

Universidade de São Paulo
Tatiana Ferreira de Almeida

Diagnóstico molecular do transtorno do espectro autista através do sequenciamento completo de exoma

**Molecular diagnosis of autism spectrum
disorder through whole exome
sequencing**

São Paulo
2018

Universidade de São Paulo
Tatiana Ferreira de Almeida

Diagnóstico molecular do transtorno do espectro autista através do sequenciamento completo de exoma

Molecular diagnosis of autism spectrum disorder through whole exome sequencing

Tese apresentada ao Instituto
de Biociências da
Universidade de São Paulo,
para a obtenção de Título de
Doutor em Ciências, na Área
de Biologia/Genética.

Orientador(a): Maria Rita dos
Santos Passos-Bueno

São Paulo
2018

Almeida, Tatiana Ferreira de
Diagnóstico molecular do transtorno do espectro
autista através do sequenciamento completo de exoma /
Tatiana Ferreira de Almeida; orientadora Maria Rita dos
Santos Passos-Bueno. -- São Paulo, 2018.

201 (volume 1 e 2) f.

Tese (Doutorado) - Instituto de Biociências da
Universidade de São Paulo, Departamento de Genética e
Biologia Evolutiva.

1. Transtorno do espectro autista. 2. Sequenciamento
completo de exoma. 3. Modelos multivariados. 4. Software
de análise. I. Passos-Bueno, Maria Rita dos Santos ,
orient. II. Título.

Comissão Julgadora

Prof(a). Dr(a)

Prof(a). Dr(a)

Prof(a). Dr(a)

Prof(a). Dr(a)

Profa. Dra Maria Rita dos Santos Passos-Bueno
Orientadora

Dedico esta tese à ciência, e a todos os seus amantes.

*O, reason not the need! Our basest beggars
Are in the poorest thing superfluous.
Allow not nature more than nature needs,
Man's life is cheap as beast's. Thou art a lady:
If only to go warm were gorgeous,
Why, nature needs not what thou gorgeous wear'st
Which scarcely keeps thee warm. But, for true need-
You heavens, give me that patience, patience I need!
You see me here, you gods, a poor old man,
As full of grief as age; wretched in both."*

King Lear, William Shakespeare, 1608

Agradecimentos

Agradeço em primeiro lugar ao meu grande amor Danilo, sem ele, há 11 anos atrás essa jornada não teria começado, a vida não teria me surpreendido e eu estaria de volta em casa, com todos os peixinhos do mar, e infeliz pelo resto da minha vida.

Agradeço minha família, principalmente minha mãe Flávia e meu pai Airton por sacrificarem com mínimos protestos, as melhores horas que passaríamos juntos para que eu cumprisse meu trabalho. Aos compadres e amigos Bruna e Wagniel e Cíntia e Fernando que não somente me apoiaram neste caminho, mas colocaram na minha vida os maiores presentes que eu poderia sonhar, Alice e Juliana, que mesmo em tão pouco tempo de vida e com tantas ausências ainda sorriem ao me ver.

À minha orientadora Maria Rita Santos Passos-Bueno que com um raciocínio feroz e implacável não mediou esforços para me ajudar a percorrer o caminho mais desafiador que eu já enfrentei. Ao professor Paulo Otto que solidificou meu amor pela matemática e me ajudou a pensar em probabilidades.

Agradeço ao Renato Puga que pegou na minha mão e me ensinou a programar (em awk!!), obrigada por ser o ursinho toda vez que eu errava uma vírgula. Jovem, “Até logo, e obrigada pelos pães!”

Aos amigos de laboratório, o Lab200, que me ensinaram a força da cooperação, a pensar em todos. Eu nunca tinha visto pessoas trabalharem com tamanha vontade de fazer o melhor por todos e espero ter aprendido com vocês e levar para a vida essa tão valiosa lição. Gerson, Lucas e Luciano obrigada pela constante inspiração. Obrigada meninas do GeneEvol, Carol, Ágatha, Camila e Clarice, e mais tarde Duda, por me deixarem entrar na vida de vocês e sempre lembrarem de me incluir nos encontros mesmo eu quase nunca podendo ir.

Aos amigos do DLE, principalmente Dayse, Karina, Cíntia e Mi, que sempre me viam correndo de um lado para o outro com as “coisas do doutorado” e foram compreensivos e pacientes com a vida dupla que eu levava. Ao Gustavo Campana que me mostrou como deve ser um exame para o diagnóstico laboratorial e me encantou com esse mundo da patologia clínica.

Às amadas bailarinas, obrigada por fazerem da minha vida melhor e mais divertida, nada disso teria sido possível se não fosse minha terapia em roupas de ballet. Lili, obrigada por me ensinar tanto na vida e no palco. Paty, a bola é sua agora!!

Table of Contents

1 Introduction	5
1.1 Clinical and epidemiological aspects of autism spectrum disorder	5
1.2 Molecular aspects of autism spectrum disorder	6
1.3 Next Generation Sequencing and the Diagnosis of Rare Diseases	10
1.4 Next Generation Sequencing and the ASD Diagnosis	11
2 Objectives	13
3 Material and Methods	14
3.1 Subjects	15
3.1.1 Autism Spectrum Disorder Subjects	15
3.1.2 Control group	15
3.1.3 Kinship and ancestry analysis	19
3.2 Gene lists	21
3.2.1 Genes with critical exons	21
3.2.2 Chromatin modifier genes	22
3.2.3 1527 genes related to ASD	22
3.2.4 FMRP targets	22
3.2.5 208 genes related to ASD	23
3.2.6 GABA pathway	23
3.2.7 Hypermutable Genes	23
3.2.8 mTOR and RAS/MAPK Pathways	24
3.2.9 IGF1 Pathway	24
3.2.10 Genes intolerant to loss-of-function mutations	24
3.2.11 Differentially expressed genes	25
3.2.12 SFARI	25
3.3 Sequencing and annotation strategies	25

3.4 Comparisons pipeline	27
3.4.1 Rare Loss-of-function and missense	29
3.4.2 Discriminating models	31
3.4.2.1 Classification Performance	31
3.4.2.2 Logistic Regression	32
3.4.2.3 Decision Tree	33
3.4.2.4 Neural network	33
3.4.2.5 Support Vector Machine	34
3.4.3 Principal Component Analysis	35
4 Results	35
4.1 Exploratory Analysis	36
4.1.1 Removing outliers	39
4.1.2 Variants distribution	39
4.1.3 Kmeans	41
4.1.4 Correlation Matrices	43
4.1.5 Variant distribution between affected and control	44
4.1.6 Comparison Analysis	47
4.1.6.1 Loss-of-Function grouping	47
4.1.6.2 Multivariate models	48
4.1.6.3 Principal Components Analysis	49
4.1.7 Analysis by clusters	53
4.1.7.1 Cluster 1	54
4.1.7.2 Cluster 2	54
4.1.8 Separate regions	55
4.2 Report Analysis	56
4.2.1 Genes with critical exons	57

4.2.2 Chromatin modifier genes.....	58
4.2.3 1,527 genes related to ASD.....	59
4.2.4 FMRP targets.....	59
4.2.5 208 genes related to ASD.....	61
4.2.6 GABA pathway.....	61
4.2.7 Hipermutable Genes.....	62
4.2.8 mTOR Pathway.....	62
4.2.9 IGF1 Pathway.....	63
4.2.10 Genes intolerant to loss-of-function mutations.....	63
4.2.11 Differentially expressed genes M1.....	64
4.2.12 Differentially expressed genes M2.....	64
4.2.13 Differentially expressed genes M8.....	65
4.2.14 Differentially expressed genes M16.....	65
4.2.15 Differentially expressed genes M18.....	67
4.2.16 RAS/MAPK Pathway.....	68
4.2.17 SFARI High Confidence.....	70
4.2.18 SFARI Strong Candidate.....	70
4.2.19 SFARI Suggestive Evidence.....	70
4.2.20 SFARI Syndromic.....	70
4.2.21 SFARI High Confidence, Strong Evidence, Suggestive Evidence and Syndromic.....	71
5 Discussion and Conclusions.....	71
6 Resumo.....	79
7 Abstract.....	81
8 Reference.....	83
9 Attachments.....	96
9.1 Attachment 1.....	95

9.2 Attachment 2.....	100
9.3 Attachment 3.....	100
9.4 Attachment 4.....	106
9.5 Attachment 5.....	108
9.6 Attachment 6.....	109
9.7 Attachment 7.....	109
9.8 Attachment 8.....	110
9.9 Attachment 9.....	110
9.10 Attachment 10.....	111
9.11 Attachment 11.....	134
9.12 Attachment 12.....	136
9.13 Attachment 13.....	138
9.14 Attachment 14.....	138
9.15 Attachment 15.....	139
9.16 Attachment 16.....	139
9.17 Attachment 17.....	140
9.18 Attachment 18.....	140
9.19 Attachment 19.....	141
9.20 Attachment 20.....	141
9.21 Attachment 21.....	142
9.22 Attachment 22.....	143
9.22.1 Prefilter.....	155
9.22.2 Spring.....	158
9.22.3 Counting.....	163
9.23 Attachment 23.....	167
9.24 Attachment 24.....	193

1 Introduction

1.1 Clinical and epidemiological aspects of autism spectrum disorder

Autism spectrum disorders (ASD) are a group of neurodevelopment disorders characterized by impairments in communication, behavior and social interaction. According to DSM-V criteria (Diagnostic and Statistical Manual of Mental Disorders, 5th edition), the diagnosis of ASD is made based on clinical findings in two different categories, namely: (1) social and communication skills disorders and (2) patterns of behavior with stereotypes and repetitive motions (la Torre-Ubieta et al. 2016; Lord and Bishop 2015). Social-communicative skill changes must have deficits in three distinct areas: (a) socio-emotional reciprocity, (b) nonverbal behaviors used in social interaction, and (c) maintaining, developing and understanding relationships, and adjusting to various social contexts (Lord and Bishop 2015). This new classification encompasses all pervasive developmental disorders previously separated into Asperger's syndrome, autism, childhood disintegrative disorders, and unspecified pervasive developmental disorders. According to the new classification criteria, although there is no age limit for the onset of symptoms, all individuals with ASD should present them from childhood, even if they are recognized only later.

Through a 2014 meta-analysis, the worldwide prevalence of ASD was estimated in 1/132 individuals (Baxter et al. 2014; Centers for Disease Control and Prevention 2012), with a rate of 3-4:1 more males than females, with a more balanced ratio if patients present with dysmorphic features (Bourgeron 2015; Manning-Courtney et al. 2013). Recent reports achieved a prevalence of 1.68% in the United States, and some places like North Carolina the number was estimated to 3.62%. It has been argued that this increase in prevalence or the difference among sites could be related to discrepancies of survey designs and the participation of parents of children that never received an ASD diagnosis (Fombonne 2018). In 1990 the prevalence of ASD was 1/1000, this 10-fold increase in prevalence is not distinguishable from the changes in ascertainment to a true increase in prevalence (Anagnostou et al. 2014).

In addition to these core characteristics, the ASD can have highly complex clinical features and harbors many comorbidities, 50% of cases have language delay, 35-50% association with intellectual disability (ID), approximately 21.5% of individuals with ASD and ID are affected by epilepsy, while in only 8% of those with average intellect (Bourgeron 2015; Geschwind and State 2015; Woodbury-Smith and Scherer 2018). Gastrointestinal problems, sleep disorders and psychiatric disorders are also common comorbidities of ASD (Anagnostou et al. 2014).

Approximately 10-25% of the ASD cases are part of a recognizable Mendelian disorder such as fragile X-Syndrome (1.9%), Rett syndrome, tuberous sclerosis (0.9%), neurofibromatosis type 1 (0.28%), Williams-Beuren syndrome and many others

(Anagnostou et al. 2014; la Torre-Ubieta et al. 2016), a great number with a known molecular basis that can be tested once it is suspected. For this reason routine screening for Rett syndrome and fragile X syndrome should be done for individuals with ASD (Schaefer and Mendelsohn 2013). This particular characteristic confers to ASD the label of being a common syndrome that comprises, at least in part, several hundred rare disorders (Geschwind and State 2015).

Being a relatively common disorder the risk for any couple in the population matches the prevalence of the disease for that population (around 1-2% worldwide). Because of its genetic component, with many population studies placing the ASD heritability higher than 0.5 (Colvert et al. 2015; Lichtenstein et al. 2010; Sandin et al. 2014; 2017) it is estimated that almost 50% of the simplex families are at high-risk class in which an inherited cause could be predominant (Ronemus et al. 2014). The recurrent risk for the families with one child diagnosed is around 18,7%, for those individuals with two older siblings this risk goes to 32.3% (Woodbury-Smith and Scherer 2018).

The available data suggest that at least some forms of ASD involve time-specific developmental deficits that might be targets for treatment, even after the first emergence of symptoms (Betancur, Sakurai, and Buxbaum 2009). Many drugs are being tested based on pathways and specific gene targets (Griesi-Oliveira et al. 2014; Lin et al. 2014; Riikonen 2016) but the only FDA-approved drugs, risperidone and aripiprazole are not directed to the core social deficits (la Torre-Ubieta et al. 2016).

Family studies show that ASD can segregate according to all possible modes of inheritance (De Rubeis and Buxbaum 2015) being compatible with what is known for complex diseases. Some authors use the term Complex Disease as synonymous for diseases with multifactorial or polygenic modes of inheritance (Mitchell 2012), even though both represent very distinct genetic basis. Both modes of inheritance necessarily involve the action of several genes together, compatible with additive or epistasis models, and the main difference between them is that the multifactorial model adds the environmental component and the polygenic only the relative expression of each gene (Ridley 2009). Complex Disease, can also be interpreted as a pattern of inheritance that can include all modes of inheritance together, which means including in the same term the mendelian, multifactorial and polygenic modes of inheritance (Mitchell 2012; Schork 1997). In the case of ASD this more inclusive definition for complex disease can be widely applied and the results of biological and statistical approaches would be more successful when analyzing together all the possibilities of inheritance for the individual in question (la Torre-Ubieta et al. 2016).

1.2 Molecular aspects of autism spectrum disorder

After the availability of high-throughput microarray and sequencing platforms much advances have been made to the understanding of ASD genetic risk, nowadays it can be considered that the genetic basis of ASD is better understood, even in the absence of a resolute molecular test for the diagnosis (Woodbury-Smith and Scherer 2018).

Copy number variation (CNV) and single nucleotide variants (SNV) are both implicated, with a strong confirmation in literature for the *de novo* highly disruptive variants (Iossifov et al. 2012; O'Roak et al. 2011; Ronemus et al. 2014; Sanders et al. 2012; Turner et al. 2017) along with rare variants in many genes related to ASD, showing pleiotropy and variable penetrance (Tordjman et al. 2014; Weiner et al. 2017; Woodbury-Smith and Scherer 2018). Autism Spectrum Disorders' heritability has been estimated from 0.5 to 0.9 (Lichtenstein et al. 2010; Sandin et al. 2014; 2017), this estimative is suggestive of a strong genetic component for ASD. For ASD no genetic variation alone accounts for more than 1% of the cases, and even individuals with identical genetic variants demonstrate a large degree of phenotypic variability. In this manner the strong genetic component should be related to a highly variable genetic background (Bourgeron 2015; Chahrour et al. 2012; la Torre-Ubieta et al. 2016).

CNVs, both deletion and duplications above 1kb, are frequent in patients with ASD (Geschwind and State 2015), mainly involving specific regions of chromosomes 1, 2, 3, 5, 7, 11, 12, 13, 15, 16, 17, 18, 21, 22 and X (Schaefer and Mendelsohn 2013), with special attention to 16p11.2 (affecting 0.8% of individuals with ASD), 15q11-13 (0.5%) and 22q11 (0.5%) (Geschwind and State 2015; Moreira et al. 2016; Woodbury-Smith and Scherer 2018). These variants were found in about ten-times excess in simplex families compared with typical controls (Bandler et al. 2015; Sebat et al. 2007), and early studies showed that CNVs disrupted the locus containing SHANK3 and other promising candidate autism risk loci (Szatmari et al. 2007). Also, 10% of the individuals with an established pathogenic CNV carried a second large one, and 72% of the second CNVs were inherited from the mother, while the primary CNV (*de novo* or inherited) had no parental transmission bias (Girirajan et al. 2013). These studies converge on an estimate that around 5-15% of individuals with a diagnosis of ASD will carry a CNV conferring substantial risk, versus 1-2% of the general population (Bourgeron 2015; Geschwind and State 2015).

Single nucleotide variants (SNV) have also been recognized, but to a lesser extent and for the majority of cases without a well described penetrance of the phenotype related to the mutations (Scherer and Dawson 2011; Woodbury-Smith and Scherer 2018).

The overall rate of *de novo* SNV in ASD was slightly elevated when compared with unaffected siblings, but two to three-fold predicted to be more disruptive than the ones carried by the siblings (De Rubeis et al. 2014; Iossifov et al. 2014; Neale et al. 2013; O'Roak et al. 2011; O'Roak, Vives, Girirajan, et al. 2012a; Ronemus et al. 2014; Samocha et al. 2014). The majority of *de novo* point mutations in ASD originate from

the father, with the correlation between paternal age and number of *de novo* events reflecting a 1.3-fold increase in the number of *de novo* events for every 10 years of paternal age (Ronemus et al. 2014). The *de novo* likely gene disruptive mutations appears more often in genes expressed in the brain and genes already classified as candidate (Cukier et al. 2014; Sanders et al. 2012; Yuen et al. 2016).

Some studies show that transmission of probably causative variants originates most often from the mother, and transmission of such variants is seen more often in offspring with lower IQ (Iossifov et al. 2015; Krumm et al. 2015; Stessman et al. 2017). These findings are compatible to what first have been described for CNV (Jacquemont et al. 2014) and is in concordance with a model of female protective background for rare moderately penetrant mutations (SNVs and CNVs). There are more inherited SNVs that disrupt protein function in conserved genes transmitted from the mother to individuals with ASD than in unaffected siblings. However, there are contradictory research regarding this same finding for the CNVs (la Torre-Ubieta et al. 2016).

The percentage of the genetic diagnosis from likely gene disruptive variants in ASD can vary depending on the study and it is around 5-15% (Iossifov et al. 2015; Sanders et al. 2012). A model for the disease with several high penetrance genes could not explain the network data found which can lead to the conclusion that a single deleterious event is unlikely to fully explain the disease in most cases (Neale et al. 2013).

The fact that many families segregating ASD also seem to share risk for other neuropsychiatric disorders suggests that this genetic risk may be more correctly ascribed to fundamental brain-based traits (Woodbury-Smith and Scherer 2018). The number of genes estimated to be responsible for “monogenic” forms of ASD is between 400-1000 (Bourgeron 2015; Iossifov et al. 2015; W. Luo et al. 2018). The pathophysiology of the disorder is related to the enrichment of *de novo* events at specific genes and pathways.

In addition to the findings described previously, evidence supports a substantial contribution of common genetic variation to ASD (Gaugler et al. 2014; Grove et al. 2017; Woodbury-Smith and Scherer 2018), nevertheless most loci identified, even those with genome-wide significance, were not replicated in subsequent studies (Abrahams and Geschwind 2008; Weiss et al. 2009). Even with the estimated 40-60% of the genetic risk for ASD from the common variants (Gaugler et al. 2014; Klei et al. 2012; Weiner et al. 2017), studies of inherited forms of ASD need a higher number of samples to have sufficient power to identify and replicate common alleles at genome-wide significance (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014).

Considering the types of variants and genes related to the etiology of ASD there are many models being proposed to explain its pathophysiology. The synaptic homeostasis hypothesis for ASD suggests that genetic variants could either increase or reduce the number or strength of the synapses’ connections and lead to the emergence of the phenotype, with common single nucleotide polymorphisms acting as a buffer for deleterious rare variants and unaffected individuals between these two extremes (Bourgeron 2015). Many studies have confirmed the role of disrupting

synaptic genes for the etiology of ASD (De Rubeis et al. 2014; Pinto et al. 2014; Rotschafer, Marshak, and Cramer 2015).

Luo et al. 2018 proposed a complex model involving two modules interconnected by the MAPK pathway. Module I include Wnt signaling, cell adhesion, junction and cytoskeleton-related pathways, while module II includes glutamatergic synapse, GABAergic synapse and related process. In this model a double hit of gene + pathway is necessary for the development of the full phenotype, which means a less importance for the burden theory, which favors single hits in genes recurrently mutated in ASD. They built a more inclusive dual-hit model, where the disease variants need to disrupt the target genes and hit the relevant pathways. This model better explains ASD risk than the well-accepted mutation burden model (W. Luo et al. 2018).

The amplification of the investigation for the whole genome sequencing did not reveal a great impact of regulatory non-coding variants (Werling et al. 2017), but instead confirming findings of candidate genes (C Yuen et al. 2017; Yuen et al. 2015). Recently Bradler et al. (2018) investigated possible roles for the regulatory regions in the etiology of ASD, like the cis-regulatory structural variants (Brandler et al. 2018).

Nevertheless, even with the main discoveries regarding ASD genetic basis there is no laboratory test that can return a reliable specific result for 70% of the families. In this sense it is possible to hypothesize what types of mutations are important for the diagnosis of ASD, and perhaps construct a more robust test to provide a molecular diagnosis for the families. It is always important to keep in mind that each of these models explains a portion of the disease and that the only certainty is that the ASD is not caused by a single model of transmission (Pinto et al. 2014). In site of these models of ASD etiology it is difficult to establish which is the scenario for the specific individual, so it can be programmed a treatment or genetic counseling.

The importance to discriminate between complex and essential ASD (essential refers to the presence of isolated characteristics of ASD, and complex refers to the individuals with the co-occurrence of other morbidities) is because the presence of specific comorbidities increases the probability of a diagnostic hypothesis that can be evaluated by a clinical geneticist to establish the most effective molecular test (Schaefer and Mendelsohn 2013; Woodbury-Smith and Scherer 2018).

The American College of Medical Genetics and Genomics (ACMG) 2013 guideline for the molecular investigation of autism is still in agreement with the recent literature (Schaefer and Mendelsohn 2013; Woodbury-Smith and Scherer 2018). Penetrant single variants (CNV, chromosomal rearrangement, short indel or SNV) is calculated to happen in 10-25% of the individuals with ASD (Bourgeron 2015). The CNV findings are believed to be responsible for the majority of the cases with a molecular diagnosis, so a molecular investigation with microarray techniques is indicated to detect these alterations as a first approach to a patient with ASD (Krumm et al. 2013; Schaefer and Mendelsohn 2013).

In this context, ACMG suggests that there is no formal indication for the specific test of each gene when suspected ASD (Schaefer and Mendelsohn 2013).

Despite the constant registration of new candidate genes, the finding of a single disrupting variant in a candidate gene is not sufficient to provide a molecular diagnosis, which encourages searches for complex relationships between variants in the same individual and opens space for exploring new hypotheses regarding their etiological bases (Vaishnavi et al. 2014). Without a highly definitive molecular diagnosis for ASD, the best predictor of sibling recurrence remains family history (Geschwind and State 2015).

1.3 Next Generation Sequencing and the Diagnosis of Rare Diseases

With the advent of the Next Generation Sequencing (NGS) techniques, it became possible to investigate gene alterations on a large scale, with the advantage of being able to observe many genes at the same time, without the need to choose which candidate gene would be preferentially researched (Mardis 2008). This methodology made the sequencing of the human whole genome cost-effective and increase the appeal to apply this technique to the rare disease investigation (Lapin et al. 2016).

To turn this methodology even more cost-effective it is possible to sequence only the coding regions of the genes, the exons (whole exome sequencing - WES). It is estimated that 85% of the mendelian diseases that has a known genetic etiology could be diagnosed with this approach (Trujillano et al. 2016), with sequencing only 2% of the genome, which means a great reduction in the amount of data sequenced (Matullo, Di Gaetano, and Guarrrera 2013).

Despite these high estimations, around 31% of the exomes sequenced in a clinical laboratory, had a positive gene finding, 25% had a variant of unknown significance, and 45% had no relevant variant, these rates can vary depending on the diseases being investigated (Trujillano et al. 2016). This difference in the numbers happens because more than 3,000 known mendelian diseases does not have a genetic etiology established (Chong et al. 2015).

A single WES generates approximately 30,000-50,000 variants different from the human reference genome (Lapin et al. 2016). Each individual carries around 150 loss of function variants and 2,500 nonsynonymous (Matullo et al. 2013), from which 281-515 predicted by *in silico* tools to be highly damaging and 40-85 in homozygous state, (Xue et al. 2012). This number of variants makes it difficult to determine which ones may be related to an effect on the development of the central nervous system to the point of generating the pathological phenotype. Theoretical and experimental studies suggest that individuals typically carry hundreds of mildly disadvantageous variants and several severe disease alleles (Xue et al. 2012).

Most living humans share large segments of genetic variation, especially because they are the progeny of rapid expansion of the human populations from a relatively small number of remote ancestors (about 10000 years ago) (Ramachandran

et al. 2005). This scenario can lead to interesting population effects, with the accumulation of new mutations and rare alleles that did not have enough time to be removed by natural selection (Matullo et al. 2013), that if not taken carefully can lead to false positive results in the clinical-laboratory context and misconduct by the physician in charge (Manrai et al. 2016).

Two processes are fundamental to the achievement of the final result in molecular investigation of genetic diseases, the gene and variant prioritization (Eilbeck, Quinlan, and Yandell 2017). Both processes can be done with the aid of bioinformatics and clinical knowledge. Information such as predicted damage to the protein, population allele frequency, functional testing, presence of the variant in other individuals with same phenotype, segregation in the family and mode of inheritance can be used to classify variants and give a final result (Richards et al. 2015). The misclassification of variants can have a high impact leading to false positives or false negatives in the clinical practice. As an example of the possible extension of those errors the ExAC database was constructed with approximately 65,000 WES from adults without severe genetic conditions. An average ExAC participant harbors close to 54 variants reported as disease-causing in two other databases, an implausibly high number, even considering that these individuals can be affected with several polygenic disorders (Lek et al. 2016). This can have a high impact on the choices of candidate genes for genetic diseases depending on the population assessed.

In rare monogenic diseases, NGS is a tool that obeys the precepts of a diagnostic examination, making a clear distinction between the group of affected and non-affected individuals and can be used as a first diagnostic choice (Trujillano et al. 2016). However, in complex diseases of high prevalence, quality parameters for WES have not yet been well established. Most of dominant or recessive mendelian disorders are caused by rare mutations (with frequency under 0.01 in the population), consistent with the low frequency of these conditions in most populations. Nonetheless, in general, for genetic conditions, common variants (with frequency above 0.01) with large disease effect sizes (>2.0) are exceptional: most known common risk alleles have been shown to increase susceptibility by less than 1.5-2.0 fold (Geschwind and State 2015). Laboratory tests for diagnosis should have high accuracy, leading to a minimum number of false positives and false negatives. The excess of variants present in each individual can become a major obstacle to research.

It is therefore important to consider the probability of damage and the probability of the phenotype penetrance of candidate variants when attempting to counsel families about risk. For children with positive genetic test results, genetic counseling offers the opportunity to plan treatment and/or prophylactic measures for any medical conditions known to be associated with the particular gene (Woodbury-Smith and Scherer 2018).

1.4 Next Generation Sequencing and the ASD Diagnosis

Analysis of the coding regions of the genome by the whole exome sequencing (WES) detected several candidate genes and etiological agents of the Autism Spectrum Disorder (ASD), and was also effective in identifying some variants with probable oligogenic inheritance (Guo et al. 2011). However, even with the current innovative techniques, the number of cases elucidated still remains very low, since the mutations found usually refer to genes not previously described in ASD, and with low recurrence in affected individuals (Yu et al. 2013).

Until now there is no evidence that WES can be used as a clinical-laboratory test for ASD, despite all the main advances that this technique has made for the understanding of ASD in the scientific context, little was done to test if WES could achieve high sensitivity or specificity for each individual in the clinical-laboratory context. Most studies focus in comprehending the pathophysiology and proposing scenarios or candidate genes (He et al. 2013; Liu et al. 2014), but few focus in discriminating the individuals and proposing a physiological explanation for the disease in a particular subject.

It is possible to observe that, since efforts to find a single cause did not bring the expected resoluteness, the research groups return their work to the search for general hypotheses involving more than one mode of inheritance and more than one associated gene, for explaining the etiology of ASD. So far these questions have brought innumerable answers, but each one with a restricted resolving power (Chahrour et al. 2012; Cukier et al. 2014; Kumar 2011; O'Roak et al. 2011; Sanders et al. 2012; Yu et al. 2013) and without taking into account that for the affected individual any of the hypotheses described may be true. Therefore, a multivariate approach is required for these hypotheses, in which individuals can be analyzed as a whole in the face of all possibilities.

Genetic models for complex diseases, such as the ASD, are very hard to elaborate, specially because the phenotype can be caused by all possible inheritances. Fully penetrant genes, like *PTEN* and *SHANK3* are responsible for less than 0.5% of the ASD cases, other genetic variants such as the 15q13 microdeletion/duplication have a higher rate, but are not fully penetrant (la Torre-Ubieta et al. 2016). Also, scenarios were the combination of rare high risk variants, along with common low risk variants can lead to the phenotype (Weiner et al. 2017).

This evidence for genetic heterogeneity among individuals and even within the same family with more than one affected, along with pleiotropy, speaks to the need for precision medicine, a personalized approach to genetics that is cognizant of the fact that assessment of an individual's complete genomic risk is required (Woodbury-Smith and Scherer 2018).

These are the challenges of any complex diseases, and why the initial analysis of a candidate group of genes can be exhaustive. The aim of this study is to facilitate

this first approach and to detect shortly if the candidate genes could be used to discriminate between the affected and control groups and how they would do it, using whole exome sequencing. The approach to count the variants and to separate them into flexible categories depending on the initial hypothesis is a way to access burden of types of variants in specific genes or pathways.

This thesis can be divided into two parts, one for the development of the pipelines and another for the interpretation of the results provided by them. The subject of interpretation can be further divided into the interpretation of each report independently, the impact this analysis has for the laboratory that interprets the whole exome sequencing, and finally the results in the context of the etiology of the autism spectrum disorders.

2 Objectives

The main objective of this study was to propose a series of tests to maximize the molecular diagnosis for autism spectrum disorder with whole exome sequencing in the clinical-laboratory reality. The tests would involve direct approaches and also multivariate analysis based on the quantity of variants presented by each individual, applying statistical tests to evaluate the discriminatory power between affected and control groups.

The specific objectives were:

- To identify potential biases for the analysis involving subjects and strategies of sequencing, and to identify good strategies to overcome the biases losing the least amount of individuals.
- To develop a pipeline to count the number and types of variants filtered for specific gene lists for each individual from a whole exome sequencing, considering the quality of the predicted variant.
- To automatically test gene lists with different mathematical approaches and display the results into a report of friendly interpretation.
- To identify potential discriminant strategies for autism spectrum disorder with whole exome sequencing that would achieve higher sensitivity and specificity than the current approach.

3 Material and Methods

This section is divided into four parts, the first describes the choice for the individuals that were used in the study and, the second the description of the gene lists used to test the different approaches for the discrimination of the ASD and controls, the third, the methods for sequencing and annotation of each individual, and fourth the description of the statistical approaches used to construct the final report for each list. The project of this thesis received the approval of the Ethics Committee from Instituto de Biociências da Universidade de São Paulo (IB-USP), under the number 62602016.3.0000.5464.

3.1 Subjects

3.1.1 Autism Spectrum Disorder Subjects

Initially we started with 71 individuals with clinical diagnosis of autism spectrum disorder (ASD), including 54 different families and 34 trios. In order to simulate the reality of a clinical genetic laboratory, and to avoid bias on the results due to the use of family members we decided to continue the analysis with the 54 probands, 14 with other family members diagnosed (multiplex), and 40 from simplex families.

From these 54 individuals, five were removed due to sequencing quality issues, the final 49 individuals were distributed in 38 simplex cases and 12 multiplex.

3.1.2 Control sample

The choice of the individuals for the control sample followed important steps to avoid bias. Two different approaches were tested, one with the 609 healthy individuals with more than 60 years old, and the other with 175 individuals with other clinical diagnosis, such as rasopathies, skeletal dysplasia and cleft lip and palate. All of this individuals were sequenced in the Centro de Estudos do Genoma Humano da Universidade de São Paulo (CEGH-USP) under other scientific projects, and are representative of the São Paulo and Brazilian ancestry (Bertola et al. 2014; Naslavsky et al. 2017; Yamamoto et al. 2014).

From the first attempts to differentiate the groups with Principal Components Analysis (PCA) (more detail in further sections of Material & Methods), it was observed that the sequencing batch had a high influence in the PCA results, as seen in

Figure 1. Changes in the bioinformatics pipeline, to produce the variant call format (VCF) files had to be established to overcome the effects that the different sequencing had on rare variants (more details in further sections). Figure 2 shows an example of the kind of effect different VCF files can have on the final result.

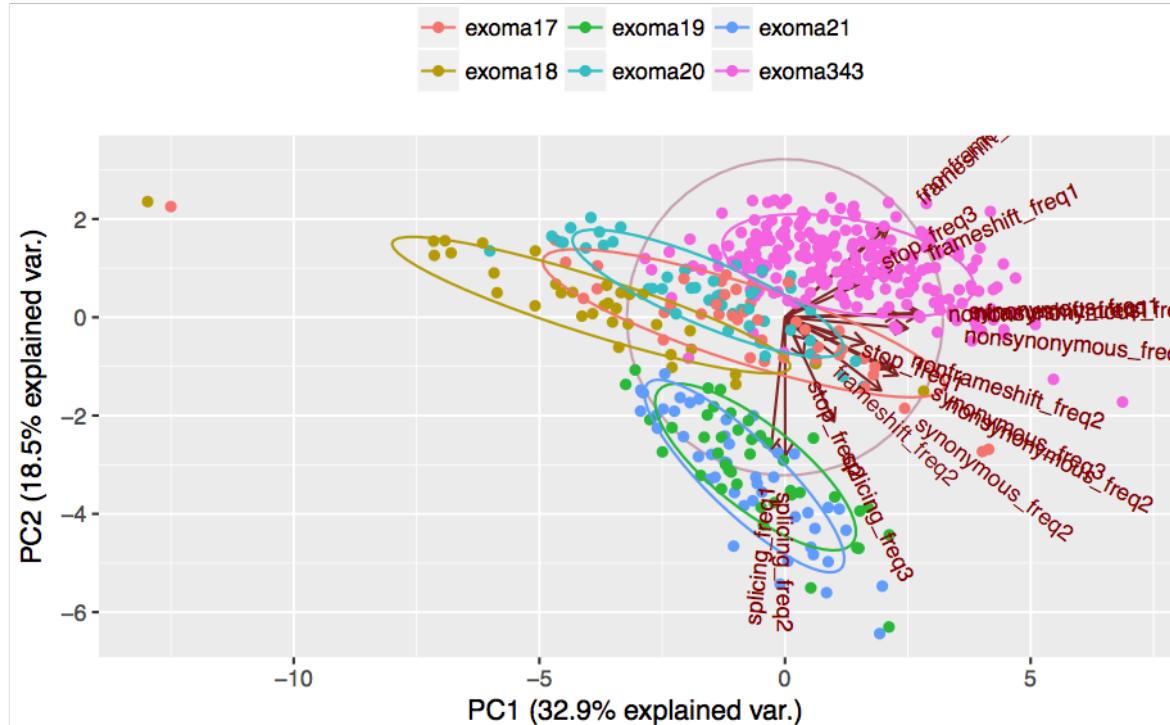


Figure 1: First two PC from different sequencing batches, from 17 to 21, and the pink (exoma343) with 343 individuals called together.

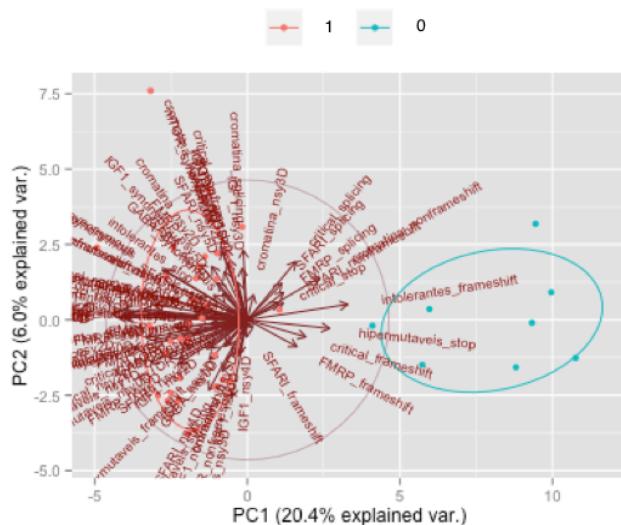


Figure 2: The first two PCs with only the probands of ASD, in red (1), and the probands with cleft lip and palate, in blue (0), using all rare variants, before the changes in the pipeline to produce a unique VCF file.

After the adjustments to make a unique VCF file, the PCA from all variants still resulted in a complete separation from the groups, specially separating the 609 elderly healthy samples, as it can be seen in Figure 3.

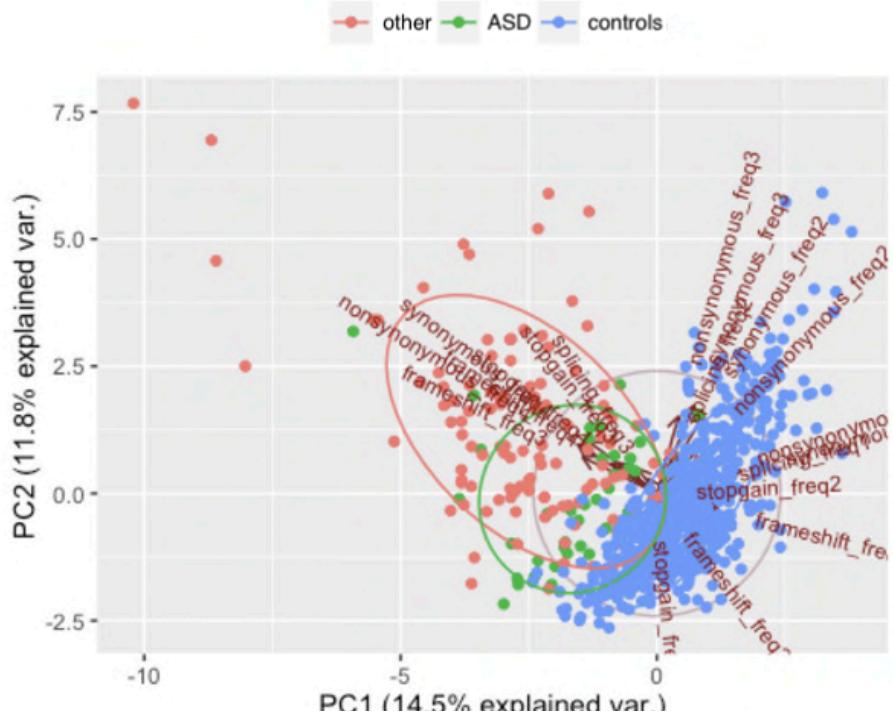


Figure 3: The first two PCs for the three groups, red is the controls with other diagnosis, green is the ASD group, and blue the 609 healthy controls.

Based on the results published in literature it was not expected to see this clear separation among the samples, specially between ASD and the healthy controls. This kind of phenomena can lead to false positive results, and overfitting models. Figures 4 and 5 are examples of overfitting models. Both shows the receiver operating characteristic (ROC) curve from a logistic regression model based on variants from VCF files after corrections in the pipeline (more detail in further sections of Material & Methods). While Figure 4 shows no discrimination between ASD and control (175 individuals with other diagnostics), Figure 5 shows the same dataset in comparison with the 609 healthy individuals, revealing a high degree of discrimination between the groups, which based on the scientific literature published so far, was not expected.

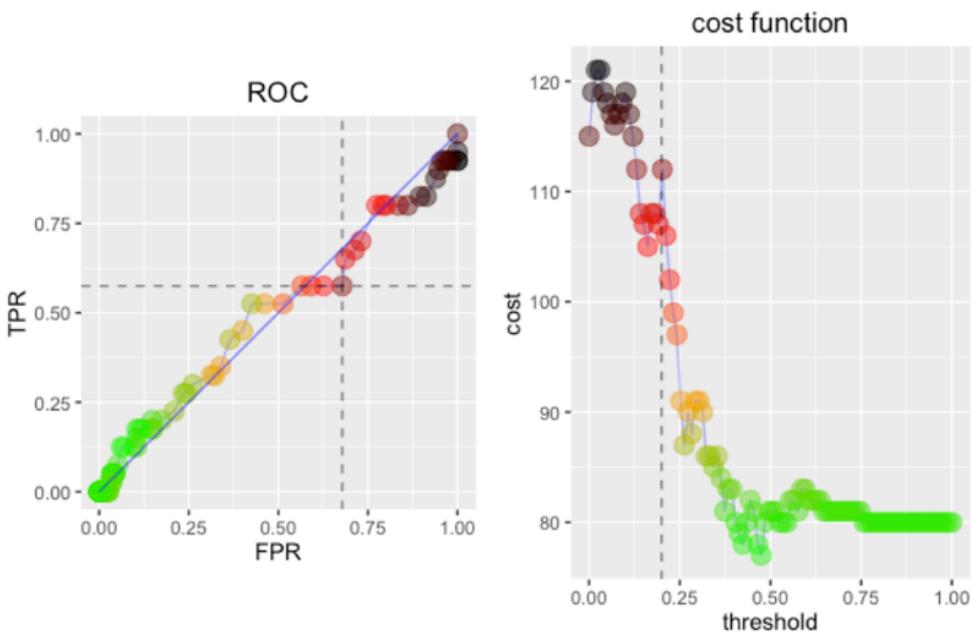


Figure 4: ROC curve of all variants comparing the results of a logistic regression between ASD and other disease groups.

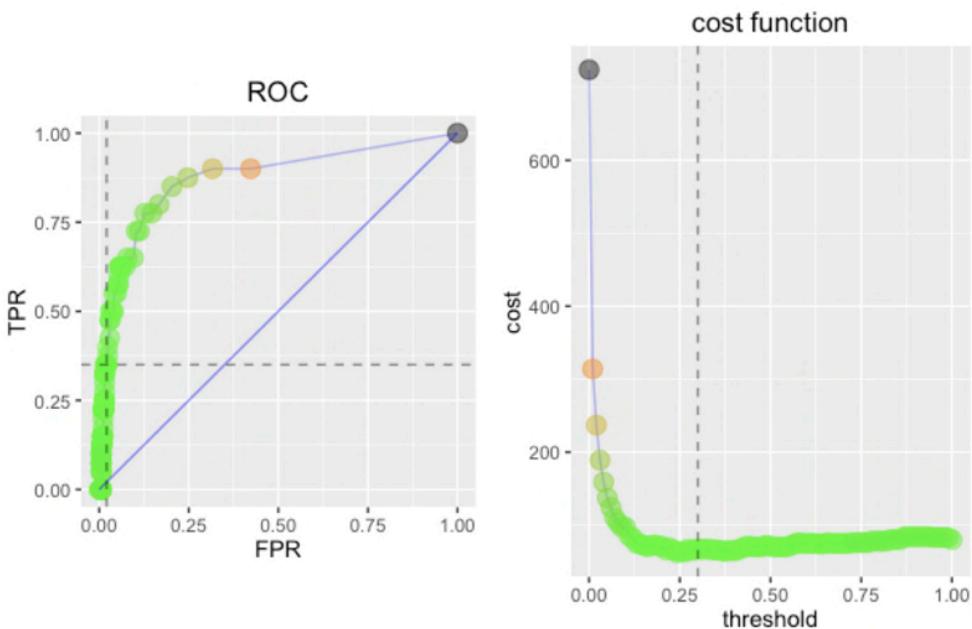


Figure 5: ROC curve of all variants comparing the results of a logistic regression between ASD and the healthy individuals.

This result can be related to a bias generated by the type of sequencing and bioinformatics the 609 healthy individuals were submitted. For this reason, only the probands of the individuals with other diagnostics, 119 individuals in total, were used for subsequent analysis. In summary, this work was done with 49 ASD cases and 119 controls, resulting in 168 individuals.

3.1.3 Kinship and ancestry analysis

Focusing on 23,556 of pre-selected single nucleotide variants (SNV), multiple fast quality and kinship control were computed for the 168 individuals in this analysis, with the software Peddy developed by Pedersen and Quinlan (2017).

The analysis of the ratio of heterozygous calls by the median depth of coverage, showed all 168 individuals with a proportion of heterozygous between 0.27 and 0.35, with no outliers with an excess of heterozygous variants that can happen in situations of sample contamination or high degree of sequencing errors (Figure 6) (J. Wang et al. 2015). All sequencing had at least 80% of the target regions covered with a minimum of 10x, a pattern considered of high sequencing quality by the ExAC consortium for population studies (Lek et al. 2016).

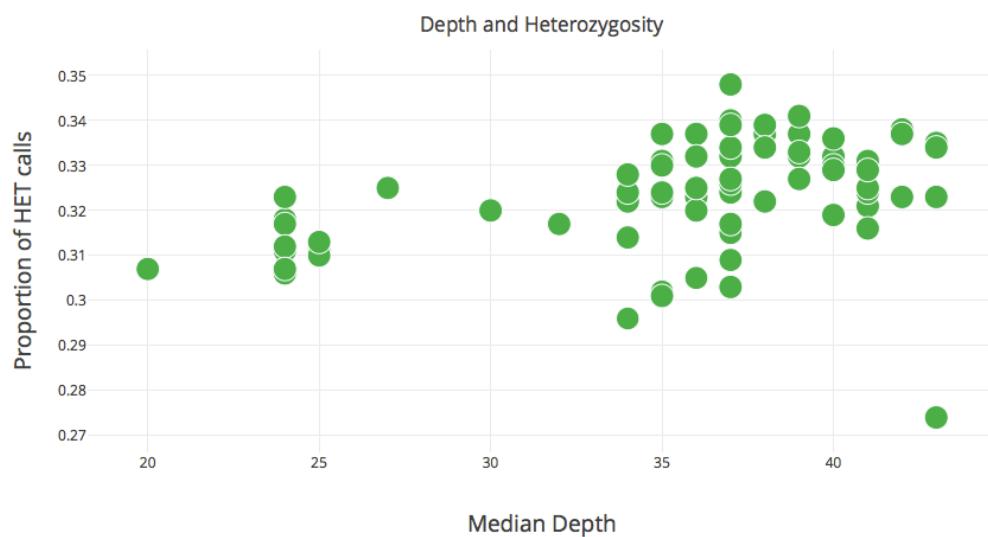


Figure 6: Scatterplot with the proportion of heterozygotes by the median depth of coverage in the WES, for each individual in the analysis.

