial bladder cancer cells’ malignancy’, *Oncology Letters*, 16(3), pp. 3013–3021. doi: 10.3892/ol.2018.9014.
- Vaccari, T. *et al.* (2019) ‘A genetic model of CEDNIK syndrome in zebrafish highlights the role of the SNARE protein Snap29 in neuromotor and epidermal development’, *Scientific Reports*, 9(1), pp. 1–13. doi: 10.1038/s41598-018-37780-4.
- Vago, P. *et al.* (2012) ‘An atypical 0.8 Mb inherited duplication of 22q11.2 associated with psychomotor impairment’, *European Journal of Medical Genetics*. Elsevier Masson SAS, 55(11), pp. 650–655. doi: 10.1016/j.ejmg.2012.06.014.
- Veres, D. V. *et al.* (2015) ‘ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis’, *Nucleic Acids Research*, 43(D1), pp. D485–D493. doi: 10.1093/nar/gku1007.
- Wang, X., Gulbahce, N. and Yu, H. (2011) ‘Network-based methods for human disease gene prediction’, *Briefings in Functional Genomics*, 10(5), pp. 280–293. doi: 10.1093/bfpg/blr024.

Weiten, R. *et al.* (2018) 'The Mediator complex subunit MED15, a promoter of tumour progression and metastatic spread in renal cell carcinoma', *Cancer Biomarkers*, 21(4), pp. 839–847. doi: 10.3233/CBM-170757.

Witten, I. H. , Frank, E., & Hall, M. A. (2011) *Data Mining : Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers. doi: 10.1016/C2009-0-19715-5.

Woodward, K. J. *et al.* (2019) 'Atypical nested 22q11.2 duplications between LCR22B and LCR22D are associated with neurodevelopmental phenotypes including autism spectrum disorder with incomplete penetrance', *Molecular Genetics and Genomic Medicine*, (September 2018), pp. 1–17. doi: 10.1002/mgg3.507.

Yamagishi, H. and Srivastava, D. (2003) 'Unraveling the genetic and developmental mysteries of 22q11 deletion syndrome', *Trends in Molecular Medicine*, 9(9), pp. 383–389. doi: 10.1016/S1471-4914(03)00141-2.

Yang, J.-H. *et al.* (2016) 'Differential regulation of the histone chaperone HIRA during muscle cell differentiation by a phosphorylation switch', *Experimental & Molecular Medicine*. Nature Publishing Group, 48(8), pp. e252–e252. doi: 10.1038/emm.2016.68.

Zeitz, M. J. *et al.* (2013) 'Implications of COMT long-range interactions on the phenotypic variability of 22q11. 2 deletion syndrome', *Nucleus*, 4(6), pp. 6–7. doi: 10.4161/nucl.27364.

Zhu, L. *et al.* (2016) 'Analysis of the gene expression profile in response to human epididymis protein 4 in epithelial ovarian cancer cells', *Oncology Reports*, 36(3), pp. 1592–1604. doi: 10.3892/or.2016.4926.