To confirm that the samples had no shared genetic background (individuals could be considered as distinct family) the pairwise coefficient of relatedness (CR) for all samples was applied. Figure 7 shows the CR for each pair for the 168 individuals, all resulted in less than 0.05 relatedness, the same score expected for random individuals in the human population (Pedersen and Quinlan 2017).

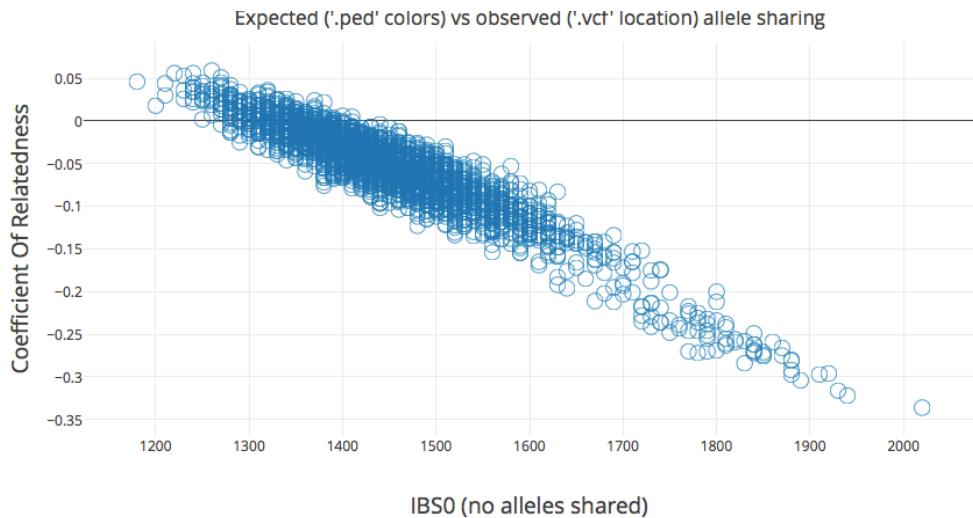


Figure 7: Scatterplot with showing the coefficient of relatedness for the 168 individuals in the analysis.

Also, an analysis of the ethnic background was done for the samples. A principal components analysis based on informative SNPs from 1000Genomes project divided the world population into five categories of European, African, East Asians, South Asians and Americans. For our sample the great majority of individuals were classified as American (Figure 8), for the individuals classified as unknown, European and African, their distributions were very close to the majority of the group and could be classified into the dispersion of Americans.

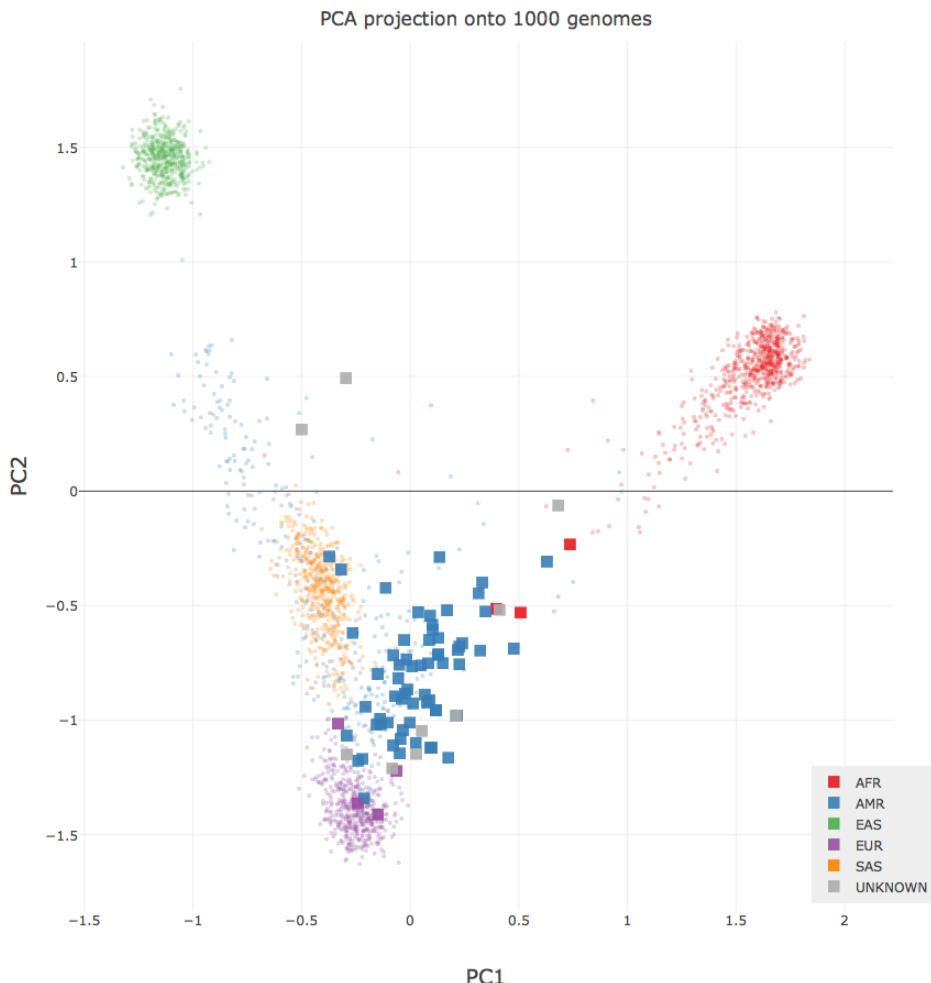


Figure 8: The first two PCs of the ancestry distribution based on 1000Genomes, the squares are the 168 individuals in this study.

3.2 Gene lists

Twenty-one gene lists were separated based on scientific publications and personal discoveries from the group of Maria Rita Passos-Bueno, the lists will be described in detail further.

3.2.1 Genes with critical exons

Uddin et al. 2014, compared the number of mutations in critical exons (exons with high expression in central nervous system and low rate of rare pathogenic polymorphisms) between individuals with ASD and controls. It has been observed that a greater proportion of these events occurs in individuals with ASD. In the study

they performed the WES and transcriptome from different regions of brain tissue and observed 1,744 genes preferentially expressed in the brain containing 3,955 critical exons. They found that there was an accumulation of de novo mutations, predicted to be deleterious in the exons considered critical in individuals with ASD compared to controls. This study was quite effective in showing a possible way to discriminate molecularly individuals with ASD and those unaffected because the results tended to be extreme for both groups (Uddin et al. 2014). This list of genes can be seen in supplement material of the article and consists in 1,744 genes (Attachment 1).

3.2.2 Chromatin modifier genes

Cotney *et al.* in 2015 described the relation between *CHD8* and other chromatin modifiers in the pathophysiology of ASD (Cotney et al. 2015). This article, along with previous and recent articles, recognized the importance of *de novo* mutations in chromatin targets (Krumm et al. 2015; O'Roak, Vives, Girirajan, et al. 2012a). These implications made the genes that are target for *CHD8* or chromatin modifiers a good candidates to test differences between ASD and control groups. The list of 263 genes was made with the genes described by Cotney et al. 2015 and the Reactome genes related to chromatin organization and signal transduction (Attachment 2).

3.2.3 1527 genes related to ASD

This list of genes was selected by our group previously based on literature for genes related to ASD, these genes were used to construct a customized array-CGH (Moreira et al. 2016) (Attachment 3).

3.2.4 FMRP targets

The most common recurrent cause of syndromic ASD is Fragile X syndrome, which is caused by a loss of the FMRP function (la Torre-Ubieta et al. 2016). Because of this strict relation between FMRP and autism it was hypothesized that mutations in FMRP targets could be related to ASD (Darnell et al. 2011; Steinberg and Webber 2013). Darnell et al. 2011 identified 842 genes that were FMRP targets and Steinberg et al. 2013 found an enrichment of disruptive mutations in FMRP targets in ASD than controls. The list of genes are in Attachment 4.

3.2.5 208 genes related to ASD

Stessman *et al.* (2017) selected 208 genes related to ASD from published sequencing studies based on the number of *de novo* recurrence, overlap with CNV morbidity map, pathway connectivity and absence of *de novo* variants in 1909 unaffected siblings (Stessman et al. 2017). They describe 91 genes with excess *de novo* mutations or private disruptive mutations in 5.7% of individuals with ASD. This work was highly criticized by Barret *et al.* (2017) because of the flaws in the statistical approaches and correction for genome-wide scan for discovery genes (Barrett et al. 2017). Because of this recent dispute about the methodology of this list we concluded that could be a good subject for further discussion. The 208 genes are in Attachment 5.

3.2.6 GABA pathway

Cellot *et Cherubini* (2014) taking into account that GABA is the main inhibitory neurotransmitter of the central nervous system and the fact that one of the most common deletions involved in the etiology of ASD is the region 15q11-q13, which is enriched for genes encoding subunits of the GABA receptor proposed a role, further confirmed for this pathway in the etiology of ASD (Cellot and Cherubini 2014; Hogart et al. 2007; la Torre-Ubieta et al. 2016; W. Luo et al. 2018). The genes related to GABA reuptake, synthesis and degradation, GABA receptor activation and GABA modulators resulted in a list of 74 genes of interest (Attachment 6).

3.2.7 Hypermutable Genes

Hypermutable genes were considered by Michaelson *et al.* (2012) as a possible source for the ASD etiology (Michaelson et al. 2012). The hypermutable sites could be hot-spots regions for disruptive mutations and could explain the high prevalence of autism. The article compared the genome sequencing in nine monozygotic twins and described many regions of hypermutability. They showed that the mutation frequencies varied widely by the genome, with a difference of up to 100 times depending on the region, this variation could be explained by the intrinsic characteristics of the DNA sequence and chromatin structure. Many of these regions were related to previously described genes for ASD. The supplemental material of this article gives a spreadsheet of the transcripts and their hypermutable rank for 1858 genes. For this work, only the genes with hypermutable scores above three in more than 50% of the transcripts were included, and resulted in a list of 181 genes (Attachment 7).

3.2.8 mTOR and RAS/MAPK Pathways

Pinto et al. 2014 used the impact of new, or inherited, CNVs on 2,446 individuals with ASD and their families, and 2,640 controls to characterize candidate genes to be associated with ASD. They chose the genes disrupted by these CNVs and identified the biological relationship between them and the pathways in common. They have shown that both deletions and duplications involving the FMRP and post synaptic density genes increase the risk for ASD and genes present in CNVs and SNVs with presumed loss of function converge on functional pathways such as PI3K-Akt-TSC-PTEN-mTOR and PI3K-RAS-MAPK (Bourgeron 2015; W. Luo et al. 2018; Pinto et al. 2014). In Reactome there were nine different pathways for mTOR. For this study only the mTOR signaling and the regulation of Rheb GTPase activity by AMPK were used. The total number of genes in this list is 39 (Attachment 8). The MAPK family signaling cascade, was also obtained from Reactome with 282 genes in total (Attachment 16).

3.2.9 IGF1 Pathway

Many recent studies shows an impact of the treatment with IGF-1 for individuals with *SHANK3* mutations (Kolevzon et al. 2014; Riikonen 2016; Shemesh et al. 2012) and the possible use as a biomarker for ASD (Krumm et al. 2015). Because of the importance of the mTOR pathway to ASD etiology, which is entirely contained in the IGF-1 pathway, and the possible roles described above, the IGF-1 pathway was taken into consideration for further study. The list was made with the Reactome signaling by IGF1R (type 1 insulin growth-factor 1 receptor) and consists of 303 genes (Attachment 9).

3.2.10 Genes intolerant to loss-of-function mutations

Many studies of *de novo* disruptive mutations in ASD shows an enrichment of this type of mutations in affected individuals (Iossifov et al. 2012; O'Roak et al. 2011; Ronemus et al. 2014; Sanders et al. 2012). The article published by Krumm et al. (2015) showed the role of loss-of-function *de novo* mutations in the etiology of ASD (Krumm et al. 2015), and concluded that genes more intolerant to loss-of-function mutation could have a major role in the etiology of ASD. From the 16,596 genes scored for intolerance by Petrovski et al. (2013), the genes with a residual variation intolerance score (RVIS) (Petrovski et al. 2013) higher than 50% were enriched in ASD individuals. For this study only 7,764 genes achieved this threshold (Attachment 10).

3.2.11 Differentially expressed genes

One researcher from the group, Karina Griesi, is working with differentially expressed genes from ASD and control iPS cells. Her results generated five different sets of differentially expressed genes from a more relaxed search (M1), to a stricter one (M18). M1 is described in Attachment 11, M2 in Attachment 12, M8 in Attachment 13, M16 in Attachment 14 and M18 in Attachment 15.

3.2.12 SFARI

The Simons Foundation Autism Research Initiative (SFARI) is a scientific initiative to improve understanding, diagnosis and treatment of ASD that involves more than 400 investigators in the U.S. and abroad. The SFARI has a gene scoring advisory panel that review literature on the genetics of autism, with emphasis on evidence from human genetics, that separates the genes involved in the ASD etiology into 6 categories: syndromic, high confidence, strong candidate, suggestive evidence, minimal evidence and hypothesized but not tested. For the purpose of this work only the first 4 categories were assessed, and the gene lists are as follows: syndromic (101 genes - Attachment 17), high confidence (23 genes - a Attachment 18), strong candidate (54 genes - Attachment 19), suggestive evidence (149 genes - Attachment 20) and all of the three together (279 genes - Attachment 21).

3.3 Sequencing and annotation strategies

All the samples were sequenced by CEGH-USP from 2014 to 2017 with an Illumina's HiSeq2500 sequencer. Twenty-one batches of exomes were sequenced, the numbering follows a time sequence from the first set of individuals to have their WES sequenced at CEGH-USP, batch 1, and the last, batch 21.

During this period some changes were made in the protocol, such as library preparation started as Illumina Nextera, and move to Agilent's SureSelect capture.

Alignment to the reference human genome GRCh37/Hg19 was done with BWA-MEM (H. Li and Durbin 2009) with a minor modification from the human reference, excluding the sequences that did not have precise chromosome information and were classified as Unknwon. The CEGH-USP bioinformatics team was responsible for the sequencing and alignment of the samples and their scripts are not present in this study, nevertheless the scripts developed by this study will be in the attachments.

Each sample was genotyped separately using the Genome Analysis Toolkit Haplotype Caller (GATK - version 3.4) (McKenna et al. 2010). The decision for calling all genotypes in a gVCF format was made to maintain the information about each base position and avoid the bias for the comparison of different VCF files, as described previously. After each calling all the gVCF files were combined for a unique call with GenotypeGVCF. With the gVCF file it is possible to combine all individuals in one VCF and maintain each base information minimizing the effect of the batches of sequencing. This effect can be seen by the proximation of batches 19 to 21 in this new approach, nevertheless some differences could not be removed, as it can be seen by the isolation of exome 18 (Figure 9).

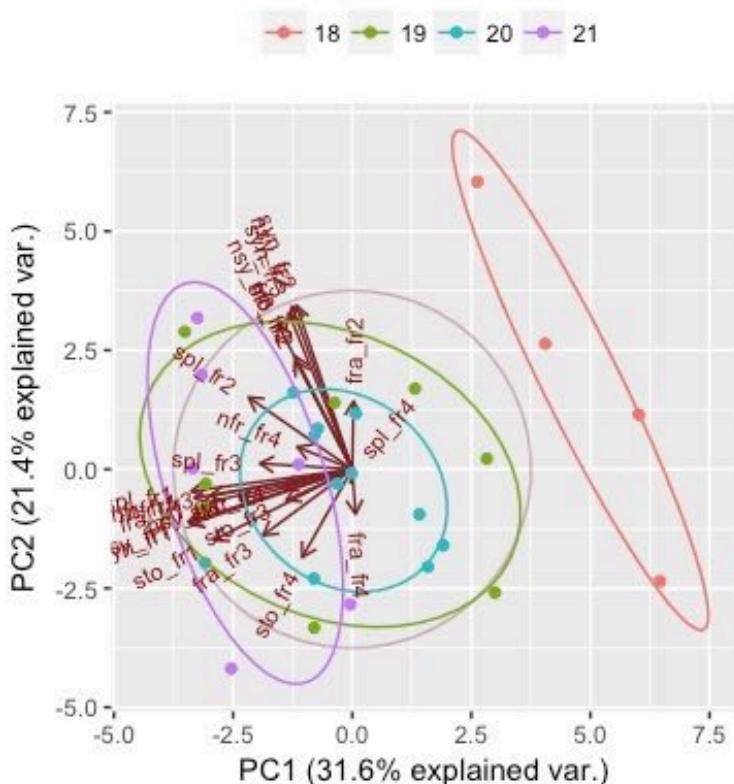


Figure 9: First two PC of the whole exome sequencing regarding only the coding variants for batches 18 to 21.

Even with the remaining differences among WES batches, after the combination of the gVCF files, all the individuals were maintained for further analysis in order to simulate the real clinical-laboratory context where most of the bioinformatics strategy were done to control for sequencing biases.

Since different gene lists are being tested the VCF file was filtered only to contain the information for that specific genes. The chromosome coordinates (BED) files for each gene list was obtained by UCSC Table Browser taking as genome reference the GRCh37/Hg19 and whole gene as parameters. The resulting variant call format (VCF) file, with all 168 individuals, where filtered based on the BED file for each gene list and converted for annotation with ANNOVAR (K. Wang, M. Li, and Hakonarson 2010), and other customized annotations developed by the CEGH-USP

bioinformatics group. Variants with genotype-likelihood below 98 score, according to GATK, with less than 10x of depth and alternative balanced alleles inferior to 0.3 were removed (Aziz et al. 2015; Matthijs et al. 2015; Pirooznia et al. 2014).

A script in R (Team 2017) was developed to count, from this annotated file, the number of variants in each class for each individual. The counting was separated in two different tiers, one based on population frequency and the other based of protein alteration. The first took into consideration three population databases, the 1000Genomes (<http://www.internationalgenome.org>), 6500 exomes (evs.gs.washington.edu/EVS) and ABraOM (<http://abraom.ib.usp.br>), the variants were classified as singleton (appear in only one individual from the 168 probands, with no frequency in the databases, rare (below 0,01 in all databases), polymorphism (between 0,01 and 0,05 in at least one database), and common (above 0,05 in at least one database). The protein alterations were classified as frameshift, stop (stopgain and stoploss), splicing (only two bases adjacent to the exon-intron junction), nonframeshift, missense, synonymous, regions of micro RNA, non-coding RNA, regulatory sites, binding sites, 3'-UTR and 5'-UTR.

All the command lines from calling to the final counting are in Attachment 22.

3.4 Comparisons pipeline

A script in R was developed to analyze different approaches to compare affected and controls based on the quantity of variants for each class, as described in the previous session. An example of the counting file is given in Attachment 23.

This report in R is a compilation of different common strategies for analyzing whole exome sequencing (WES) with four multivariate models. The report starts with an exploratory analysis of the total variants, first with the distribution of the common synonymous variants (frequency above 0.05) to address possible outliers in the sample, using the rational that most common synonymous variants have a neutral pattern of selection and are uniformly distributed in the populations worldwide, expecting similar numbers of variants in this category for all individuals. Second with a histogram of all the variants, to observe the different pattern of distribution for each variant and incorporate this information to the analysis. Also, it allows the observation wether a particular variant had a different pattern from what was expected, indicating possible errors in the process.

The next step of the report is to evaluate if the whole sample can be separated into different clusters. Due to the genetic evidence that considers ASD not as a single disease but as a number of etiologically distinct conditions (la Torre-Ubieta et al. 2016), or both extremes of a normal continuum (Bourgeron 2015), also the factors contributing to differences in the sequencing process, and particularities of each individual, it is prudent to separate groups by a clusterization method to have more homogeneous sets for the final comparison.

A non-hierarchical method (K-means) was chosen because they are less susceptible to atypical samples, to the distance measures and to the possibility of irrelevant variables (Hair et al. 2009). The method used for the k-means was the Hartigan-Wong (Hartigan and Wong 1979). The seeds were chosen at random based on the initial number of clusters, and were repeated 50 times each. In a case of two clusters, for example, the k-means were calculated 50 times based on two initial individuals. We understood the limitations of not providing a matrix of centers for the initial seeding process, and we chose to continue the analysis in this format for two reasons. First, we considered 50 random choices of two pairs to be sufficient to represent the possibility for the best result because it would involve almost two-thirds of the whole sample. This could generate, at random, a better result than to start with a preconceived tree from other methods. The second reason was because with the R package stats (Team 2017), that has the function kmeans, used in this work, would be possible that no point would be closest to one or more centres, returning an error. The ideal number of clusters were chosen between a number from one to 15, comparing the total within-clusters sum of squares. The last point with a high difference from the previous (the inflection point) was considered the best option. For the purpose of the analysis made in this study all the reports had two clusters as the best choice.

The results obtained from the different clusters could be interpreted in three ways. First, findings present in the whole sample and confirmed in the clusters can be considered a robust general behavior of the disease under investigation. Second, findings present in only one cluster can be the result of a specific relation evident in a more homogenous grouping, and will require further investigations. Third, for characteristics present in the whole sample, but not confirmed in the clusters, can be interpreted as a general characteristic that requires a higher quantity of samples, or this characteristic is a bias from the presence of heterogeneous grouping and representations of the affected and controls.

After the cluster determination, in order to access possible deviations in sequencing and annotation a correlation matrix of all variants was calculated. For coding regions it was expected a smaller number of disruptive mutations than the variants considered neutral, therefore we expect a smaller number of frameshift mutations than synonymous mutations. Following this same line of thought we expect the number of common mutations to be higher than the number of rare mutations within the same type of variant (Lek et al. 2016).

Despite this properties of frequency the mutations occurs at random and different individuals should have similar amounts of each type of mutations, therefore it is not expected significant correlations between different types of variants. Therefore, deviations from this pattern can be interpreted as characteristic of the sequencing and annotation workflow and should be taken in consideration for future interpretations. Nevertheless, if there is difference between the correlation matrix of affected and controls, it could be possible that, for complex diseases, certain types of mutations may occur together, what can lead to this deviation from the expected. All the correlations were calculated with the R package psych (Revelle 2017), function corr.test, with Pearson correlation test as the correlation method, the graphical display was done with the corrplot package (Wei and Simko 2017). the p-value matrix shows

in the lower diagonal the p-values for the correlations, and in the upper diagonal the p-value corrected for multiple tests with the holm method, considered less conservative than normally applied tests, such as Bonferroni test (Team 2017).

After the initial exploratory analysis of the variants to determine degrees of possible bias in the sample, the first attempt to identify a possible mechanism to discriminate between the groups of interest was to investigate if a particular variant can have a major effect for the affected. All comparisons were done with the Wilcoxon rank sum test, considered the same as the Mann-Whitney test (Field 2017), with the R package stats, function wilcox.test. This test is the equivalent of Student t test for non-parametric samples, and is specific for comparing two groups using ordinal arrangements. This ordinal grouping can be applied to the ASD model because of the additive effect expected, for example, one is different from two and could be considered one degree below, if one was for male and two for female, this same ordinal relation would not apply. The same approach was done for the clusters, to observe if some of the previous findings could be attributed to a deviation in the number of affected individuals within each cluster. The graphical display was done with the ggplot2 package from R (Wickham 2009).

The data the investigation continues with more elaborate questions about the possible models that could explain, and discriminate between affected and controls. There are three main approaches that will be described in further detail, one with loss-of-function and missense mutations, other with Multivariate Models, and a last with Principal Component Analysis.

3.4.1 Rare Loss-of-function and missense

Previous studies found an excess of rare disruptive mutations (LoF and missense) in individuals with autism spectrum disorders (ASD) than controls (Devlin and Scherer 2012; Iossifov et al. 2012; Krumm et al. 2015; Stein, Parikshak, and Geschwind 2013). According to these findings the first approach for whole exome sequencing analysis in the clinical laboratory is to look for likely gene disruptive variants and report as related to the clinical phenotype. For this reason, the amount and severity of this type of variants were compared between ASD and controls.

All the loss-of-function (LoF) mutations (frameshift, canonical splicing sites and stopgain/loss) were summed into one category of LoF mutations and further separated by population frequency, and rare and singleton variants were tested separately. A contingency table was calculated in three different manners, in order to test different categorization approaches, instead of separating the individuals in two categories of presence and absence of LoF mutations. First the individuals were divided based on frequency, with the attempt to have the same number of individuals in each category, second divided for presence of any number of mutations versus absence, and third the groups were divided by a non-hierarchical clusterization method. This was done separately for the rare and singleton variants, with the

calculation of the Pearson χ^2 test (PX2), the best choice for unconditional trials (Ludbrook 2008). Only the p-values below 0.05 were taken into consideration for further discussion. Since the contingency table is limited to grouping the individuals into two categories, in order to test for each number of LoF in particular, a receiver operating characteristic (ROC) curve was plotted for each number of mutations.

The ROC curve is a graphical approach to select classifiers based on their performance (Fawcett 2006). It is constructed based on the proportion of true positives (TP) (positives correctly classified as positives divided by the total number of positives) by false positives (FP) (negatives incorrectly classified as positives divided by the total number of negatives) (Fawcett 2006). Each quantity of LoF was considered as a threshold to classify the individuals, a contingency table was calculated for each quantity and the results of TP and FP were plotted, the more a point deviates from the 45 degrees line, more this particular point is capable of discriminating between the groups (Swets, Dawes, and Monahan 2000).

The missense approach was different, because of the high number of non-synonymous variants in each subject and therefore necessity to discriminate the mutations with higher probability of disrupting the protein function. Also, missense mutations have a highly specific pattern and tend to hit residues extremely close to each other and in the same domain, this occurs specially in Wnt and synapse pathways (W. Luo et al. 2018). The prioritization of the variants was conducted with the software SPRING (Snv PRioritization via the INtegration of Genomic data) (J. Wu, Y. Li, and R. Jiang 2014). This software is “based on six functional effect scores calculated by the existing methods (SIFT, PolyPhen2, LRT, MutationTaster, GERP and PhyloP) and five association scores derived from a variety of genomic data sources (gene ontology, protein-protein interactions, protein sequences, protein domain annotations and gene pathway annotations)” (J. Wu et al. 2014). The software is valid for diseases whose genetic bases is poorly known and was tested for the capability to detect pathogenic *de novo* mutations in ASD, using 34 gene seeds to prioritize the query.

For this study, the variants of interest were separated from the annotated VCF file, already filtered for the list of genes. For the seeding genes we used the same list to test whether affected and controls would have different patterns of distribution just with the damaging predictions from the software. After the SPRING prioritization, all the variants received three numbers, the first concerns to the rank for their strength of association with the query disease, the second the p-value, which is a statistical significance that the variant is causative for the disease and the third, the q-value which is the converted p-values for addressing the multiple test correction problem.

The missense variants were separated into two groups, of rare and singleton variants, then each of these groups were evaluated by the SPRING software to rank all the variants based on the specific gene list and the pathogenicity, only variants with rank one were used for further analysis. The q-values were log transformed and the samples were compared in two different patterns. First the distribution of the number of variants for each log(q-value) for affected and controls, and second, the average of each log(q-value) per individual, also separating affected and controls. It is expected that if there was an enrichment of likely gene disruptive mutations in the ASD group

it would appear as a deviation towards the higher log(q-value) for the ASD group, in both comparisons.

Both approaches, with LoF and missense variants, can be considered as proxy for the interpretation of whole exome sequencing in the laboratory routine. The analysis is focused on the individual, searching for the most disruptive rare mutations and associating them with genes previously described as related to ASD.

3.4.2 Discriminating models

Four multivariate discriminating methods were chosen to initially test the difference between the groups of interest, (a) logistic regression, (b) decision tree, (c) neural network and (d) support vector machine. These methods were chosen because they are adequate for binary data and use machine learning techniques in different degrees of depth.

3.4.2.1 Classification Performance

All the multivariate approaches applied in this work will provide a discriminate model, which stress the importance of evaluating the model in its power to discriminate between the groups of interest. The evaluating approaches chosen in this study were: the area under the ROC curve (AUC), the Kolmogorov-Smirnov (KS), and the Hand metric (H). All the metrics were obtained with the function HMeasure from the package hmeasure in R (Anagnostopoulos and Hand 2012).

In situations when the discriminant model returns a probability of the event it is possible to construct two normal distributions for the true positive and false positive rates plotting at the x axis the probability estimated by the model for each individual (Swets et al. 2000). If each probability point, the x axis determined above, were plotted to a graphic were the x axis is the rate of false positive and the y axis is the rate of true positives, the union of all the points for each probability is the ROC curve for that model. The AUC is the area under this curve, and it measures how much the model classifier is superior to a classifier chosen at random (Fawcett 2006). The higher the AUC more accurate is the model in discriminating between groups, the value to consider a good model will be attributed depending on the specific problem and commonly the threshold is established by a previous model for the same individuals (Fawcett 2006). At the reports the threshold established was 0.8 because of a deviation in the number of initial individuals from control and ASD groups would bring an initial discriminating model by chance of around 0.66 (two-thirds of the samples in this study is control individuals).

The KS is a non-parametric statistic developed to test whether the population has a specific distribution, because the test compares the distribution seen in the population with a hypothetical known distribution (Field 2017). A modification of the original use can be done to compare between two distributions, allowing the test to be used as a measure of goodness of fit for predictive models (Qin 1997). There is no established threshold for this metric, because it is a comparison metric, differently from the AUC that starts from the 0.5, which is equivalent of choosing at random, this metric starts from zero, therefore we established the threshold of 0.4 for this study purposes.

H measure was described in 2009 by David J. Hand as an alternative to previous metrics such as AUC and KS, mainly because it takes into consideration different penalties for misclassifying classes 0 or 1 (Hand 2009). This is particularly important if we are dealing with a test that has different consequences to each group. For the purpose of this study we set the worse misclassification as loosing a true positive case, in other words, the misclassification as false negative was more problematic. This decision was made because most individuals that come for a genetic test already have a clinical ASD diagnosis, therefore the probability of this individual to be a true ASD case is higher. On the other hand, individuals that come for different diagnosis cannot exclude ASD, which increases the probability of having an underlying ASD diagnosis. For this reason the parameter severity.ratio is set to 0.75. The same threshold of 0.4 as KS was admitted.

The number of individuals classified as false positive, false negative, true positive and true negative are calculated for each prediction model.

3.4.2.2 Logistic Regression

From all the regression models the logistic regression was chosen for this study because the variable response, considered here as affected or not affected, follows the Bernoulli distribution of zero and one (BARRETO 2011). The logistic regression will construct a predictive model with the final result being an log odds, which can be interpreted as the probability of the response variable designated as one, in this case affected. The logistic regression is not as sensitive to scale and distribution of the original independent variables as the linear models are (BARRETO 2011) allowing the data to be used in their original format with no further manipulation. For the logistic regression, the `glm` function from the package `stats` in R was used. This function calculates the generalized linear models and can be applied to many different models, to specify that it concerns a logistic regression the parameter `family=binomial(link='logit')` is set in the formulae.

The samples were divided in two, half of the individuals for the learning group and half to the test group, using a random sampling obeying the original proportions of the ASD and control individuals, with function `createDataPartition` from the package `caret` in R {Kuhn:C8_eJz_I}. From this first model a backward approach was applied

to select the variables for the best model with the function step from the package stats in R. The backward approach consists in the removal of the variable with less effect and the recalculation of the logistic regression and the Akaike Information Criteria (AIC). The AIC measures the goodness of fit and simplicity of the models using the concept of entropy, when comparing between two different models, the one with the smaller number of AIC will be a better model (Akaike 1985). The step function tests all models until it finds the one with the smallest AIC, which will be interpreted as the best model. This final model was applied to the testing sample and from this result was obtained the ROC curve and the discriminating metrics AUC, H and KS. The true positive, false positive, true negative and false negative were calculated with the best discriminant threshold found.

3.4.2.3 Decision Tree

The method of classification and regression trees (CART) was described in 1984 by Breiman et al. as an improvement of the current tree methods, and an alternative to linear regression models (Breiman 2017). The CART is a binary recursive partitioning procedure that handles data in their raw form. Beginning in the root node, with all the individuals, the data are split in two children with less entropy from the parents node, and this process is continued until a stop rule is applied, or they can grow to a maximum size and then be pruned back (Hastie, Tibshirani, and Friedman 2013; X. Wu et al. 2007).

In this study we applied the classification tree, which was let to grow to the maximum size, and then applied the cross-validation with a testing sample. The individuals were divided into learning and test groups, as described for the logistic regression. The variables used to construct the tree and the information about each split were printed. From this testing result we obtained the fitting of the model, making it possible to select the best tree for that dataset. The error columns have been scaled so that the first node has an error 1.

The best model is the one with the minor error rate, that avoids overfitting (Hastie et al. 2013), for this reason pruning the tree could be necessary, to obtain the best possible model without overfitting. The *xerror* is an important metric to support the pruning decision. The *xerror* is a summary measure of the fit of the model to the testing sample, the better the fit the smaller *xerror* is. The pruning can follow the rule of stopping in the split which has the small improvements from the anterior. All the decision tree calculations were done with the rpart package (Therneau, Atkinson, and Ripley 2017).

3.4.2.4 Neural network

The neural network is a nonlinear statistical model that consists in a two-stage regression or classification model, typically presented by a network diagram (Hastie et al. 2013). The neural network is a common approach to machine learning, with many modifications to improve specific questions (Anastasiadis, Magoulas, and Vrahatis 2005; Riedmiller, undefined author, 1993 1993). The one applied in this study is a resilient backpropagation with weight backtracking (Riedmiller et al. 1993). The backpropagation is the most widely used algorithm for supervised learning and it consists in the repeated application of a specific rule to compute the influence of each coefficient in the network regarding an arbitrary error function (Riedmiller et al. 1993). The resilient backpropagation performs a direct adaptation of the weight step (which determines the coefficients) based on local gradient information. This modification reduces the number of learning steps with less computational effort (Riedmiller et al. 1993).

Neural network computations were done with the neuralnet package in R (Fritsch and Guenther 2016). For better results the original dataset was scaled from 0 to 1, two hidden layers were chosen with 15 and 5 neurons, respectively. The final result is applied to a testing sample (in the same mode as already described) to compute the model goodness of fit.

3.4.2.5 Support Vector Machine

The support vector machine is a machine learning approach that searches for the optimal hyperplane to separate the groups of interest. The difference of the support vector machine from other linear boundary classifications, is that it produces nonlinear boundaries by constructing a linear boundary in a large, transformed version of the initial space (Hastie et al. 2013). Each data is plotted in an n-dimensional space (being n the number of variables computed). The support vector machine will construct hyperplanes that best separate the groups not restricted to three dimensions.

The choices of the hyperplane are very similar to the process of the decision tree, the support vectors have to separate the samples to have groups with minimal intersection, this is a particular challenge to avoid overfitting (Hastie et al. 2013). All these computations were done with the e1071 R package (Meyer et al. 2017). We applied the linear kernel because we did not have previous information for what would be the best method to compute the hyperspace, we used the classification approach for the support vector machine, calculating the probabilities of each classification, and plotting the probability associated with the target. The model was applied to a testing sample to determine the discriminant metrics.

3.4.3 Principal Component Analysis

The principal component analysis is a multivariate approach that reduces the variance of the samples from a correlation or covariance matrix to vectors of complex variables that is related to the information of more than one variable of the original dimension. In summary, the information within one principal component aggregates the information of more than one variable regarding the correlation or covariance between them (Mingoti 2005).

For this study two approaches were used, the first with the counting matrix and the second with the raw VCF file. Principal components were calculated for the counting matrix with the prcomp function in R stats package, for this purpose the covariance matrix with all variables shifted to be zero centered and scale to a unit variance. These modifications had to be done due to the highly different amount of variants in each class.

The raw VCF approach was done with the function snpgdsPCA from the R package SNPRelate (Zheng et al. 2012). This package was created to accelerate the computations on SNP data for principal component analysis and identity-by-descent measures to decrease the computational time of the GWAS studies. To compute the principal components they construct the covariance matrix from all variants in the raw VCF file, with the main goal of finding and correcting population structure that could interfere with the GWAS analysis. This functionality was co-opted to this study and the population structure investigated was the difference between ASD and control groups.

The R script for the formulation of the report with all the analysis above is in Attachment 24.

4 Results

All the results described in this chapter will be related to the files created with the scripts detailed in Material & Methods, because of the great number of pages generated by each report, they are not being printed with the thesis, instead can be accessed in the link below.

https://www.dropbox.com/sh/cqyhdrygfhoy3b/AADXrpaMS_emgnWyL-l7mnSFa?dl=0

The name of the respective file will be the same as the name of each chapter. For example, the file related to the exploratory analysis will be: exploratory-analysis.html (Figure 10). All the files are in html extension which can be open in any device, desktop or mobile. In the eventuality of an issue with the browser all the files are also available in pdf format.

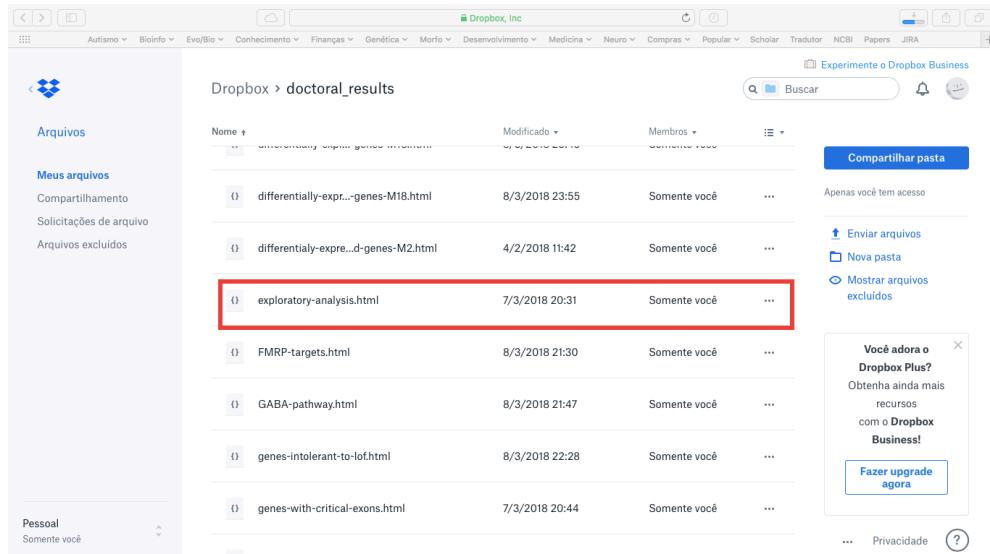


Figure 10: Visualization of the arquives in the reports folder, the report inside the red box is the exploratory-analysis.html.

4.1 Exploratory Analysis

This description concerns the first report done with the whole exome sequencing and no gene filtering (Figure 11). This is an example of the steps taken for the interpretation of the full report. For the subsequent reports concerning specific gene filtering only the results that are significant at a level of p-value inferior to 0.05 (marked in red in the report - Figure 12), or, for the multivariate models, three discriminant metrics, H, AUC and K, above the thresholds: H > 0.4, AUC > 0.8 and K > 0.4 (marked in blue in the report - Figure 13) will be discussed. The principal

components analysis will be described if there is a notable difference between the dispersion of the groups after the removal of the outliers.

The screenshot shows a web browser window with the title "Target X Control Analysis" and the author "Tatiana Almeida". The date "07 Março, 2018" is also displayed. Below the title is a bulleted list of analysis components:

- Exploratory Analysis of the variants
 - Variants distribution
 - Kmeans
 - Correlation matrices
 - Correlation matrix for all individuals
 - P-value matrix for all individuals
 - Correlation matrix for target individuals
 - Correlation matrix for control individuals
 - Comparison for each variant between affected and control
 - Comparison for each variant for each kmeans cluster
- Comparison analysis
 - Loss-of-function grouping
 - Loss-of-function ROC curve
 - Logistic Regression
 - Decision Tree
 - Neural Network
 - Support Vector Machine
 - Classification Performance
 - PCA with counting variants
 - PCA with raw vcf
- Cluster 1
 - LoF
 - Loss-of-function grouping
 - Loss-of-function ROC curve
 - Logistic Regression
 - Decision Tree
 - Neural Network
 - Support Vector Machine
 - Classification Performance

Figure 11: Screenshot of the index of the exploratory analysis.html report.

The screenshot shows a table titled "Logistic Regression" with columns "term", "estimate", and "p.value". A red box highlights the "p.value" column. The table includes rows for various terms such as (Intercept), ur5_fr1, ur5_fr2, ur3_fr3, ncRNA_fr1, ncRNA_fr3, ncRNA_fr4, miRNA_fr3, miRNA_fr4, bnd_fr1, bnd_fr2, reg_fr1, and reg_fr3. The "p.value" column contains values like 0.005, 0.032, 0.034, 0.007, 0.145, 0.002, 0.064, 0.02, 0.016, 0.053, 0.031, 0.126, and 0.076.

term	estimate	p.value
(Intercept)	-7.029	0.005
ur5_fr1	-0.004	0.032
ur5_fr2	0.058	0.034
ur3_fr3	-0.013	0.007
ncRNA_fr1	0.003	0.145
ncRNA_fr3	0.028	0.002
ncRNA_fr4	0.152	0.064
miRNA_fr3	0.972	0.02
miRNA_fr4	-3.421	0.016
bnd_fr1	0.001	0.053
bnd_fr2	-0.009	0.031
reg_fr1	-0.027	0.126
reg_fr3	0.132	0.076

Below the table is a section titled "Decicion Tree" with a small diagram showing a node labeled "Control 100%".

Figure 12: Screenshot of an example of p-values under 0.05 in the report.

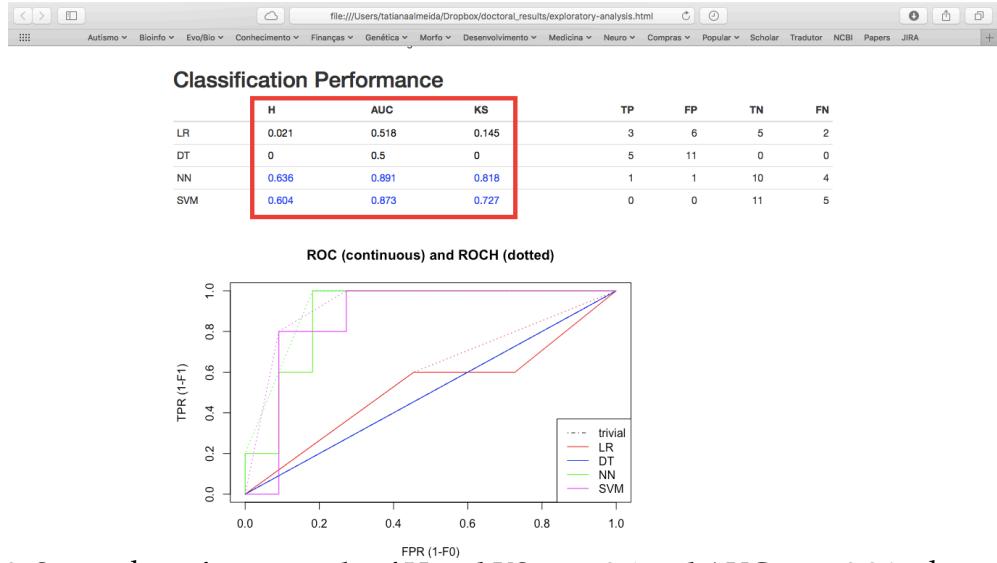


Figure 13: Screenshot of an example of H and KS over 0.4 and AUC over 0.8 in the report.

Because of the volume of the names of the variants' classes the abbreviation was conducted in two parts, separated by underline. The first part is the name of the variant and the second part the population frequency such as follows:

Type of variants	
fra	Frameshift
sto	Stopgain/loss
spl	Canonical splicing (+1,+2,-1,-2)
nfr	Nonframeshift
nsy	Nonsynonymous
syn	Synonymous
utr3	3'-UTR
utr5	5'-UTR
ncRNA	Noncoding RNA
miRNA	Micro RNA
bnd	Transcription binding sites
reg	Regulatory sites
lof	Frameshift+Splicing+Stopgain/loss
Frequencies	
fr1	Above 0.05
fr2	Between 0.05 and 0.01
fr3	Below 0.01
fr4	Singletons

As an example, if we refer to a frameshift variant present in only one individual and absent in population data (singleton), the abbreviation will be fra_fr4, if we are

referring to a regulatory site with common frequency above 0.05, it will be reg_fr1. This table will be printed in the first page of attachments 25.

4.1.1 Removing outliers

The synonymous distribution graphic shows that despite a non-homogeneous composition of the samples there are no evident outliers, therefore all the 168 individuals were maintained in the analysis.

4.1.2 Variants distribution

This analysis shows that the frameshift, stopgain/loss and nonframeshift distributions have a normal distribution (Shapiro-Wilk $p=0.98, 0.35, 0.10$, respectively). However, the splicing, nonsynonymous, synonymous, 3'-UTR, 5'-UTR, binding sites, miRNA, ncRNA and regulatory sites, contrary to what would be expected do not have a normal distribution (Shapiro-Wilk $p=0.03, 1.05e-07, 1.6e-07, 6.6e-16, 7.5e-11, 9.9e-05, 3.6e-07, 1.7e-13, 0.0001$, respectively). The histogram in the report reveals a tendency to a bimodal distribution for these variants (Figure 14).

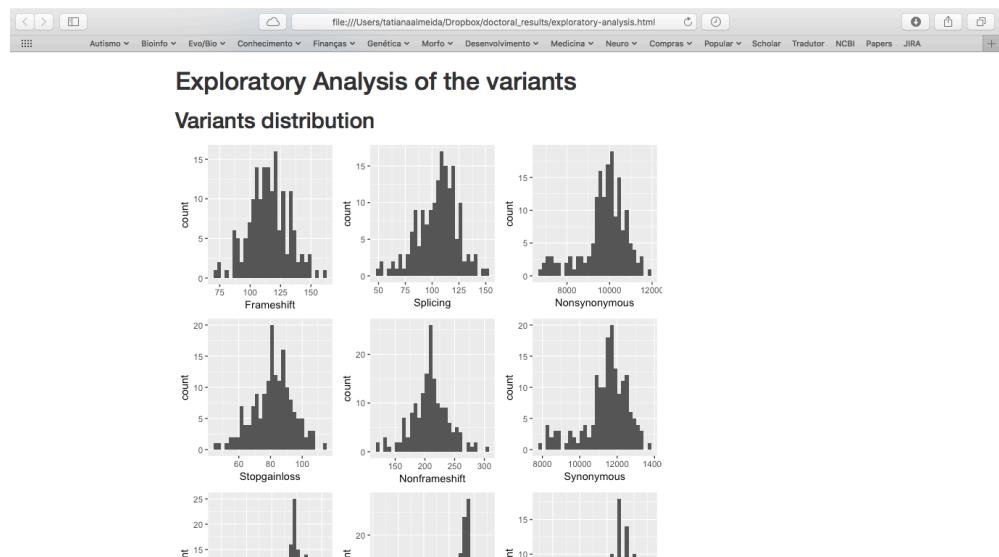


Figure 14: Screenshot of the distribution of each type of variant for all the individuals in the report.

One hypothesis for this bimodal distribution is the difference among the capture strategy and the alignment pipelines along the years. Figure 15 shows the distribution of the quantity of common nonsynonymous and synonymous variants along different batches of WES and there is no clear distinction between the number of variants and the period the WES were sequenced. The same occurred for the splicing, binding sites and regulatory sites. Therefore, the non-normal distribution for

these variants cannot be attributed to the differences in the capture and alignments over the years.

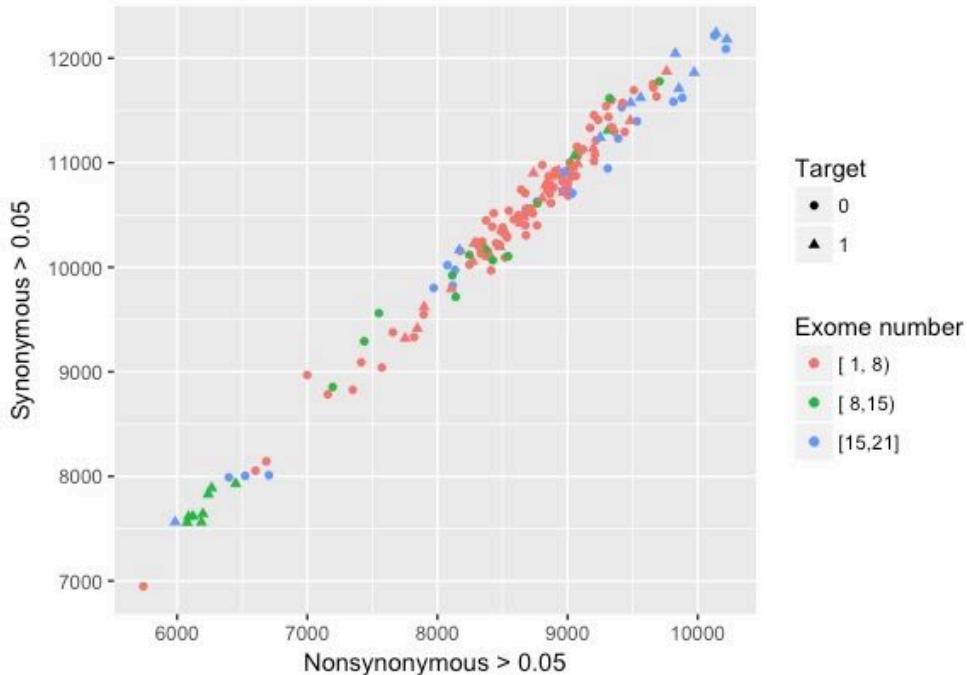


Figure 15: Synonymous and nonsynonymous distribution of the common variants for each individual separated by the number of the WES batch.

However, for the non-coding variants 5'-UTR and 3'-UTR, the distribution has a clear difference explained by the different batches of WES (Figure 16), this pattern is the same for the miRNA and ncRNA. The batches above number 14, sequenced most recently, showed less regulatory variants.

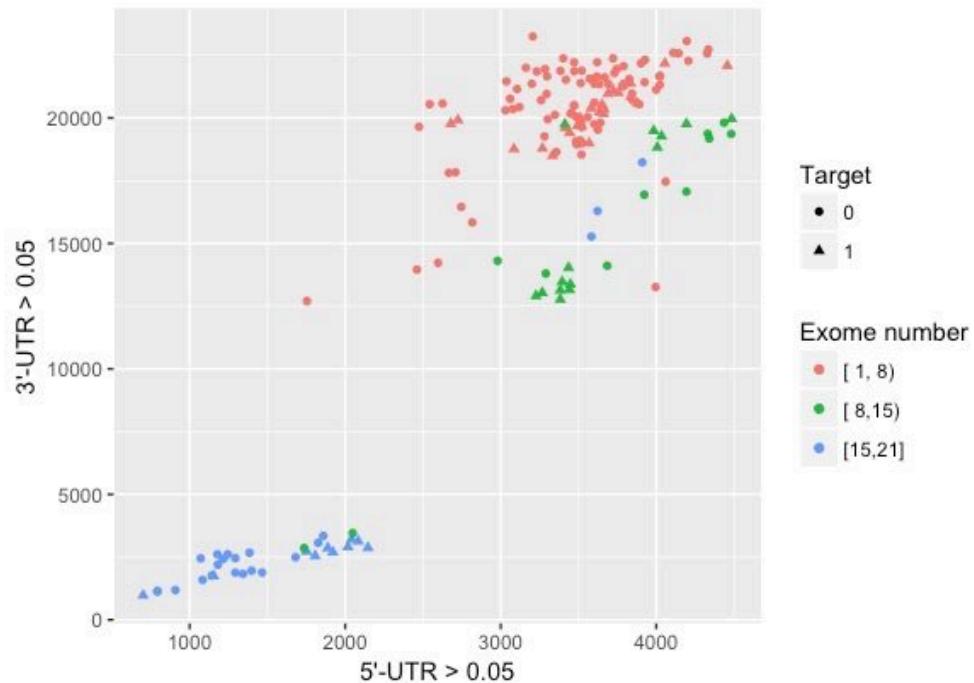


Figure 16: 3'-UTR and 5'-UTR distribution of the common variants for each individual separated by the number of the WES.

These graphics show that the capture and alignment strategies can have a strong effect, especially in the analysis of non-coding variants. Because of these differences the results concerning these regions will not be further analyzed. However, since the pipelines were constructed to be applied in any scenario including different diseases they were maintained in the reports.

4.1.3 Kmeans

The screeplot shows the inflection for the optimal number of clusters. Two clusters were chosen here and used in all subsequent analysis because it is the point where there is still a major difference in the sum-of-squares compared to three clusters (Figure 17).

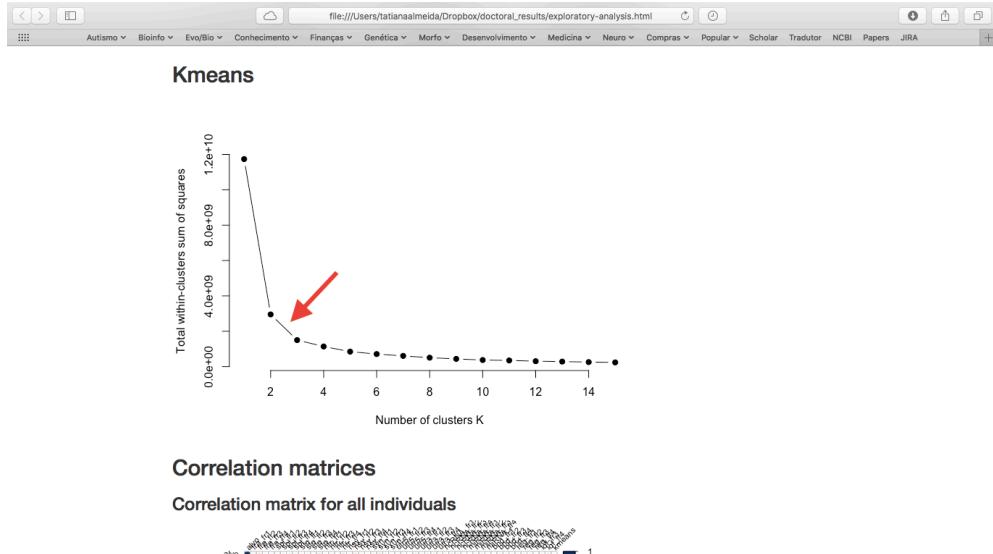


Figure 17: Screenshot of the screeplot for the kmeans sum-of-squares, the arrow marking between the points with most inflection.

The best explanation for this division, as shown in Figure 18, is the number of 3'-UTR and 5'-UTR variants for each subject, which can be a proxy for the WES capture and alignment strategy, as discussed above, strongly because the majority of individuals of the second cluster is present in more recent WES batches.

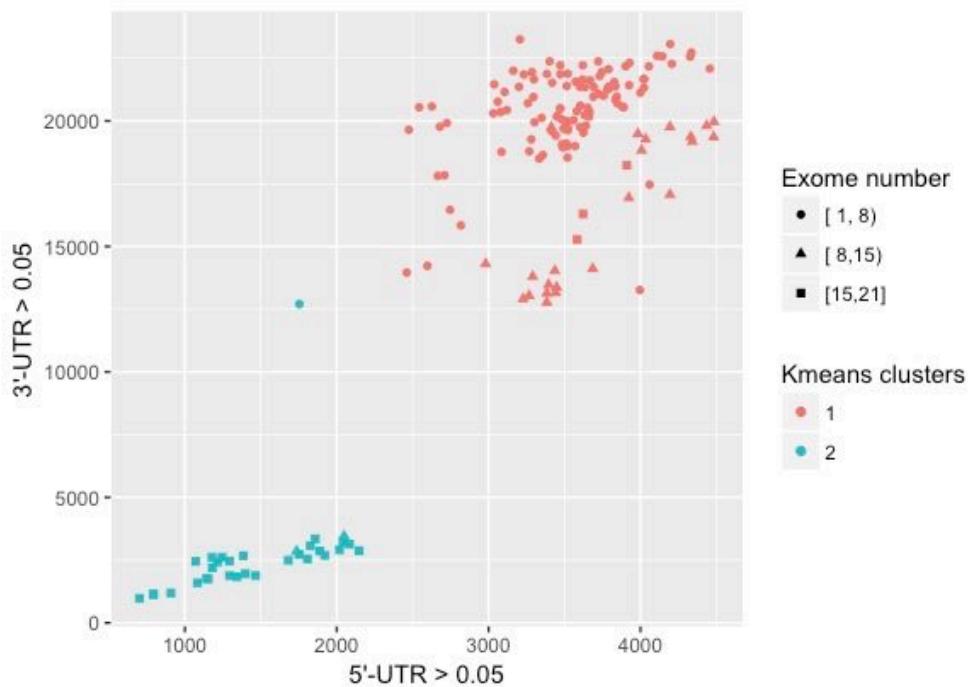


Figure 18: 3'-UTR and 5'-UTR distribution by Kmeans cluster and WES number.

4.1.4 Correlation Matrices

The *Correlation matrix for all individuals* shows the pairwise correlations among all variants (Figure 19) and *P-value matrix for all individuals* shows in lower diagonal the p-value for all the correlations and in the upper diagonal the p-value corrected for multiple tests (Figure 20).

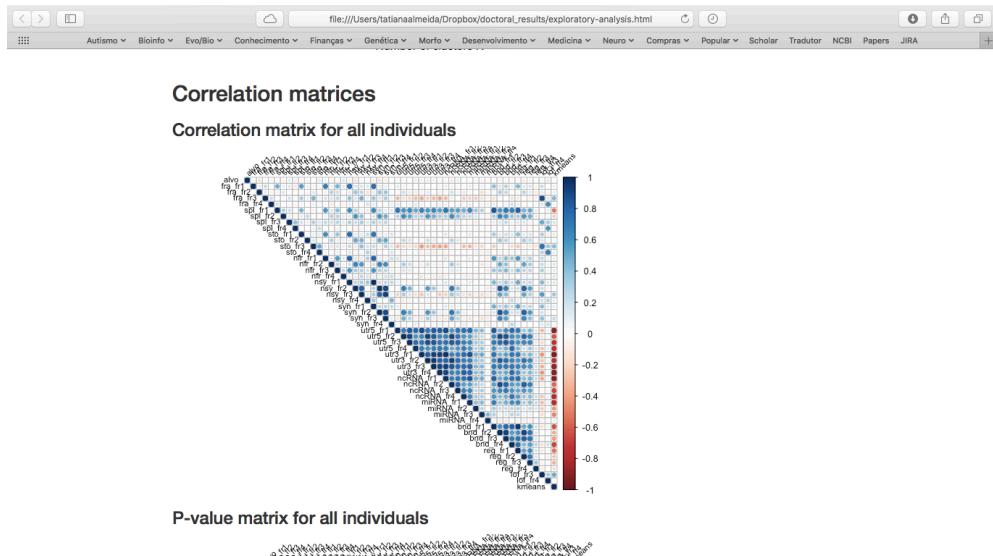


Figure 19: Screenshot of the plot for the correlation among the variants, the lighter the color, closest to zero (no correlation), the blue is a positive correlation and red a negative correlation.

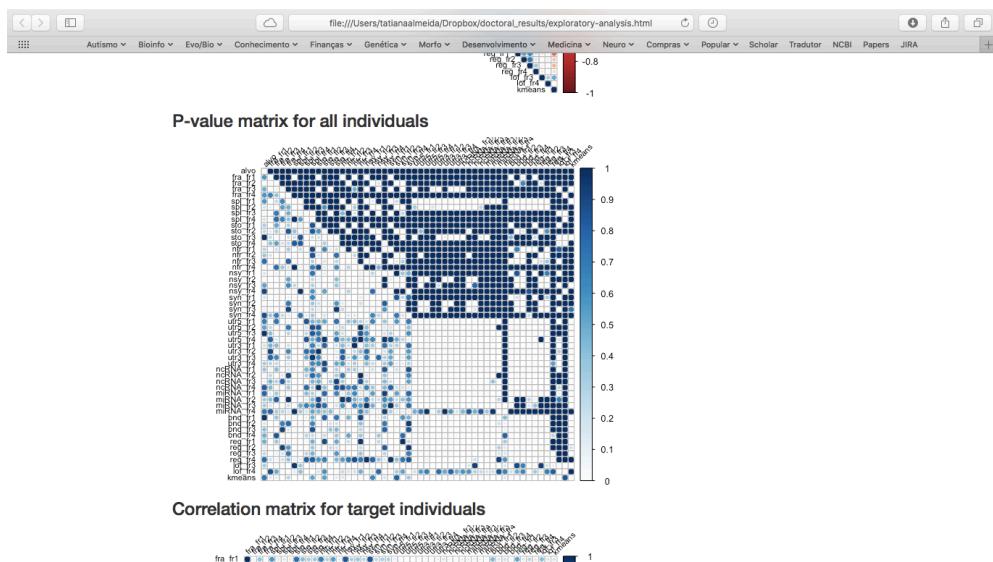


Figure 20: Screenshot of the plot for the p-values associated with the correlations above, the upper diagonal is the p-values corrected for multiple tests, the darker blue are the values closer to one.

The most prominent correlations occur among splicing, 3'UTR, 5'UTR, ncRNA, miRNA and regulatory regions. Most of these correlations can be attributed to the differences from the capture and alignments, discussed previously. Other correlations that drives attention and cannot be explained by the difference in alignment are seen in Table 1.

Table 1: Significant correlations not explained by the capture and alignment of WES.

		Correlation	P-value
Fra_fr2	Nfr_fr2	0.30792824	5.160595e-05
	Nsy_fr2	0.37571758	5.632275e-07
	Syn_fr2	0.38959341	1.951014e-07
Fra-fr3	Sto_fr1	0.33166109	1.195594e-05
	Sto_fr3	0.45095636	9.611695e-10
	Nfr_fr3	0.26410295	5.631498e-04
Sto_fr1	Nfr_fr1	0.618509682	5.266167e-19
	Nfr_fr3	0.374817077	6.023304e-07
Sto_fr3	Sto_fr4	0.50811621	2.392915e-12
Sto_fr4	Nsy_fr4	0.30196937	7.309479e-05
Nfr_fr1	Nfr_fr3	0.55806064	4.687232e-15
	Syn_fr4	0.22796829	3.046679e-03
Nfr_fr2	Nsy_fr2	0.65734783	4.993796e-22
	Syn_fr2	0.65438485	8.804785e-22
	NcRNA_fr2	0.37613606	5.458922e-07
	Nfr_fr3	0.27564389	3.116095e-04
Nfr_fr3	Nsy_fr2	0.32334838	2.023643e-05
	Syn_fr2	0.29198808	1.287941e-04
Nsy_fr2	Syn_fr2	0.97484634	1.716169e-109
	NcRNA_fr2	0.69934092	7.663895e-26
Nsy_fr4	Syn_fr4	0.453322335	7.661126e-10
	Lof_fr4	0.303181013	6.814103e-05
Syn_fr2	NcRNA_fr2	0.71913446	7.039138e-28

Since there is no gene filtering in this result it is possible to consider these correlations as a result of intrinsic factors of exome sequencing with no correlation to the discrimination of the groups of interest. The two subsequent graphics in the report show the difference in correlation for control and ASD groups, the correlations are overall similar in both groups, but stronger in ASD group.

These findings will not be further considered in future analysis because they may be related to some specific characteristics of the WES and not with the discriminating parameters of the samples.

4.1.5 Variant distribution between affected and control

The *comparison for each variable between affected and control* consists in the comparison of the quantity of variants for each individual in each group of interest (ASD and controls). The variants that showed difference between ASD and control at a level of significance lower than 0.05, were fra_fr2, nfr_fr4, nsy_fr2, syn_fr2 and lof_fr4. Variants with a higher mean in affected group were nfr_fr4 and lof_fr4, all the others had higher mean values in control group. The p-values for the Wilcoxon test are inside each graphic in the report (Figure 21).

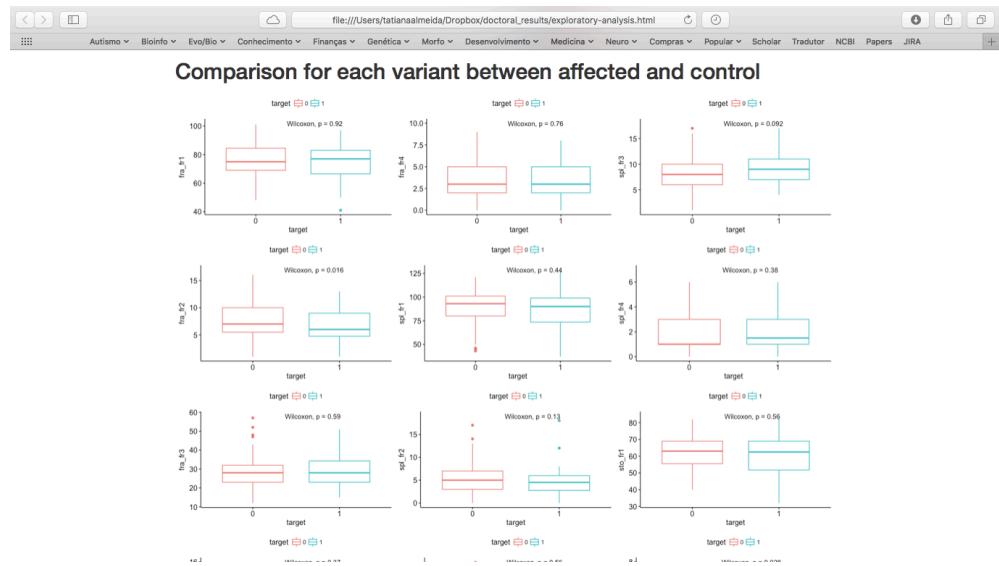


Figure 22: Screenshot of the first six graphics showing the comparison for each variant between affected and control with the respective Wilcoxon test inside each plot.

The sum of the variables with lower mean in ASD group (fra_fr2, nsy_fr2 and syn_fr2) maintained the discrimination between ASD and controls (Wilcoxon $p=0.0047$). Despite this significant p-value, the distributions of the two groups are very similar (Figure 23).

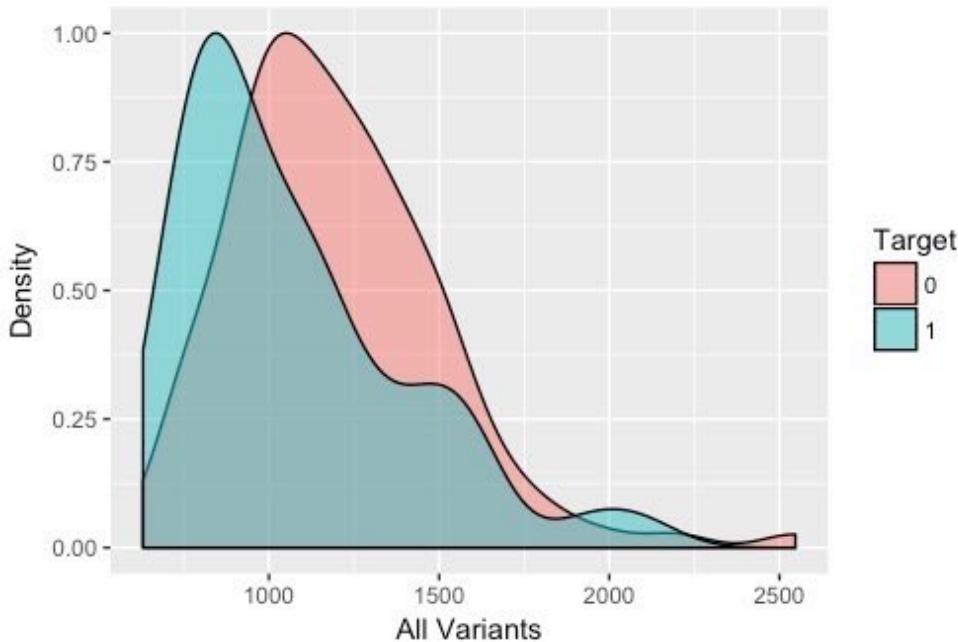


Figure 23: ASD and control (target 1 and 0, respectively) distribution of the sum of the variants fra_fr2, nsy_fr2, syn_fr2.

The sum of the variants nfr_fr4 and lof_fr4 maintains the difference between the groups (Wilcoxon $p=0.002$). The ASD group has a normal distribution (Shapiro-Wilk $p=0.47$), while the control group does not have (Shapiro-Wilk $p=0.009$), despite this difference the controls are entirely included in the ASD distribution suggesting that this is not a good discriminant between the groups.

The importance of evaluating not only the p-value for the discriminant test, but also the pattern of the distributions is mostly because significance tests have limitations for heterogeneous distributions. For this reason, when analyzing both distributions, the peak, or the number of variants for the most representative part of the groups can be different, but most of the individuals are overlapped and this difference is caused by few individuals. For this reason all significant differences will be analyzed concerning their distribution, and with both information it will be decided if the model could reveal a good discriminant.

The variables were also analyzed between the two Kmeans clusters and the ones that were statistically different were: 3'-UTR, 5'-UTR, ncRNA, binding sites and regulatory sites in all frequencies, splicing, miRNA in all frequencies except singletons, nsy_fr1, nsy_fr3, fra_fr3, sto_fr3, syn_fr3 and lof_fr3, nfr_fr2, syn_fr1, sto_fr4 (Figure 24). Because K-means separated the individuals mainly by the timeline of WES it can be argued that those differences are due to the effects discussed before. The distribution of affected and controls within each cluster is not statistically different (FET $p=0.829$). For the subsequent analysis only the variants that showed a significant difference in the first analysis will be described if they are also different between the clusters.

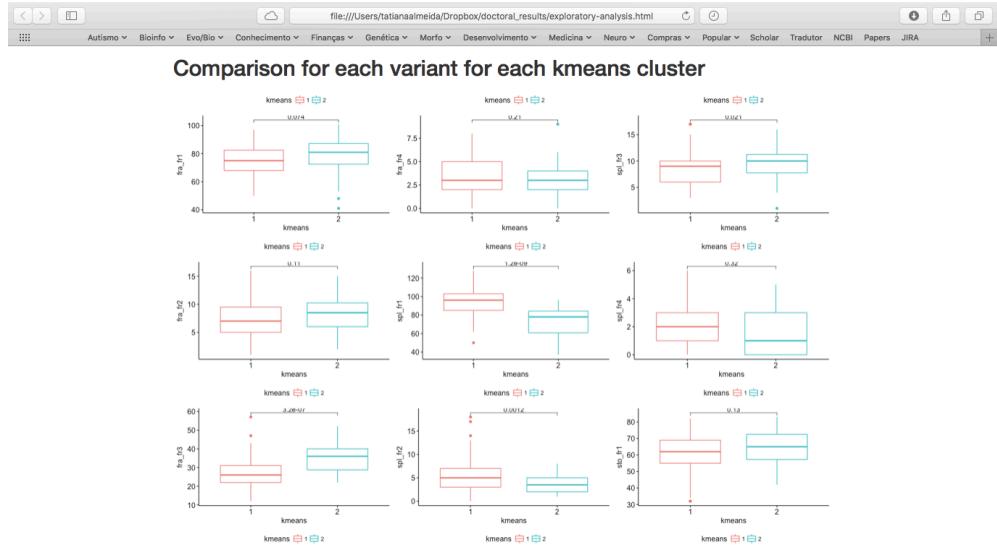


Figure 24: Screenshot of the first six graphics showing the comparison for each variant between cluster 1 and cluster 2 with the respective Wilcoxon test inside each plot.

4.1.6 Comparison Analysis

4.1.6.1 Loss-of-Function grouping

The three approaches of grouping loss-of-function (LoF) mutations, explained in detail in Material & Methods, did not result in any statistically differences between ASD and controls, in either rare or singleton frequencies (Figure 25). The ROC curve for both frequencies did not show a better discrimination than the random choice. This corroborates the findings that there is no significant differences in the total number of LoF mutation between ASD and controls (Figure 26).

Number of LoF	Affected	Control	ChiSquare
Grouping Method: Frequency			
Rare variants			
[25,48]	20.000	66.000	
[48,92]	28.000	53.000	0.149
[25,48]	0.417	0.555	
[48,92]	0.583	0.445	
Singletons			
[0, 8)	24.000	67.000	
[8,17]	24.000	52.000	0.57
[0, 6)	0.500	0.563	
[8,17]	0.500	0.437	
Grouping Method: Presence versus Absence			
Rare variants			
0	0.000	0.000	
[1,92]	48.000	119.000	99
0	0.000	0.000	
[1,92]	1.000	1.000	
Singletons			

Figure 25: Screenshot of the first comparison for LoF variants in the report.

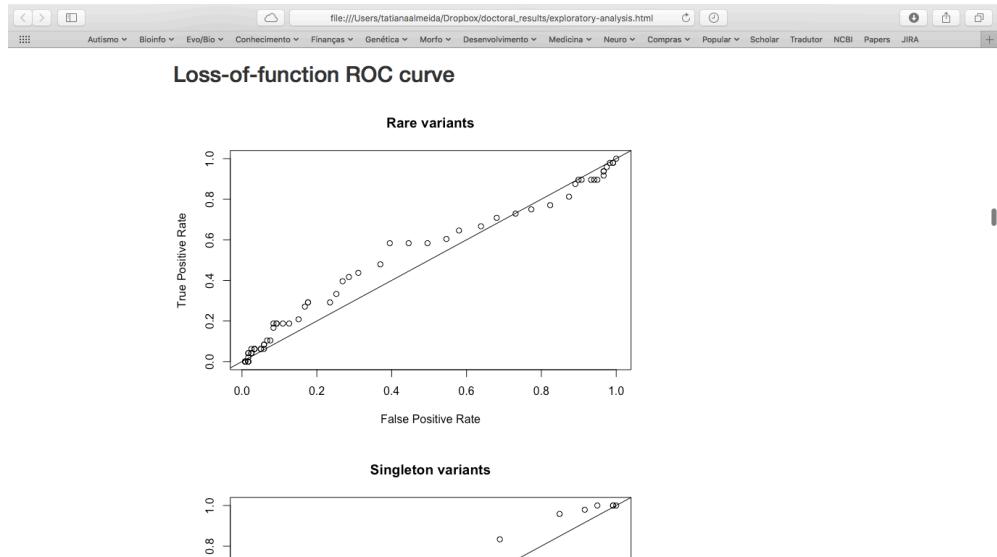


Figure 26: Screenshot of the ROC curve for all the numbers of LoF variants for this analysis.

This finding is different from the one found in the *comparison for each variable between affected and control* for the *lof_fr4*. The p-value for the *lof_fr4* is nearly non significant in the first scenario, one of the reasons why this finding was not confirmed in this analysis.

4.1.6.2 Multivariate models

In this section it will be discussed the Logistic Regression, the Decision Tree, the Neural Network and the Support Vector Machine. No models were efficient in separating affected from controls (Figure 27).

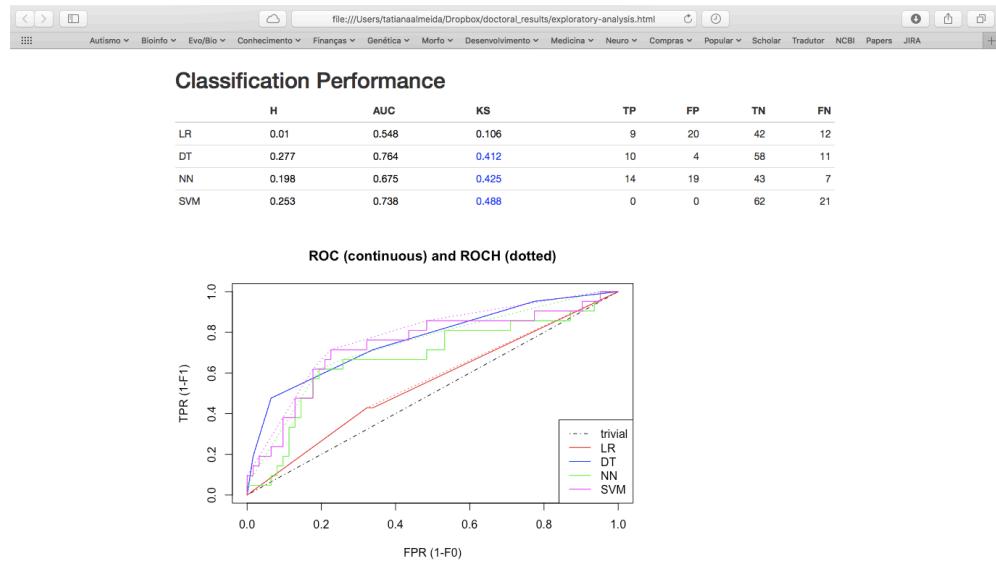


Figure 27: Screenshot of the final comparison among all four multivariate approaches to the discriminant model. This session shows the performance for the good-of-fitness metrics and the ROC curve comparison.

The best model from logistic regression had 24 variants showed at the report, for the decision tree the best model included the variants utr3_fr3, bnd_fr1 and ncRNA_fr3. The neural network and the support vector machine were the best models, but they did not reach the threshold for the H and AUC metrics.

4.1.6.3 Principal Components Analysis

The section - *PCA with counting variants* - represents the first three principal components in two graphics (PCA1 vs. PCA2 and PCA2 vs. PCA3) extracted from the matrix of the number of variants for each category, example of the matrix in Attachment 22. Both graphics show no separation for the control and ASD groups, the variables responsible for the first component are mainly regulatory variants (utr3, utr5, bnd, miRNA, ncRNA, bnd and reg), this could be explained by the major differences in capture and alignment technique (Figure 28). This effect is removed with the analysis of the coding variants (described in a subsequent section), and with the visualization of the second graphic with the second and third components.

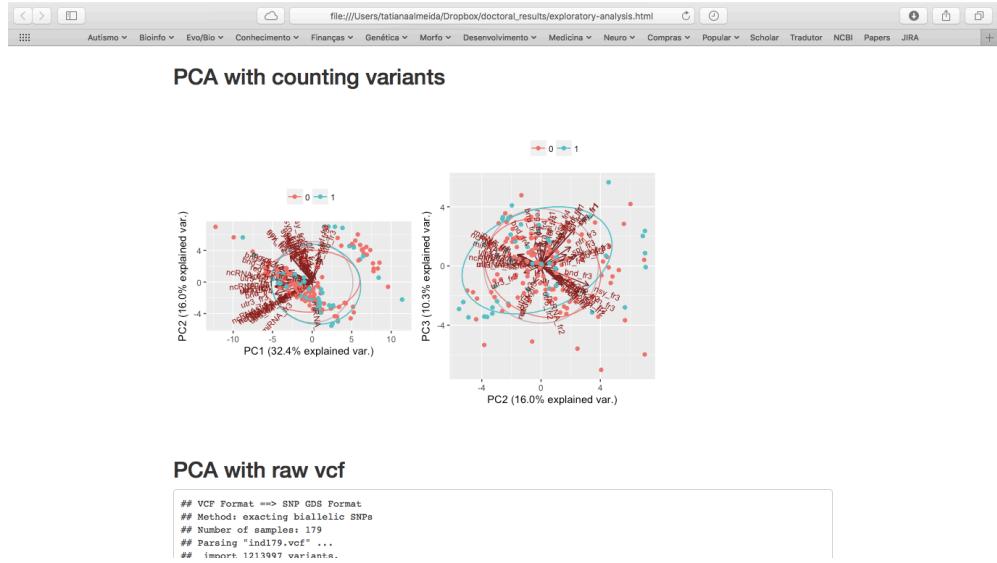


Figure 28: Screenshot of the graphic of PCA with counting variants in the report.

The section - *PCA with raw vcf* - refers to the principal component analysis showing the first three components constructed with the VCF file directly. The graphic in the report shows two clear outliers the CC5098, and the CC5200 (Figure 29). Figure 30 shows the PCA after the two outliers' removal, now there is three new outliers (C5054, F7935-1 and F8442-1). Figure 31 shows the first three PCs after the removal of the outliers, where four other outliers remained CC5064, CC962, F8385-1 and CC5305.

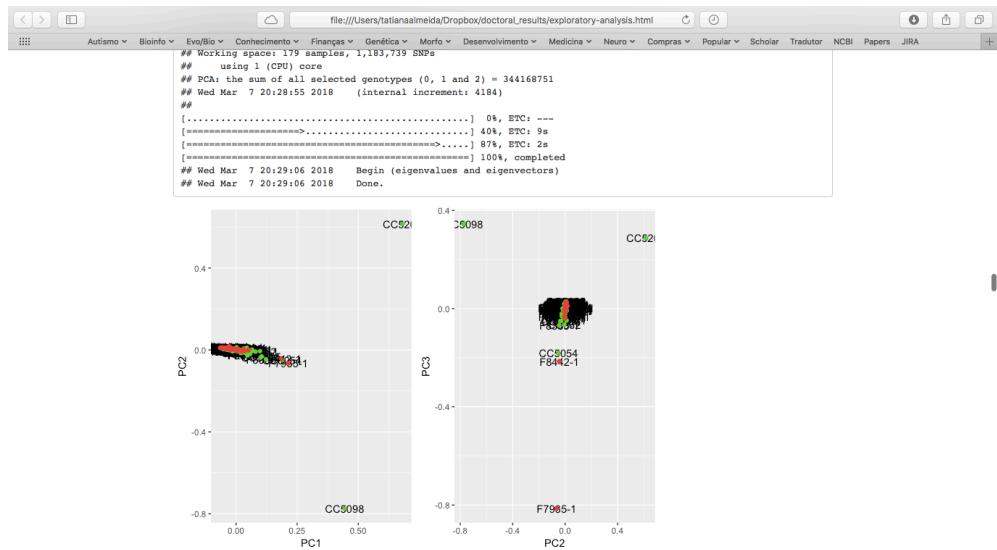


Figure 29: Screenshot of the PCA with raw VCF graphic in the report. The green dots are the control group and red dots are the ASD subjects.

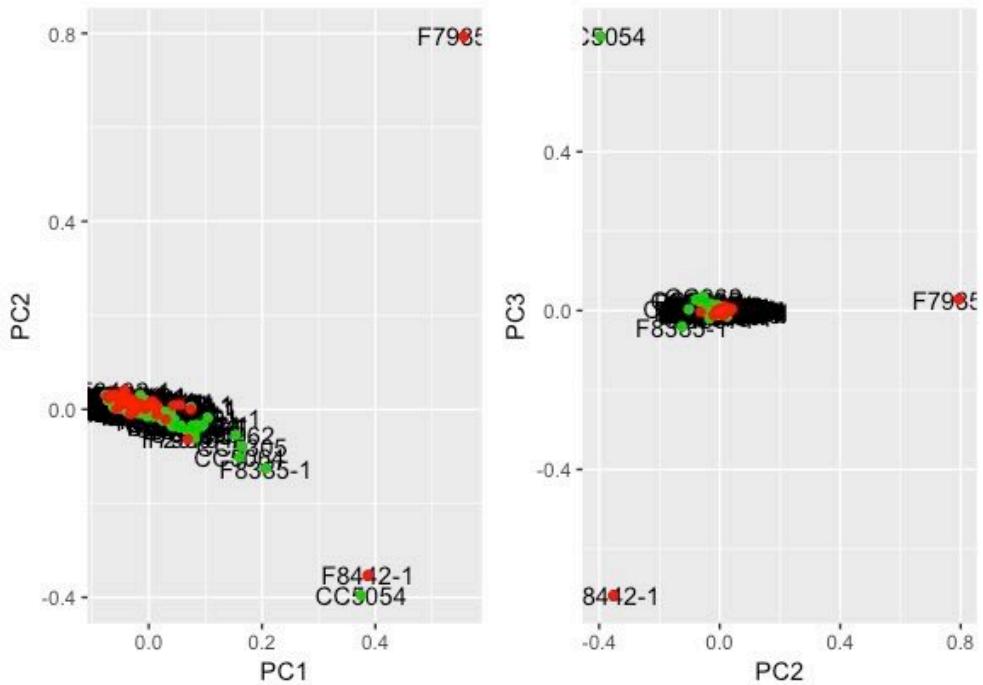


Figure 30: First three principal components from the variant call file, after removal of the CC5054, F7935-1 and F-8442-1. The green dots are the control group and red dots are the ASD subjects.

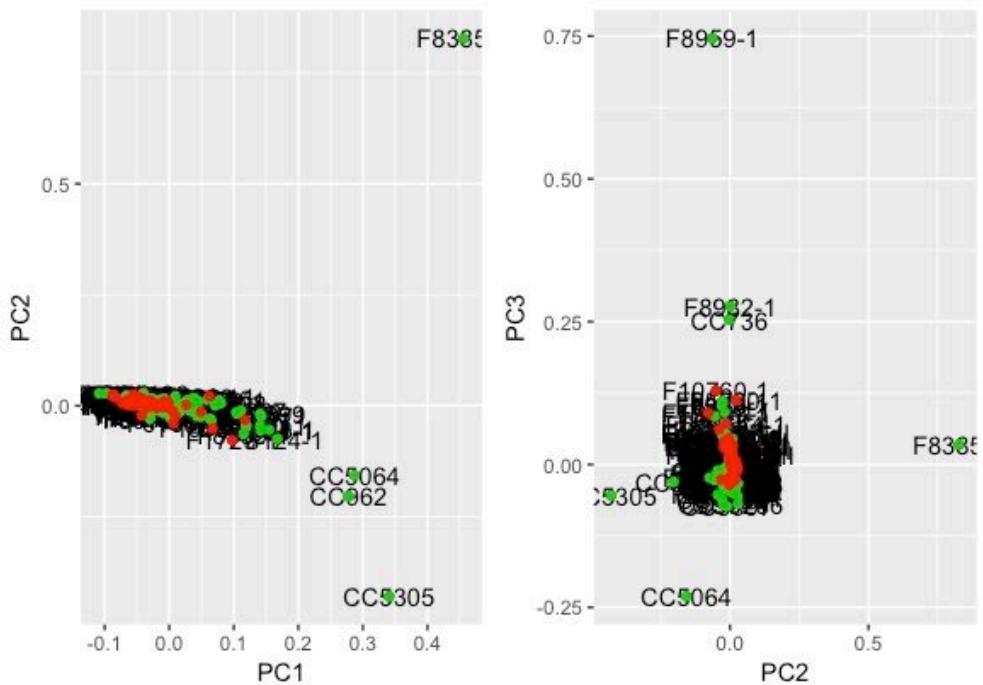


Figure 31: First three principal components from the raw vcf, after the removal of CC5054, F7935-1 and F8442-1. The green dots are the control group and red dots are the ASD subjects.

After the removal of CC5064, CC962, CC5305 and F8385-1 there is a higher dispersion among the samples (Figure 32). For this reason the outliers, F8959-1, F8982-1 and CC736, were removed, as it can be seen in Figure 33.

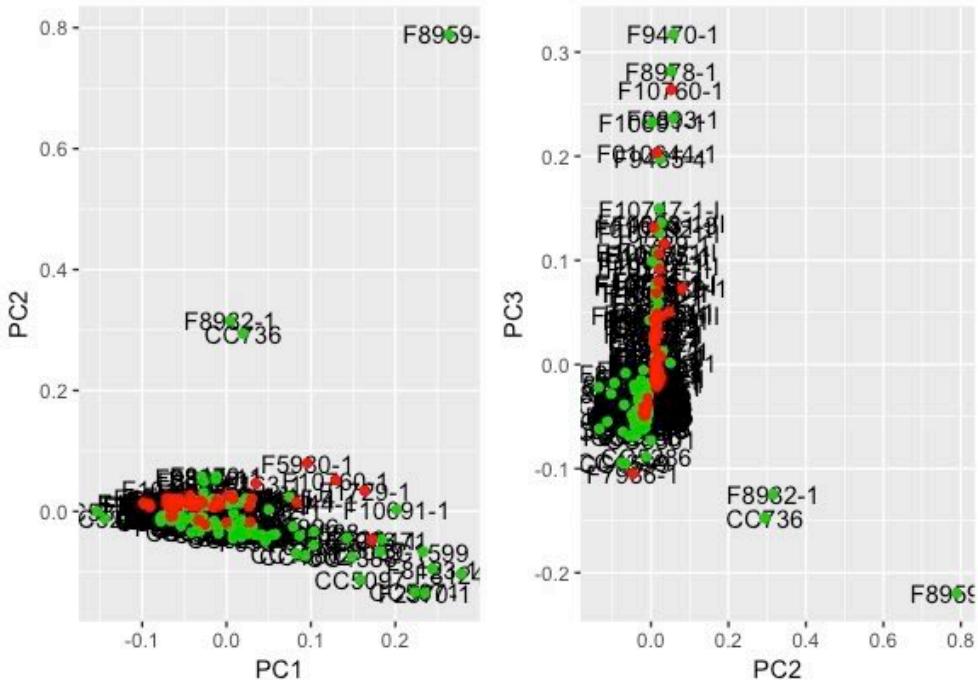


Figure 32: First three principal components from the raw VCF, after the removal of CC5064, CC962, CC5305 and F8385-1. The green dots are the control group and red dots are the ASD subjects.

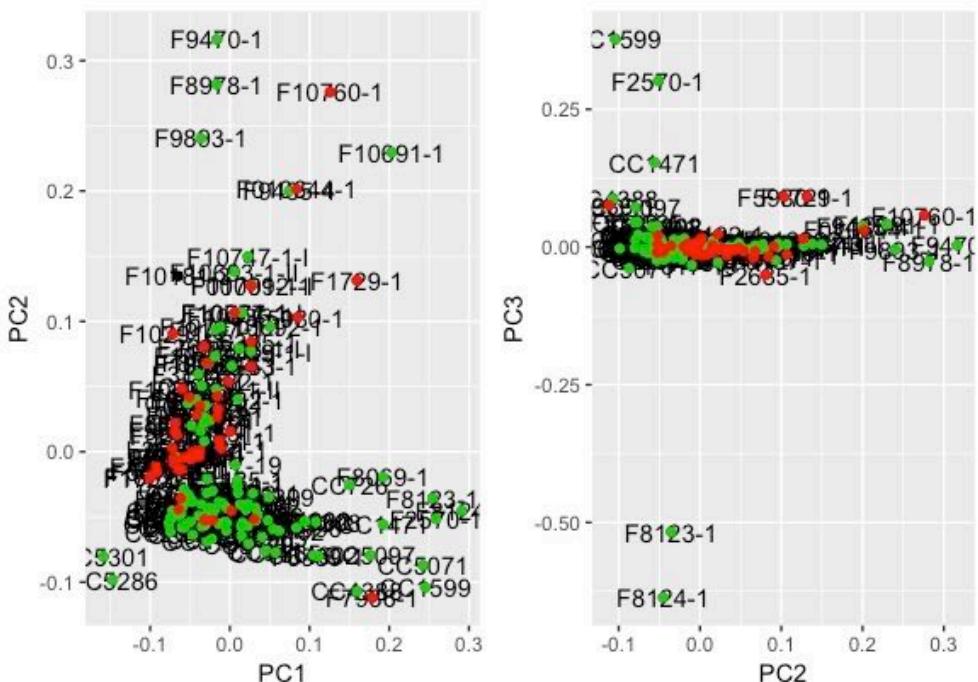


Figure 33: First three principal components from the raw VCF, after the removal of F8959-1, F8982-1 and CC736. The green dots are the control group and red dots are the ASD subjects.

Figure 33 reveals a tendency of different dispersion for both groups. Because of the presence of many control individuals inside the ASD dispersion this raised the hypothesis that this difference is explained by the different WES batches. Figure 34 shows the groups separated by the batch number, green is lower than 15 and red is higher.

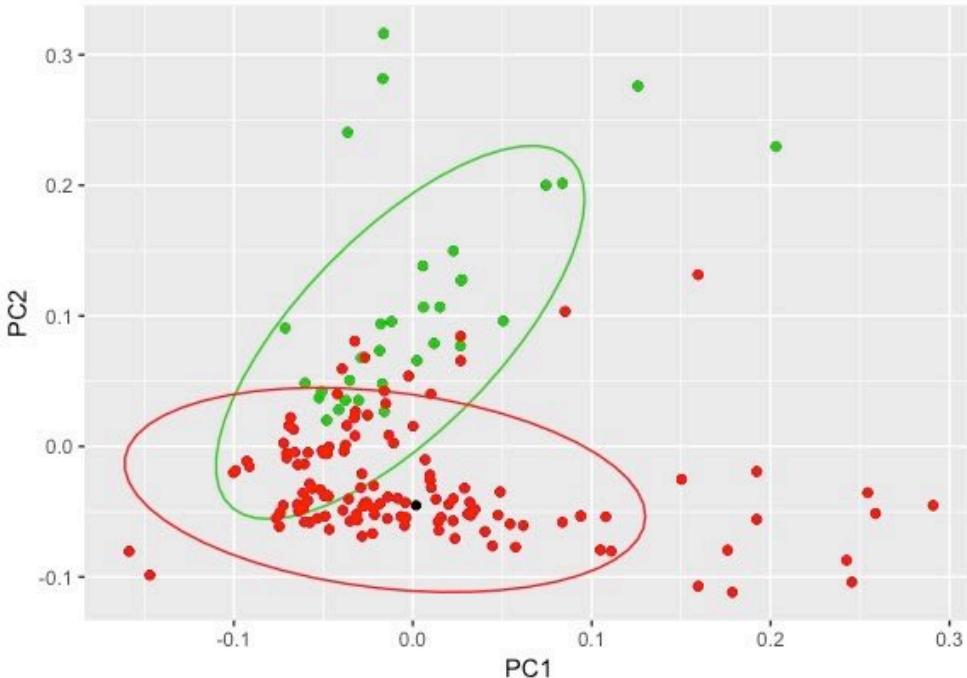


Figure 34: First two principal components from the raw VCF, the samples in green were sequenced before exome batch 15 and in red after batch 15.

For this reason, if the gene list's report have a similar pattern found in this analysis all the outliers described here will be removed, if, after the removal the distribution pattern is similar to this, or there is more outliers the investigation will stop, the investigation will only be move forward if after the removal of the individuals there is a clear different pattern between ASD and control groups.

4.1.7 Analysis by clusters

The kmeans separated the dataset into two clusters as previously described in Material & Methods, each cluster was analyzed separately with all the approaches above. Figure 35 shows the beginning of the analysis for cluster 1, subsequently in the report is cluster 2.

Number of LoF	Affected	Control	ChiSquare
Grouping Method: Frequency			
Rare variants			
[25,46]	19.000	50.000	
[46,92]	19.000	47.000	
[25,46]	0.500	0.515	
[46,92]	0.500	0.485	1
Singletons			
[0, 8]	20.000	52.000	
[8,17]	18.000	45.000	
[0, 8]	0.526	0.536	
[8,17]	0.474	0.464	
Grouping Method: Presence versus Absence			
Rare variants			
0	0.000	0.000	
[1,92]	38.000	97.000	99
~	~	~	99

Figure 35: Screenshot of the first analysis for cluster 1 in the exploratory analysis' report.

4.1.7.1 Cluster 1

None of the analysis performed resulted in a good model.

4.1.7.2 Cluster 2

For cluster 2 both neural network and support vector machine had good models, with the AUC, KS and H metrics above the threshold. All the individuals from the testing samples were classified according to the model generated by the learning samples and received a probability value of being classified as 1 (in this case as part of the affected group). Figure 36 shows the neural network prediction value for all the testing individuals, and almost all of the individuals received a prediction under 0.25. The best threshold to differentiate affected from control for the neural network model is 0.1, with this value, there were three true positives, one false positive, ten true negatives and two false negatives for the testing set. In this cluster there were 22 controls and 10 ASD, this low number of individuals could have interfered in the overall results. The distributions of the ASD and controls were very similar for this sample, for this reason this cannot be considered as a good model.

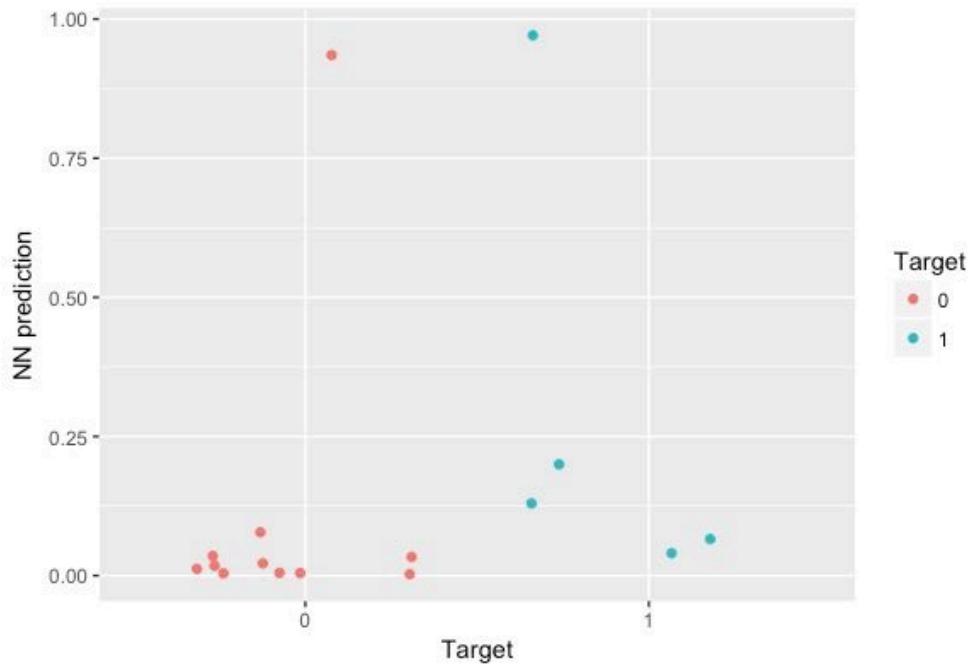


Figure 36: The neural network predicted probability for the groups, blue the ASD and red the control group.

The best threshold for the support vector machine model is 0.25, with this value, there are five true positives, three false positives, eight true negatives and none false negatives, for the testing set. This model had a 100% sensitivity and 64% specificity. Due to the limited number of samples it is possible that this effect is a bias for this particular sample, and further investigations are needed.

4.1.8 Separate regions

These sections are divided into two groups of analysis, one with *only exonic regions* (frameshift, canonical splicing, stopgain/loss, nonframeshift, nonsynonymous and synonymous), and the *regulatory regions* (3'-UTR, 5'-UTR, miRNA, ncRNA, regulatory sites and transcription binding sites) (Figure 37).

Only Exonic Regions

LoF

Loss-of-function grouping

Number of LoF	Affected	Control	ChiSquare
Grouping Method: Frequency			
Rare variants			
[25,48]	20.000	66.000	
[48,92]	28.000	53.000	
[25,48]	0.417	0.555	0.149
[48,92]	0.583	0.445	
Singletons			
[0, 8)	24.000	67.000	
[8,17]	24.000	52.000	
[0, 8)	0.500	0.563	0.57
[8,17]	0.500	0.437	
Grouping Method: Presence versus Absence			
Rare variants			
0	0.000	0.000	
[1,92]	48.000	119.000	99
~	~	~	

Figure 37: Screenshot of the first analysis for the *Only Exonic Regions* in the exploratory analysis' report.

Neither exonic or regulatory regions individually had a result that could distinguish between ASD and controls. The Decision Tree of the regulatory variants had an improvement on the final error, but only the KS metric achieved the threshold, for this reason this tree will not be considered a good discriminant model and will not be further discussed.

In summary, for this first analysis with all the genes in the whole exome sequencing, conclusions can be made that is relevant for the rest of the interpretations. The cluster division is based mainly on the number of regulatory variants, which are very distinct for the lower batches of WES. This means that all the results concerning the different clusters should be taken carefully. The correlation among variants did not reveal any difference between the groups, and the correlations found could be interpreted as a characteristic of the samples. After further analysis of the variants with a possible contribution to a discriminative model, neither loss-of-function or multivariate approaches rendered good models.

4.2 Report Analysis

Sections 3.2.1 to 3.2.21 are the analysis of the generated report for each of the hypothesis described previously. The description will obey the same order as done in section 3.1, but will not be separated by topics. All the results within the criteria established (p-value below 0.05, H and KS above 0.4 and AUC above 0.8) in the previous session will be described further, the results that do not meet the criteria will only have the presentation on the reports, link in chapter 3.

4.2.1 Genes with critical exons

The variants that showed difference between the groups, with a significant p-value below 0.05, as it can be seen in the reports, were: nsy_fr2, nsy_fr3, syn_fr2, syn_fr3 for all these variants the ASD group showed a lower quantity of variants per individuals (lower mean). None of these variants were different between clusters (the comparison between the clusters designated during the exploratory analysis).

Variants from the same classes and all variants were summed to test if the significance remains for all cases. Results are presented in Figure 38.

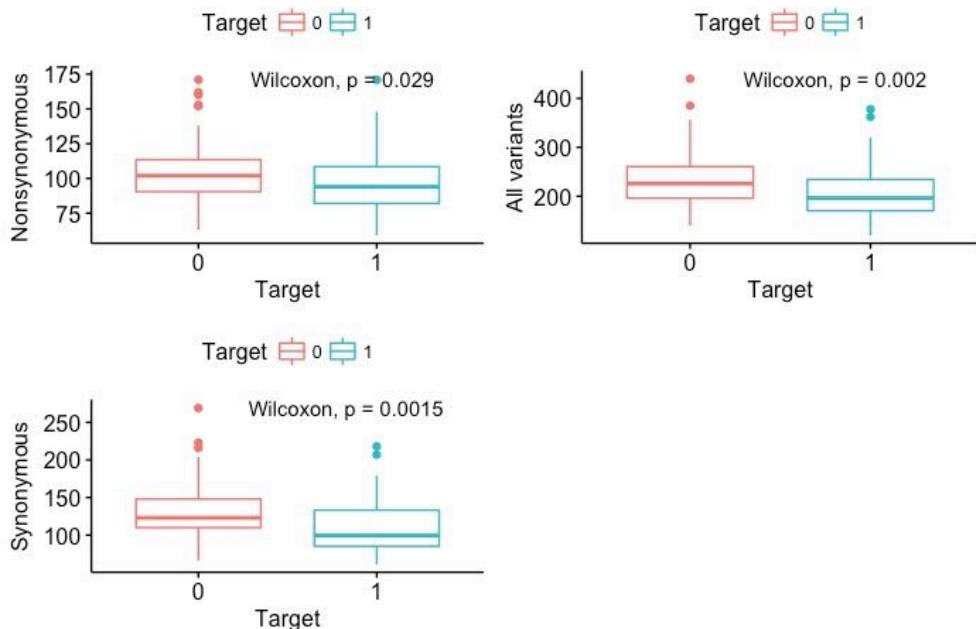


Figure 38: Boxplot of the sum of the variants and the Wilcoxon p-value. Target 0 is the control group, 1 is the ASD.

The distribution for both groups were very similar (Figure 39), for this reason, even with a significant difference it cannot be considered a good start for the discriminant model.

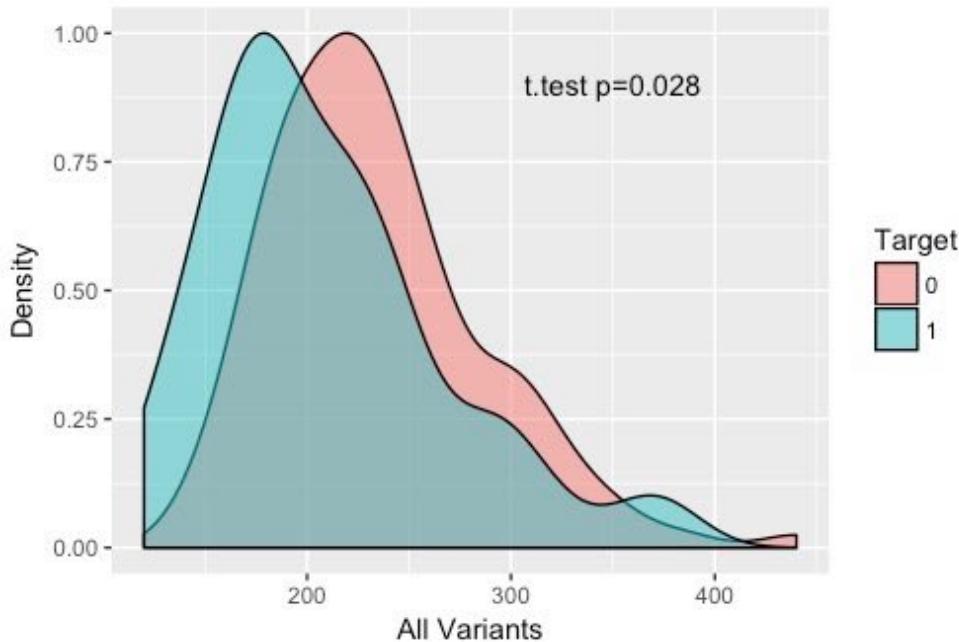


Figure 39: Distribution of all variants with a significant difference by the Wilcoxon test (nsy_fr2, nsy_fr3, syn_fr2, syn_fr3), and a t-test. Target 0 is control and 1 is ASD.

The LoF and missense distributions were very similar for both groups. The multivariate approaches also did not return a good discriminating model.

The principal component analysis for the variants and raw VCF showed a similar pattern found in the exploratory analysis and were not effective in separating the groups.

4.2.2 Chromatin modifier genes

The variants that revealed a significant difference between affected and controls were sto_fr1 and spl_fr4, showing a higher mean in ASD group. From these variants only sto_fr1 was different between the clusters (p-values can be seen in the report for all the variants).

The sum of sto_fr1 and spl_fr4 was also different between the groups (Wilcoxon $p=9e-05$). Twenty-seven ASD individuals (57%) had one or more variants in the sum of sto_fr1 and spl_fr4 while only 30 control individuals (25%), the PX2 $p=0.0001$.

The support vector machine for cluster 2 had all the metrics above the threshold, and could discriminate the individuals for the best threshold (0.3) with 60% of sensitivity and 67% specificity.

4.2.3 1,527 genes related to ASD

The variants with difference between both groups were nsy_fr2, nsy_fr3, syn_fr2. Only nsy_fr3 was different between the clusters. All variants had a lower mean in ASD group. The sum of all the variants maintained the difference (Wilcoxon p-value = 0.0052). The distribution of the two groups are very similar which indicates that despite the low p-value this parameter is not a good discriminant for the groups (Figure 40).

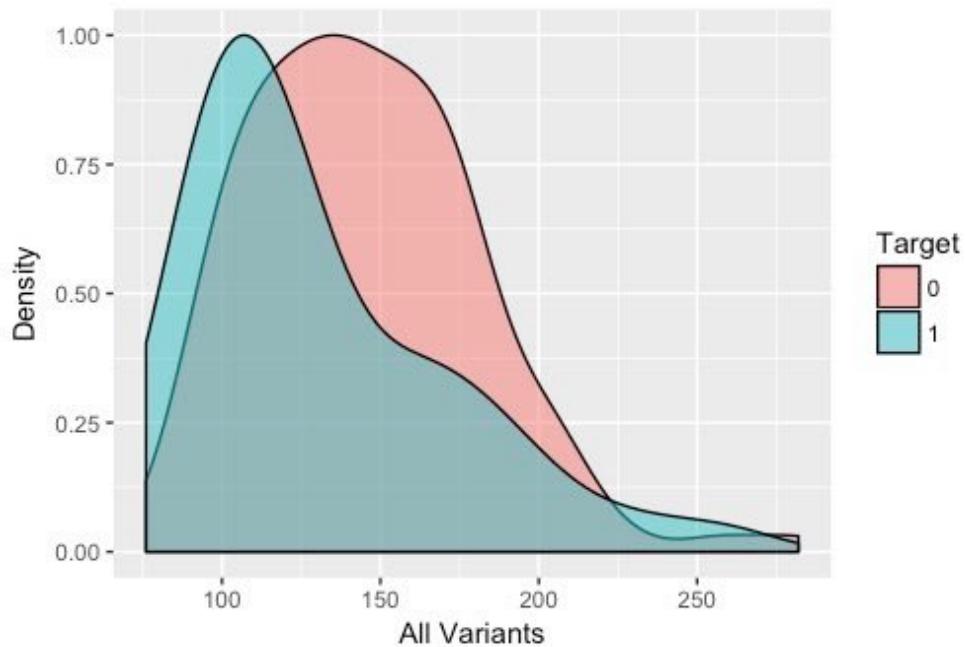


Figure 40: The distribution of the sum of nsy_fr2, nsy_fr3, syn_fr2 variants for both groups.
Target 0 is control, 1 is ASD.

The support vector machine for cluster 2 was a good discriminant approach with an AUC above 0.8, KS and H above 0.4. With a threshold of 0.328 the sensitivity was 100% and the specificity was 90%.

4.2.4 FMRP targets

The variants that separated the groups of interest were fra_fr1, sto_fr1, nsy_fr2, syn_fr2, syn_fr3. The variants fra_fr1, nsy_fr2, syn_fr2, syn_fr3 showed a lower mean for the ASD group, and sto_fr1 a higher mean for ASD group. From these variants the only ones that remained with this difference between the clusters were fra_fr1, sto_fr1 and syn_fr3.

The sum of fra_fr1, nsy_fr2, syn_fr2, syn_fr3 variants maintained the difference between the ASD and control groups (Wilcoxon p=0.011). Nevertheless the

distribution of both groups is very similar and cannot be considered a good discriminant model (Figure 41).

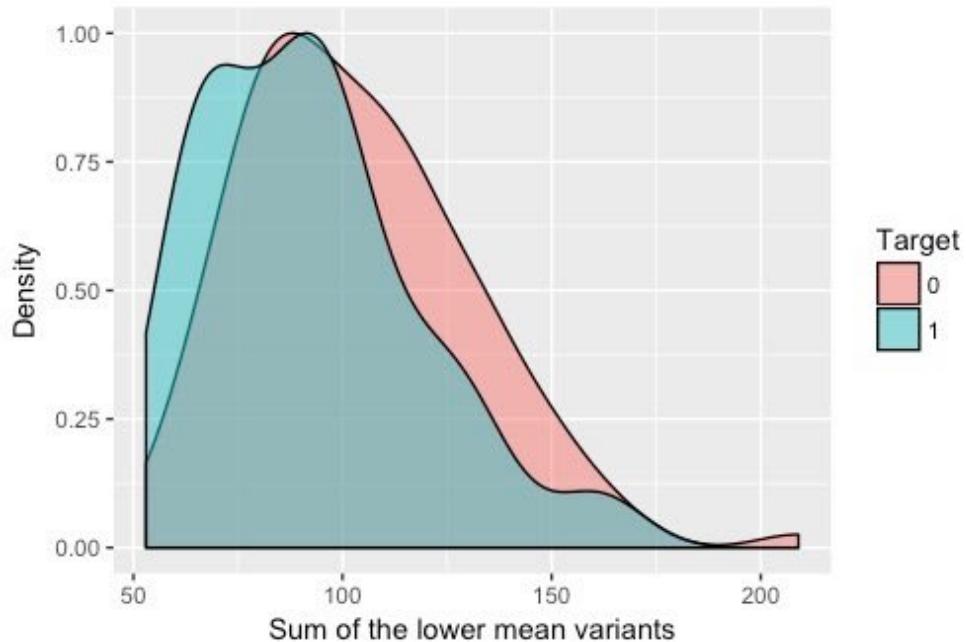


Figure 41: The distribution of the sum of lower mean variants in WES. Target 0 is control, and 1 is ASD.

Figure 42 shows the sto_fr1 distribution for both groups, the analysis of the number of individuals with more or less than two variants returned a non-significant p-value (PX2 p=0.09).

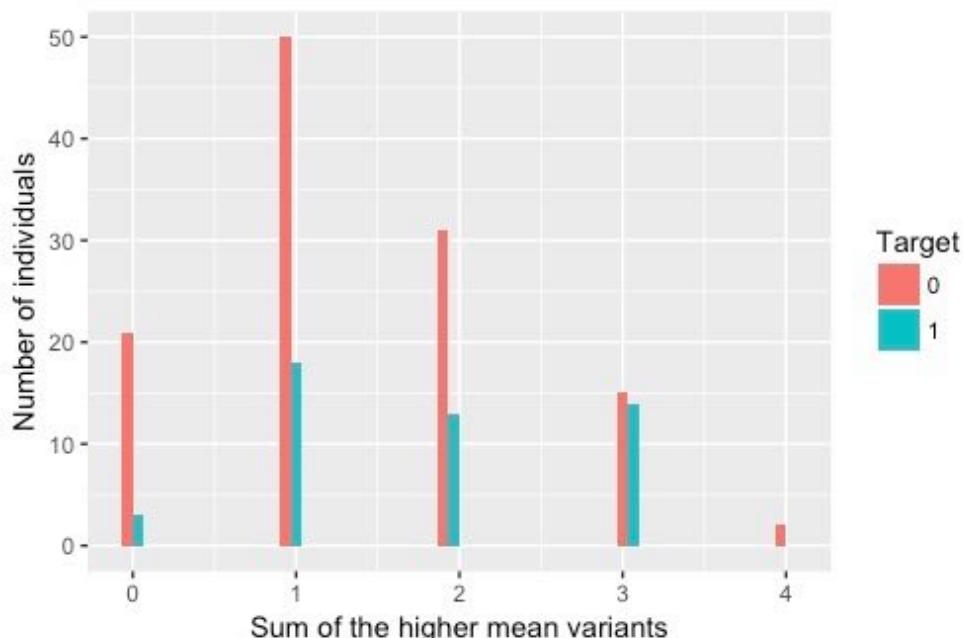


Figure 42: The histogram of the sto_fr1 in both groups. Target 0 is control, 1 is ASD.

Cluster 2 showed a significant difference between the number of rare LoF mutations, with 77% of the control group with no mutations versus 80% of the ASD group with one or more mutations. This result confirms the pattern expected in the literature, but when applied to the whole number of individuals this signal is no longer significant.

4.2.5 208 genes related to ASD

The variants that revealed a significant difference between ASD and controls were sto_fr1, sto_fr4 and nsy_fr3. Variants nsy_fr3 showed a lower mean for the ASD group, and sto_fr1 and sto_fr4 a higher mean in ASD group. The variant that remained with significant difference between the clusters was sto_fr1.

Despite the significance in the Wilcoxon test for nsy_fr3, the distribution of the variants between both groups is very similar and further analysis will not be pursued.

The sum of sto_fr1 and sto_fr4 maintains the higher mean in ASD samples (Wilcoxon $p=0.0002$). In this sum of sto_fr1 and sto_fr4, the controls have 75% of the individuals with no variant, whereas the ASD group has 44% with one or more variants per individual, PX2 $p=0.0002$ (Figure 43).

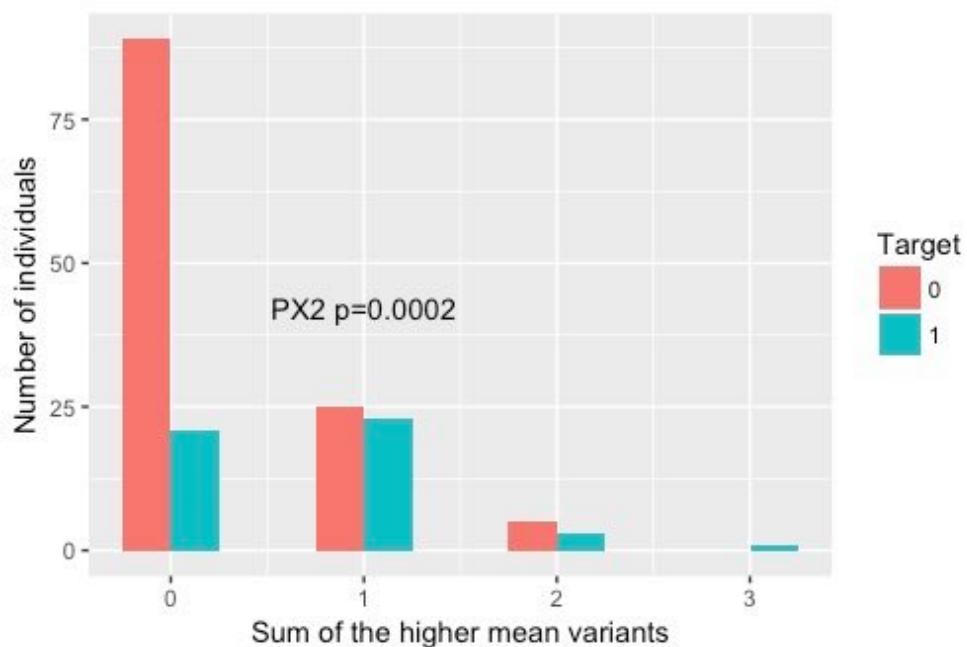


Figure 43: The histogram of the sum of sto_fr1 and sto_fr4.

4.2.6 GABA pathway

No analysis showed a significant result for this gene list.

4.2.7 Hipermutable Genes

The variants that revealed a significant difference between affected and controls were nfr_fr3 and nfr_fr4, with a higher mean for the ASD group. None of the variants were different in the clusters.

The sum of the nfr_fr3 and nfr_fr4 maintained the difference (Wilcoxon p=0.019). Figure 44 shows the distribution of these variants in both groups, the ASD group has 31% of its individuals with one or more variants, while the control group have only 16%, but the significance was only marginal (PX2 p=0.045).

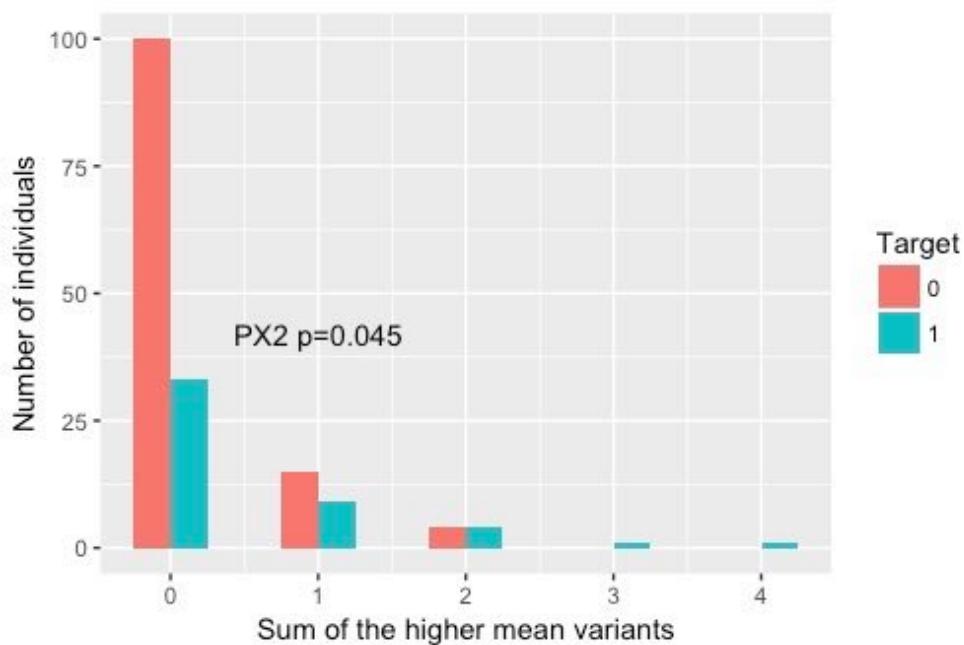


Figure 44: Histogram of the sum of nfr_fr3 and nfr_fr4 for the control and ASD groups.

4.2.8 mTOR Pathway

The variants that revealed a significant difference between affected and controls were nsy_fr2, with a lower mean for the ASD group.

Regarding nsy_fr2, the control group had 46% of its individuals with one or more nsy_fr2 variants, versus 25% of ASD group (PX2 p=0.014).

Despite the values of H, KS and AUC were above threshold for the cluster 1 logistic regression, none of the parameters of the regression had a p-value under 0.05, which means that this model will not result in a good discriminant model.

4.2.9 IGF1 Pathway

The support vector machine for cluster 2 was a good discriminant approach with an AUC above 0.8, KS and H above 0.4. With a threshold of 0.5 the sensitivity was 60% and the specificity was 90%.

4.2.10 Genes intolerant to loss-of-function mutations

The variants that revealed a significant difference between affected and controls were sto_fr1, nsy_fr2, syn_fr2 and syn_fr3, all with a lower mean for the ASD group. For the cluster comparison, from the variants above, the one that showed a significant difference was syn_fr3.

The sum of all these variants maintained the difference (Wilcoxon $p=0.0021$), the distribution pattern, as seen in previous analysis by the graphic dispersion, of both groups is very similar which will not return a good discriminant model.

The neural network for cluster 2 was a good discriminant approach with an AUC above 0.8, KS and H above 0.4. With a threshold of 0.03 the sensitivity was 100% and the specificity was 80%. Figure 45 shows the probabilities distributions for both groups, because of the proximity of ASD and control this model has to be considered carefully.

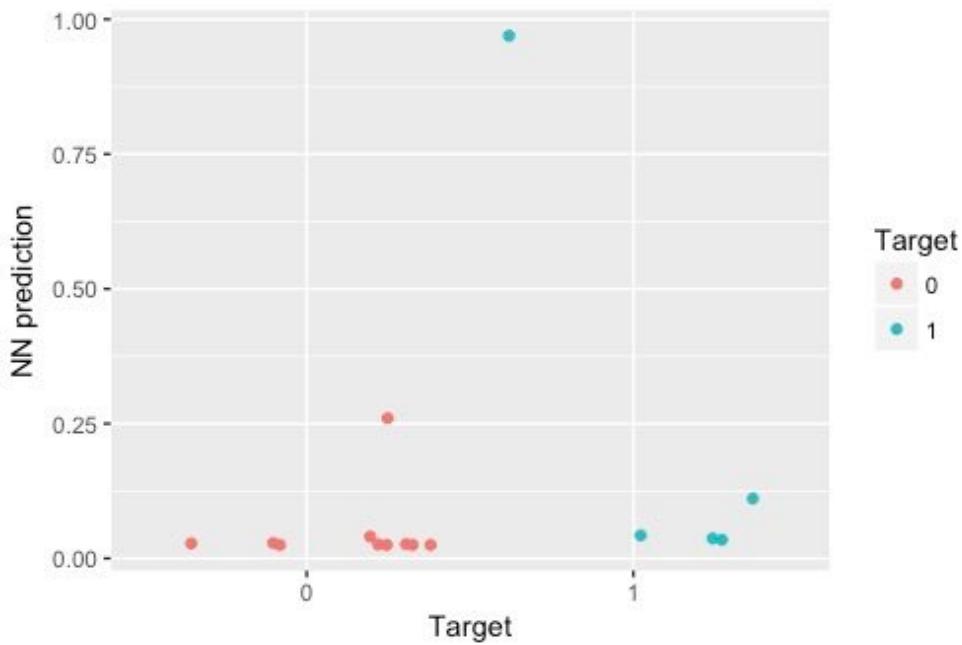


Figure 45: The neural network predicted probability for both groups, blue the ASD and red the control group.

4.2.11 Differentially expressed genes M1

The variants that revealed a significant difference between affected and controls were sto_fr3, with a lower mean for the ASD group. Despite the significant difference, the two distributions are very similar and no discriminant model can be constructed with this data.

The support vector machine for cluster 2 had a discriminant model with an AUC above 0.8, KS and H above 0.4. With a threshold of 0.5 the sensitivity and specificity were 80%.

4.2.12 Differentially expressed genes M2

The variants that revealed a significant difference between affected and controls were fra_fr4 and nsy_fr2. Variants nsy_fr2 with a lower mean for the ASD group, and fra_fr4 a higher mean for the ASD group. None of them were different in the cluster comparison.

Despite the difference between ASD and controls for the nsy_fr2 variants, the distribution of those variants are very similar which will not return a good discriminant model.

Regarding the fra_fr4 only one control individual showed one singleton frameshift variant (CC245 (0.8%)), in comparison to six ASD with one or more variants (12%), PX2 p=0.002 (Figure 46). If we add the other LoF variants like stopgain/loss and/or splicing this signal decreases. With only the addition of splicing singleton variants (PX2 p=0.04) and disappears with the addition of stopgain/loss (PX2 p=0.05). All three variants together is seen in LoF table with PX2 p=0.11.

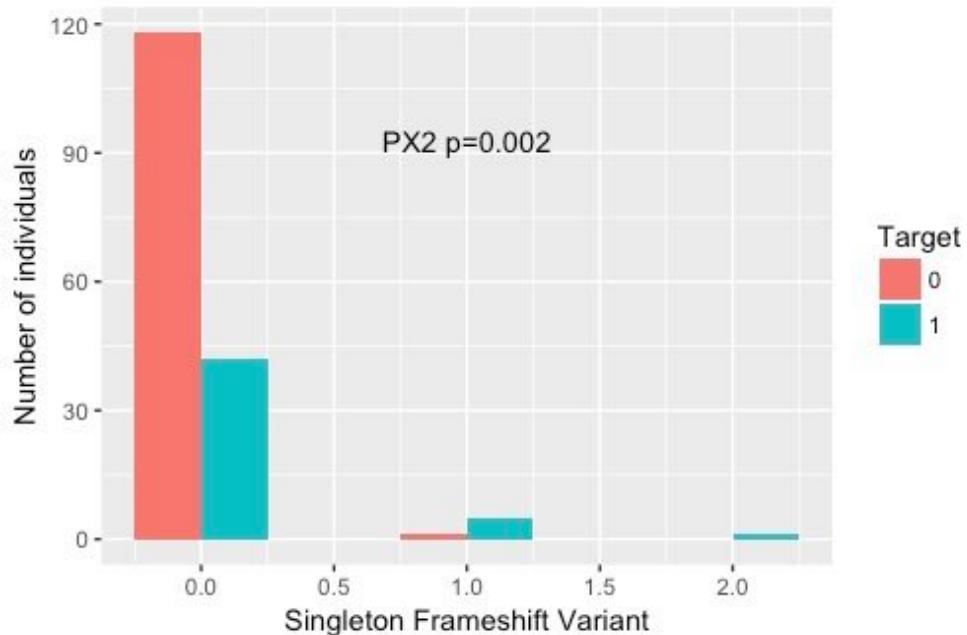


Figure 46: Histogram of Singleton frameshift distribution in the GOI.

The support vector machine for cluster 2 was a good discriminant approach with an AUC close to 0.8, KS and H above 0.4. With a threshold of 0.55 the sensitivity was 80% and the specificity was 92%.

4.2.13 Differentially expressed genes M8

The variants that revealed a significant difference between affected and controls were nsy_fr3, with a higher mean for the ASD group.

The distribution of the rare nonsynonymous variants showed 93% of the controls with three variants or less while 73% of the ASD group were in that category, with the other 27% with four to ten variants per individual (PX2 p=0.0008).

4.2.14 Differentially expressed genes M16

The variants that revealed a significant difference between affected and controls were nfr_fr4, with a higher mean for the ASD group.

For the singleton nonframeshift variant two individuals in the ASD group had one variant while no individual in the control group, this difference draws attention because of the uniqueness of the ASD group, but it is not sufficient to generate a significant difference (PX2 p=0.08), and could be the result of a sequencing bias.

The support vector machine for cluster 2 was a good discriminant approach with an AUC close to 0.8, KS and H above 0.4. With a threshold of 0.55 the sensitivity was 100% and the specificity was 58%.

The principal component analysis for cluster 2 revealed a slight difference in PC1 and PC2 between the groups. After the removal of five outliers the pattern seems to be consistent with two different groups mainly for PC1 (Figure 47).

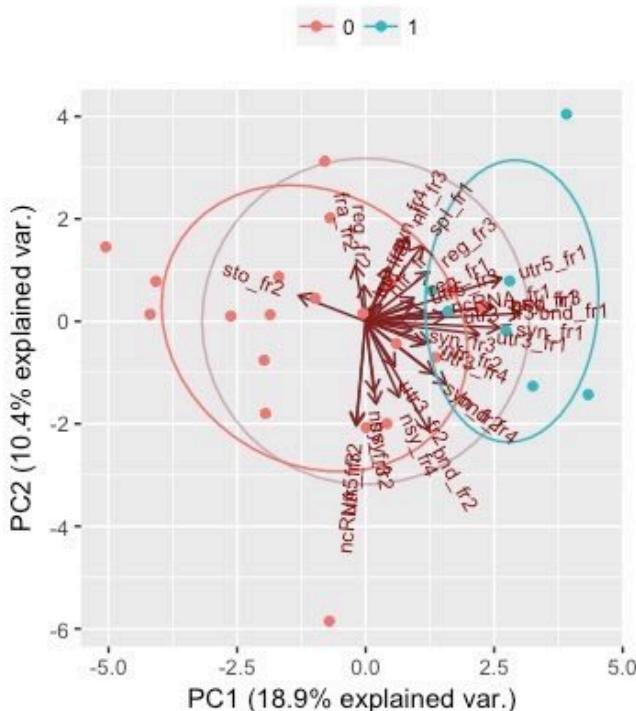


Figure 47: First two PC showing the dispersion for both GOI. The blue is ASD and red control.

The variants mainly responsible for this discrimination are: bnd_fr1, utr5_fr1, syn_fr1, nsy_fr1, bnd_fr3, utr3_fr1 and sto_fr2. These variants were used to build a principal component analysis for cluster 1, as seen in Figure 48 this same variants are not responsible for the differentiation in PC1, and cannot be applied to this group.

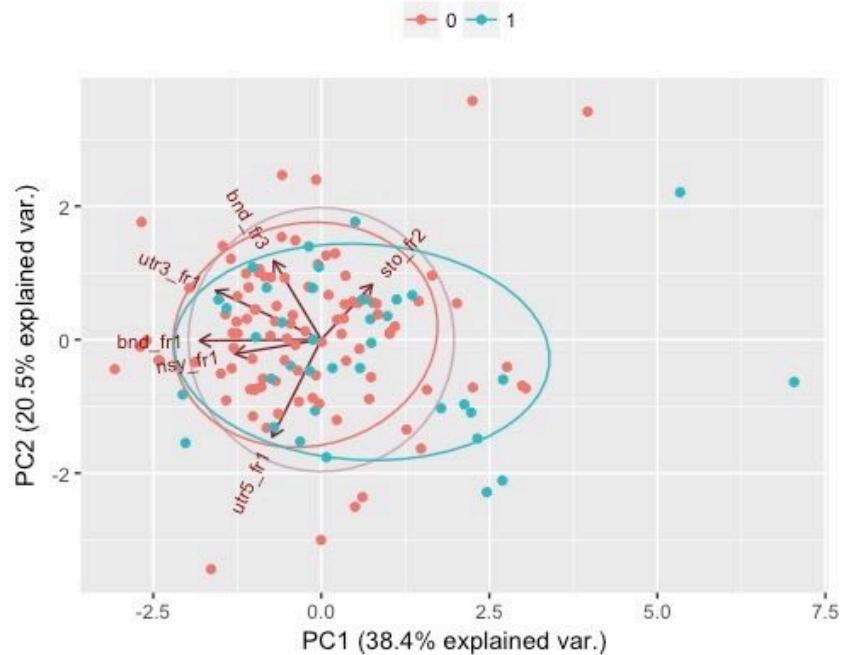


Figure 48: First two PC for the variants bnd_fr1, utr5_fr1, syn_fr1, nsy_fr1, bnd_fr3, utr3_fr1 and sto_fr2 in individuals within cluster 1. The blue is ASD and red control.

4.2.15 Differentially expressed genes M18

Only nfr_fr1 showed a significant higher mean in the ASD group, and was significantly different between the clusters.

Figure 49 shows more control individuals with no nfr_fr1 variants compared to ASD. There are 54 controls with no nfr_fr1 variant, which means 45% of the total individuals, against 12 ASD individuals, 23%, with PX2 p=0.01.

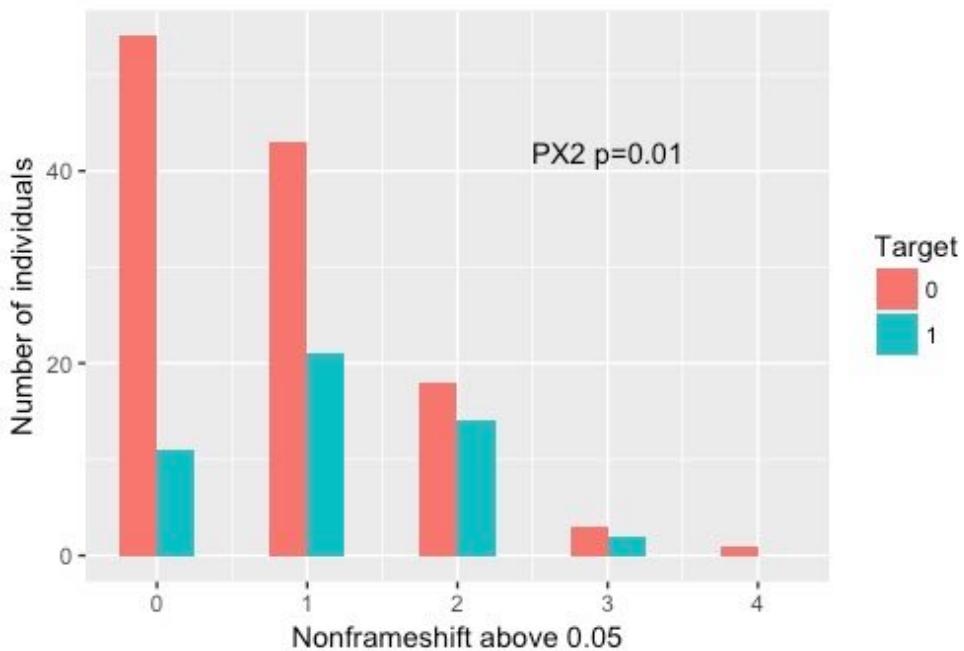


Figure 49: Histogram of Nonframeshift variants with frequency above 0.05 for both GOI.
Target 0 is control, 1 is ASD.

The support vector machine for cluster 2 was a good discriminant approach with an AUC above 0.8, KS and H above 0.4. Nevertheless, the threshold was inverted, and the ASD individuals received smaller numbers of probability than the control group. Because of this pattern this result was not considered consistent with a discriminating model created with the learning sample.

4.2.16 RAS/MAPK Pathway

The variants that revealed a significant difference between affected and controls were nsy_fr3 with a lower mean for the ASD group.

Because the control group is formed by several individuals with Noonan syndrome it is expected that they have more rare nonsynonymous variants in RAS/MAPK pathway than the ASD group. After the removal of the patients with clinical diagnosis of Noonan syndrome, cardiofaciocutaneous syndrome and neurofibromatosis, the lower mean for the ASD group still maintains (Wilcoxon $p=0.007$).

This could be an indication that this pattern is actually a characteristic of the ASD individuals, the distribution of these variants are seen in Figure 50, between the two groups. It is possible to see that the distributions are quite similar, and only the control group has a normal distribution (Shapiro-Wilk $p=0.51$), ASD (Shapiro-Wilk $p=0.002$).

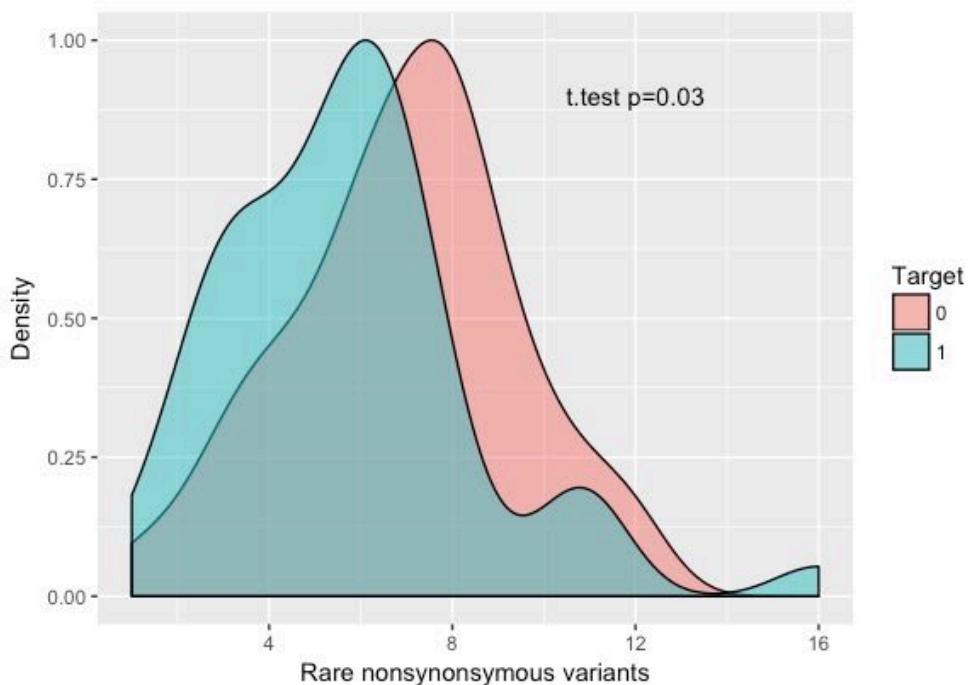


Figure 50: The distribution of the rare nonsynonymous variants between ASD and controls without the individuals with Rasopathies.

The neural network for cluster 2 was a good discriminant approach with an AUC close to 0.8, KS and H above 0.4. With a threshold of 0.2 the sensitivity and specificity were 80%. Figure 51 shows the probabilities distributions for both groups, in this case the ASD group tended to have higher probability scores than the control group, which can be considered as a good model, and to be further tested with more individuals.

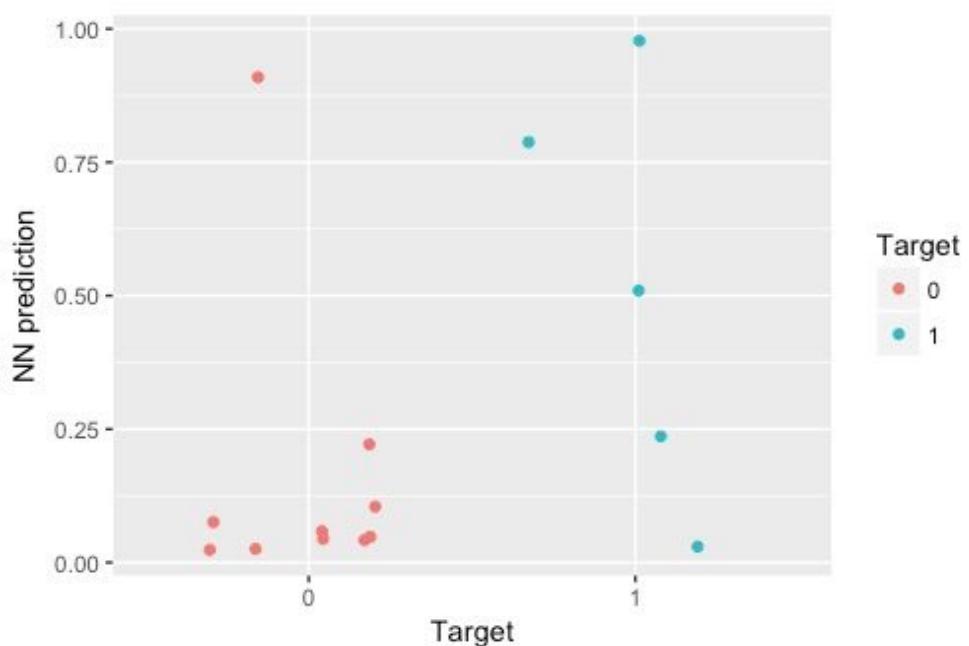


Figure 51: The NN predicted probability for the GOI, blue the ASD and red the control group.

4.2.17 SFARI High Confidence

The variants that revealed a significant difference between affected and controls were fra_fr1, with a higher mean in ASD group.

For the ASD group 19% of individuals had one mutation, while only 6% of control groups had it (PX2 p=0.02).

4.2.18 SFARI Strong Candidate

No analysis showed a significant result for this gene list.

4.2.19 SFARI Suggestive Evidence

The variants that revealed a significant difference between affected and controls were fra_fr1 and syn_fr2. Variants syn_fr2 showed a lower mean for the ASD group, and fra_fr1 a higher mean. None of them had differences in the cluster comparison.

The distribution of the synonymous variants between 1-5% of population frequency are very similar, with the complete inclusion of the ASD distribution in the control group. This pattern does not allow a good discriminant model from these variants.

Despite the marginally significant difference between the quantity of common frameshift variants for the groups of interest, the distribution in both are very similar with no significant different comparing contingency tables for different quantity of variants (PX2 p=0.13).

4.2.20 SFARI Syndromic

The variants that revealed a significant difference between affected and controls were sto_fr1 and nfr_fr2. Variants sto_fr1 showed a lower mean for the ASD

group, and nfr_fr2 a higher mean. None of them were different in the cluster comparison.

The common stopgain/loss variants have a similar distribution between the groups but with a significant difference between the individuals bearing one or more variants, 68% of the control group has at least one variant, while 56% of the ASD met the same condition (PX2 p=0.006).

For the nonframeshift between 1-5% of population frequency had a marginally significant difference between the groups with 18% of the control group with one or more variants, and 33% of the ASD group (PX2 p=0.045).

4.2.21 SFARI High Confidence, Strong Evidence, Suggestive Evidence and Syndromic (SFARI 3S)

The variants that revealed a significant difference between affected and controls were fra_fr1, with a higher mean for the ASD group.

For the ASD group 81% of individuals had two or more fra_fr1 variants, while the control group had 62%, with a PX2 p=0.019.

5 Discussion and Conclusions

The primary goal of the reports was to detect good models to discriminate two groups of interest using next generation sequencing results. The proposed tests are more suitable for complex diseases compatible with additive model and burden effects.

The first three analysis of the reports, can be considered a *proxy* of burden tests, and focused in discovering properties directly from the observation of the variants: (1) compare each type of variant to investigate a direct contribution for the disease, (2) compare the role of loss-of-function (LoF) variants, and (3) whether missense variants with higher pathogenic scores by SPRING could be discriminative. With these analyzes, it is possible to find direct reasons that confers the differences between the groups and to elaborate physiologic models for further testing. As an example, in a scenario of a very rare, severe, autosomal dominant full penetrant genetic condition caused by gain-of-function mutations, like Thanatophoric Dysplasia, Type I, it is expected a higher number of individuals with singleton likely gene disruptive missense mutations in candidate genes on affected group than in controls. This signal should be seen in the analysis of singleton missense variants ranked by SPRING with different peaks on the q-values distribution curve, a peak with lower q-values for the control group and one with higher q-values for the affected. This type of analysis, with different approaches, is part of the current procedures for most whole exome sequencing (WES) in the genomic diagnosis laboratories, where it uses a gene list based on the phenotype of the subject under investigation and search for highly disruptive variants (LoF, missense highly damaging and extremely rare variants) that could explain the development of the condition (Trujillano et al. 2016).

On the other hand, the multivariate approaches (logistic regression, decision tree, neural network, support vector machine and principal components) are attempts to understand the interactions among different types of variants to explain complex diseases (Zhao, Pan, and Du 2018). As an example of a disease with an oligogenic mode of inheritance that in order to develop the affected phenotype, two phenomena must happen, first the individual has to lack common variants, that normally confers protection against the development of the disease, and secondly a rare LoF mutation in any gene of the same pathway. In this scenario, for the affected group it should be expected a lower number of common variants with a higher number of rare LoF compared to the control group. The characteristic that differs from the first scenario is that both events should happen together in order to develop the phenotype since the analysis of only the common or rare LoF variants wouldn't be sufficient to differentiate the groups of interest. Multivariate approaches could suggest more complex relationships.

The 21 reports attempted to test these two scenarios with different gene lists that were previously described in the literature as related to autism spectrum disorder (ASD). Making it clear that the main goal of the reports was to create models based on the knowledge already produced to distinguish affected and controls and not to propose new pathways to the disease. It is mandatory to understand the difference

between discovering candidate genes or new pathways from determining which model can be applied for the diagnosis of a particular individual.

Many different methodologies were developed to understand the genetic background of ASD, the most present in literature is the burden analysis, that investigates which type of variant is most representative among individuals with ASD. From this type of investigation many conclusions already presented in the introduction were made, including the role of *de novo* and rare inherited likely gene disruptive mutations (Iossifov et al. 2012; Krumm et al. 2015; Sanders et al. 2012).

Efforts were made to develop integrated models to explain ASD genetics, the Transmission And De novo Association, known as TADA, is one of those largely used models, this approach integrates the information from *de novo* and inherited variants in affected and controls. With this method it is possible to access candidate genes that confers higher risk for the development of ASD being more inclusive than a simple comparison for the presence of *de novo* LoF (He et al. 2013). This method was used by Li et al. (2018) in the context of discovering different genes and pathways for the subcategories of ASD, and suggested that the three subcategories of ASD (Asperger syndrome, autism and pervasive developmental disorder not otherwise specified) have more convergences than divergences concerning the type of genes and mutations related to the different phenotypes (J. Li et al. 2018). Liu et al. (2014) developed a method, called DAWN, that uses the TADA scores to identify candidate genes in networks based in co-expression in mid-fetal prefrontal and motor somatosensory neocortex (Liu et al. 2014). Luo et al. (2018) developed a multilevel genetic prioritization and association analysis to construct a better model to the development of the disease.

The importance of a decision about the statistical approach in the investigation of common complex diseases is well illustrated by the publications of Stessman et al. 2017 and Barret et al. 2017. Stessman et al. 2017 published an article where they described 91 neurodevelopment risk genes from 208 candidate genes from the literature and sequenced more than 10,000 affected individuals with a broad spectrum of neurodevelopment disorders. They did a simulation by shuffling the labels of private case and control observations 100,000 times and calculated the probability of observing at least the number of likely gene disruptive (LoF or missense variants with CADD prediction over 30) events seen among the cases. These P values were corrected (Benjamini-Hochberg) for the number of genes in the study ($n=208$) (Stessman et al. 2017). Barret et al. 2017 draw the attention to the statistical challenges for studies involving gene finding and confirmatory hypothesis of complex diseases such as ASD by a reply to the Stessman et al. 2017 publication. Barret's article described which were the possible statistical flaws that could lead to a misconception of the results. The main statistical issue was the threshold applied for the p-value, that did not correct for the effects of scanning all 20,000 genes, and despite the significance of the replication sample, the rate of *de novo* LoF mutations was dramatically lower (Barrett et al. 2017), showing the importance of determining the statistical methodology very carefully. Also the work raises up the importance of having population matched controls, sequencing and analysis in identical ways and the average depth of coverage for the

target regions, without these parameters set correctly even the use of permutation test to assess significance does nothing to eliminate these problems.

Until now all the knowledge acquired about the pathophysiology was not applied to the clinical-laboratory context to test whether a stepwise and statistical approach to analyze WES of probands without parents sequencing information could result in an efficient discrimination of individuals and bring information about the genetic development of the phenotype to that specific subject. The main goal of the reports was to test which hypothesis for the development of the ASD could generate good models that could be applied as a routine test in the clinical-laboratory context. Regarding this context all but two reports returned a preliminary result that could be further investigated concerning the strength of the models. These results and implications for the reality of future WES testing will be discussed in sequence.

With the exception of the GABA pathway and SFARI Strong Candidate, all of the reports resulted (19 in 21) in at least one significant difference between the groups that could be further investigated. From these 19 reports, the analysis with the highest number of positives was the *Comparison for each variant between affected and control*, this is in accordance with the burden model for the development of ASD, however many significant results were concentrated in variants not considered likely gene disruptive, such as common or synonymous and a higher number of mutations in the control groups, such as rare missense RAS/MAPK variants. These findings were not expected based on the majority of the ASD candidate genes publication, which have prioritized loss of function mutations (la Torre-Ubieta et al. 2016; Woodbury-Smith and Scherer 2018).

The *Comparison for each variant between affected and control* was significant for 19 reports, eleven maintained at least one variant with significant difference after further investigation, from this, eight with variants with higher mean in ASD groups (chromatin target, 208 genes, hypermutable, differentially expressed M2, M8 and M18, SFARI High Confidence (SFARI1) and SFARI all (SFARIS3)), one with both higher and lower mean in ASD group (SFARI Syndromic (SFARIS)) and two with only lower mean in ASD group (mTOR and RAS/MAPK). The sole analysis of the level of significance from each finding is not the best approach to eliminate or consider a specific scenario, the investigation of the distribution, along with the critical interpretation of physiological and genetic pathways are fundamental for the final decision to pursue a theory and confirming in other scenarios.

The gene lists mTOR, SFARI Syndromic (SFARIS) and RAS/MAPK had variants that showed a significant decrease in number for the ASD group. The nonsynonymous rare variants from the RAS/MAPK group had different distributions in the ASD and control groups, with only the first showing a normal distribution, this pattern remained even after the removal of the individuals with the diagnosis of rasopathies. Even with a marginal p-value (0.03), this result could suggest a certain protective effect of the missense rare mutations in this pathway that remained after the removal of the individuals. The mTOR gene list also had a marginally significant (p-value = 0.014) decrease of the missense mutations for the ASD group, but in this case not for the rare variants, instead for the ones with frequency between 0.01 and

0.05, suggesting also a protective role for the missense mutations in different pathways. The exploratory analysis without gene filtering also revealed a significant lower mean in ASD group for this type of variant, for that reason this is still a preliminary result that requires further testing, specially to detect if this result is not biased by the set of individuals in this analysis. Nevertheless, this specific protective effect of missense variants was not described in the literature until now and could be a promising model to pursue with more individuals and tested in different samples.

The SFARIS genes and chromatin modifier had somewhat conflicting results since the first showed a significant decrease for the common stopgain variants in ASD, while the latter showed an increase, with special attention to the fact that the chromatin result, in addition with singleton splicing mutations, is the most significative of all reports. The SFARIS genes consists, in most cases, of fully penetrant recognizable syndromes that follows the LoF model, and was the only list that had both lower and higher mean variants for the ASD group with significant results. The common nonframeshift mutations had a higher mean in the ASD group with 33% of the individuals with one or more variants *versus* 18% from the control group. One hypothesis that could be pursued with these results is to test in a larger cohort a combined model of inheritance that are modifying the phenotypic outcome for the patients. However, since this group of genes belongs to very rare clinically recognizable diseases further investigations to test multivariate models with common mutations may not be the best approach.

On the other hand, the chromatin genes with an increase of common nonsense and singleton canonical splicing variants could reveal a combination of events that predispose to ASD, and therefore should be further investigated. Publications about chromatin modifier genes focused mainly in rare or *de novo* highly gene disruptive variants (Cotney et al. 2015; De Rubeis et al. 2014; Gallagher et al. 2014; O'Roak, Vives, Fu, et al. 2012b). The finding that common nonsense variants could be acting together with highly rare canonical splicing events is in accordance with the model proposed by Luo et al. 2018 with one gene hit and a second pathway hit. Because these results are preliminary, further testing is needed to confirm this finding on a larger cohort, and to test the second hit model for these genes.

Besides the chromatin modifiers, other gene lists showed a higher number of variants in the ASD group like: 208 genes, hypermutable, differentially expressed M2, M8 and M18, SFARI1 and SFARIS3. For the other lists, different variants were responsible for the findings. The descriptions above highlighted the most prominent results that should be first addressed, but each of these variants could set off investigations in larger cohorts.

The role of machine learning predictive models in medicine is getting larger for areas such as imaging and cancer treatment (Greenlaw et al. 2017; Letai 2017; Machiela et al. 2011), and even other aspects of ASD diagnostics such as electroencephalogram, movement patterns and brain imaging (Bosl, Tager-Flusberg, and Nelson 2018; B. Li et al. 2017; Zare, Rezvani, and Benasich 2016), for this reason the application of predictive models was an approach considered in all gene lists. Neither logistic regression, neural network, support vector machine or decision tree returned

consistent good discriminating models. This result was not expected taking into consideration the arising models of multiple causes for the development of the ASD phenotype (Betancur et al. 2009; W. Luo et al. 2018; Pinto et al. 2014). These approaches were chosen because they consist of different strategies that can use quantitative variables to construct discriminant models and differentiate the groups finding one of the possible resolutions, and in many times are not straightforward to interpret. All of them depend on learning the patterns from individuals considered as examples of the discriminant behavior in search for a pattern that best separate them (Zhao et al. 2018).

Differently from the multivariate approaches described above, the principal component analysis is not based on a learning and testing sets to construct a discrimination model, instead, it divides the individuals in a graphical space based on the variance within the variables, revealing which variables are most related to the difference among groups (Aschard et al. 2014). This approach was chosen because it can evaluate all the data available, without the bias of having to choose different sets of individuals. For this reason, it was tested in two different ways, one with the counting matrix, as a *proxy* of the additive model, and the other with the raw VCF, so the variants alone, and their frequency only in the cohort, could be tested for patterns, and not related to their effect in population or protein. If a set of specific variants were responsible for the disease it would be seen in this approach. None of the raw VCF principal components analysis resulted in a specific variant or sets of variants that could discriminate between the groups. This finding is in accordance with a recent systematic reconstruction of the ASD genetic profiles that revealed significant recurrence only for genes and pathways, but not for specific variants (X. Luo and Sorock 2008).

All of the multivariate approaches are susceptible to type and quantity of variables, in this sense is always a hard choice between gaining and loosing power with the addition of variables. In this pipeline many possible variants are taken into account, and when analyzed all together, specially with a small number of individuals, the signal of good discriminant models could be lost (Hastie et al. 2013). To determine the best set of variants all possible combinations should be tested, a scenario computationally exhaustive, to address this issue we chose to decrease the number of variables by separating into two groups, even with the risk of loosing the relations among them: (1) variants that affects protein directly and (2) variants in regulatory regions.

Other important matter to address when dealing with complex diseases is the heterogeneity of the groups (Betancur 2011; Bora et al. 2016), because of that, in addition to the direct analysis of affected and controls, the separation into clusters based on the final counting matrix adds power to the discovery process by two main paths, the first one is to detect major differences in the sequencing process and separate the groups to avoid bias, an example of what has been found in this study. The second goal is related to the heterogeneity of distinct diagnosis for complex diseases, the aim was to separate the individuals into more homogeneous groups (clusters), what could represent individuals with similar backgrounds for the development of the disease.

In general, none of the multivariate approaches returned good models for further testing. The presence of significant results for the multivariate approaches only in cluster 2 draws attention to an issue concerning this type of investigation. Two main reasons could explain this pattern, if the individuals in the cluster were separated because of intrinsic sequencing issues, and not related to their direct genetic background, which seems to be the case in this study, then this result can be the consequence of a particular separation of the samples, and not an indication of the genetic structure of the disease. The second scenario is that, if the individuals were separated by their genetic architecture, the presence of such a signal in only one architecture indicates that not all individuals should be analyzed equally, even if the phenotype is the same.

Other contribution of this study to the ASD research was to identify that individuals, sequenced in different time lapses and with data obtained from different bioinformatics pipeline can harbor significant differences that cannot be removed during the investigation process, and are not evident with classical quality assessments such as number of variants per individual, transition/transversion proportions and other metrics. The choice of the individuals that will form the affected and control groups have to be made very carefully, and the interpretation of the results with databases limitations in mind along with confirming the findings in different sets could lead to a more reliable conclusion. Research in the ASD field is very common, and every time new hypothesis are being made and tested, this consequent large number of publications, from many regions of the world, sometimes confirms an important finding, such as the FMRP target and *CHD8* involvement in the disease (Barnard, Pomaville, and O'Roak 2015; Cotney et al. 2015; Darnell et al. 2011; Steinberg and Webber 2013). But other times reveals largely different candidate genes, with the exclusion of important genes such as *SHANK3* and *PTEN* (Cukier et al. 2014; He et al. 2013; Krishnan et al. 2016; Woodbury-Smith et al. 2015).

This study was very effective in showing that even with the limitations on the number of individuals many different results were obtained from a heterogeneous, relatively common complex disease. Most of the gene lists had at least one significant result, and p-values varied from borderline ($p=0.045$) to very low ($p=9e-05$). But none of these results alone should be considered sufficient proof of a specific contribution to the pathophysiology of the disease, instead, testing the hypothesis in other samples with matched controls, analyzing the distribution of the results and how the statistics is counting the p-values, and doing further testing, such as cellular and animal models, looking if similar results could be obtained from different candidate genes (Barnard et al. 2015; Iossifov et al. 2015; Krishnan et al. 2016). This could be some approaches that would make a more reliable final result concerning ASD.

All the reports have many statistical tests which should be corrected for multiple testing, for this reason the level of significance for each result should be accounted carefully. The choice of not making the correction already in the report was because the level of significance accepted could change for each research with the chance of losing interesting relations that could be confirmed in other manners.

One approach that can be proposed to improve the statistical confidence of the results of ASD research (Barrett et al. 2017), is to divide the sample in learning and testing (even if the main strategy does not involve machine learning). With the learning samples all the research protocol is made, then the testing samples is resampled in smaller sets how many times depending on the strength of the initial finding. The sets should be formed by three configurations, (1) individuals classified a priori as affected and control, (2) individuals classified a priori as control, but randomly separated in affected and controls and (3) individuals classified as affected, but randomly separated as affected and control. The protocol of the principal hypothesis is made in each sample and the proportion that the same result appears will be relevant to the importance of this model to the disease. With this strategy it is possible to confirm a finding as a general characteristic of the disease, or a particular characteristic of some individuals with the disease, or a characteristic that could be found for the population (with the results from the two homogeneous groups). This approach is consistent with the one applied by Iossifov et al. 2015, in order to test for burden of rare and de novo mutations in ASD, they did a permutation swapping genes between sets and permuting the “affected” labels (Iossifov et al. 2015).

Other important limitation of our study is that only specific gene lists could be accessed for the tests, nevertheless, with the pipeline at hand and further knowledge of the necessary steps to develop the disease. Studies such as Luo et al. 2018 that showed that most variants disrupt extremely close residues within the same gene, or affect similar pathway-related gene functions, or Uddin et al. 2014, that showed a higher concentration of variants in brain expressed exons (X. Luo and Sorock 2008; Uddin et al. 2014), can be added a second layer of interpretation to the variants and provide a more reliable model.

The final annotated file from a WES has around 50,000 variants per individual and its analysis consists mainly in experts prioritizing variants based on clinical features, population frequency, *in silico* predictor and scientific literature (Eilbeck et al. 2017). Some parts of the WES analysis can be automatized, such as filtering for variant quality, population frequency, and in some cases, phenotype (Mughal et al. 2017; Sobreira et al. 2015). Even with these modifications the initial parameters have to be set by experts and the remaining variants are classified and analyzed also by them. For the mendelian disorders these steps increase the chance of finding the variant related to the phenotype described (Eilbeck et al. 2017).

The addition of WES in the clinical laboratory practice as a manner for diagnosing rare genetic diseases was well accepted due to its high diagnostic yield (around 25-30%), and lower final costs for the investigation (Trujillano et al. 2016). For the genetic diseases with a clear mendelian inheritance the advantages of the WES are very straightforward, but for the complex diseases, such as ASD, there are remaining questions whether or not to use this approach for the molecular diagnosis (Retterer et al. 2016).

The importance of differentiating the clinical report from the scientific literature is that for the last one significant burden results adds to the understanding of the disease’s global etiology, however, is not made to change particular medical

decisions and to affirm the diagnosis, while laboratory results will be seen by the doctors and family as a result for the specific patient (Matullo et al. 2013). Even with the advances of metabolomics, biochemistry, brain imaging and electroencephalography, there is no clear biomarkers to distinguish between ASD and controls, that can be used in the clinical practice or can be related to an individualized therapy (Dieme et al. 2015; Hampton 2017; la Torre-Ubieta et al. 2016). For this reason the models could still be tested in a context of more samples.

The ASD is a complex disease, and although its pathophysiology is clearer, thousands of genes and mutations are described, making it hard to select the few disease causing events and to replicate it in other studies (X. Luo and Sorock 2008). In most cases ASD does not follow the mendelian inheritance, for that reason the American College of Medical Genetics and Genomics (ACMG), followed by other institutions, does not indicate the WES as one of the genetic strategies to molecular diagnosis, but recognizes that in the future this could bring more evidence to the diagnosis (Schaefer and Mendelsohn 2013).

Even without the formal indication of the institutions for the WES testing in ASD, the physicians are requesting this test, in the hope that the patient has a pathogenic variant that could explain the phenotype and could aid further investigations and genetic counseling. This demand associated with scientific reports that shows an increase of rare or *de novo* loss-of-function (LoF) variants in genes highly related to autism (such as genes of the SFARI High Confidence, or other lists, many of them tested in this study) (Cotney et al. 2015; Y.-H. Jiang et al. 2013; la Torre-Ubieta et al. 2016; Sanders et al. 2012) are leading the WES analysis to search for highly gene disruptive variants in such genes.

In conclusion, the aim of the study to develop a pipeline to investigate different approaches to discriminate individuals from a complex disease was completed, but none of the 21 different gene lists revealed positive results that could be used in the clinical practice without further investigations. The preliminary results suggest that different types of variants can have distinct degrees of contribution depending on the pathway or genes analyzed and that this different contribution for each variant is more prominent than the role of different variants in the same pathway or genes, represented by the presence of significant results in the variant analysis alone *versus* the non significant results in the multivariate approaches. Nevertheless, the primary achievement done by this work was to reveal the weakness for the main assumptions when analyzing a WES for ASD in clinical context, concluding that the institutions should be more careful with the possibility of the false positives, and for now remaining concentrated on the ASD mendelian syndromes with LoF *de novo* variants in fully penetrant genes.

6 Resumo

O transtorno do espectro autista (TEA) é um distúrbio do neurodesenvolvimento caracterizado por uma incapacidade de comunicação comportamento e interações sociais que afeta em torno de 1-2% da população mundial. Até o momento a etiologia do TEA ainda não é totalmente compreendida, mas nos últimos 18 anos muitos avanços foram feitos para entender o componente genético relacionado ao desenvolvimento do quadro clínico. Com o advento das análises de varredura genômica como a análise cromossômica por microarray e o sequenciamento completo de exoma (SCE) muitos avanços foram feitos para a compreensão da fisiopatologia da doença. Em torno de 10-15% dos casos podem ser explicados por grandes perdas ou ganhos (deleções ou duplicações superiores a 1000 pares de bases) do material genético, que geralmente envolvem a disruptão de um ou mais genes. As metodologias de sequenciamento de nova geração foram fundamentais para a descrição das mutações de ponto e pequenas inserções e deleções associadas ao TEA. O SCE permitiu que muitas descobertas fossem feitas sobre novos genes candidatos e mecanismos para o desenvolvimento da doença. Atualmente afirma-se que as alterações *de novo* (não herdadas) e de maior probabilidade de ruptura gênica, como as mutações de perda-de-função e as alterações não-sinônimas com alta predição de dano por ferramentas computacionais, em genes de susceptibilidade a doenças do neurodesenvolvimento sejam um grande contribuidor para o mecanismo da doença. Entretanto essas mutações, além de não explicar a totalidade dos casos raramente são recorrentes na população, o que dificulta o estabelecimento de um diagnóstico molecular definitivo para a maioria dos pacientes. O SCE já é uma prática nos laboratórios clínicos de genética e demonstra uma alta efetividade para as doenças que seguem um padrão de herança mendeliano, e têm uma causa genética estabelecida. Na prática clínica o SCE é solicitado para os casos de TEA, apesar de ter diferentes modos de herança e terem mais de 1,000 genes associados à doença. Devido a estas características o SCE para os casos de TEA são um grande desafio para o laboratório clínico. Este estudo propõem a construção de uma rotina computacional de análise do SCE que possa testar diferentes genes candidatos quanto à sua sensibilidade e especificidade para a detecção dos indivíduos afetados. A abordagem proposta é a contagem de variantes separadas por seu possível dano à proteína e frequência populacional para cada indivíduo de grupos afetado e controle em 168 indivíduos com SCE, sendo 49 com TEA e 119 controles. Após a formulação da contagem esses valores são submetidos a uma sequência de testes estatísticos, buscando diferença significativa em quantidade de mutações de todas as variantes isoladamente, das mutações de perda-de-função, ou não-sinônimas danosas como um conjunto e a aplicação de modelos de análise multivariada como: regressão logística, árvore de decisão, rede neural, máquinas de suporte de vetor e análise de componente principal para a elaboração de modelos mais complexos para o desenvolvimento na doença. Ao todo foram testadas 21 listas de genes, destas, 19 apresentaram ao menos um resultado significativo, sendo a análise de variantes isoladamente a que obteve maior número de eventos significativos. Desde variantes aparentemente protetoras (maior número no grupo controle), como as variantes não-sinônimas em via de RAS/MAPK quanto variantes de perda de códon de parada com

frequência populacional acima de 0.05 em genes de cromatina em maior número nos indivíduos com TEA. Nenhum dos modelos de análise multivariada obteve resultados significativos na discriminação entre os dois grupos. Devido ao pequeno número amostral os resultados deste estudo devem ser interpretados com limitações, sendo necessária a replicação deste cenário em outros bancos de dados. Entretanto, estes achados sugerem que diferentes tipos e frequências de variantes podem ter contribuições distintas para o desenvolvimento da doença a depender dos genes analisados, mais de que relações complexas entre as variantes de uma mesma lista de genes.

7 Abstract

Autism spectrum disorder (ASD) is a neurodevelopment disorder characterized by impairment in communication skills, behavior and social interactions that affects around 1-2% worldwide. To date the etiology of ASD has not yet been fully understood, but in the last 18 years many advances have been made to understand the genetic component related to the development of the clinical phenotype. With the advent of genomic scan analysis such as chromosome analysis by microarray and whole exome sequencing (WES) many advances have been made to understand the pathophysiology of the disease. About 10-15% of the cases can be explained by large losses or gains (deletions or duplications greater than 1000 base pairs) of the genetic material, which generally involve the disruption of one or more genes. Next generation sequencing methodologies were fundamental in the description of point mutations and small insertions and deletions associated with ASD. The WES has allowed many discoveries to be made about new candidate genes and mechanisms for the development of the disease. It is now claimed that *de novo* (non-inherited) and likely gene disruptive mutations, such as loss-of-function and non-synonymous changes with high prediction of damage by computational tools, in genes related to neurodevelopment are a major contributor to the disease mechanism. However, these mutations, in addition to not explaining the majority of cases, are rarely recurrent in the population, which makes it difficult to establish a definitive molecular diagnosis for most patients. WES is already a practice in clinical genetics laboratories and demonstrates high effectiveness for diseases that follow a Mendelian pattern of inheritance, and have an established genetic cause. In clinical practice WES is requested for cases of ASD, despite having different modes of inheritance and having more than 1,000 genes associated with the disease. Due to these characteristics the analysis of WES for ASD is a major challenge for the clinical laboratory. This study proposes the construction of a computerized WES analysis routine that can test different candidate genes for their sensitivity and specificity for the detection of affected individuals. The proposed approach consists in the counting of variants separated by their possible protein damage and population frequency for each individual from affected and control groups, this study analyzed 168 WES, being 49 with ASD and 119 controls. After counting formulation, these values are subjected to a sequence of statistical tests, seeking a significant difference in the amount of mutations of all the variants alone, loss-of-function or damaging missense mutations, and the application of models of multivariate analysis such as: logistic regression, decision tree, neural network, vector support machine and principal component analysis for the elaboration of more complex models for disease development. A total of 21 lists of genes were tested, of which 19 presented at least one significant result, and the analysis of variants alone was the one that obtained the largest number of significant events. From apparently protective variants (higher number in the control group), such as the missense variants in RAS/MAPK pathway as variants of stopgain with population frequency above 0.05 in chromatin genes in greater number in individuals with ASD. None of the multivariate analysis models had significant discrimination results between the two groups. Due to the small sample size, the results of this study should be interpreted with limitations, and it is necessary to

replicate these scenarios in other databases. However, these findings suggest that different types and frequencies of variants may have distinct contributions to disease development depending on the genes analyzed, rather than complex relationships between variants of the same gene list.

8 Reference

- Abrahams, Brett S. and Daniel H. Geschwind. 2008. "Advances in Autism Genetics: on the Threshold of a New Neurobiology." *Nature Reviews Genetics* 9(5):341–55.
- Akaike, Hirotugu. 1985. "Prediction and Entropy." Pp. 1–24 in *A Celebration of Statistics*. New York, NY: Springer, New York, NY.
- Anagnostopoulos, Christoforos and David J. Hand. 2012. "Hmeasure: the H-Measure and Other Scalar Classification Performance Metrics.." Retrieved (<https://CRAN.R-project.org/package=hmeasure>).
- Anagnostou, Evdokia et al. 2014. "Autism Spectrum Disorder: Advances in Evidence-Based Practice.." *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 186(7):509–19.
- Anastasiadis, Aristoklis D., George D. Magoulas, and Michael N. Vrahatis. 2005. "New Globally Convergent Training Scheme Based on the Resilient Propagation Algorithm." *Neurocomputing* 64:253–70.
- Aschard, Hugues et al. 2014. "Maximizing the Power of Principal-Component Analysis of Correlated Phenotypes in Genome-Wide Association Studies." *The American Journal of Human Genetics* 94(5):662–76.
- Aziz, Nazneen et al. 2015. "College of American Pathologists' Laboratory Standards for Next-Generation Sequencing Clinical Tests.." *Archives of pathology & laboratory medicine* 139(4):481–93.
- Barnard, Rebecca A., Matthew B. Pomaville, and Brian J. O'Roak. 2015. "Mutations and Modeling of the Chromatin Remodeler CHD8 Define an Emerging Autism Etiology." *Frontiers in Neuroscience* 9.
- BARRETO, ALEXANDRE S. 2011. *Modelos De Regressao*.
- Barrett, Jeffrey C. et al. 2017. "New Mutations, Old Statistical Challenges." *bioRxiv* 115964.
- Baxter, A J. et al. 2014. "The Epidemiology and Global Burden of Autism Spectrum Disorders." *Psychological Medicine* 1–13.
- Bertola, Débora R. et al. 2014. "Further Evidence of the Importance of RIT1 in Noonan Syndrome." *American Journal of Medical Genetics Part A* n/a–n/a.
- Betancur, Catalina. 2011. "Etiological Heterogeneity in Autism Spectrum Disorders: More Than 100 Genetic and Genomic Disorders and Still Counting.." *Brain Research* 1380:42–77.
- Betancur, Catalina, Takeshi Sakurai, and Joseph D. Buxbaum. 2009. "The Emerging Role of Synaptic Cell-Adhesion Pathways in the Pathogenesis of Autism Spectrum Disorders.." *Trends in neurosciences* 32(7):402–12.

- Bora, Emre, Aydan Aydin, Tuğba Sarac, Muhammed T. Kadak, and Sezen Köse. 2016. "Heterogeneity of Subclinical Autistic Traits Among Parents of Children with Autism Spectrum Disorder: Identifying the Broader Autism Phenotype with a Data-Driven Method." *Autism research : official journal of the International Society for Autism Research*.
- Bosl, William J., Helen Tager-Flusberg, and Charles A. Nelson. 2018. "EEG Analytics for Early Detection of Autism Spectrum Disorder: a Data-Driven Approach." *Scientific Reports* 8(1):6828.
- Bourgeron, Thomas. 2015. "From the Genetic Architecture to Synaptic Plasticity in Autism Spectrum Disorder.." *Nature Reviews Neuroscience* 16(9):551–63.
- Brandler, William M. et al. 2015. "Frequency and Complexity of De Novo Structural Mutation in Autism." *bioRxiv* 030270.
- Brandler, William M. et al. 2018. "Paternally Inherited Cis-Regulatory Structural Variants Are Associated with Autism.." *Science* 360(6386):327–31.
- Breiman, Leo. 2017. *Classification and Regression Trees*. Routledge.
- C Yuen, Ryan K. et al. 2017. "Whole Genome Sequencing Resource Identifies 18 New Candidate Genes for Autism Spectrum Disorder.." *Nature Neuroscience* 20(4):602–11.
- Cellot, Giada and Enrico Cherubini. 2014. "GABAergic Signaling as Therapeutic Target for Autism Spectrum Disorders." *Frontiers in Pediatrics* 2.
- Centers for Disease Control and Prevention. 2012. "Prevalence of Autism Spectrum Disorders--Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008.." *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002)* 61(3):1–19.
- Chahrour, Maria H. et al. 2012. "Whole-Exome Sequencing and Homozygosity Analysis Implicate Depolarization-Regulated Neuronal Genes in Autism." *PLoS Genetics* 8(4):e1002635.
- Chong, Jessica X. et al. 2015. "The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities." *The American Journal of Human Genetics* 97(2):199–215.
- Colvert, Emma et al. 2015. "Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample." *JAMA Psychiatry* 72(5):415–23.
- Cotney, Justin et al. 2015. "The Autism-Associated Chromatin Modifier CHD8 Regulates Other Autism Risk Genes During Human Neurodevelopment.." *Nature communications* 6:6404–4.
- Cukier, Holly N. et al. 2014. "Exome Sequencing of Extended Families with Autism Reveals Genes Shared Across Neurodevelopmental and Neuropsychiatric Disorders.." *Molecular autism* 5(1):1.
- Darnell, Jennifer C. et al. 2011. "FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism." *Cell* 146(2):247–61.

- De Rubeis, S and J D. Buxbaum. 2015. "Genetics and Genomics of Autism Spectrum Disorder: Embracing Complexity." *Human Molecular Genetics*.
- De Rubeis, Silvia et al. 2014. "Synaptic, Transcriptional and Chromatin Genes Disrupted in Autism." *Nature* 515(7526):209–15.
- Devlin, Bernie and Stephen W. Scherer. 2012. "Genetic Architecture in Autism Spectrum Disorder.." *Current Opinion in Genetics & Development* 22(3):229–37.
- Dieme, Binta et al. 2015. "Metabolomics Study of Urine in Autism Spectrum Disorders Using a Multiplatform Analytical Methodology.." *Journal of proteome research* acs.jproteome.5b00699.
- Eilbeck, Karen, Aaron Quinlan, and Mark Yandell. 2017. "Settling the Score: Variant Prioritization and Mendelian Disease." *Nature Reviews Genetics* 18(10):599–612.
- Fawcett, Tom. 2006. "An Introduction to ROC Analysis." *Pattern Recognition Letters* 27(8):861–74.
- Field, Andy. 2017. *Discovering Statistics Using IBM SPSS Statistics*. SAGE.
- Fombonne, Eric. 2018. "Editorial: the Rising Prevalence of Autism." *Journal of Child Psychology and Psychiatry* 59(7):717–20.
- Fritsch, Stefan and Frauke Guenther. 2016. "Neuralnet: Training of Neural Networks.." Retrieved (<https://CRAN.R-project.org/package=neuralnet>).
- Gallagher, D, A Voronova, M A. Zander, and G I. Cancino. 2014. "Ankrd11 Is a Chromatin Regulator Involved in Autism That Is Essential for Neural Development." *Developmental cell*.
- Gaugler, Trent et al. 2014. "Most Genetic Risk for Autism Resides with Common Variation." *Nature Genetics* 46(8):881–85.
- Geschwind, Daniel H. and Matthew W. State. 2015. "Gene Hunting in Autism Spectrum Disorder: on the Path to Precision Medicine." *The Lancet Neurology* 14(11):1109–20.
- Girirajan, Santhosh et al. 2013. "Refinement and Discovery of New Hotspots of Copy-Number Variation Associated with Autism Spectrum Disorder.." *American journal of human genetics* 92(2):221–37.
- Greenlaw, Keelin, Elena Szefer, Jinko Graham, Mary Lesperance, and Farouk S. Nathoo. 2017. "A Bayesian Group Sparse Multi-Task Regression Model for Imaging Genetics." *Bioinformatics* 33(16):2513–22.
- Griesi-Oliveira, K et al. 2014. "Modeling Non-Syndromic Autism and the Impact of TRPC6 Disruption in Human Neurons.." *Molecular psychiatry*.
- Grove, Jakob et al. 2017. "Common Risk Variants Identified in Autism Spectrum Disorder." *bioRxiv* 224774.

- Guo, Hui, Zhengmao Hu, Jingping Zhao, and Kun Xia. 2011. "Genetics of Autism Spectrum Disorders.." *Zhong nan da xue xue bao. Yi xue ban = Journal of Central South University. Medical sciences* 36(8):703–11.
- Hair, J F., W C. Black, B J. Babin, R E. Anderson, and R L. Tatham. 2009. *Análise Multivariada De Dados*.
- Hampton, Tracy. 2017. "Early Brain Imaging in Infants May Help Predict Autism." *JAMA* 318(13):1211–12.
- Hand, David J. 2009. "Measuring Classifier Performance: a Coherent Alternative to the Area Under the ROC Curve." *Machine Learning* 77(1):103–23.
- Hartigan, J A. and M A. Wong. 1979. "Algorithm as 136: a K-Means Clustering Algorithm." *Applied Statistics* 28(1):100.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2013. *The Elements of Statistical Learning*. New York, NY: Springer Science & Business Media.
- He, Xin et al. 2013. "Integrated Model of De Novo and Inherited Genetic Variants Yields Greater Power to Identify Risk Genes." edited by S. M. Williams. *PLoS Genetics* 9(8):e1003671.
- Hogart, Amber, Raman P. Nagarajan, Katherine A. Patzel, Dag H. Yasui, and Janine M. Lasalle. 2007. "15q11-13 GABA_A Receptor Genes Are Normally Biallelically Expressed in Brain Yet Are Subject to Epigenetic Dysregulation in Autism-Spectrum Disorders.." *Human Molecular Genetics* 16(6):691–703.
- Iossifov, Ivan et al. 2012. "De Novo Gene Disruptions in Children on the Autistic Spectrum.." *Neuron* 74(2):285–99.
- Iossifov, Ivan et al. 2014. "The Contribution of De Novo Coding Mutations to Autism Spectrum Disorder." *Nature* 515(7526):216–21.
- Iossifov, Ivan et al. 2015. "Low Load for Disruptive Mutations in Autism Genes and Their Biased Transmission." *Proceedings of the National Academy of Sciences of the United States of America* 112(41):E5600–E5607.
- Jacquemont, Sébastien et al. 2014. "A Higher Mutational Burden in Females Supports a 'Female Protective Model' in Neurodevelopmental Disorders." *American journal of human genetics*.
- Jiang, Yong-hui et al. 2013. "Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing." *The American Journal of Human Genetics*.
- Klei, Lambertus et al. 2012. "Common Genetic Variants, Acting Additively, Are a Major Source of Risk for Autism." *Molecular autism* 3(1):9.
- Kolevzon, Alexander et al. 2014. "A Pilot Controlled Trial of Insulin-Like Growth Factor-1 in Children with Phelan-McDermid Syndrome.." *Molecular autism* 5(1):54.

- Krishnan, Arjun et al. 2016. "Genome-Wide Prediction and Functional Characterization of the Genetic Basis of Autism Spectrum Disorder." *Nature Neuroscience*.
- Krumm, Niklas et al. 2013. "Transmission Disequilibrium of Small CNVs in Simplex Autism." *The American Journal of Human Genetics*.
- Krumm, Niklas et al. 2015. "Excess of Rare, Inherited Truncating Mutations in Autism." *Nature Genetics* 47(6):582–88.
- Kumar, R A. 2011. "Autism, Mutations, and the Environment: Insights From Exome Sequencing.." *Clinical Genetics* 80(4):331–33.
- la Torre-Ubieta, de, Luis, Hyejung Won, Jason L. Stein, and Daniel H. Geschwind. 2016. "Advancing the Understanding of Autism Disease Mechanisms Through Genetics.." *Nature medicine* 22(4):345–61.
- Lapin, Valentina et al. 2016. "Regulating Whole Exome Sequencing as a Diagnostic Test.." *Human Genetics* 135(6):655–73.
- Lek, Monkol et al. 2016. "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." *Nature* 536(7616):285–91.
- Letai, Anthony. 2017. "Functional Precision Cancer Medicine—Moving Beyond Pure Genomics." *Nature medicine* 23(9):1028–35.
- Li, Baihua, Arjun Sharma, James Meng, Senthil Purushwalkam, and Emma Gowen. 2017. "Applying Machine Learning to Identify Autistic Adults Using Imitation: an Exploratory Study" edited by M. Sakakibara. *PLOS ONE* 12(8):e0182652.
- Li, Heng and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.." *Bioinformatics* 25(14):1754–60.
- Li, Jinchen et al. 2018. "A Comparative Study of the Genetic Components of Three Subcategories of Autism Spectrum Disorder." *Molecular psychiatry* 383:896.
- Lichtenstein, Paul, Eva Carlstrom, Maria Rastam, Christopher Gillberg, and Henrik Anckarsater. 2010. "The Genetics of Autism Spectrum Disorders and Related Neuropsychiatric Disorders in Childhood." *American Journal of Psychiatry* 167(11):1357–63.
- Lin, I-Fan et al. 2014. "The Effect of Intranasal Oxytocin Versus Placebo Treatment on the Autonomic Responses to Human Sounds in Autism: a Single-Blind, Randomized, Placebo-Controlled, Crossover Design Study." *Molecular autism* 5(1):20.
- Liu, Li et al. 2014. "DAWN: a Framework to Identify Autism Genes and Subnetworks Using Gene Expression and Genetics.." *Molecular autism* 5(1):22.
- Lord, Catherine and Somer L. Bishop. 2015. "Recent Advances in Autism Research as Reflected in DSM-5 Criteria for Autism Spectrum Disorder.." *Annual review of clinical psychology* 11(1):53–70.

- Ludbrook, John. 2008. "Analysis of 2 X 2 Tables of Frequencies: Matching Test to Experimental Design.." *International journal of epidemiology* 37(6):1430–35.
- Luo, Weijun, Chaolin Zhang, Yong-hui Jiang, and Cory R. Brouwer. 2018. "Systematic Reconstruction of Autism Biology From Massive Genetic Mutation Profiles." *Science Advances* 4(4):e1701799.
- Luo, Xianghua and Gary S. Sorock. 2008. "Analysis of Recurrent Event Data Under the Case-Crossover Design with Applications to Elderly Falls." *Statistics in Medicine* 27(15):2890–2901.
- Machiela, Mitchell J. et al. 2011. "Evaluation of Polygenic Risk Scores for Predicting Breast and Prostate Cancer Risk.." *Genetic epidemiology* 35(6):506–14.
- Manning-Courtney, Patricia et al. 2013. "Autism Spectrum Disorders." *Current Problems in Pediatric and Adolescent Health Care* 43(1):2–11.
- Manrai, Arjun K. et al. 2016. "Genetic Misdiagnoses and the Potential for Health Disparities." *New England Journal of Medicine* 375(7):655–65.
- Mardis, E R. 2008. "Next-Generation DNA Sequencing Methods." *Annu Rev Genomics Hum Genet.*
- Matthijs, Gert et al. 2015. "Guidelines for Diagnostic Next-Generation Sequencing." *European Journal of Human Genetics* 24(1):2–5.
- Matullo, Giuseppe, Cornelia Di Gaetano, and Simonetta Guarnera. 2013. "Next Generation Sequencing and Rare Genetic Variants: From Human Population Studies to Medical Genetics.." *Environmental and molecular mutagenesis* 518–32.
- McKenna, Aaron et al. 2010. "The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data.." *Genome Research* 20(9):1297–1303.
- Meyer, David, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. 2017. "E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)." Retrieved (<https://CRAN.R-project.org/package=e1071>).
- Michaelson, Jacob J. et al. 2012. "Whole-Genome Sequencing in Autism Identifies Hot Spots for De Novo Germline Mutation." *Cell* 151(7):1431–42.
- Mingoti, S A. 2005. *Análise De Dados Através De Métodos De Estatística Multivariada: Uma Abordagem Aplicada*.
- Mitchell, Kevin J. 2012. "What Is Complex About Complex Disorders?." *Genome Biology* 13(1):237.
- Moreira, Eloisa S. et al. 2016. "Detection of Small Copy Number Variations (CNVs) in Autism Spectrum Disorder (ASD) by Custom Array Comparative Genomic Hybridization (aCGH)." *Research in Autism Spectrum Disorders* 23:145–51.
- Mughal, Sajid et al. 2017. "Pheno4J: a Gene to Phenotype Graph Database." *Bioinformatics* 33(20):3317–19.

- Naslavsky, Michel S. et al. 2017. "Exomic Variants of an Elderly Cohort of Brazilians in the ABraOM Database.." *Human Mutation* 38(7):751–63.
- Neale, Benjamin M. et al. 2013. "Patterns and Rates of Exonic De Novo Mutations in Autism Spectrum Disorders." *Nature News* 485(7397):242–45.
- O'Roak, Brian J. et al. 2011. "Exome Sequencing in Sporadic Autism Spectrum Disorders Identifies Severe De Novo Mutations.." *Nature Genetics* 43(6):585–89.
- O'Roak, Brian J., Laura Vives, Santhosh Girirajan, et al. 2012a. "Sporadic Autism Exomes Reveal a Highly Interconnected Protein Network of De Novo Mutations.." *Nature News* 485(7397):246–50.
- O'Roak, Brian J., Laura Vives, Wenqing Fu, et al. 2012b. "Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders.." *Science* 338(6114):1619–22.
- Pedersen, Brent S. and Aaron R. Quinlan. 2017. "Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy." *The American Journal of Human Genetics* 100(3):406–13.
- Petrovski, Slave, Quanli Wang, Erin L. Heinzen, Andrew S. Allen, and David B. Goldstein. 2013. "Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes" edited by S. M. Williams. *PLoS Genetics* 9(8).
- Pinto, Dalila et al. 2014. "Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders." *The American Journal of Human Genetics* 94(5):677–94.
- Pirooznia, Mehdi et al. 2014. "Validation and Assessment of Variant Calling Pipelines for Next-Generation Sequencing." *Human Genomics* 8(1):14.
- Qin, J. 1997. "A Goodness-of-Fit Test for Logistic Regression Models Based on Case-Control Data." *Biometrika* 84(3):609–18.
- Ramachandran, Sohini et al. 2005. "Support From the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa.." *Proceedings of the National Academy of Sciences of the United States of America* 102(44):15942–47.
- Retterer, Kyle et al. 2016. "Clinical Application of Whole-Exome Sequencing Across Clinical Indications.." *Genetics in Medicine* 18(7):696–704.
- Revelle, W. 2017. "Psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA." Retrieved (<https://CRAN.R-project.org/package=psych> Version = 1.7.8).
- Richards, Sue et al. 2015. "Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.." *Genetics in Medicine*.
- Ridley, Mark. 2009. *Evolution*. Wiley-Blackwell.

- Riedmiller, M, H B., 1993. 1993. "A Direct Adaptive Method for Faster Backpropagation Learning: the RPROP Algorithm." *Neural Networks, IEEE International Conference on*.
- Riikonen, Raili. 2016. "Treatment of Autistic Spectrum Disorder with Insulin-Like Growth Factors." *European Journal of Paediatric Neurology*.
- Ronemus, Michael, Ivan Iossifov, Dan Levy, and Michael Wigler. 2014. "The Role of De Novo Mutations in the Genetics of Autism Spectrum Disorders.." *Nature Publishing Group* 15(2):133–41.
- Rotschafer, Sarah E., Sonya Marshak, and Karina S. Cramer. 2015. "Deletion of Fmr1 Alters Function and Synaptic Inputs in the Auditory Brainstem" edited by O. Birmingham-McDonogh. *PLoS ONE* 10(2):e0117266.
- Samocha, Kaitlin E. et al. 2014. "A Framework for the Interpretation of De Novo Mutation in Human Disease." *Nature Genetics* 46(9):944–.
- Sanders, Stephan J. et al. 2012. "De Novo Mutations Revealed by Whole-Exome Sequencing Are Strongly Associated with Autism.." *Nature* 485(7397):237–41.
- Sandin, Sven et al. 2014. "The Familial Risk of Autism.." *JAMA* 311(17):1770–77.
- Sandin, Sven et al. 2017. "The Heritability of Autism Spectrum Disorder." *JAMA* 318(12):1182–84.
- Schaefer, G B. and Nancy J. Mendelsohn. 2013. "Clinical Genetics Evaluation in Identifying the Etiology of Autism Spectrum Disorders: 2013 Guideline Revisions." *Genetics in Medicine* 15(5):399–407.
- Scherer, Stephen W. and Geraldine Dawson. 2011. "Risk Factors for Autism: Translating Genomic Discoveries Into Diagnostics." *Human Genetics* 130(1):123–48.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. "Biological Insights From 108 Schizophrenia-Associated Genetic Loci.." *Nature News* 511(7510):421–27.
- Schork, N J. 1997. "Genetics of Complex Disease: Approaches, Problems, and Solutions.." *American Journal of Respiratory and Critical Care Medicine* 156(4 Pt 2):S103–9.
- Sebat, Jonathan et al. 2007. "Strong Association of De Novo Copy Number Mutations with Autism.." *Science* 316(5823):445–49.
- Shemesh, Elad, Assaf Rudich, Ilana Harman-Boehm, and Tali Cukierman-Yaffe. 2012. "Effect of Intranasal Insulin on Cognitive Function: a Systematic Review." *The Journal of clinical endocrinology and metabolism* 97(2):366–76.
- Sobreira, Nara, François Schiettecatte, Corinne Boehm, David Valle, and Ada Hamosh. 2015. "New Tools for Mendelian Disease Gene Identification: PhenoDB Variant Analysis Module; and GeneMatcher, a Web-Based Tool for Linking Investigators with an Interest in the Same Gene.." *Human Mutation* 36(4):425–31.

- Stein, Jason L., Neelroop N. Parikshak, and Daniel H. Geschwind. 2013. "Rare Inherited Variation in Autism: Beginning to See the Forest and a Few Trees.." *Neuron* 77(2):209–11.
- Steinberg, Julia and Caleb Webber. 2013. "The Roles of FMRP-Regulated Genes in Autism Spectrum Disorder: Single- and Multiple-Hit Genetic Etiologies.." *American journal of human genetics* 93(5):825–39.
- Stessman, Holly A. F. et al. 2017. "Targeted Sequencing Identifies 91 Neurodevelopmental-Disorder Risk Genes with Autism and Developmental-Disability Biases." *Nature Genetics* 49(4):515–26.
- Swets, J A., R M. Dawes, and J Monahan. 2000. "Better Decisions Through Science.." *Scientific American* 283(4):82–87.
- Szatmari, Peter et al. 2007. "Mapping Autism Risk Loci Using Genetic Linkage and Chromosomal Rearrangements." *Nature Genetics* 39(3):319–28.
- Team, R C. 2017. "R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.." Retrieved (<https://www.R-project.org/>).
- Therneau, Terry, Beth Atkinson, and Brian Ripley. 2017. "Rpart: Recursive Partitioning and Regression Trees.." Retrieved (<https://CRAN.R-project.org/package=rpart>).
- Tordjman, Sylvie et al. 2014. "Gene × Environment Interactions in Autism Spectrum Disorders: Role of Epigenetic Mechanisms." *Schizophrenia* 5.
- Trujillano, Daniel et al. 2016. "Clinical Exome Sequencing: Results From 2819 Samples Reflecting 1000 Families." *European Journal of Human Genetics*.
- Turner, Tychele N. et al. 2017. "Genomic Patterns of De Novo Mutation in Simplex Autism." *Cell* 171(3):710–12.
- Uddin, Mohammed et al. 2014. "Brain-Expressed Exons Under Purifying Selection Are Enriched for De Novo Mutations in Autism Spectrum Disorder.." *Nature Genetics*.
- Vaishnavi, Varadharajan et al. 2014. "Mining the 3'UTR of Autism-Implicated Genes for SNPs Perturbing MicroRNA Regulation." *Genomics, Proteomics & Bioinformatics*.
- Wang, Jing, Leon Raskin, David C. Samuels, Yu Shyr, and Yan Guo. 2015. "Genome Measures Used for Quality Control Are Dependent on Gene Function and Ancestry." *Bioinformatics* 31(3):318–23.
- Wang, Kai, Mingyao Li, and Hakon Hakonarson. 2010. "ANNOVAR: Functional Annotation of Genetic Variants From High-Throughput Sequencing Data.." *Nucleic Acids Research* 38(16):e164–64.
- Wei, Taiyun and Viliam Simko. 2017. "R Package 'Corrplot': Visualization of a Correlation Matrix (Version 0.84)." Retrieved (<https://github.com/taiyun/corrplot>).

- Weiner, Daniel J. et al. 2017. "Polygenic Transmission Disequilibrium Confirms That Common and Rare Variation Act Additively to Create Risk for Autism Spectrum Disorders." *Nature Genetics* 450:86–11.
- Weiss, Lauren A. et al. 2009. "A Genome-Wide Linkage and Association Scan Reveals Novel Loci for Autism." *Nature News* 461(7265):802–8.
- Werling, Donna M. et al. 2017. "Limited Contribution of Rare, Noncoding Variation to Autism Spectrum Disorder From Sequencing of 2,076 Genomes in Quartet Families." *bioRxiv* 127043.
- Wickham, H. 2009. "Ggplot2: Elegant Graphics for Data Analysis.." *Springer-Verlag New York*.
- Woodbury-Smith, Marc and Stephen W. Scherer. 2018. "Progress in the Genetics of Autism Spectrum Disorder.." *Developmental Medicine & Child Neurology* 60(5):445–51.
- Woodbury-Smith, Marc et al. 2015. "Using Extended Pedigrees to Identify Novel Autism Spectrum Disorder (ASD) Candidate Genes." *Human Genetics* 134(2):191–201.
- Wu, Jiaxin, Yanda Li, and Rui Jiang. 2014. "Integrating Multiple Genomic Data to Predict Disease-Causing Nonsynonymous Single Nucleotide Variants in Exome Sequencing Studies" edited by G. Gibson. *PLoS Genetics* 10(3).
- Wu, Xindong et al. 2007. "Top 10 Algorithms in Data Mining." *Knowledge and Information Systems* 14(1):1–37.
- Xue, Yali et al. 2012. "Deleterious-and Disease-Allele Prevalence in Healthy Individuals: Insights From Current Predictions, Mutation Databases, and Population-Scale Resequencing." *American journal of human genetics* 91(6):1022–32. Retrieved (<http://www.sciencedirect.com/science/article/pii/S0002929712005381>).
- Yamamoto, Guilherme L. et al. 2014. "Mutations in PCYT1A Cause Spondylometaphyseal Dysplasia with Cone-Rod Dystrophy.." *American journal of human genetics* 94(1):113–19.
- Yu, Timothy W. et al. 2013. "Using Whole-Exome Sequencing to Identify Inherited Causes of Autism." *Neuron* 77(2):259–73.
- Yuen, Ryan K. C. et al. 2015. "Whole-Genome Sequencing of Quartet Families with Autism Spectrum Disorder.." *Nature medicine*.
- Yuen, Ryan K. et al. 2016. "Genome-Wide Characteristics of De Novo Mutations in Autism." *npj Genomic Medicine* 1:16027.
- Zare, Marzieh, Zahra Rezvani, and April A. Benasich. 2016. "Automatic Classification of 6-Month-Old Infants at Familial Risk for Language-Based Learning Disorder Using a Support Vector Machine." *Clinical Neurophysiology* 127(7):2695–2703.

Zhao, Yunpeng, Qing Pan, and Chengan Du. 2018. "Logistic Regression Augmented Community Detection for Network Data with Application in Identifying Autism-Related Gene Pathways." *Biometrics*.

Zheng, Xiuwen et al. 2012. "A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data." *Bioinformatics* 28(24):3326–28.

9 Attachments

9.1 Attachment 1

AAK1 ANP32E ATRNL1 C3orf17 CDC5L CNOT2 DEPDC5 EIF4G3
ABCA2 AP1G1 ATXN1 C4orf21 CDH10 CNOT6 DGCR8 EIF6
ABCA3 AP1S1 ATXN2 C5orf42 CDH18 CNTN1 DGKB ELAVL1
ABCC5 AP3B1 ATXN3 C5orf45 CDH20 CNTN2 DGKD ELAVL3
ABCC8 AP3B2 ATXN7L3B C8orf34 CDH22 CNTN4 DGKG ELAVL4
ABCD2 AP3M2 AUTS2 C8orf46 CDH23 CNTN6 DGKH ELK4
ABCG2 AP4B1 AVIL C9orf72 CDH24 CNTNAP1 DGKI ELMO1
ABCG4 AP5M1 B3GAT1 CA10 CDH7 CNTNAP2 DHX36 ELMO2
ABI2 APBA1 B4GALNT1 CA11 CDHR3 CNTNAP4 DICER1 ELMOD1
ACACA APBA2 B4GALNT4 CA8 CDIP1 CNTNAP5 DIRAS2 ELOVL4
ACAP3 APBB1 B4GALT6 CABP1 CDK13 COA5 DISP2 ELP4
ACBD3 APBB2 BAALC CACHD1 CDK14 COL13A1 DIXDC1 EMC4
ACSL3 APC BAI1 CACNA1A CDK17 COL19A1 DLG1 EML5
ACSL6 APLP1 BAI2 CACNA1B CDK5R1 COL27A1 DLG2 ENAH
ACTL6B APLP2 BAI3 CACNA1D CDK8 CORO2B DLG3 ENO2
ADAM11 APOO BARHL2 CACNA1E CDKL5 CPD DLG4 ENOX1
ADAM22 APP BAZ2B CACNA1G CDON CPLX1 DLGAP1 ENY2
ADAM23 APPL1 BBS7 CACNA2D1 CDR2L CPLX2 DLGAP4 EOMES
ADAMTS16 AQP4 BCAN CACNA2D2 CDS1 CPLX3 DLK2 EP400
ADAMTS18 AQR BCAS1 CACNB3 CDYL2 CPNE6 DMTF1 EPB41
ADARB1 ARGLU1 BCAS3 CACNB4 CECR2 CPNE9 DMXL1 EPB41L1
ADCY1 ARHGAP19-SLIT1 BCL11A CACNG2 CELF3 CPT1C DMXL2
EPB41L3
ADCY8 ARHGAP22 BCL2L2 CACNG7 CELF4 CRAMP1L DNAJB2
EPB41L4B
ADCYAP1R1 ARHGAP26 BCL2L2-PABPN1 CADM1 CELF5 CRB1
DNAJC17 EPB41L5
ADD1 ARHGAP29 BDP1 CADM3 CELF6 CRHR1 DNAJC6 EPC2
ADD2 ARHGAP33 BEGAIN CADM4 CELSR2 CRMP1 DNAJC7 EPHA6
ADD3 ARHGAP44 BEND6 CADPS CELSR3 CRTAM DNER EPHA7
AEBP2 ARHGDIG BIN1 CADPS2 CENPT CRTC1 DNM1 EPHB1
AES ARHGEF11 BIVM CALB2 CEP120 CRY1 DNMT1 EPHB2
AFF2 ARHGEF25 BIVM-ERCC5 CALN1 CEP170 CSMD1 DNMT3A ERC1
AFG3L1P ARHGEF4 BLCAP CAMK2B CEP192 CSPG5 DOCK3 ERC2
AFTPH ARHGEF7 BMPER CAMK2G CEP95 CSRNP3 DOCK4 ERCC3
AGAP1 ARID4A BMPR2 CAMK4 CERCAM CSTF3 DOCK9 ERCC6
AGAP2 ARID4B BPTF CAMKK1 CERKL CTIF DOK6 ERGIC2
AGAP3 ARL1 BRD8 CAMKK2 CERS5 CTNNB1 DOPEY1 ERH

AGPAT4 ARL8A BRD9 CAMKV CERS6 CTNND2 DOPEY2 ERP44
AGTPBP1 ARMC8 BRSK1 CAMSAP1 CFL1 CTPS2 DPF1 ESYT3
AHI1 ARMC9 BRSK2 CAMSAP2 CHD2 CTTNBP2 DPF3 ETV1
AJAP1 ARMCX5 BRWD1 CAMTA1 CHD3 CUL9 DPP10 EVL
AKAP10 ARMCX5-GPRASP2 BRWD3 CAND1 CHD6 CWC27 DPP6
EXOC1
AKAP11 ARNT2 BSDC1 CAPRIN2 CHD9 CWF19L2 DPYSL2 EXOC2
AKAP9 ARPP21 BSN CASD1 CHL1 CXorf23 DPYSL3 EXOSC10
AKD1 ARVCF BTBD1 CASK CHM CXXC1 DPYSL5 EXT2
AKT3 ASAP1 BTBD10 CASKIN1 CHN1 CYP26B1 DRG1 EZH1
ALS2 ASB3 BTBD9 CATSPER2 CHRD CYP46A1 DROSHA FADS2
AMER2 ASIC1 BTN2A2 CBLN1 CHRNB2 DAAM1 DTNA FAIM2
AMIGO1 ASIC2 BTRC CBX1 CHRNB3 DAB1 DTNB FAM120A
AMPH ASPHD1 C10orf35 CCAR1 CHST10 DAB2IP DTX3 FAM120C
AMZ2 ASPHD2 C11orf30 CCBL1 CIC DACT1 DUSP11 FAM126B
ANAPC5 ASTN1 C12orf29 CCDC111 CIT DAGLA DVL2 FAM131B
ANGEL2 ASTN2 C12orf76 CCDC115 CKAP5 DBC1 SYNC1H1 FAM135A
ANK1 ASXL1 C14orf132 CCDC132 CLASP1 DCAF17 SYNC1LI2 FAM13B
ANK2 ATCAY C14orf2 CCDC136 CLASP2 DCAF8 DYRK1A FAM160A2
ANK3 ATF2 C14orf37 CCDC6 CLCN3 DCC DZIP3 FAM168A
ANKRD12 ATG4C C15orf27 CCDC88A CLCN6 DCLK1 E4F1 FAM171A2
ANKRD13B ATL1 C17orf85 CCDC92 CLIP2 DCLK2 EBAG9 FAM171B
ANKRD13C ATP1B2 C18orf32 CCNC CLIP3 DCTN1 ECE2 FAM179B
ANKRD17 ATP2B1 C1orf126 CCNG2 CLK4 DCTN2 EDIL3 FAM188A
ANKRD26 ATP2B2 C1orf51 CCNL CLSTN3 DCUN1D4 EFHA2 FAM193A
ANKRD6 ATP2B3 C1orf61 CCNT2 CLTC DDX26B EFR3B FAM199X
ANKS1B ATP2C1 C1QTNF3 CCNY CLUHP3 DDX42 EHBP1 FAM19A5
ANKS6 ATP6V1G2 C1QTNF3-AMACR CCP110 CLVS1 DDX6 EHMT1
FAM216A
ANKZF1 ATP6V1G2-DDX39B C20orf112 CCSAP CMC2 DEAF1 EIF2B5
FAM217B
ANLN ATP8A1 C20orf24 CD200 CNDP1 DEF8 EIF4A2 FAM228B
ANO10 ATP8A2 C20orf27 CDC42BPA CNIH2 DENND4A EIF4E3 FAM5B
ANO4 ATRN C2CD5 CDC42BPB CNKS2 DENND5B EIF4G2 FAM5C
FAM73B GAP43 HAUS3 KCNAB3 KRT222 MAP4K3 MTERFD2 NKAIN3
FAM98B GARNL3 HCN1 KCNC3 KSR2 MAP4K4 MTHFD1L NKIRAS1
FAN1 GDAP1L1 HDAC2 KCND2 L1CAM MAP4K5 MTMR9 NKTR
FANCL GDPD1 HECTD2 KCND3 LANCL1 MAP6D1 MTPAP NLGN1
FAT2 GEMIN7 HECTD4 KCNH1 LAS1L MAP7D2 MTSS1L NLGN2
FAXC GFAP HECW1 KCNH7 LATS1 MAPK1 MTUS1 NLGN3
FBRSL1 GFOD1 HERC1 KCNIP1 LCOR MAPK10 MUM1 NLGN4X
FBXL14 GHITM HERC3 KCNIP4 LDLRAD3 MAPK4 MYCBP2 NLK
FBXL16 GIGYF2 HERPUD2 KCNJ6 LEMD3 MAPK8 MYH10 NMNAT2
FBXL20 GIT2 HINFP KCNK1 LENG8 MAPK8IP1 MYO15A NNAT

FBXL3 GLCCI1 HIPK1 KCNK10 LENG9 MAPK8IP3 MYO5A NOL4
FBXO11 GLCE HIPK2 KCNK12 LEO1 MAPK9 MYT1 NOP58
FBXO16 GLMN HIST1H2BC KCNK9 LGALSL MAPRE3 MYT1L NOVA1
FBXO27 GLRA2 HIST1H4E KCNMB4 LGI1 MAPT N4BP2L2 NOVA2
FBXO3 GLRB HIVEP2 KCNN1 LGI3 MARCH8 NAA38 NPAS3
FBXO9 GLT25D2 HMG20A KCNN2 LHFPL4 MARK1 NAB1 NPAT
FBXW11 GMFB HMGCLL1 KCNQ2 LIMK1 MARK4 NAE1 NPFFR1
FBXW7 GNAI1 HMGCR KCNQ3 LIN7C MAST1 NALCN NPM2
FER GNAL HMP19 KCTD13 LIX1 MBD5 NAP1L2 NPPA-AS1
FEZ1 GNAO1 HNRNPA0 KCTD3 LLGL1 MBOAT2 NAP1L5 NPR2
FGF14 GNAZ HNRNPH3 KCTD6 LMAN1L MBTPS2 NAPB NPTN
FGF17 GNB3 HNRNPR KCTD8 LMBR1 MCF2L NAV2 NPTX1
FGF5 GNB5 HOOK3 KDM2B LMBRD1 MCF2L2 NAV3 NR1D1
FGFR2 GNL1 HP1BP3 KDM4C LMNB1 MCTP1 NBAS NRBP1
FIG4 GNPDA2 HPCA KHDRBS2 LMTK2 MCU NBEA NRCAM
FLT3 GNPTAB HPCAL4 KHDRBS3 LOC100131825 MDGA1 NCAM1

NRD1

FMN2 GOLM1 HR KIAA0319 LOC100379224 MDN1 NCAM2 NRIP2
FMNL2 GPM6A HRH3 KIAA0408 LONRF2 MEA1 NCAN NRSN1
FMR1 GPM6B HS3ST1 KIAA0430 LPHN1 MED12L NCBP1 NRSN2
FN1 GPR107 HSF4 KIAA0513 LPHN3 MED13L NCDN NRXN2
FNBP4 GPR158 ICA1L KIAA0895L LPPR2 MEGF10 NCKAP1 NRXN3
FOCAD GPR158-AS1 IDUA KIAA0930 LRCH1 MEGF11 NCOA1 NSD1
FOXJ3 GPR180 IFT46 KIAA1033 LRCH2 MEGF8 NCOA2 NSF
FOXK2 GPR37L1 IGFBP7 KIAA1109 LRCH3 METTL10 NCOA6 NSG1
FOXN2 GPR52 IGSF21 KIAA1191 LRP1B METTL21A NCOA7 NTM
FOXP1 GPR75-ASB3 IGSF8 KIAA1199 LRRC40 METTL21D NCOR1

NTRK2

FRMD5 GPR83 IKBKAP KIAA1324L LRRC7 MGARP NCS1 NTRK3
FRS2 GPR98 IL16 KIAA1456 LRRFIP2 MGAT3 NDFIP1 NUAK1
FRY GPRASP1 INPP4A KIAA1468 LRRTM3 MGAT4A NDP NUDT3
FSD1 GPRASP2 INPP5F KIAA1598 LRRTM4 MGAT5B NDRG2 NUFIP1
FSTL4 GRAMD1B INTS3 KIAA1704 LSAMP MID1 NDRG3 NUFIP2
FUBP3 GRIA1 INTS4 KIAA1841 LSS MINK1 NDRG4 NUP133
FUT8 GRIA2 INTU KIF1A LTN1 MIR497HG NDST3 NUP153
FUT9 GRIA3 IRX2 KIF1B LUC7L MLL NDUFA6 NUP93
FXYD7 GRIA4 ITFG1 KIF21A LY6G5B MLL2 NEDD4L NXPH3
FZD3 GRID1 ITM2C KIF2A LY6H MLL3 NEFH OCIAD1
FZR1 GRID2 ITPR1 KIF3A LYPLAL1 MLL5 NEFL OGFRL1
G3BP2 GRIK1 ITSN1 KIF3C LYST MLST8 NEFM OLFM1
GABBR1 GRIK2 IWS1 KIF5A MACF1 MMP16 NEGR1 OLFM2
GABBR2 GRIK5 JAK2 KIF5B MACROD2 MMP24 NEK1 OLFM3
GABRA1 GRIN1 JAKMIP2 KIF5C MADD MMS19 NELL2 OPCML
GABRA2 GRIN2A JAM2 KIFAP3 MAG MOG NEMF OPHN1

GABRA3 GRIN2B JARID2 KIFC2 MAGEE1 MON2 NEO1 OSBP
GABRA6 GRIN2C JMJD1C KIN MAGI1 MORN1 NETO2 OSBP2
GABRB1 GRM1 JMJD7 KLC2 MAGI2 MPDZ NEURL4 OSBPL2
GABRB2 GRM3 JMJD7-PLA2G4B KLHL18 MAGI3 MPHOSPH6
NEUROD2 OSBPL6
GABRB3 GRM4 JMY KLHL2 MAGOHB MPHOSPH9 NEUROD6 OSBPL7
GABRD GRM5 JPH3 KLHL24 MAML3 MPP6 NF1 OSBPL8
GABRG1 GSTA4 JPH4 KLHL3 MANEAL MREG NF2 OTUB1
GABRG2 GTF2H4 KALRN KLHL32 MAP1A MRPL50 NFASC OTUD6B
GAD1 GTPBP10 KANSL2 KLHL5 MAP2 MRPS21 NFAT5 OTUD7A
GAD2 GUSBP4 KAT2A KLHL8 MAP2K5 MRPS30 NINL P2RX7
GALNT13 HABP4 KAT6B KNDC1 MAP3K10 MSANTD3-TMEFF1 NIPA1
P4HTM
GALNT5 HACE1 KBTBD3 KPNA3 MAP3K12 MSANTD4 NIPAL3 PAAF1
GALNT9 HAPLN2 KCNA2 KREMEN2 MAP4 MSH2 NKAIN1 PACS1
GAN HAUS2 KCNAB1 KRIT1 MAP4K2 MTA3 NKAIN2 PACSIN1
PAGR1 PLD5 PSIP1 RFC1 SEC31B SMAD4 STX1B TMC2
PAK1 PLEKHA6 PSME3 RFTN2 SEC62 SMARCA2 STX6 TMCC2
PAK3 PLEKHB1 PSPC1 RFWD2 SECISBP2L SMARCAD1 STXBP1 TMEFF1
PAK7 PLEKHG5 PTCHD1 RFX3 SEL1L3 SMARCD3 STXBP5 TMEFF2
PALM PLK2 PTK2 RGS11 SEMA4F SMC1A STXBP5-AS1 TMEM116
PANK4 PLXDC2 PTK2B RGS12 SEMA6A SMG1 SULT4A1 TMEM130
PAPD5 PLXNA3 PTN RGS7 SENP6 SMG7 SUPT3H TMEM132A
PAPL PLXNB3 PTPN13 RGS8 SENP7 SNAP25 SUPT7L TMEM135
PAPSS1 PMP2 PTPN5 RHBDL1 SEP15 SNAPC3 SUV420H1 TMEM14C
PAQR6 PMPCB PTPRA RHOT1 SEPT3 SNCA SUV420H2 TMEM170A
PARD6A PMVK PTPRD RICTOR SEPT4 SNCAIP SV2A TMEM175
PARP2 POGZ PTPRG-AS1 RIF1 SEPT8 SNCB SV2B TMEM180
PAX6 POLA1 PTPRN RIMS1 SERF2 SNTG1 SYBU TMEM189-UBE2V1
PAXIP1 POLB PTPRR RIMS2 SERF2-C15ORF63 SNX14 SYN1 TMEM196
PCBP3 POLE PTPRS RIMS3 SERGEF SNX2 SYN2 TMEM200B
PCCA POLR3A PTPRZ1 RIT2 SEZ6 SNX25 SYN3 TMEM241
PCDH9 POLR3F PUM2 RLF SEZ6L SNX27 SYNDIG1 TMEM25
PCDHGC3 POU2F1 PVALB RNF112 SEZ6L2 SOGA1 SYNE1 TMEM30A
PCF11 PPARGC1B PYURF RNF145 SF3B1 SOGA2 SYNGR1 TMEM50B
PCGF3 PPFIA1 QSER1 RNF146 SFPQ SOGA3 SYNJ1 TMEM59L
PCIF1 PPFIA2 QSOX2 RNF165 SGIP1 SORBS2 SYNPR TMEM74
PCLO PPFIA3 R3HDM1 RNF182 SGK196 SORL1 SYP TMEM9B
PCNXL2 PPID R3HDM2 RNFT2 SGTB SORT1 SYT1 TMTC2
PCYOX1L PPIP5K1 RAB11FIP4 RNGTT SH3D21 SOS1 SYT11 TNK2
PDE1A PPME1 RAB15 RPAIN SH3GL2 SOS2 SYT12 TNK2
PDE3B PPP1R11 RAB1A RPAP3 SH3GLB2 SOX2 SYT13 TNKS2
PDE4B PPP1R12A RAB26 RPH3A SHANK1 SPAG9 SYT14 TNPO2
PDIA2 PPP1R16B RAB37 RPL17-C18orf32 SHANK2 SPAST SYT2 TNR

PDP1 PPP1R9B RAB39B RPRD1A SHC3 SPIN3 SYT3 TNRC6A
PDS5A PPP2CA RAB3A RPRD2 SHF SPNS1 SYT4 TNRC6C
PDS5B PPP2R2B RAB3C RPS10-NUDT3 SHISA6 SPOCK2 SYT5 TOP2B
PDXP PPP2R2C RAB3GAP2 RPS6KA5 SHPRH SPOCK3 SYT9 TOP3A
PDZD4 PPP2R2D RAB5A RPS6KL1 SIPA1L1 SPRYD3 TAB2 TP53BP1
PEA15 PPP2R5B RAB6B RSBN1 SLAIN1 SPTAN1 TAF15 TPCN1
PEG10 PPP3CA RABGAP1L RTEL1 SLBP SPTBN2 TAF1C TPD52L2
PEG3 PPP3R1 RALGAPA1 RTEL1-TNFRSF6B SLC12A2 SPTBN4 TAF2

TPGS2

PELI3 PPP6R2 RALGAPB RTF1 SLC12A5 SRCIN1 TAGLN3 TRA2A
PER3 PRDM10 RALGDS RTN1 SLC17A7 SRGAP3 TAOK3 TRAK2
PEX1 PRDM15 RALGPS1 RUFY2 SLC1A6 SRI TARBP1 TRAPP10
PGAP1 PRDM2 RALYL RUFY3 SLC22A15 SRRM4 TBC1D19 TRAPP11
PGBD5 PRDX3 RAP1GAP2 RUNDC3A SLC22A17 SRSF11 TBC1D5

TRAPP12

PGM2L1 PREP RAP1GDS1 RUNDC3B SLC23A2 SRSF7 TBC1D9B

TRAPP18

PHACTR3 PREX2 RAP2A RUNX1T1 SLC24A2 SRSF9 TBCEL TRIM23
PHF14 PRKACB RAPGEF2 RUSC1 SLC24A3 SS18L1 TBCK TRIM3
PHF21A PRKAG2 RAPGEF4 RYBP SLC25A27 SSTR2 TBRG1 TRIM33
PHKB PRKAR2B RAPGEF6 SALL2 SLC25A40 ST18 TCEAL2 TRIM46
PHTF1 PRKCE RASGEF1A SAMD14 SLC26A10 ST6GAL2 TCEB2 TRIM67
PHYHIP PRKCG RASGEF1C SAR1A SLC26A5 ST8SIA1 TCF12 TRIM8
PHYHIPL PRMT7 RASGRF1 SART3 SLC30A9 ST8SIA3 TCF4 TRIM9
PI4KA PRMT8 RASGRP1 SATB2 SLC32A1 ST8SIA5 TCTN1 TRIO
PIGY PROSAPIP1 RB1 SBF2 SLC36A4 STAM2 TDRKH TRIP12
PIGZ PROSER1 RBFOX1 SCAI SLC4A10 STAMBPL1 TEF TRNAU1AP
PIK3C3 PRPF18 RBM15 SCAPER SLC4A7 STARD4 TELO2 TRO
PIK3R2 PRPF31 RBM39 SCG3 SLC4A8 STAU1 TENM1 TRPC1
PIP4K2B PRPF38B RCAN2 SCN1A SLC6A1 STK19 TFDP1 TRPC3
PIP5K1C PRPF39 RCBTB1 SCN2A SLC6A15 STK32C TGIF2-C20orf24

TRPC4AP

PITPNA PRPF6 RCOR3 SCN2B SLC6A7 STK39 THAP10 TRPM2
PITPNC1 PRR3 RDH13 SCN3A SLC7A14 STMN1 THOC7 TRPM3
PITPNM1 PRRC2C REC8 SCN3B SLC9A5 STMN2 THRA TRRAP
PITPNM3 PRRT1 REEP2 SCN4B SLC9A6 STMN3 THUMPD2 TSC1
PKD1 PRRT2 REEP5 SCN8A SLC01A2 STMN4 TIAM1 TSEN2
PLA2G4B PRTFDC1 RELL2 SCRNN3 SLIT1 STRC TIAM2 TSPAN5
PLCB1 PRUNE2 RELN SDC3 SLITRK3 STRIP1 TLE2 TSPYL4
PLCB4 PSD REPS1 SDCCAG8 SLITRK5 STRN4 TLK1 TTC21B
PLCG1 PSD2 REV1 SEC24B SLK STX16 TLL1 TTC33
PLCL2 PSEN1 REV3L SEC31A SLX4 STX1A TM2D3 TTC37
TTC39B ULK2 VAMP2 WDR27 YTHDF1 ZFYVE27 ZNF25 ZNF532
TTLL11 UNC5C VAPB WDR41 YTHDF2 ZFYVE28 ZNF250 ZNF536

TTLL7 UNC79 VAT1L WDR47 YY1 ZFYVE9 ZNF256 ZNF540
TUB UNC80 VOPP1 WDR67 ZBTB41 ZIC1 ZNF280D ZNF566
TUBB2A UPF3B VPS13C WDR7 ZC2HC1A ZIC2 ZNF287 ZNF568
TUBB2B URI1 VPS18 WDSUB1 ZC3H15 ZIC3 ZNF335 ZNF585A
TUBB3 USP16 VPS29 WHSC1L1 ZCCHC12 ZIC4 ZNF343 ZNF589
TUBGCP6 USP20 VPS33B WRAP73 ZDHHC17 ZIM2 ZNF362 ZNF605
TULP4 USP21 VPS37A WSCD1 ZDHHC21 ZMYM2 ZNF365 ZNF702P
TUSC3 USP31 VPS41 WSCD2 ZDHHC4 ZMYM4 ZNF382 ZNF721
TYRO3 USP32 VPS53 WWC3 ZDHHC8 ZMYND11 ZNF385D ZNF738
UBAP2L USP33 VPS54 XKR6 ZEB1 ZMYND8 ZNF418 ZNF740
UBE2K USP34 VSX1 XPO7 ZFAND1 ZNF133 ZNF461 ZNF75A
UBE2O USP45 VWC2 XRCC1 ZFC3H1 ZNF136 ZNF471 ZNF764
UBE2V1 USP46 WASF1 XYLT1 ZFHX2 ZNF138 ZNF483 ZNF767
UBP1 USP48 WASF3 YEATS2 ZFP1 ZNF175 ZNF490 ZNF778
UBQLN1 USP9Y WASL YLPM1 ZFP112 ZNF211 ZNF493 ZNF793
UBTD2 UTP18 WDFY3 YME1L1 ZFP30 ZNF224 ZNF521 ZNF85
UBTF UTP23 WDR17 YPEL4 ZFPM2 ZNF236 ZNF528 ZP2
UCK1 UXS1 WDR19 YTHDC2 ZFYVE16 ZNF238 ZNF529 ZSWIM8

9.2 Attachment 2

AKT1 BTRC HDAC1 PYGO2 SOX2 SOX9 TLE1 UBB
AKT2 CBY1 KMT2D RBBP5 SOX3 SRY TLE2 UBC
APC CHD8 LEF1 RPS27A SOX4 TCF7 TLE3 XIAP
ASH2L CTBP1 MEN1 SOX13 SOX6 TCF7L1 TLE4 XPO1
BCL9 CTNNB1 PYGO1 SOX17 SOX7 TCF7L2 UBA52 YWHAZ
BCL9L CTNNBIP1

9.3 Attachment 3

5S_rRNA AMPD1 BCAS1 CD2BP2 CTNNB1 DNMT1 DRD1 FER
7SK AMT BCKDK CD38 CTTNBP2 DNMT3A DRD2 FEZF2
AANAT ANK2 BCL11A CD44 CUBN DNMT3B DRD3 FGA
AB209061 ANK3 BCL7B CD99 CUL2 DOC2A DRP2 FGD1
ABAT ANKRD11 BCL7C CD99L2 CUL3 DOCK10 DST FGFBP3
ABCA1 ANXA1 BCL9 CDC42BPB CUL4B DOCK11 DTNA FHIT
ABCA7 ANXA5 BDNF CDC45 CX3CR1 DOCK4 DUSP22 FKBP6
ABCC1 ANXA7 BDNF-AS CDH10 CXCR3 DOLK DYDC1 FLJ00278
ABCC6 AP1S2 BGN CDH11 CXorf36 DPP10 DYDC2 FLJ00285
ABHD11 APBA2 BIN1 CDH22 CYBB DPP3 DYRK1A FLJ00322
ABI2 APC BIRC6 CDH8 CYFIP1 DPP6 EEF1A2 FLNA
ABL1 APH1A BOLA2 CDH9 CYP11B1 DPYD EFR3A FLT1
abParts APOBR BRAF CDIPT CYP19A1 DQ571461 EGR2 FMO5
ACP6 APP BRCA2 CDIPT-AS1 DAB1 DQ572979 EHMT1 FMR1
ACR AQP4 BRSK2 CDK2 DAGLA DQ574263 EIF1AY FOLH1
ACTN4 AR BRWD1 CDK4 DAPK1 DQ575242 EIF2S3 FOPNL
ACY1 ARHGAP11B BRWD3 CDKL5 DAPP1 DQ575284 EIF3C FOXG1
ADA ARHGAP24 BX648073 CDKN1B DAZ1 DQ575576 EIF3CL FOXP1
ADAMTS18 ARHGAP32 BZRAP1 CECR2 DAZ2 DQ575741 EIF4E FOXP2
ADARB1 ARHGDIA C11orf30 CELF4 DAZ3 DQ576952 EIF4EBP2 FRK
ADCY5 ARHGEF9 C12orf57 CELF6 DAZ4 DQ577333 EIF4H FRMPD4
ADK ARID1B C15orf43 CEP290 DCTN5 DQ577902 ELN FUS
ADNP ARID4B C15orf49 CEP41 DCTPP1 DQ578199 EML1 FZD9
ADORA2A ARNT2 C16orf45 CERS4 DCUN1D1 DQ578369 EN2 GABBR1
ADORA3 ARSA C16orf54 CHAT DCX DQ578700 ENO2 GABBR2
ADRB2 ARVCF C16orf92 CHD1L DDC DQ578838 EP300 GABRA1
ADSL ARX C16orf93 CHD2 DDX11 DQ579907 EP400 GABRA2
AF072097 ASH1L C16orf98 CHD7 DDX3X DQ582071 EPC2 GABRA3
AFF2 ASMT C22orf29 CHD8 DDX3Y DQ582073 EPHA6 GABRA4
AGAP1 ASMTL C22orf39 CHL1 DDX53 DQ582448 EPHB2 GABRA5
AGAP2 ASPHD1 C3orf58 CHRFAM7A DDX54 DQ582641 EPHB6
GABRA6
AGBL4 ASS1 C4B CHRM3 DEAF1 DQ582939 EPS8 GABRB1
AGMO ASTN1 CA6 CHRNA7 DEPDC5 DQ583847 ERBB2 GABRB2
AGTR2 ASTN2 CACNA1B CHRN3 DGCR10 DQ585847 ERBB2IP
GABRB3
AHI1 ASXL3 CACNA1C CIB2 DGCR11 DQ586415 ERBB3 GABRD
AIFM3 ATG7 CACNA1D CLDN3 DGCR14 DQ586919 ERBB4 GABRE
AK023548 ATOH1 CACNA1F CLDN4 DGCR2 DQ587888 ERCC6 GABRG1
AK023809 ATP10A CACNA1G CLDN5 DGCR5 DQ588687 ERG GABRG2
AK056973 ATP1B1 CACNA1H CLIP2 DGCR6L DQ588973 ESR1 GABRG3
AK075019 ATP2A1 CACNA1I CLN3 DGCR8 DQ589249 ESR2 GABRP

AK091688 ATP2B2 CACNA2D1 CLSTN3 DGKZ DQ589803 ESRRB
GABRQ AK093758 ATRNL1 CACNA2D2 CLTCL1 DHCR7 DQ590322 ETFB GAD1
AK097453 ATXN2L CACNA2D3 CMIP DIAPH3 DQ590616 ETS1 GAD2
AK097472 ATXN7 CACNB2 CNOT3 DIP2A DQ593032 EXOC3 GALNT13
AK097527 AUTS2 CACNB4 CNR1 DISC1 DQ593239 EXOC5 GALNT14
AK124131 AVPR1A CADM1 CNR2 DKFZP434L187 DQ593342 EXOC6
GALNTL4 AK124673 AX747189 CADM2 CNTN3 DLG4 DQ593983 EXOC6B GAN
AK125489 AX747846 CADPS2 CNTN4 DLGAP1 DQ594309 EXT1 GAP43
AK129567 AXL CAMK2G CNTN5 DLGAP2 DQ594669 F13A1 GARNL1
AK302306 BAIAP2 CAMK4 CNTN6 DLGAP3 DQ595055 FABP3 GAS2
AK302545 BANK1 CAMSAP2 CNTNAP2 DLGAP4 DQ595631 FABP5
GATA4 AK303607 BAZ1B CAMTA1 CNTNAP3 DLL1 DQ595648 FAM135B GATM
AK307870 BBS4 CAPN12 CNTNAP4 DLX1 DQ596229 FAM189A1 GBX2
AK309255 BC029255 CAPRIN1 CNTNAP5 DLX2 DQ596303 FAM190A
GCH1 AK310030 BC033281 CAPS2 COL25A1 DLX3 DQ596319 FAM57B GDA
AK310228 BC035867 CARKD COMT DLX6 DQ596685 FAM91A1 GDI1
AK310526 BC039500 CASC4 CORO1A DMD DQ596686 FAN1 GDPD3
AK311660 BC041003 CASK CREBBP DMPK DQ597560 FARP2 GGNBP2
AKT1 BC041466 CBS CRKL DNAH10 DQ597873 FAT1 GIGYF1
AKT2 BC043570 CBX4 CSDE1 DNAJB1 DQ599532 FBRS GJA1
AKT3 BC050343 CC2D1A CSMD1 DNAJC19 DQ599733 FBXL19 GJA5
ALDH1A3 BC050599 CC2D1B CSNK1E DNAJC30 DQ599785 FBXL19-AS1
GJA8 ALDH5A1 BC070376 CCDC101 CSTF2T DNER DQ600136 FBXO15 GLO1
ALDOA BC073928 CCDC64 CTCF DNHD1 DQ600342 FBXO33 GLRA2
ALOX5AP BC127858 CD163L1 CTF1 DNM1 DQ786262 FBXO40 GLRA3
AMELY BCAP31 CD19 CTNNA3 DNM1L DQ786280 FCRL6 GLRB
GLUD2 HDAC6 IL1RAPL2 KMT2C MAPK3 MSR1 NRP2 PDE4A
GNAS HDAC9 IL27 KMT2E MAPK7 MTF1 NRXN1 PDE4B
GNB1L HDHD1 IMMP2L KPTN MAPK8 MTHFR NRXN2 PDIA3P
GOLGA6L1 HDLBP INADL KRT34 MAPK8IP2 MTMR10 NRXN3 PDK1
GOLGA6L2 HEPACAM INO80E LAMA1 MARK1 MTMR12 NSD1
PDXDC1 GOLGA6L7P HERC2 INPP1 LAMB1 MAX MTOR NSF PDZD4
GOLGA8EP HERC2P10 IPW LAMC3 MAZ MTR NTAN1 PDZK1P1
GOLGA8F HERC2P2 IQGAP2 LAT MBD1 MTX2 NTNG1 PEGR
GOLGA8G HERC2P7 IQSEC2 LAT2 MBD2 MVP NTRK1 PER1
GOLGA8I HERC2P9 ITGA4 LEO1 MBD3 MYH10 NTRK3 PEX7
GOLGA8S HGF ITGA5 LEP MBD3L1 MYH11 NUPR1 PGR
GOLGA8T HIRA ITGAL LIMK1 MBD3L2 MYH4 NXF5 PHF15

GPC1 HIRIP3 ITGAM LIN7B MBD3L3 MYLPF NXPH1 PHF2
GPC6 HIST1H1A ITGB3 LINC00035 MBD3L4 MYO16 OCA2 PHF8
GPD2 HIST1H2BB ITGB7 LINC00624 MBD3L5 MYO1A ODF3L2
PHKG2
GPHN HIST1H2BC JARID2 LINC00895 MBD4 MYO9B OFD1 PI4KA
GPLD1 HIST1H2BK JB175342 LINC00896 MBD5 MYOC OGT PI4KAP1
GPR139 HIST1H3F JMJD1C LINC00929 MBD6 MYOZ1 OPHN1 PIK3CB
GPR37 HIST1H4A JPX LMCD1 MC4R MYT1L OPRM1 PIK3CG
GPR6 HIST1H4B JUP LMX1B MCC NAA15 OR1C1 PIK3R2
GPR85 HIST1H4C JX456220 LOC100128714 MCF2 NBEA ORAI2
PIKFYVE
GPR89B HIST1H4D KAL1 LOC100288332 MCPH1 NBPF13P ORAI3
PINX1
GPR89C HIST1H4F KANK1 LOC100288637 MDGA2 NBPF24 OTUD7A
PIR
GPX1 HLA-A KAT2B LOC100289092 MDM2 NBPF8 OTX1 PITX1
GRIA1 HLA-DRB1 KAT8 LOC100289656 MECP2 NCKAP1 OXTR PKD1P1
GRIA2 HMGN1 KATNAL2 LOC100862671 MED13L NCKAP5 P2RX4
PLA1A
GRIA3 HNRNPH2 KCNAB2 LOC23117 MED15 NCKAP5L P2RX6 PLAUR
GRIA4 HOMER1 KCND2 LOC283683 MEF2C NCOR1 P2RX6P PLCB1
GRID1 HOXA1 KCNJ10 LOC283685 MET NCOR2 PAFAH1B1 PLCD1
GRID2 HOXB1 KCNJ12 LOC283710 Metazoa_SRP NDE1 PAGR1 PLD1
GRID2IP HRAS KCNJ15 LOC283767 MFRP NDN PAH PLD2
GRIK2 HS3ST5 KCNJ2 LOC284865 MIB1 NDNL2 PAK3 PLN
GRIK3 HSD11B1 KCNMA1 LOC388849 MIB2 NDUFA5 PAR1 PLXNA3
GRIN1 HSD3B7 KCNQ2 LOC400891 MIR1286 NEFL PAR4 PLXNA4
GRIN2A HSP90AA1 KCNQ3 LOC440243 MIR1306 NELL1 PAR5 PLXNB3
GRIN2B HSPH1 KCNT1 LOC440354 MIR137 NF1 PARD3 PNPLA4
GRIN2C HTR1A KCTD13 LOC595101 MIR185 NFATC2IP PARD3B POGZ
GRIN2D HTR1B KCTD4 LOC606724 MIR211 NFIA PARK2 POM121L4P
GRIN3A HTR1D KDM3A LOC613037 MIR3179-2 NHLRC1 PAR-SN
POMGNT1
GRIN3B HTR1E KDM4B LOC613038 MIR3179-3 NIPA1 PASK PON1
GRIP1 HTR1F KDM5B LOC643699 MIR3180-3 NIPA2 PAX5 POT1
GRM1 HTR2A KDM5C LOC646278 MIR3180-4 NIPBL PCDH10 PPP1A1
GRM2 HTR2B KDM5D LOC729444 MIR3618 NLGN1 PCDH11X PPM1D
GRM3 HTR2C KDM6B LOC730183 MIR4284 NLGN2 PCDH11Y PPP1R1B
GRM4 HTR3A KHDRBS2 LPL MIR4508 NLGN3 PCDH15 PPP1R3F
GRM5 HTR3B KIAA0100 LRFN5 MIR4509-1 NLGN4X PCDH19 PPP2R5D
GRM6 HTR3C KIAA0182 LRP2 MIR4517 NLGN4Y PCDH8 PPP3CB
GRM7 HTR3D KIAA0232 LRP2BP MIR4518 NOMO1 PCDH9 PPP4C
GRPR HTR3E KIAA0430 LRPPRC MIR4519 NOMO3 PCDHA1 PRICKLE1
GSC2 HTR4 KIAA1586 LRRC1 MIR4715 NONO PCDHA10 PRICKLE2

GSDMC HTR5A KIAA1653 LRRC14 MIR4721 NOS1 PCDHA11 PRKAA1
GSK3B HTR6 KIAA2022 LRRC7 MIR4761 NOS1AP PCDHA12 PRKAA2
GSN HUWE1 KIF13B LSAMP MIR484 NOS2A PCDHA13 PRKAB2
GSS HV593096 KIF1A LZTR1 MIR5087 NPAP1 PCDHA2 PRKCA
GSTM1 HV593110 KIF22 LZTS2 Mir_548 NPAS2 PCDHA3 PRKCB
GTF2I HV593127 KIF4A MACROD2 MIR590 NPIP PCDHA4 PRKD1
GTF2IRD1 HV593134 KIF5C MAFK Mir_649 NPIPL1 PCDHA5 PRKY
GUCY1A2 HV593135 KIRREL3 MAGED1 MIR762 NPIPL3 PCDHA6

PRODH

HAUS7 HV593178 KIT MAGEL2 MKL2 NPY1R PCDHA7 PRPF39
HBT8 HV593183 KLC1 MAL MKRN3 NPY5R PCDHA8 PRR14
HCFC1 HYDIN KLC2 MAOA MLST8 NR3C2 PCDHA9 PRRG4
HCN1 ICA1 KLF13 MAOB MLXIPL NRCAM PCDHAC1 PRRT2
hCPE-R IKBKG KLHL13 MAP2 MPP6 NRG1 PCDHAC2 PRSS36
HDAC1 IL16 KLHL22 MAPK1 MPV17L NRG2 PCDHGA11 PRSS38
HDAC2 IL1R2 KMO MAPK11 MRPL40 NRG3 PCOLCE PRSS53
HDAC4 IL1RAPL1 KMT2A MAPK12 MSN NRG4 PDE1C PRSS8
PRUNE2 RIMS1 SHANK3 SND1 SNORD116-5 TBC1D10B TRPC1 VPS13B
PRY RIMS3 SIN3A SNORA30 SNORD116-8 TBC1D23 TRPC3 VPS37D
PSD3 RNF216P1 SIX2 SNORD109B SNRPN TBC1D5 TRPC4 VSIG4
PSMB1 RNF40 SKI SNORD115-1 SNTG2 TBC1D7 TRPC5 WAC
PSMC2 RNPS1 SLC16A7 SNORD115-10 SNURF TBCK TRPC6 WBP5
PSMD10 ROBO1 SLC17A1 SNORD115-11 SNURF-SNRPN TBL1X TRPM1

WBSCR22

PTCHD1 ROBO2 SLC17A2 SNORD115-13 SNX19 TBL1XR1 TSC1

WBSCR27

PTEN RORA SLC17A6 SNORD115-14 SNX29P2 TBL2 TSC2 WBSCR28
PTGER3 RPL10 SLC17A7 SNORD115-15 SNX9 TBP TSN WDFY3
PTGS2 RPP25 SLC17A8 SNORD115-16 SOD1 TBR1 TSPAN6 WDR55
PTPN11 RPS4Y1 SLC18A1 SNORD115-17 SOD2 TBX1 TSPAN7 WDR93
PTPRB RPS6KA2 SLC18A2 SNORD115-2 SOX1 TBX6 TSPYL2 WHAMMP2
PTPRC RPS6KA3 SLC18A3 SNORD115-22 SOX5 TCF20 TSSK2

WHAMMP3

PTPRD RPTOR SLC1A1 SNORD115-24 SOX9 TCF3 TTI2 WHSC1
PTPRM RRN3 SLC1A2 SNORD115-25 SPARCL1 TCF4 TTLL3 WNK3
PTPRT RRN3P2 SLC1A4 SNORD115-26 SPAST TCTE3 TTTY10 WNT1
PWRN1 RTN4R SLC1A6 SNORD115-27 SPATA13 TDO2 TTTY14 WNT2
PWRN2 S100G SLC1A7 SNORD115-28 SPN TENM1 TTTY15 WT1
PXDN SAE1 SLC22A15 SNORD115-3 SPNS1 TGM1 TUBA3FP XIST
PYCARD SAP130 SLC24A2 SNORD115-30 SRCAP TH TUBGCP5 XPC
PYDC1 SAP30 SLC25A1 SNORD115-31 SRD5A2 THAP7 TUFM XPO1
QPRT SATB2 SLC25A12 SNORD115-32 SREBF1 THAP7-AS1 TXLNG

YEATS2

QRICH1 SBF1 SLC25A14 SNORD115-33 SRPK2 THAP8 TXLNG2P YPEL3

RAB11FIP5 SCAMP5 SLC25A24 SNORD115-35 SRY THBS1 TXNRD2

Y_RNA

RAB19 SCARA3 SLC25A27 SNORD115-37 ST7 THRA TYR YTHDC1

RAB2A SCARF2 SLC29A4 SNORD115-38 ST8SIA2 TIGD4 U84523

YTHDC2

RAB39B SCFD2 SLC30A5 SNORD115-39 STAU2 TIMP1 UBA6 YWHAE

RAB3A SCN1A SLC32A1 SNORD115-4 STK39 TJP1 UBE2A ZBTB16

RAB3IL1 SCN2A SLC35A3 SNORD115-40 STS TLK2 UBE2H ZBTB20

RAB5A SCN5A SLC38A1 SNORD115-41 STX1A TLS/FUS-ERG UBE3A

ZDHHC8

RABEP2 SCN7A SLC38A10 SNORD115-44 STX1B TM4SF20 UBE3B ZFX

RAC1 SCN8A SLC39A11 SNORD115-45 STX4 TMEM191A UBE3C ZFY

RAC3 SCRIB SLC4A10 SNORD115-47 STXBP1 TMEM191B UBN2 ZG16

RAI1 SDC2 SLC6A1 SNORD115-48 STXBP5 TMEM219 UBR1 ZMYND11

RANBP1 SDK1 SLC6A11 SNORD115-5 STXBP6 TMEM231 UBR3 ZNF18

RANBP17 SEMA3A SLC6A12 SNORD115-6 SULT1A1 TMLHE UBR5

ZNF292

RAP1GDS1 SEMA5A SLC6A13 SNORD115-7 SULT1A2 TMLHE-AS1

UBR7 ZNF385B

RAPGEF4 SEPHS2 SLC6A2 SNORD115-8 SULT1A4 TMSB4Y UFD1L

ZNF407

RASD1 SEPT1 SLC6A3 SNORD115-9 SUV420H1 TMX2 ULK4P1 ZNF433

RASSF5 SEPT5-GP1BB SLC6A4 SNORD116-1 SVIL TNIP2 ULK4P3 ZNF48

RB1CC1 SERPIND1 SLC6A8 SNORD116-10 SYAP1 TNK2 UPB1 ZNF517

RbBP7 SERPINE1 SLC7A4 SNORD116-11 SYN1 TNN UPF2 ZNF559

RBFOX1 SETBP1 SLC7A5P1 SNORD116-12 SYN2 TNRC6B UPF3B ZNF629

RBM8A SETD1A SLC9A6 SNORD116-13 SYN3 TOMM20 USP41 ZNF646

RBMS3 SETD2 SLC9A9 SNORD116-15 SYNCRI P TOP1 USP9X ZNF668

RBMX SETD5 SLCO1B1 SNORD116-16 SYNE1 TOP3B USP9Y ZNF688

RBMX2 SETDB1 SLCO1B3 SNORD116-18 SYNGAP1 TPO UTP6 ZNF689

RBMY1B SETDB2 SLIT3 SNORD116-19 SYNJ2 TPRX1 UTRN ZNF713

RBMY1F SEZ6L2 SLITRK5 SNORD116-2 SYP TRIM17 UTY ZNF74

REEP3 SF1 SLX1A SNORD116-20 SYT1 TRIM33 VAMP1 ZNF747

RELN SGSH SLX1B-SULT1A4 SNORD116-22 SYT11 TRIM72 VAMP2

ZNF764

RER1 SGSM3 SMARCC2 SNORD116-23 SYT17 TRIO VASH1 ZNF768

RERE SH2B1 SMC2 SNORD116-24 SYT3 TRIP12 VCP ZNF771

RFC2 SH3GL3 SMCX SNORD116-25 SYT4 TRMT2A VCX ZNF774

RFWD2 SH3KBP1 SMCY SNORD116-26 TAAR2 TRNA_Asn VCX2

ZNF778

RGS7 SHANK1 SMG6 SNORD116-28 TAF1C TRNA_Gln VGF ZNF785

RHEB SHANK2 SMURF1 SNORD116-29 TAF4 TRNA_Glu VIL1 ZNF8

RHOXF1 SHANK2-AS1 SNAP25 SNORD116-3 TANGO2 TRNA_His VIP

ZNF804A

RIBC1 SHANK2-AS3 SNAP29 SNORD116-4 TAOK2 TRNA_Pseudo
VKORC1 ZNF827
RIMBP3

9.4 Attachment 4

AAK1 ARID1A CALM3 DDN FAM190B HK1 LLGL1 MYO5A
AATK ARID1B CAMK2A DDX24 FAM21A HMGXB3 LMTK2 MYST3
ABCA2 ARID2 CAMK2B DENND5A FAM5B HNRNPUL1 LMTK3
MYT1L
ABCA3 ARNT2 CAMK2N1 DGCR2 FAM65A HSP90AB1 LPHN1 NACAD
ABCG1 ARPP21 CAMKK2 DGKZ FAM91A1 HTT LPHN3 NAT8L
ABR ARRB1 CAMSAP1 DHX30 FASN HUWE1 LPIN2 NAV1
ACCN1 ARVCF CAMSAP1L1 DICER1 FAT1 IDS LPPR4 NAV2
ACLY ASH1L CAMTA1 DIDO1 FAT2 IGSF9B LRP1 NAV3
ACO2 ATF7IP CAMTA2 DIP2A FAT3 INPP4A LRP3 NBEA
ACTB ATG2A CAND1 DIP2B FAT4 INTS1 LRP8 NCAM1
ADAP1 ATG2B CASKIN1 DIP2C FBXL16 IPO13 LRRC41 NCAN
ADARB1 ATG9A CBX6 DIRAS2 FBXL19 IPO4 LRRC4B NCDN
ADCY1 ATMIN CBX6-NPTXR DISP2 FBXO41 IPO5 LRRC68 NCKAP1
ADCY5 ATN1 CDC42BPA DLC1 FCHO1 IQSEC2 LRRC7 NCOA1
ADD1 ATP13A2 CDC42BPB DLG2 FKBP8 IQSEC3 LRRC8B NCOA2
ADNP ATP1A1 CDK16 DLG4 FOXX2 IRS2 LRRN2 NCOA6
ADRBK1 ATP1A2 CDK17 DLG5 FOXO3 ITPR1 LYNX1 NCOR1
AFF3 ATP1A3 CDK5R1 DLGAP1 FRMPD4 ITSN1 MACF1 NCOR2
AFF4 ATP1B1 CDK5R2 DLGAP2 FRY JAK1 MADD NCS1
AGAP1 ATP2A2 CDKL5 DLGAP3 FSCN1 JPH3 MAGED1 NDRG2
AGAP2 ATP2B2 CELF5 DLGAP4 FYN JPH4 MAGI2 NDRG4
AGAP3 ATP2B4 CELSR2 DMWD GABBR1 KALRN MAN2A2 NDST1
AGPAT3 ATP5A1 CELSR3 DMXL2 GABBR2 KCNA2 MAP1A NEDD4
AGRN ATP5B CHD3 DNAJC6 GARNL3 KCNB1 MAP1B NELF
AGTPBP1 ATP6V0A1 CHD4 DNM1 GAS7 KCNC3 MAP2 NEURL
AHDC1 ATP6V0D1 CHD5 DOCK3 GBF1 KCND2 MAP3K12 NEURL4
AKAP6 ATP6V1B2 CHD6 DOCK4 GCN1L1 KCNH1 MAP4 NF1
AKAP9 ATP9A CHD8 DOCK9 GIT1 KCNH3 MAP4K4 NFIC
AKT3 ATXN1 CHN1 DOPEY1 GLUL KCNH7 MAP7D1 NFIX
ALDOA AUTS2 CHN2 DOPEY2 GM996 KCNMA1 MAPK1 NGEF
ALDOC B3GAT1 CHST2 DOT1L GNAL KCNQ2 MAPK4 NHSL1
ALS2 BAI1 CIC DPP8 GNAO1 KCNQ3 MAPK8IP1 NISCH
AMPH BAI2 CIT DPYSL2 GNAS KCNT1 MAPK8IP3 NLGN2
ANAPC1 BAP1 CKAP5 DSCAM GNAZ KDM4B MAPKBP1 NLGN3
ANK1 BAT2 CKB DSCAML1 GNB1 KDM5C MAST1 NOMO1

ANK2 BAT2L1 CLASP1 DST GPAM KDM6B MAST2 NPAS2
ANK3 BAT2L2 CLASP2 DTNA GPM6A KIAA0090 MAST4 NPTXR
ANKRD11 BAT3 CLCN3 DTX1 GPR158 KIAA0100 MAZ NR2F1
ANKRD17 BAZ2A CLEC16A DUSP8 GPR162 KIAA0226 MBD5 NRGN
ANKRD52 BCAN CLIP3 DYNC1H1 GPRIN1 KIAA0284 MBP NRIP1
AP1B1 BCL9L CLSTN1 EEF1A2 GRAMD1B KIAA0317 MED13 NRXN1
AP2A1 BCR CLTC EEF2 GRIK3 KIAA0430 MED13L NRXN2
AP2A2 BIRC6 CNP EGR1 GRIK5 KIAA0664 MED14 NRXN3
AP2B1 BMPR2 COBL EHMT1 GRIN1 KIAA0802 MED16 NSD1
AP3D1 BPTF COPG EHMT2 GRIN2A KIAA0913 MEF2D NSF
APBA1 BRD4 CPE EIF2C1 GRIN2B KIAA0947 MFHAS1 NTRK2
APBB1 BRSK1 CPLX1 EIF2C2 GRLF1 KIAA1045 MGAT5B NTRK3
APC BRSK2 CPLX2 EIF4G1 GRM4 KIAA1109 MIB1 NUP98
APC2 BSN CPT1C EIF4G2 GRM5 KIAA1244 MICAL2 NWD1
APLP1 BZRAP1 CREBBP EIF4G3 GSK3B KIAA1688 MINK1 ODZ2
APOE C11ORF41 CRMP1 ELFN2 GTF3C1 KIAA2018 MKL2 ODZ3
APP C14ORF4 CRTIC1 ELMO2 GTF3C2 KIF1A MLL ODZ4
ARAP2 C19ORF26 CTBP1 EML2 HCFC1 KIF1B MLL2 OGDH
ARF3 C20ORF117 CTNNB1 ENC1 HCN2 KIF21A MLL3 OLFM1
ARFGEF1 C20ORF12 CTNND2 EP300 HDAC4 KIF21B MLL5 OXR1
ARHGAP20 C2CD2L CUL9 EP400 HDAC5 KIF3C MMP24 PACS1
ARHGAP21 CABIN1 CUX1 EPB41L1 HDLBP KIF5A MON2 PACS2
ARHGAP23 CACNA1A CUX2 EPB49 HEATR5B KIF5C MPPIP PAK6
ARHGAP32 CACNA1B CYFIP2 EPHA4 HERC1 KIFC2 MTMR4 PCDH1
ARHGAP33 CACNA1E DAB2IP EPN1 HERC2 KLC1 MTOR PCDH10
ARHGEF11 CACNA1G DAGLA EXTL3 HIPK1 KLHL22 MTSS1L PCDH7
ARHGEF12 CACNA1I DAPK1 FAM115A HIPK2 KNDC1 MYCBP2
PCDH9
ARHGEF17 CACNB1 DBC1 FAM120A HIPK3 LARGE MYH10 PCDHA4
ARHGEF2 CACNB3 DCAF6 FAM160A2 HIVEP1 LARS2 MYO10
PCDHAC2
ARHGEF4 CADPS DCLK1 FAM171B HIVEP2 LHFPL4 MYO16
PCDHGA12
ARHGEF7 CALM1 DCTN1 FAM179B HIVEP3 LINGO1 MYO18A
PCDHGC3
PCLO POLR2A RALGAPB SCD2 SMARCC2 SYNGR1 TRIM37 USP22
PCNX PPARGC1A RALGDS SCN2A SMG1 SYNJ1 TRIM9 USP32
PCNXL2 PPFLA3 RAP1GAP SCN8A SMPD3 SYNPO TRIO USP34
PCNXL3 PPM1E RAP1GAP2 SEC14L1 SNAP25 SYT1 TRIP12 USP5
PDE2A PPP1R9B RAPGEF1 SEC16A SNAP91 SYT7 TRO USP9X
PDE4B PPP2R1A RAPGEF2 SEC23A SNPH TANC2 TRPC4AP VAMP2
PDE4DIP PPP2R2C RAPGEF4 SECISBP2L SOBP TAOK1 TRPM3 VPS13D
PDE8B PPP2R5B RAPGEFL1 SEPT3 SON TAOK2 TRRAP VPS41
PDS5B PPP3CA RASGRF1 SEPT5 SORBS2 TBC1D9 TSC2 WASF1

PDZD2 PREX1 RASGRP1 SETD5 SORL1 TBC1D9B TSC22D1 WDFY3
PDZD8 PREX2 RC3H1 SETX SORT1 TCF20 TSHZ1 WDR13
PEG3 PRICKLE2 RC3H2 SEZ6L2 SPAG9 TCF25 TSPAN7 WDR6
PER1 PRKACB RELN SGIP1 SPARCL1 TCF4 TSPYL4 WNK1
PFKM PRKCB RERE SGSM2 SPEG TEF TTBK1 WNK2
PGM2L1 PRKCE REV3L SH3BP4 SPEN THRA TTBK2 WWC1
PHACTR1 PRKCG RGS7BP SHANK1 SPHKAP TIAM1 TTC3 XPO6
PHF12 PROSAPIP1 RHOB SHANK2 SPIRE1 TLE3 TTC7B XPO7
PHF20 PRPF8 RHOBTB2 SHANK3 SPRED1 TLN2 TTLL7 YWHAG
PHLDDB1 PRR12 RICH2 SIPA1L1 SPRN TMEM132A TTYH1 ZC3H4
PHYHIP PSAP RIMBP2 SIPA1L2 SPTAN1 TMEM151A TTYH3 ZC3H7B
PI4KA PSD RPH3A SIPA1L3 SPTB TMEM151B TUBB3 ZCCHC14
PIGQ PTCH1 RPRD2 SKI SPTBN1 TMEM63B TUBB4 ZEB2
PIKFYVE PTEN RPTOR SLC12A5 SPTBN2 TMEM8B TULP4 ZER1
PINK1 PTK2 RTN1 SLC12A6 SRCIN1 TMOD2 UBA1 ZFHGX2
PIP5K1C PTK2B RTN3 SLC17A7 SREBF2 TNK UBAP2L ZFP106
PITPNM1 PTPN11 RTN4 SLC1A2 SRGAP3 TNK2 UBE2O ZFR
PITPNM2 PTPN5 RTN4R SLC22A17 SRRM2 TNKS UBE3B ZFYVE1
PJA2 PTPRD RUSC1 SLC24A2 STK25 TNPO2 UBE3C ZHX3
PKD1 PTPRF RUSC2 SLC25A23 STOX2 TNRC18 UBQLN1 ZMIZ1
PKP4 PTPRG RYR2 SLC4A3 STRN4 TNRC6B UBQLN2 ZMIZ2
PLCB1 PTPRJ SALL2 SLC4A4 STXBP1 TNS3 UBR3 ZNF238
PLCH2 PTPRN2 SAMD4B SLC4A8 STXBP5 TPPU UBR5 ZNF365
PLD3 PTPRS SAP130 SLC6A1 SUPT6H TRAK1 UHRF1BP1L ZNF462
PLEC PTPRT SAPS1 SLC6A17 SV2A TRAK2 ULK1 ZNF521
PLP1 PUM1 SAPS2 SLC8A1 SV2B TRAPPC10 ULK2 ZNF536
PLXNA1 PUM2 SASH1 SLC8A2 SYMPK TRIL UNC13A ZNF704
PLXNA2 QKI SBF1 SLITRK5 SYN1 TRIM2 UNC13C ZNF827
PLXNA4 R3HDM1 SCAF1 SMARCA2 SYNE1 TRIM3 UNC5A ZNFX1
PLXNB1 R3HDM2 SCAP SMARCA4 SYNGAP1 TRIM32 UQCRC1
ZYG11B
PLXND1 RALGAPA1

9.5 Attachment 5

ABI2 CCDC88C DNAJC6 ILF2 NAA15 PLXNA3 SLC6A1 TCF4
ADCY5 CDC42BPB DOCK8 IQGAP3 NAT8L PLXNB1 SLC6A10P TCF7L2
ADCY9 CDC45 DSCAM IQSEC2 NAV2 POGZ SLC6A8 TNRC18
ADGRL2 CDH10 DVL3 ITPR1 NBEA PPM1D SLC9A3 TNRC6B
ADNP CDKL5 DYRK1A KAT2B NCKAP1 PPP2R5D SLITRK5 TRIO
AGAP2 CHD1 EPHB2 KATNAL2 NCOR1 PROX2 SMAD2 TRIP12
AGO4 CHD2 ERBIN KCNH1 NEMF PSEN2 SMAD4 TRRAP

AHNAK CHD4 FAM8A1 KCNH3 NFIA PTEN SMARCC2 TSC2
AK8 CHD7 FOXP1 KIRREL3 NISCH PTK7 SMC3 TUBA1A
ALDH5A1 CHD8 FOXP3 KMT2A NLGN3 PTPN11 SPAST UBN2
ANK2 CHRNA7 GABRB3 KMT2C NLGN4X RAB2A SPP2 UIMC1
ANK3 COL4A3BP GALNT18 KMT2E NR2E3 RALGAPB SRCAP UNC80
ANKRD11 CREBBP GIGYF2 KMT5B NRXN1 RANBP2 SRGAP2 VCP
ANP32A CSMD1 GLI2 LAMC3 NSD2 RELN SRGAP3 WAC
APH1A CSTF2T GPRIN1 LHX1 NTM RIMBP2 SRM WDFY3
ARHGAP30 CTNNB1 GPS1 LZTR1 NTNG1 RIMS1 ST3GAL6 WDR45
ARHGAP32 CTNND2 GRB10 MAGEL2 NUAK1 RNF38 STAG1 WNT7B
ARHGAP5 CTTNBP2 GRIA2 MAP3K1 PARD3B RPL19 STXBP1 WNT9A
ARID1B CUL3 GRIN1 MECP2 PASK SCN1A STXBP5 YES
ARIH1 CUL7 GRIN2A MED13L PAX5 SCN2A SUPT16H YTHDC1
ARVCF DCX GRIN2B MFRP PCOLCE SETBP1 SVIL ZBTB45
ARX DDX3X GTPBP4 MIP PHF2 SETD2 SYNCRIP ZC3H4
ASH1L DHX9 HDAC9 MKI67 PHF3 SETD5 SYNGAP1 ZFHX3
ASXL3 DIP2A HECTD1 MOV10 PHF7 SHANK1 TANC2 ZMYM2
CACNA2D3 DLG2 HIVEP3 MYH9 PHIP SHANK2 TBL1XR1 ZMYND11
CAPRIN1 DLG4 HUWE1 MYT1L PIK3CA SIN3A TBR1 ZNF292
CASK DLGAP1

9.6 Attachment 6

ABAT ALDH5A1 GABRA5 GABRR2 GNB1 GNG7 KCNJ2 SLC6A1
ADCY1 ARHGEF9 GABRA6 GABRR3 GNB2 GNG8 KCNJ3 SLC6A11
ADCY2 CPLX1 GABRB1 GAD1 GNB3 GNGT1 KCNJ4 SLC6A12
ADCY3 DNAJC5 GABRB2 GAD2 GNG10 GNGT2 KCNJ5 SLC6A13
ADCY4 GABBR1 GABRB3 GNAI1 GNG12 HSPA8 KCNJ6 SNAP25
ADCY5 GABBR2 GABRG2 GNAI2 GNG2 KCNJ10 KCNJ9 STX1A
ADCY6 GABRA1 GABRG3 GNAI3 GNG3 KCNJ12 RAB3A STXBP1
ADCY7 GABRA2 GABRQ GNAL GNG4 KCNJ15 RIMS1 SYT1
ADCY8 GABRA3 GABRR1 GNAT3 GNG5 KCNJ16 SLC32A1 VAMP2
ADCY9 GABRA4

9.7 Attachment 7

AASS CRYGD GADD45G KBTBD11 MKRN1 PHLDA1 RPS11 TAF7
ACAT2 CTU1 GCA KCNK12 MRPS10 PNISR RPS26P15 TCP1
AGBL5 CYYR1 GDF10 KDR MTM1 PODXL SALL3 TDP2
ALDH7A1 DCUN1D2 GNB3 KIT MYEF2 POLR3G SCPEP1 TENM3
ALOX5 DCUN1D4 GPR27 KLF4 MYL12A POP4 SEMA6A TFG
ARPC5 DDIS4L GUF1 KRR1 MYL12B PRKDC SF3B6 TLE3
ATOH1 DHX9 HERC2 L1TD1 NACC1 PSMA7 SGCB TMEM126A
ATP5G2 DLX5 HIST1H2AE LAMA1 NAP1L1 PYGL SLC6A1 TMEM141
BAGE2 DLX6 HIST1H2BG LAMP1 NDUFA8 RAB6B SLC6A6 TMEM261
BEND4 DNMT3B HIST1H3G LGALS8 NFE2L3 RAD21 SMCO4 TMEM63B
BEND5 DUSP23 HIST1H4C LHPP P3H3 RAPGEF5 SNORA18 TRIM41
C12orf57 E2F5 HIST1H4F LHX1 PAPSS1 RBM28 SNRPB2 TWIST1
C14orf1 EIF4A3 HIST1H4I LIN7C PCDH8 RBP4 SP5 U2AF1
CACHD1 EIF4EBP3 HMGB3 LINC00094 PCDHB5 RHOB SPATA18
UBE2H
CADPS2 EMBP1 HMGN3 LMO1 PCNP RIDA SPSB2 UCHL1
CAND2 ENO1 HNRNPA2B1 LONRF1 PCTP RMRP SRPRB USP1
CCDC174 FAM189A1 HNRNPF LRIG1 PDCD10 RNF126P1 SRSF9 VRTN
CCKBR FEZF1 HSPD1 LRRCC1 PDCD11 RPL13A SSR1 XKR7
CDCA7 FGF19 HTR1B LSM4 PDXP RPL13P5 STC2 ZFAND4
COX5A FGF4 ID4 LTA4H PHAX RPL22 SUPT6H ZFP42
CPZ FOS IFI6 MAPRE1 PHGDH RPRM SYT17 ZNF488
CRABP1 FZD5 IGDCC4

9.8 Attachment 8

AKT1 EEF2K LAMTOR1 MLST8 PRKAB1 RHEB RRAGB STRADB
AKT1S1 EIF4B LAMTOR2 MTOR PRKAB2 RPS6 RRAGC TSC1
AKT2 EIF4E LAMTOR3 PPM1A PRKAG1 RPS6KB1 RRAGD TSC2
CAB39 EIF4EBP1 LAMTOR4 PRKAA1 PRKAG2 RPTOR STK11 YWHAB
CAB39L EIF4G1 LAMTOR5 PRKAA2 PRKAG3 RRAGA STRADA

9.9 Attachment 9

ACTB DUSP5 FGFR3 IRS4 NRTN PSMA2 PSMD8 RRAGB
ACTG1 DUSP6 FGFR4 ITGA2B PAQR3 PSMA3 PSMD9 RRAGC
ACTN2 DUSP7 FGG ITGB3 PDE3B PSMA4 PSME1 RRAGD
AKAP9 DUSP8 FN1 JAK1 PDGFA PSMA5 PSME2 SHC1
AKT1 DUSP9 FRS2 JAK2 PDGFB PSMA6 PSME3 SHC2
AKT1S1 EEF2K FRS3 JAK3 PDGFRA PSMA7 PSME4 SHC3
AKT2 EGF FYN KBTBD7 PDGFRB PSMA8 PSMF1 SOS1
ANGPT1 EGFR GAB1 KIT PDPK1 PSMB1 PSPN SPRED1
APBB1IP EIF4B GDNF KITLG PEA15 PSMB10 PTK2 SPRED2
ARAF EIF4E GFRA1 KL PEBP1 PSMB11 PTPN11 SPRED3
ARRB1 EIF4EBP1 GFRA2 KLB PHB PSMB2 PTPRA SPTA1
ARRB2 EIF4G1 GFRA3 KRAS PIK3C3 PSMB3 RAF1 SPTAN1
ARTN ERBB2 GFRA4 KSR1 PIK3CA PSMB4 RAP1A SPTB
BRAF ERBB3 GRB2 KSR2 PIK3CB PSMB5 RAP1B SPTBN1
BRAP ERBB4 GRIN1 LAMTOR1 PIK3R1 PSMB6 RAPGEF2 SPTBN2
BTC EREG GRIN2A LAMTOR2 PIK3R2 PSMB7 RASA1 SPTBN4
CAB39 FGA GRIN2B LAMTOR3 PIK3R4 PSMB8 RASA2 SPTBN5
CAB39L FGB GRIN2C LAMTOR4 PPM1A PSMB9 RASA3 SRC
CALM1 FGF1 GRIN2D LAMTOR5 PPP2CA PSMC1 RASA4 STK11
CAMK2A FGF10 HBEGF LAT PPP2CB PSMC2 RASAL1 STRADA
CAMK2B FGF16 HRAS MAP2K1 PPP2R1A PSMC3 RASAL2 STRADB
CAMK2D FGF17 IGF1 MAP2K2 PPP2R1B PSMC4 RASAL3 SYNGAP1
CAMK2G FGF18 IGF1R MAP3K11 PPP2R5A PSMC5 RASGEF1A TEK
CNKSR1 FGF19 IGF2 MAPK1 PPP2R5B PSMC6 RASGRF1 THEM4
CNKSR2 FGF2 IL17RD MAPK3 PPP2R5C PSMD1 RASGRF2 TLN1
CSF2 FGF20 IL2 MARK3 PPP2R5D PSMD10 RASGRP1 TLR9
CSF2RA FGF22 IL2RA MLST8 PPP2R5E PSMD11 RASGRP3 TRIB3
CSF2RB FGF23 IL2RB MTOR PPP5C PSMD12 RASGRP4 TSC1
CSK FGF3 IL2RG NCAM1 PRKAA1 PSMD13 RBX1 TSC2
CUL3 FGF4 IL3 NEFL PRKAA2 PSMD14 RET UBA52
DAB2IP FGF5 IL3RA NF1 PRKAB1 PSMD2 RHEB UBB
DLG4 FGF6 IL5 NRAS PRKAB2 PSMD3 RPS27A UBC
DUSP1 FGF7 IL5RA NRG1 PRKAG1 PSMD4 RPS6 VCL
DUSP10 FGF8 IQGAP1 NRG2 PRKAG2 PSMD5 RPS6KB1 VWF
DUSP16 FGF9 IRS1 NRG3 PRKAG3 PSMD6 RPTOR WDR83
DUSP2 FGFR1 IRS2 NRG4 PSMA1 PSMD7 RRAGA YWHAB
DUSP4 FGFR2

9.10 Attachment 10

A1BG ACADS ADAM9 AHCY AMT AP2M1 ARHGAP24 ARRB1
A1CF ACADSB ADAMTS10 AHCYL1 ANAPC13 AP3B2 ARHGAP26
ARRB2
A4GALT ACAP2 ADAMTS14 AHCYL2 ANAPC2 AP3D1 ARHGAP28
ARRDC3
AADAC ACAP3 ADAMTS17 AHNAK ANAPC4 AP3M1 ARHGAP29
ARRDC5
AADACL2 ACAT1 ADAMTS18 AHR ANAPC5 AP3M2 ARHGAP32 ARSJ
AADACL4 ACAT2 ADAMTS19 AHSA1 ANGEL2 AP3S2 ARHGAP33
ARSK
AADAT ACBD5 ADAMTS2 AHSA2 ANGPT1 AP4B1 ARHGAP35 ARVCF
AAK1 ACBD6 ADAMTS20 AIDA ANGPT2 AP4E1 ARHGAP36 ASAP1
AANAT ACCSL ADAMTS3 AIFM1 ANGPTL1 AP4M1 ARHGAP39
ASAP2
AAR2 ACD ADAMTS6 AIFM2 ANGPTL2 AP4S1 ARHGAP44 ASAP3
AARS ACE ADAMTS9 AIFM3 ANGPTL3 AP5Z1 ARHGAP5 ASB1
AARSD1 ACHE ADAMTSL1 AIG1 ANK1 APAF1 ARHGAP6 ASB13
AASDHPP1 ACIN1 ADAP2 AIMP2 ANK2 APBA1 ARHGAP9 ASB14
AASS ACLY ADAR AIP ANK3 APBA2 ARHGDIA ASB2
AATF ACMSD ADARB1 AIRE ANKDD1A APBB1 ARHGDIB ASB4
ABAT ACOT7 ADARB2 AK1 ANKFY1 APBB1IP ARHGDIG ASB5
ABC1 ACOX2 ADCK3 AK4 ANKH APBB2 ARHGEF1 ASB6
ABC12 ACOX3 ADCY1 AK5 ANKHD1 APC ARHGEF10L ASB7
ABC2 ACPL2 ADCY3 AK7 ANKHD1-EIF4EBP3 APCDD1 ARHGEF11
ASB8
ABC3 ACPP ADCY4 AKAP17A ANKIB1 APCS ARHGEF12 ASB9
ABC7 ACPT ADCY5 AKAP8L ANKLE2 APEH ARHGEF17 ASF1B
ABC10 ACR ADCY6 AKR1B1 ANKMY1 APEX2 ARHGEF3 ASGR1
ABC4 ACRBP ADCY7 AKR1C1 ANKRA2 API5 ARHGEF4 ASGR2
ABC6 ACRC ADCY8 AKR1C2 ANKRD11 APITD1-CORT ARHGEF40
ASH1L
ABC9 ACSBG1 ADCY9 AKR1D1 ANKRD12 APLNR ARHGEF9 ASH2L
ABCC1 ACSF2 ADCYAP1 AKT1S1 ANKRD13A APLP1 ARID1A ASIC1
ABCC10 ACSF3 ADCYAP1R1 AKT2 ANKRD13B APLP2 ARID1B ASIC2
ABCC5 ACSL1 ADD2 AKT3 ANKRD13C APOA1 ARID2 ASIC5
ABCC8 ACSL4 ADH1A ALAD ANKRD17 APOA1BP ARID3A ASNA1
ABCC9 ACSL6 ADH6 ALAS1 ANKRD27 APOBEC3C ARID3B ASPA
ABCD1 ACSM4 ADHFE1 ALB ANKRD28 APOBEC3D ARID4A ASPH
ABCD2 ACSS2 ADI1 ALDH1A1 ANKRD29 APOC1 ARID4B ASPHD1
ABCD3 ACSS3 ADIPOR1 ALDH1A2 ANKRD32 APOL6 ARID5A ASPHD2
ABCE1 ACTA1 ADK ALDH1A3 ANKRD34A APOOL ARIH1 ASPRV1
ABCF1 ACTA2 ADNP ALDH1L1 ANKRD39 APPBP2 ARIH2 ASPSCR1

ABCF2 ACTB ADNP2 ALDH2 ANKRD40 APRT ARL10 ASRGL1
ABCF3 ACTC1 ADORA1 ALDH3A2 ANKRD44 AQP2 ARL14EP ASS1
ABCG1 ACTG1 ADORA2A ALDH6A1 ANKRD49 AQP3 ARL15 ASTN2
ABCG2 ACTG2 ADORA3 ALDH7A1 ANKRD5 AQP4 ARL2 ASUN
ABCG4 ACTL6B ADPRH ALDH8A1 ANKRD50 AQP5 ARL3 ASXL3
ABHD1 ACTL7A ADPRHL2 ALDOA ANKRD52 AQP9 ARL4A ASZ1
ABHD11 ACTL7B ADRA1B ALDOC ANKS1A AR ARL4D ATAD1
ABHD12 ACTL8 ADRA2B ALG10 ANKS1B ARAF ARL5B ATAD5
ABHD13 ACTN1 ADRBK2 ALG11 ANKS6 ARAP1 ARL6 ATAT1
ABHD15 ACTN2 ADSL ALG12 ANO1 ARAP2 ARL6IP1 ATCAY
ABHD16A ACTN4 AES ALG3 ANO2 ARAP3 ARL6IP5 ATE1
ABHD3 ACTR10 AFAP1 ALK ANO3 ARC ARL8A ATF2
ABHD5 ACTR1A AFAP1L1 ALKBH4 ANO4 ARCN1 ARMC1 ATF7
ABHD6 ACTR1B AFF1 ALKBH5 ANO6 ARF1 ARMC12 ATG13
ABHD8 ACTR2 AFF3 ALOX12B ANO8 ARF6 ARMC2 ATG14
ABI1 ACTR3B AFF4 ALOX15 ANO9 ARFGAP1 ARMC6 ATG16L2
ABI2 ACTR6 AFG3L2 ALOX5AP ANP32A ARFGEF1 ARMC8 ATG4D
ABL1 ACTR8 AGAP1 ALPI ANP32B ARFGEF2 ARMCX1 ATG7
ABLM1 ACVR1 AGAP2 ALS2 ANTXR1 ARFIP2 ARMCX2 ATG9A
ABLM2 ACVR1B AGFG1 ALS2CR12 ANXA1 ARFRP1 ARMCX5 ATHL1
ABP1 ACVR2A AGGF1 ALX3 ANXA7 ARG1 ARNT ATL1
ABRA ACVR2B AGPAT3 ALX4 AOC2 ARGLU1 ARNT2 ATL2
ABT1 ACY3 AGPAT4 AMBP AOX1 ARHGAP1 ARNTL ATL3
ACAA2 ADA AGPAT5 AMD1 AP1B1 ARHGAP11A ARPC1A ATMIN
ACACA ADAD1 AGPAT6 AMDHD2 AP1G1 ARHGAP12 ARPC1B

ATOH1
ACACB ADAM10 AGPAT9 AMIGO2 AP1G2 ARHGAP15 ARPC2

ATP11A
ACAD10 ADAM15 AGPS AMMECR1 AP1M1 ARHGAP17 ARPC4-TLLL3

ATP11B
ACAD8 ADAM20 AGR2 AMN1 AP1M2 ARHGAP18 ARPC5 ATP11C
ACAD9 ADAM21 AGR3 AMOTL1 AP1S1 ARHGAP20 ARPC5L ATP12A
ACADL ADAM22 AGRN AMPD2 AP2A1 ARHGAP21 ARPP21 ATP13A1
ACADM ADAM30 AGTRAP AMPD3 AP2A2 ARHGAP22 ARR3 ATP13A4
ATP1A2 B3GNT1 BEX1 BTF3 C1orf172 C8orf34 CAMK1D CCDC115
ATP1A3 B3GNT2 BFSP2 BTG2 C1orf189 C8orf40 CAMK1G CCDC121
ATP1B1 B4GALT1 BGN BTG3 C1orf226 C8orf46 CAMK2A CCDC124
ATP1B2 B4GALT2 BHLHE40 BTG4 C1orf51 C8orf76 CAMK2B CCDC132
ATP1B3 B4GALT3 BICD1 BTK C1orf52 C9orf117 CAMK2G CCDC134
ATP2A1 B4GALT7 BICD2 BTN3A3 C1orf61 C9orf152 CAMK4 CCDC138
ATP2A2 B9D1 BID BTNL9 C1orf65 C9orf156 CAMKK1 CCDC14
ATP2A3 B9D2 BIN1 BTRC C1orf74 C9orf171 CAMKK2 CCDC147
ATP2B2 BABAM1 BIN3 BUB1 C1QA C9orf174 CAMKMT CCDC149
ATP2B3 BACE1 BIRC2 BUB1B C1QC C9orf3 CAMKV CCDC15

ATP2B4 BACE2 BIRC3 BUB3 C1QTNF1 C9orf41 CAMLG CCDC155
ATP2C1 BACH1 BIRC6 BZW1 C1QTNF2 C9orf72 CAMSAP1 CCDC167
ATP2C2 BACH2 BIVM BZW2 C1QTNF3 C9orf78 CAMSAP2 CCDC171
ATP5A1 BAD BLK C10orf107 C1QTNF7 C9orf89 CAMTA1 CCDC172
ATP5C1 BAG2 BLM C10orf116 C1RL C9orf9 CANX CCDC3
ATP5D BAG5 BLMH C10orf118 C2 CA10 CAP1 CCDC36
ATP5F1 BAG6 BLNK C10orf12 C20orf111 CA11 CAPN1 CCDC41
ATP5G3 BAI2 BLOC1S1 C10orf137 C20orf194 CA12 CAPN10 CCDC43
ATP5H BAI3 BMF C10orf2 C20orf195 CA14 CAPN12 CCDC47
ATP5L BAIAP2 BMI1 C10orf32 C20orf20 CA2 CAPN13 CCDC56
ATP6V0A1 BAIAP2L1 BMP1 C10orf57 C20orf24 CA5B CAPN3 CCDC58
ATP6V0A2 BAIAP2L2 BMP2 C10orf76 C20orf26 CA7 CAPN5 CCDC6
ATP6V0A4 BAIAP3 BMP6 C10orf88 C20orf27 CAB39 CAPN7 CCDC64
ATP6V0B BAK1 BMP7 C11orf30 C20orf43 CAB39L CAPNS1 CCDC75
ATP6V0C BAMBI BMPR1A C11orf41 C20orf96 CABIN1 CAPRIN1
CCDC80
ATP6V0D1 BANP BMPR1B C11orf49 C21orf59 CABLES2 CAPRIN2
CCDC84
ATP6V1A BARHL1 BMX C11orf53 C21orf62 CABP7 CAPSL CCDC88A
ATP6V1B1 BARHL2 BNC1 C11orf57 C22orf13 CACFD1 CAPZB CCDC89
ATP6V1B2 BARX2 BNIP1 C11orf63 C22orf28 CACHD1 CARD11 CCDC93
ATP6V1C1 BAX BNIP2 C11orf70 C22orf32 CACNA1A CARD9 CCDC94
ATP6V1H BAZ1A BNIP3L C11orf85 C2CD2L CACNA1C CARHSP1
CCDC96
ATP7B BAZ1B BOC C12orf29 C2orf18 CACNA1D CARS CCDC97
ATP8A1 BAZ2A BOD1L1 C12orf4 C2orf29 CACNA1E CASC1 CCKAR
ATP8B2 BAZ2B BPGM C12orf44 C2orf47 CACNA1F CASC3 CCKBR
ATP8B3 BBOX1 BPIFA1 C12orf56 C2orf56 CACNA1G CASD1 CCL2
ATP9A BBS1 BPIFB2 C12orf57 C2orf62 CACNA1H CASK CCL21
ATPAF1 BBS10 BPTF C12orf76 C2orf65 CACNA1I CASKIN2 CCL5
ATPAF2 BBS2 BRAF C13orf33 C2orf78 CACNA1S CASP10 CCL7
ATR BBS7 BRAP C14orf101 C3 CACNA2D1 CASP14 CCM2
ATRIP BCAS3 BRD1 C14orf126 C3AR1 CACNA2D2 CASP2 CCNA1
ATRNL1 BCCIP BRD3 C14orf166 C3orf22 CACNA2D3 CASP3 CCNB1
ATXN1 BCKDHB BRD4 C14orf166B C3orf23 CACNB1 CASQ2 CCNB1IP1
ATXN2L BCKDK BRD9 C14orf43 C3orf37 CACNB2 CASR CCNB3
ATXN7L2 BCL2 BRF1 C14orf79 C3orf38 CACNB3 CASS4 CCNC
ATXN7L3 BCL2L1 BRF2 C14orf93 C3orf45 CACNG1 CASZ1 CCND1
AUH BCL2L11 BRI3BP C15orf26 C3orf58 CACNG2 CAT CCND2
AURKAIP1 BCL2L2-PABPN1 BRIP1 C15orf38 C5 CACNG3 CATSPER1
CCNDBP1
AURKB BCL6 BRMS1 C15orf38-AP3S2 C5AR1 CACNG4 CATSPERB
CCNE1
AURKC BCL6B BROX C15orf44 C5orf25 CACNG5 CATSPERG CCNF

AUTS2 BCL7B BRPF1 C15orf52 C5orf28 CACNG7 CAV1 CCNJ
AVIL BCL7C BRS3 C16orf5 C5orf30 CACTIN CBFA2T2 CCNK
AVL9 BCL9 BRSK1 C16orf62 C5orf34 CACUL1 CBLB CCNL1
AVPR1A BCL9L BRSK2 C16orf7 C5orf42 CAD CBLL1 CCNT2
AWAT1 BCOR BRWD1 C16orf86 C6orf106 CADM1 CBLN2 CCNY
AXIN1 BCORL1 BRWD3 C16orf93 C6orf130 CADM2 CBLN3 CCNYL1
AZGP1 BCR BSCL2 C17orf49 C6orf203 CADM3 CBS CCP110
AZI2 BCS1L BSDC1 C17orf62 C6orf222 CADPS CBX2 CCR2
AZIN1 BECN1 BSN C18orf1 C6orf62 CADPS2 CBX7 CCR4
B3GALNT1 BEGAIN BSND C18orf54 C7 CALB1 CBY1 CCR6
B3GALNT2 BEND5 BST2 C18orf8 C7orf25 CALB2 CCBL1 CCR7
B3GALT1 BEND6 BTAF1 C19orf10 C7orf26 CALCBB CCDC101 CCS
B3GALT2 BEND7 BTBD1 C19orf54 C7orf43 CALHM1 CCDC103 CCT2
B3GALT4 BEST1 BTBD10 C19orf66 C7orf53 CALHM3 CCDC106 CCT3
B3GALT5 BEST4 BTBD11 C19orf70 C7orf59 CALM3 CCDC108 CCT5
B3GAT1 BET1L BTBD2 C1orf112 C7orf60 CALR CCDC112 CCT6A
B3GAT3 BET3L BTBD6 C1orf159 C7orf61 CAMK1 CCDC113 CCT7
CCT8 CDH23 CFL1 CHURC1-FNTB CMTM1 COPS2 CREBFZ CTNND1
CCZ1 CDH3 CFTR CIAO1 CMTM3 COPS3 CREM CTNND2
CD14 CDH4 CGA CIB1 CMTM7 COPS4 CRHBP CTNS
CD151 CDH6 CH25H CIB2 CNDP2 COPS5 CRHR1 CTPS1
CD160 CDH7 CHAC2 CIB3 CNFN COPS6 CRIP3 CTPS2
CD163 CDH8 CHAMP1 CIC CNGA3 COPS7A CRIP3 CTSA
CD164 CDH9 CHCHD2 CIDEV CNIH2 COPS7B CRISP1 CTSC
CD19 CDIPT CHCHD7 CIITA CNIH3 COPZ1 CRISPLD1 CTSD
CD1C CDK1 CHD1 CILP CNKSR2 COQ10A CRK CTSK
CD22 CDK14 CHD2 CIR1 CNN1 COQ5 CRKL CTSL2
CD247 CDK17 CHD3 CIRBP CNN2 COQ7 CRLF1 CTSW
CD28 CDK19 CHD4 CIRH1A CNOT1 CORIN CRLF3 CTTN
CD2AP CDK2 CHD5 CISD2 CNOT10 CORO1A CRMP1 CTTNBP2
CD300A CDK20 CHD7 CISH CNOT2 CORO1B CROT CTTNBP2NL
CD300LG CDK3 CHD8 CIT CNOT3 CORO1C CRP CUEDC1
CD36 CDK4 CHD9 CITED2 CNOT6 CORO2A CRTAC1 CUL1
CD37 CDK5R1 CHDH CIZ1 CNOT6L CORO2B CRTC1 CUL2
CD38 CDK5RAP1 CHEK1 CKAP4 CNOT7 CORO6 CRTC2 CUL3
CD3D CDK5RAP2 CHEK2 CKAP5 CNOT8 CORT CRTC3 CUL4A
CD47 CDK8 CHFR CKB CNP COTL1 CRY1 CUL4B
CD53 CDK9 CHI3L1 CKLF CNRIP1 COX15 CRY2 CUL5
CD59 CDKAL1 CHIC2 CKM CNTFR COX16 CRYAA CUL7
CD63 CDKL2 CHID1 CKMT2 CNTN2 COX18 CRYAB CUL9
CD70 CDKL5 CHKA CLASRP CNTN4 COX4I1 CRYBA2 CUTC
CD74 CDKN2AIP CHKB CLCA4 CNTN5 COX5A CRYBB1 CUX1
CD79B CDKN3 CHL1 CLCN3 CNTNAP1 COX6A1 CRYGS CUX2
CD81 CDO1 CHML CLCN6 CNTNAP2 COX7A2L CRYL1 CWC25

CD82 CDR2L CHMP1B CLCN7 CNTNAP4 CPA2 CSAD CWC27
CD9 CDS2 CHMP2A CLDN1 CNTRL CPA4 CSDC2 CX3CL1
CD93 CDYL CHMP2B CLDN10 COA1 CPB2 CSGALNACT1 CXCL17
CD96 CDYL2 CHMP3 CLDN11 COASY CPEB1 CSGALNACT2 CXCR3
CD99L2 CEACAM16 CHMP4B CLDN14 COBRA1 CPEB3 CSH2 CXCR4
CDA CEBPE CHMP5 CLDN15 COG2 CPEB4 CSK CXCR5
CDADC1 CECR1 CHMP6 CLDN16 COG3 CPLX1 CSMD1 CXCR6
CDAN1 CELA2A CHMP7 CLDN25 COG7 CPLX4 CSMD2 CXorf21
CDC123 CELA3A CHN1 CLDN4 COG8 CPN1 CSMD3 CXorf23
CDC14A CELF1 CHN2 CLDN9 COIL CPNE4 CSNK1A1 CXorf26
CDC14B CELF2 CHODL CLDND1 COL11A1 CPNE6 CSNK1D CXorf57
CDC16 CELF3 CHP1 CLEC2L COL12A1 CPNE7 CSNK1E CXXC5
CDC20 CELF4 CHPF CLIC1 COL13A1 CPNE8 CSNK1G1 CYB561
CDC25A CELF5 CHRD CLIC4 COL17A1 CPS1 CSNK1G2 CYB561D2
CDC27 CELSR1 CHRM1 CLINT1 COL19A1 CPSF1 CSNK1G3 CYB5B
CDC34 CELSR2 CHRM3 CLIP2 COL1A1 CPSF2 CSNK2A1 CYB5D2
CDC37 CELSR3 CHRNA1 CLIP3 COL20A1 CPSF3 CSNK2B CYB5R1
CDC37L1 CEND1 CHRNA2 CLIP4 COL25A1 CPSF3L CSPP1 CYBASC3
CDC40 CENPB CHRNA4 CLK1 COL2A1 CPSF4 CSRNP1 CYC1
CDC42 CENPE CHRNA5 CLK2 COL3A1 CPSF4L CSRNP2 CYFIP1
CDC42BPA CENPM CHRNA6 CLK3 COL4A1 CPSF6 CSRNP3 CYGB
CDC42BPB CEP104 CHRNB1 CLMN COL4A2 CPSF7 CSRP2 CYHR1
CDC42BPG CEP112 CHRNB2 CLN3 COL4A3BP CPT1A CSRP3 CYP11B1
CDC42EP1 CEP128 CHRNB3 CLN6 COL5A1 CPT1B CST11 CYP17A1
CDC42EP3 CEP170 CHRND CLN8 COL5A2 CPT1C CSTF1 CYP19A1
CDC42EP4 CEP250 CHRNE CLNS1A COL6A1 CPXM2 CSTF2T CYP1A2
CDC42SE2 CEP63 CHRNG CLOCK COL6A3 CPZ CSTF3 CYP26A1
CDC45 CEP72 CHST1 CLPP COL6A6 CRABP1 CTAGE1 CYP27A1
CDC5L CEP78 CHST11 CLPTM1L COL7A1 CRABP2 CTBP1 CYP27B1
CDC7 CEP85 CHST14 CLPX COL8A1 CRAMP1L CTC1 CYP2E1
CDC73 CEP95 CHST15 CLRN1 COL9A1 CRB1 CTCF CYP2J2
CDCA7 CEP97 CHST3 CLRN2 COLEC10 CRB3 CTDP1 CYP2R1
CDCA8 CERS1 CHST6 CLSTN2 COLEC11 CRBN CTDSP2 CYP2U1
CDCP2 CERS2 CHST8 CLSTN3 COLQ CREB3 CTGF CYP3A4
CDH10 CES2 CHSY1 CLTB COMMD1 CREB3L2 CTHRC1 CYP3A5
CDH12 CES4A CHSY3 CLTC COMMD4 CREB3L3 CTNNA1 CYP3A7
CDH17 CETN2 CHTF8 CLVS1 COMMD7 CREB5 CTNNA2 CYP46A1
CDH18 CETP CHTOP CMAS COMP CREBBP CTNNB1 CYP4F11
CDH20 CFDP1 CHUK CMIP COMT CREBRF CTNNBL1 CYP4F22
CYP4X1 DDR1 DHRS11 DNAJB12 DRAP1 ECHDC1 EIF3E ENTPD4
CYP51A1 DDR2 DHRS13 DNAJB14 DRD1 ECHS1 EIF3F EP300
CYP7B1 DDX1 DHRS3 DNAJB2 DRD2 ECM1 EIF3G EP400
CYP8B1 DDX10 DHX15 DNAJB4 DRD3 ECSIT EIF3H EPAS1
CYR61 DDX17 DHX30 DNAJB5 DRG1 ECT2 EIF3I EPB41

CYSLTR1 DDX19A DHX32 DNAJB6 DROSHA EDAR EIF3L EPB41L1
CYSTM1 DDX21 DHX33 DNAJB8 DRP2 EDC3 EIF3M EPB41L4B
CYTH1 DDX25 DHX35 DNAJC1 DSC2 EDC4 EIF4A1 EPB42
CYTH2 DDX26B DHX40 DNAJC11 DSCAM EDEM2 EIF4A2 EPB49
CYTH3 DDX27 DHX58 DNAJC14 DSCAML1 EDEM3 EIF4A3 EPC1
CYTIP DDX39A DIABLO DNAJC16 DSCR3 EDIL3 EIF4B EPC2
D2HGDH DDX39B DIAPH1 DNAJC18 DSE EDN1 EIF4E2 EPG5
D4S234E DDX3X DICER1 DNAJC19 DSEL EDN2 EIF4E3 EPGN
DAAM1 DDX4 DIDO1 DNAJC2 DSG3 EDN3 EIF4EBP2 EPHA1
DAB2IP DDX49 DIMT1 DNAJC21 DSN1 EDNRA EIF4G1 EPHA2
DACH2 DDX5 DIO1 DNAJC22 DSP EEA1 EIF5 EPHA3
DAD1 DDX50 DIO3 DNAJC27 DST EED EIF5A EPHA4
DAG1 DDX55 DIP2A DNAJC3 DTD1 EEF1A1 EIF5B EPHA5
DAGLA DDX56 DIP2B DNAJC5 DTNA EEF1A2 ELAC2 EPHA6
DAK DDX6 DIP2C DNAJC6 DTNB EEF1E1 ELAVL1 EPHA7
DALRD3 DDX60 DIRAS1 DNAJC7 DTWD2 EEF1G ELAVL2 EPHB1
DAO DEAF1 DIRAS2 DNAJC8 DTX1 EEF2 ELAVL3 EPHB2
DAP3 DECR2 DIRAS3 DNAL4 DTX2 EEF2K ELAVL4 EPHB3
DAPK1 DEDD DIRC2 DNASE1L2 DTX3 EEFSEC ELF4 EPHB4
DAPK3 DEDD2 DIS3L DNM1 DTX3L EEPD1 ELF5 EPHB6
DAPP1 DEFB1 DIS3L2 DNM2 DTX4 EFCAB1 ELFN2 EPHX3
DAXX DEFB110 DKK4 DNMBP DTYMK EFCAB9 ELK3 EPHX4
DAZAP1 DEFB113 DLC1 DNMT1 DUOX1 EFEMP1 ELK4 EPN1
DBN1 DEFB121 DLD DNMT3B DUOXA2 EFEMP2 ELL EPN2
DBNDD1 DEFB134 DLG2 DNMT3L DUS1L EFHA2 ELMO1 EPRS
DBR1 DEFB136 DLG3 DNTT DUS2L EFHD1 ELMO2 EPS15L1
DCAF10 DEGS1 DLG4 DNTTIP2 DUS3L EFHD2 ELMOD1 EPS8
DCAF11 DEK DLG5 DOC2A DUSP10 EFNA4 ELN EPS8L2
DCAF12 DENND1A DLGAP1 DOCK11 DUSP12 EFNA5 ELOVL2 EPT1
DCAF12L1 DENND1B DLGAP2 DOCK3 DUSP14 EFNB2 ELOVL6 ERAS
DCAF12L2 DENND2A DLGAP3 DOCK4 DUSP16 EFR3A EMC10 ERBB2
DCAF15 DENND2D DLGAP4 DOCK5 DUSP18 EFR3B EMC2 ERBB3
DCAF17 DENND3 DLGAP5 DOCK7 DUSP26 EGFR EMC8 ERBB4
DCAF6 DENND4B DLK2 DOCK8 DUSP3 EGLN2 EMD ERC1
DCAF8 DENND5A DLL4 DOCK9 DUSP4 EGLN3 EMILIN1 ERC2
DCBLD1 DENND5B DLST DOK1 DUSP7 EGR1 EMILIN2 ERCC1
DCC DEPDC1 DLX1 DOK2 DVL2 EGR2 EMILIN3 ERCC2
DCDC1 DEPDC1B DLX2 DOK3 DYM EGR3 EML1 ERCC3
DCHS1 DEPDC4 DLX3 DOK4 DYNC1H1 EHBP1 EML3 ERCC4
DCLK1 DEPDC7 DLX5 DOK5 DYNC1I1 EHD1 EML4 ERCC6L
DCLK2 DERA DLX6 DOLK DYNC1I2 EHD4 EML5 ERCC8
DCP2 DERL1 DMAP1 DOLPP1 DYNC1LI2 EHF EML6 ERF
DCPS DERL2 DMBX1 DOPEY2 DYNC2H1 EHMT1 EMP2 ERG
DCT DES DMC1 DOT1L DYNLT1 EHMT2 EMX2 ERGIC1

DCTD DESI1 DMD DPEP1 DYRK1A EID1 ENAH ERGIC2
DCTN1 DFFB DMPK DPEP2 DYRK1B EIF2A ENAM ERGIC3
DCTN4 DFNA5 DMRT1 DPF1 DYRK2 EIF2AK2 ENC1 ERI1
DCTN5 DFNB59 DMRTC2 DPF2 DYSF EIF2AK4 ENDOU ERI3
DCTN6 DGAT1 DMTF1 DPF3 DZIP1 EIF2B1 ENG ERLEC1
DCUN1D1 DGAT2L6 DMWD DPH2 E2F4 EIF2B2 ENGASE ERLIN1
DCUN1D3 DGKD DMXL1 DPP8 E2F8 EIF2B3 ENO1 ERLIN2
DCUN1D4 DGKE DMXL2 DPP9 E4F1 EIF2B4 ENO2 ERN1
DCX DGKG DNAH1 DPY19L1 EBAG9 EIF2B5 ENO3 ERN2
DDAH1 DGKH DNAH2 DPY19L3 EBF1 EIF2C1 ENOPH1 ERO1L
DDAH2 DGKQ DNAH3 DPYSL2 EBF2 EIF2C2 ENOX1 ERP29
DDHD1 DGUOK DNAH5 DPYSL3 EBF3 EIF2C4 ENOX2 ERP44
DDHD2 DHCR24 DNAI1 DPYSL4 EBI3 EIF2D ENPP1 ERVFRD-1
DDI2 DHCR7 DNAJA1 DPYSL5 EBP EIF2S1 ENPP6 ESCO1
DDIT4 DHDDS DNAJA2 DQX1 EBPL EIF3A ENTHD1 ESF1
DDIT4L DHPS DNAJA3 DRAM1 ECE1 EIF3B ENTPD1 ESRP1
DDOST DHRS1 DNAJB11 DRAM2 ECE2 EIF3D ENTPD3 ESRP2
ESRRA FAM117A FAM73A FBXW11 FLOT1 FTL GAPDH GIPC3
ESRRG FAM118B FAM76A FBXW2 FLOT2 FTO GAPDHS GIT1
ESYT1 FAM120A FAM78A FBXW4 FLRT1 FTSJ2 GAPVD1 GIT2
ESYT2 FAM120C FAM78B FBXW5 FLRT3 FUBP1 GARS GJA1
ETF1 FAM122A FAM81A FBXW7 FLT1 FUBP3 GAS6 GJA10
ETFDH FAM123B FAM82A2 FBXW8 FLT3LG FURIN GAS7 GJA5
ETHE1 FAM125B FAM82B FCF1 FLYWCH1 FUT1 GATA1 GJA8
ETS1 FAM126B FAM83B FCGR3B FMOD FUT8 GATA2 GJC1
ETV1 FAM127C FAM83C FCGRT FMR1NB FXR1 GATA3 GJD2
ETV3 FAM129B FAM84A FCHSD2 FN1 FXR2 GATAD2B GK
ETV5 FAM131A FAM91A1 FCN1 FN3K FYB GATC GK2
ETV6 FAM131B FAM96A FCRL2 FNBP1 FYN GATM GK5
EVC2 FAM134C FAM98B FEM1B FNBP1L FZD3 GATS GKAP1
EVI5L FAM135A FAM9C FEM1C FNDC3A FZD4 GATSL3 GKN1
EVL FAM135B FANCA FEN1 FNDC3B FZD5 GBA2 GLB1L
EWSR1 FAM13A FANCB FER FNDC4 FZD7 GBF1 GLB1L2
EXD1 FAM160B1 FANCE FER1L6 FNDC5 FZD8 GBP1 GLB1L3
EXOC1 FAM160B2 FANCI FERD3L FNDC8 FZD9 GBX2 GLCCI1
EXOC3 FAM166A FANCL FERMT1 FNIP1 FZR1 GCC2 GLDC
EXOC3L2 FAM168A FAR1 FERMT2 FNIP2 G3BP1 GCDH GLE1
EXOC4 FAM168B FAR2 FERMT3 FNTA G3BP2 GCET2 GLG1
EXOC6B FAM169A FARS2 FEZ2 FNTB G6PC GCFC1 GLI1
EXOC8 FAM171A1 FARSB FEZF2 FOCAD G6PD GCH1 GLI2
EXOG FAM172A FASLG FGA FOPNL GABARAPL2 GCHFR GLI3
EXOSC1 FAM175B FASN FGD1 FOSB GABBR1 GCK GLIPR1
EXOSC10 FAM184A FASTK FGD4 FOSL2 GABBR2 GCKR GLIPR1L1
EXOSC2 FAM189A2 FASTKD1 FGF11 FOXA3 GABPA GCLC GLIPR2

EXOSC4 FAM18A FASTKD5 FGF12 FOXB1 GABRA1 GCLM GLIS2
EXOSC5 FAM196A FAT1 FGF13 FOXG1 GABRA2 GCM1 GLO1
EXOSC9 FAM198B FAT4 FGF17 FOXI1 GABRA3 GCN1L1 GLOD4
EXT2 FAM19A1 FATE1 FGF18 FOXJ2 GABRA5 GCNT2 GLRA1
EXTL2 FAM204A FAXC FGF21 FOXK2 GABRB1 GDAP1 GLRA2
EXTL3 FAM208A FBF1 FGF8 FOXN1 GABRB2 GDAP1L1 GLRA3
EYA1 FAM20A FBLN1 FGFR1 FOXN2 GABRB3 GDAP2 GLRX
EYA3 FAM20B FBLN5 FGFR1OP2 FOXN3 GABRD GDE1 GLRX3
EYA4 FAM210A FBLN7 FGFR2 FOXN4 GABRG1 GDF11 GLS2
EZH1 FAM212B FBN1 FGFR3 FOXO1 GABRG2 GDF2 GLT25D1
EZH2 FAM214B FBN2 FGFR1 FOXO3 GABRG3 GDF5 GLT25D2
EZR FAM217B FBXL12 FGL2 FOXO4 GABRQ GDI1 GLT8D1
F10 FAM222B FBXL14 FGR FOXP1 GAD1 GDI2 GLT8D2
F11R FAM227B FBXL16 FH FOXP2 GAD2 GDPD1 GLTP
F2 FAM24A FBXL18 FHIT FOXP3 GAK GDPD5 GLUL
F2R FAM26D FBXL19 FHOD1 FOXRED1 GALE GEMIN6 GLYR1
F2RL1 FAM35A FBXL2 FHOD3 FPGS GALK1 GEMIN7 GMCL1
F8 FAM3A FBXL20 FIBIN FPGT-TNNI3K GALK2 GET4 GMDS
FA2H FAM3C FBXL3 FIBP FPR2 GALM GFOD1 GMEB1
FAAH FAM40A FBXL6 FICD FPR3 GALNT1 GFOD2 GMEB2
FAAH2 FAM40B FBXL7 FIG4 FREM1 GALNT10 GFPT1 GMPPA
FABP7 FAM46B FBXO10 FIGF FREM2 GALNT11 GFPT2 GMPPB
FADS1 FAM46C FBXO11 FILIP1 FRG1 GALNT12 GFRA1 GMPR
FADS2 FAM48A FBXO15 FILIP1L FRK GALNT13 GFRA3 GMPR2
FADS3 FAM49A FBXO18 FIP1L1 FRMD4A GALNT14 GGA3 GMPS
FADS6 FAM49B FBXO21 FITM1 FRMD6 GALNT2 GGCX GNA11
FAF1 FAM53B FBXO24 FKBP10 FRMD8 GALNT3 GGPS1 GNA12
FAF2 FAM53C FBXO25 FKBP11 FRMPD2 GALNT5 GGT1 GNA14
FAH FAM54B FBXO3 FKBP14 FRMPD4 GALNT6 GGT5 GNA15
FAHD1 FAM57A FBXO30 FKBP1B FRRS1 GALNT9 GGT7 GNAI1
FAHD2A FAM57B FBXO31 FKBP3 FRS2 GALNTL1 GHITM GNAI2
FAIM2 FAM59A FBXO33 FKBP7 FRY GALNTL2 GHRHR GNAI3
FAM102A FAM63B FBXO38 FKBP8 FRYL GALNTL4 GHSR GNAL
FAM102B FAM64A FBXO40 FLCN FSD1 GALNTL6 GID8 GNAO1
FAM108B1 FAM65A FBXO42 FLI1 FSD2 GALT GIGYF1 GNAQ
FAM110B FAM65C FBXO46 FLII FSHR GAMT GINS2 GNAT1
FAM111A FAM69B FBXO48 FLNA FSTL1 GAN GINS3 GNAZ
FAM111B FAM70A FBXO8 FLNB FSTL4 GANAB GINS4 GNB1
FAM114A2 FAM71D FBXO9 FLNC FTCD GAP43 GIPC2 GNB2
GNB2L1 GPR18 GSDMD HCAR2 HIST1H2AD HNMT HSPA13 IGF2R
GNB3 GPR182 GSK3A HCFC1 HIST1H2AE HNRNPA1 HSPA14 IGFALS
GNB5 GPR20 GSK3B HCN1 HIST1H2AG HNRNPA2B1 HSPA1L IGFL2
GNG13 GPR21 GSPT2 HDAC1 HIST1H2AH HNRNPA3 HSPA4 IGJ
GNL1 GPR22 GSR HDAC10 HIST1H2AI HNRNPAB HSPA5 IGLON5

GNL3L GPR37 GSS HDAC11 HIST1H2AJ HNRNPC HSPA8 IGSF1
GNPAT GPR37L1 GSTA1 HDAC2 HIST1H2AL HNRNPF HSPA9 IGSF21
GNPDA1 GPR39 GSTA4 HDAC4 HIST1H2AM HNRNPH1 HSPB3 IGSF3
GNPDA2 GPR4 GSTCD HDAC5 HIST1H2BA HNRNPH2 HSPB7 IGSF8
GNPNAT1 GPR45 GSTM2 HDAC8 HIST1H2BC HNRNPK HSPBP1 IHH
GNPTAB GPR50 GSTP1 HDC HIST1H2BD HNRNPL HSPD1 IKBKB
GNPTG GPR52 GTDC1 HDDC2 HIST1H2BF HNRNPM HSPG2 IKBKE
GNRHR GPR56 GTDC2 HDGFRP2 HIST1H2BG HNRNPR HSPH1 IKZF3
GNS GPR61 GTF2A1 HDHD1 HIST1H2BH HNRNPU HTATIP2 IKZF4
GOLGA3 GPR63 GTF2F1 HDHD2 HIST1H2BJ HNRNPUL1 HTR1A IKZF5
GOLGB1 GPR65 GTF2F2 HDHD3 HIST1H2BK HNRNPUL2 HTR1B

IL13RA1

GOLIM4 GPR68 GTF2H1 HDLBP HIST1H2BL HNRPDL HTR1D IL16
GOLPH3 GPR75 GTF2H3 HEATR2 HIST1H2BM HNRPLL HTR1E IL17A
GOLT1A GPR77 GTF2H4 HEATR5B HIST1H2BN HOMER1 HTR1F IL17B
GON4L GPR83 GTF2H5 HEBP1 HIST1H2BO HOMER3 HTR2C IL17C
GOPC GPR85 GTF2IRD1 HECA HIST1H3B HOOK1 HTR5A IL17RD
GORAB GPR97 GTF3C1 HECTD1 HIST1H3C HOOK3 HTR6 IL1B
GORASP2 GPRASP1 GTF3C2 HECTD2 HIST1H3D HOXA11 HTRA3

IL1R2

GOT1 GPRC5A GTF3C3 HECTD3 HIST1H3E HOXA2 HTT IL1RAP
GOT1L1 GPRC5B GTF3C4 HECW1 HIST1H3G HOXA5 HUNK IL1RAPL1
GP2 GPS1 GTPBP1 HECW2 HIST1H3H HOXA6 HUWE1 IL1RAPL2
GPAM GPS2 GTPBP2 HELT HIST1H3I HOXB13 HYAL2 IL1RL2
GPATCH1 GPSM1 GTSF1 HELZ HIST1H4A HOXB2 HYOU1 IL1RN
GPATCH3 GPT2 GUCA1A HEMGN HIST1H4B HOXB3 IAPP IL21
GPATCH8 GPX3 GUCY1B3 HEPACAM HIST1H4C HOXB5 IBTK IL22RA2
GPC2 GPX4 GUCY2D HEPHL1 HIST1H4D HOXC10 ICA1 IL23A
GPC3 GPX7 GULP1 HERC1 HIST1H4E HOXC11 ICAM2 IL25
GPC4 GRAMD1B GUSB HERC2 HIST1H4H HOXC6 ICAM5 IL26
GPC5 GRB10 GXYL1 HERC3 HIST1H4I HOXC8 ICK IL27RA
GPCPD1 GRB14 GYG1 HERC5 HIST1H4J HOXD13 ID1 IL28RA
GPD1L GRB2 GYLTL1B HERPUD1 HIST2H2AB HOXD4 ID2 IL2RA
GPER GREB1 GYS1 HES1 HIST2H2AC HP IDE IL2RG
GPHA2 GRHL1 GZF1 HEXA HIST3H2A HP1BP3 IDH2 IL31
GPI GRHL2 GZMA HEXIM1 HIST3H2BB HPCA IDH3B IL3RA
GPKOW GRHPR GZMK HEY1 HIST3H3 HPCAL1 IDI1 IL4
GPM6A GRIA1 H1F0 HEY2 HIST4H4 HPCAL4 IFFO1 IL5
GPM6B GRIA2 H1FX HEYL HIVEP2 HPDL IFI27L2 IL6R
GPN3 GRIA3 H2AFX HFE2 HIVEP3 HPGD IFI30 IL8
GPR107 GRID1 H2AFY HFM1 HK1 HPN IFI44 ILDR2
GPR116 GRID2 H2AFY2 HGD HK3 HPS3 IFIT1 ILK
GPR123 GRIK1 H3F3B HGF HLCS HRAS IFITM1 ILKAP
GPR124 GRIK3 H6PD HGS HLF HRASLS IFNA14 IMMP2L

GPR125 GRIK4 HABP2 HIAT1 HM13 HRH2 IFNGR1 IMMT
GPR126 GRIN2A HABP4 HIATL1 HMBOX1 HRH3 IFNW1 IMP4
GPR132 GRIN2B HACL1 HIBADH HMBS HRSP12 IFRD1 IMPAD1
GPR133 GRINA HADHA HIC2 HMCN1 HS3ST1 IFT122 IMPDH1
GPR137 GRIP1 HADHB HIF1A HMG20A HS3ST2 IFT140 IMPDH2
GPR137B GRIPAP1 HAGH HIF3A HMG20B HS6ST3 IFT172 IMPG2
GPR137C GRK6 HAL HINT1 HMGA1 HSD11B1 IFT20 INA
GPR139 GRM1 HAO2 HINT2 HMGB1 HSD11B2 IFT43 ING1
GPR141 GRM2 HAPLN1 HIPK1 HMGCL HSD17B1 IFT52 ING2
GPR142 GRM3 HAPLN3 HIPK3 HMGCR HSD17B3 IFT57 ING3
GPR146 GRM4 HARBI1 HIPK4 HMGCS1 HSD3B7 IFT80 ING5
GPR149 GRM5 HARS2 HIRA HMGCS2 HSDL1 IFT81 INHBA
GPR157 GRM6 HAS2 HIST1H1A HMGXB4 HSF1 IFT88 INHBE
GPR158 GRM7 HAS3 HIST1H1C HMOX1 HSF2 IGDCC3 INO80B
GPR161 GRM8 HAUS4 HIST1H1D HMOX2 HSF2BP IGF1R INO80D
GPR17 GRN HAUS8 HIST1H1E HMP19 HSF4 IGF2 INO80E
GPR171 GRPEL2 HAX1 HIST1H2AA HMSD HSP90AB1 IGF2BP1 INPP4A
GPR173 GRSF1 HBP1 HIST1H2AB HNF1A HSP90B1 IGF2BP2 INPP4B
GPR176 GRWD1 HCAR1 HIST1H2AC HNF1B HSPA12A IGF2BP3
INPP5A
INPP5B ITFG3 KATNA1 KCNV1 KIAA1217 KLHL24 LARP1 LIX1L
INPP5F ITGA1 KATNAL1 KCNV2 KIAA1244 KLHL25 LARP4 LLGL1
INPP5J ITGA11 KATNB1 KCTD1 KIAA1279 KLHL26 LARP6 LMAN2
INPP5K ITGA2 KAZN KCTD10 KIAA1324 KLHL29 LARS LMAN2L
INPPL1 ITGA2B KBTBD12 KCTD13 KIAA1324L KLHL3 LARS2 LMBR1
INS ITGA3 KBTBD2 KCTD14 KIAA1429 KLHL30 LAS1L LMBR1L
INSIG1 ITGA4 KBTBD3 KCTD15 KIAA1432 KLHL32 LASP1 LMBRD1
INSIG2 ITGA5 KBTBD4 KCTD16 KIAA1467 KLHL34 LAT LMBRD2
INS-IGF2 ITGA8 KBTBD5 KCTD19 KIAA1468 KLHL36 LBR LMCD1
INSR ITGAD KBTBD6 KCTD2 KIAA1598 KLHL4 LBX1 LMF2
INSRR ITGAL KCMF1 KCTD21 KIAA1644 KLHL5 LCA5 LMLN
INTS1 ITGAM KCNA1 KCTD3 KIAA1737 KLHL6 LCAT LMNB1
INTS10 ITGAV KCNA10 KCTD4 KIAA1841 KLHL7 LCE3E LMO1
INTS12 ITGB1 KCNA2 KCTD5 KIAA1967 KLHL8 LCLAT1 LMO4
INTS2 ITGB1BP1 KCNA4 KCTD6 KIAA2018 KLHL9 LCMT1 LMTK2
INTS3 ITGB3BP KCNA6 KCTD7 KIAA2022 KLK13 LCP2 LMX1A
INTS4 ITGB4 KCNA7 KCTD8 KIDINS220 KLK5 LCT LMX1B
INTS6 ITGB5 KCNAB1 KCTD9 KIF12 KLK7 LDB1 LNX1
INTS7 ITGB6 KCNAB2 KDELC1 KIF13A KLK8 LDB2 LNX2
INTS8 ITGB7 KCNAB3 KDELRL1 KIF13B KLK9 LDB3 LOC81691
INTS9 ITGB8 KCNC1 KDELRL2 KIF1A KLRC1 LDHA LONP1
INTU ITIH1 KCNC2 KDM1A KIF1B KLRK1 LDHAL6A LONP2
INVS ITIH4 KCNC4 KDM2A KIF1C KPNA1 LDHB LONRF2
IP6K1 ITK KCND1 KDM2B KIF20A KPNA3 LDHD LONRF3

IPO11 ITLN1 KCND2 KDM3A KIF21A KPNA4 LDLR LOXL2
IPO13 ITM2B KCND3 KDM3B KIF21B KPNA6 LDLRAD3 LOXL4
IPO4 ITM2C KCNF1 KDM4B KIF22 KPNB1 LDOC1L LPAR1
IPO5 ITPK1 KCNG1 KDM4C KIF23 KPTN LECT1 LPAR2
IPO7 ITPKB KCNG3 KDM5A KIF26B KRAS LEPR LPAR3
IPO8 ITPKC KCNH2 KDM5B KIF3A KRIT1 LEPREL1 LPCAT1
IPO9 ITPR1 KCNH3 KDM5C KIF3B KRR1 LEPREL4 LPCAT4
IPP ITPR2 KCNH4 KDM6A KIF3C KRT14 LGALS4 LPHN1
IQCD ITPR3 KCNH5 KDM6B KIF4A KRT17 LGALS9C LPHN2
IQCH ITPRIP KCNH6 KDM8 KIF5A KRT18 LGI1 LPHN3
IQGAP1 ITSN1 KCNH7 KDSR KIF7 KRT19 LGI2 LPIN1
IQGAP3 ITSN2 KCNH8 KEL KIFAP3 KRT222 LGR4 LPIN2
IQSEC1 IVD KCNIP2 KERA KIFC3 KRT38 LGR6 LPO
IQSEC2 IVNS1ABP KCNIP4 KHDRBS1 KIN KRT6B LHFP LPPR1
IQSEC3 IZUMO1 KCNJ1 KHDRBS2 KIRREL KRT6C LHFPL2 LPPR2
IRAK3 IZUMO2 KCNJ13 KHDRBS3 KIRREL3 KRT73 LHFPL4 LPPR4
IRAK4 JAG1 KCNJ14 KHNYN KIT KRT85 LHFPL5 LPPR5
IREB2 JAG2 KCNJ2 KHSRP KITLG KRT9 LHX2 LRBA
IRF2 JAK1 KCNJ4 KIAA0101 KLC1 KRTAP13-3 LHX4 LRCH3
IRF4 JAK2 KCNJ5 KIAA0146 KLC4 KRTAP19-5 LHX9 LRCH4
IRF5 JAKMIP1 KCNJ6 KIAA0182 KLF12 KRTAP19-7 LIAS LRFN2
IRF6 JAKMIP2 KCNJ9 KIAA0195 KLF15 KRTAP21-1 LIFR LRFN3
IRF7 JAKMIP3 KCNK1 KIAA0196 KLF3 KRTCAP2 LIG3 LRFN5
IRF8 JARID2 KCNK10 KIAA0232 KLF4 KRTDAP LIG4 LRIF1
IRF9 JAZF1 KCNK17 KIAA0240 KLF6 KY LILRA5 LRIG2
IRGC JDP2 KCNK2 KIAA0247 KLF7 KYNU LIM2 LRIG3
IRS1 JHDM1D KCNK4 KIAA0284 KLHDC1 L1CAM LIMD1 LRIT1
IRX6 JKAMP KCNK6 KIAA0317 KLHDC10 L3MBTL3 LIMK1 LRIT2
ISG15 JMJD1C KCNK9 KIAA0319L KLHDC2 LACRT LIMS2 LRP1
ISG20 JMJD6 KCNMA1 KIAA0408 KLHDC3 LACTB2 LIN28A LRP10
ISL1 JPH3 KCNMB2 KIAA0430 KLHDC5 LAMA3 LIN54 LRP12
ISL2 JRKL KCNN2 KIAA0494 KLHDC8B LAMA4 LIN7A LRP1B
ISLR JUN KCNN3 KIAA0513 KLHL1 LAMB2 LIN7B LRP2BP
ISLR2 JUP KCNN4 KIAA0528 KLHL11 LAMB4 LIN7C LRP4
ISOC1 KAAG1 KCNQ1 KIAA0564 KLHL12 LAMC1 LIN9 LRP5
ISOC2 KALRN KCNQ2 KIAA0664 KLHL13 LAMP1 LINGO1 LRP6
IST1 KANSL1 KCNQ3 KIAA0895 KLHL15 LAMP3 LINGO2 LRP8
ISY1 KAT2B KCNQ4 KIAA0907 KLHL17 LAMTOR1 LINGO4 LRPPRC
ISY1-RAB43 KAT5 KCNQ5 KIAA0913 KLHL18 LAP3 LIPC LRRC1
ISYNA1 KAT6A KCNT1 KIAA0947 KLHL20 LAPT M4A LIPH LRRC10
ITFG1 KAT6B KCNT2 KIAA1045 KLHL22 LAPT M4B LIPT1 LRRC15
ITFG2 KAT8 KCNU1 KIAA1109 KLHL23 LARGE LIX1 LRRC16A
LRRC16B MAFB MAPKAPK3 MDF1 MFAP5 MMP11 MRPS11 MYCN
LRRC20 MAFG MAPKAPK5 MDH1 MFF MMP15 MRPS12 MYD88

LRRC23 MAG MAPKBP1 MDH1B MFGE8 MMP16 MRPS14 MYEF2
LRRC27 MAGEA3 MAPRE1 MDH2 MFHAS1 MMP17 MRPS15 MYF5
LRRC28 MAGEB4 MAPRE2 MDM2 MFI2 MMP19 MRPS18A MYH10
LRRC32 MAGEC2 MAPRE3 MDP1 MFN1 MMP2 MRPS23 MYH11
LRRC33 MAGED1 MARC2 ME1 MFN2 MMP21 MRPS35 MYH13
LRRC39 MAGED2 MARCH3 ME3 MFNG MMP24 MRPS5 MYH14
LRRC3B MAGEE2 MARCH4 MEAF6 MFRP MMP25 MRTO4 MYH2
LRRC4 MAGI1 MARCH5 MECP2 MFSD11 MMP3 MS4A1 MYH3
LRRC41 MAGI2 MARCH6 MED1 MFSD12 MMRN2 MS4A3 MYH6
LRRC42 MAGOHB MARCKSL1 MED10 MFSD2A MNAT1 MS4A8B
MYH7
LRRC45 MAK MARCO MED12L MFSD5 MND1 MSANTD2 MYH7B
LRRC47 MAK16 MARK1 MED13 MFSD7 MNT MSANTD3 MYH8
LRRC49 MAL MARK2 MED13L MGA MOB1B MSANTD4 MYH9
LRRC4B MALSU1 MARK3 MED14 MGAT3 MOB2 MSH2 MYL10
LRRC4C MALT1 MARS2 MED15 MGAT4C MOB3A MSH3 MYL2
LRRC52 MAMSTR MARVELD2 MED16 MGAT5 MOCS3 MSH6 MYL3
LRRC55 MAN1A1 MAS1 MED17 MGEA5 MON1A MSI1 MYL6B
LRRC57 MAN1C1 MASP1 MED18 MGLL MON1B MSL1 MYL9
LRRC59 MAN2A1 MAST1 MED20 MGRN1 MON2 MSL2 MYLK
LRRC7 MAN2A2 MAST2 MED22 MGST1 MORC2 MSL3 MYLK2
LRRC8A MAN2C1 MAST3 MED24 MGST3 MORC3 MSN MYLK3
LRRC8B MANBA MAST4 MED25 MIB1 MORC4 MSRB3 MYLPF
LRRC8D MANBAL MAT1A MED26 MICAL1 MORF4L1 MST1 MYNN
LRRFIP2 MANEAL MAT2A MED27 MICALL1 MORF4L2 MST1R MYO10
LRRK1 MAOA MAT2B MED28 MICU1 MORN1 MST4 MYO15A
LRRN1 MAOB MATK MED29 MID1 MORN3 MTA3 MYO18A
LRRN2 MAP1B MATN1 MED6 MID2 MORN4 MTAP MYO1B
LRRN3 MAP2 MATR3 MED8 MIDN MOSPD2 MTCH1 MYO1D
LRRTM2 MAP2K1 MAU2 MEF2A MIER1 MOV10 MTDH MYO1F
LRRTM3 MAP2K2 MAX MEF2B MIER3 MOV10L1 MTF2 MYO1G
LRRTM4 MAP2K4 MB MEF2D MIF4GD MOXD1 MTFP1 MYO1H
LRSAM1 MAP2K5 MB21D2 MEGF10 MINK1 MPG MTFR1 MYO5A
LRTM2 MAP2K6 MBD1 MEGF11 MINPP1 MPHOSPH8 MTG1 MYO5B
LSM1 MAP2K7 MBD2 MEGF8 MIOX MPHOSPH9 MTHFD1L MYO6
LSM14B MAP3K1 MBD3 MEGF9 MIPOL1 MPI MTHFD2 MYO7A
LSM3 MAP3K11 MBD5 MEI1 MIS12 MPL MTHFD2L MYO9A
LSM4 MAP3K12 MBD6 MEIS1 MIS18A MPO MTIF2 MYO9B
LSR MAP3K13 MBLAC2 MEIS2 MITD1 MPP1 MTM1 MYOC
LTBP1 MAP3K2 MBNL1 MEIS3 MITF MPP3 MTMR1 MYOCD
LTBP2 MAP3K5 MBNL2 MEN1 MKI67IP MPP4 MTMR3 MYOD1
LTK MAP3K7 MBNL3 MEPCE MKL1 MPP6 MTMR4 MYOF
LUM MAP3K8 MBOAT7 MESDC2 MKL2 MPP7 MTMR8 MYOM2
LY6H MAP3K9 MBTPS1 MET MKLN1 MPPED1 MTMR9 MYOT

LYG1 MAP4K1 MBTPS2 METAP1 MKNK2 MPPED2 MTO1 MYOZ1
LYN MAP4K3 MC3R METAP1D MKRN2 MPRIP MTOR MYOZ2
LYNX1 MAP4K4 MC4R METAP2 MLANA MPV17 MTPAP MYRIP
LYPD6 MAP7D1 MCAM METRNL MLEC MPZ MTSS1 MYT1
LYPD6B MAPK1 MCC METTL1 MLF2 MPZL1 MTSS1L MYT1L
LYPLA2 MAPK10 MCCC1 METTL13 MLL MRAP2 MUCL1 N4BP1
LYRM5 MAPK11 MCCC2 METTL14 MLL2 MRAS MUM1 N4BP3
LYST MAPK12 MCF2L METTL15 MLL3 MRFAP1 MURC NAA15
LZIC MAPK13 MCFD2 METTL16 MLL4 MRGPRF MUTED NAA16
LZTR1 MAPK14 MCHR2 METTL2B MLL5 MRPL10 MVK NAA20
LZTS2 MAPK1IP1L MCL1 METTL3 MLLT1 MRPL11 MVP NAA25
M6PR MAPK3 MCM3AP METTL5 MLLT10 MRPL14 MX1 NAA30
MAB21L1 MAPK4 MCM5 METTL6 MLLT3 MRPL19 MX2 NAA35
MAB21L2 MAPK6 MCM6 METTL9 MLLT4 MRPL2 MXD1 NAA40
MAB21L3 MAPK7 MCM7 MEX3A MLST8 MRPL22 MXD4 NAA60
MACF1 MAPK8 MCMBP MEX3B MMAA MRPL27 MXI1 NAB1
MAD1L1 MAPK8IP1 MCOLN3 MEX3C MMACHC MRPL3 MYADM
NABP1
MAD2L1BP MAPK8IP3 MCRS1 MFAP2 MMD MRPL36 MYBL1 NABP2
MADD MAPK9 MCTP1 MFAP3 MMD2 MRPL42 MYBPC3 NACC1
MAEA MAPKAP1 MCTP2 MFAP3L MME MRPL44 MYBPH NADK
MAF1 MAPKAPK2 MCU MFAP4 MMEL1 MRPL54 MYCL1 NADSYN1
NAE1 NDUFAF3 NGEF NPBWR2 NUAK1 OLA1 OTOP2 PARN
NAGA NDUFAF5 NGFR NPDC1 NUAK2 OLFM1 OTUD5 PARP11
NAGK NDUFAF6 NGLY1 NPEPL1 NUB1 OLFM3 OTX1 PARP12
NAGLU NDUFB10 NHS NPL NUBP1 OLFML2A OTX2 PARP14
NAGS NDUFB11 NICN1 NPLOC4 NUBP2 OLIG3 OVCA2 PARP16
NALCN NDUFB2 NIF3L1 NPM1 NUCB1 OLR1 OVOL1 PARP2
NAMPT NDUFB4 NIM1 NPR1 NUCB2 ONECUT1 OVOL2 PARP6
NANP NDUFB7 NIP7 NPRL2 NUCKS1 ONECUT2 OXCT1 PARP9
NAP1L1 NDUFB8 NIPA1 NPTN NUDC OPA1 OXGR1 PARS2
NAP1L2 NDUFS1 NIPAL3 NPTX1 NUDCD2 OPCML OXSR1 PARVA
NAP1L4 NDUFS3 NIPBL NPY1R NUDT10 OPN1SW P2RX1 PATE2
NAP1L5 NDUFS4 NIPSNAP1 NPY2R NUDT11 OPN5 P2RX2 PATZ1
NAPB NDUFS5 NIPSNAP3A NPY5R NUDT14 OPRD1 P2RX6 PAX2
NAPG NDUFS6 NISCH NR0B1 NUDT17 OPRK1 P2RY1 PAX6
NARFL NDUFS7 NIT2 NR1D1 NUDT4 OPRL1 P2RY12 PAX8
NARG2 NDUFS8 NKAIN2 NR1H2 NUDT5 OPTN P2RY6 PAXIP1
NARS NDUFV1 NKAP NR1H3 NUF2 OR10J5 P2RY8 PBRM1
NARS2 NECAB1 NKD1 NR1I2 NUFIP1 OR13J1 P4HA1 PBX2
NASP NECAB2 NKIRAS1 NR1I3 NUMA1 OR1E1 P4HB PBX3
NAT10 NECAP1 NKIRAS2 NR2C1 NUMB OR1J4 P4HTM PBX4
NAT16 NEDD1 NKRF NR2C2 NUP107 OR1K1 PA2G4 PC
NAV1 NEDD4L NKX2-5 NR2C2AP NUP155 OR2B3 PABPC1 PCBP1

NAV2 NEDD9 NLGN1 NR2E1 NUP160 OR2Z1 PABPC5 PCBP2
NAV3 NEGR1 NLGN3 NR2F1 NUP188 OR4A15 PACRG PCBP3
NBAS NEK1 NLGN4X NR2F2 NUP210 OR4C3 PACS2 PCBP4
NBEA NEK2 NLK NR3C1 NUP210L OR4D10 PACSIN1 PCDH10
NBEAL2 NEK4 NLN NR3C2 NUP214 OR4X2 PACSIN3 PCDH11X
NCAM2 NEK6 NLRC4 NR4A1 NUP35 OR51T1 PADI1 PCDH17
NCAN NEK8 NLRP11 NR4A2 NUP37 OR56A1 PADI2 PCDH18
NCAPD2 NEK9 NLRP12 NR4A3 NUP43 OR56A3 PAF1 PCDH19
NCAPD3 NELF NLRP9 NR5A2 NUP50 OR56A4 PAFAH1B1 PCDH7
NCAPG NELL2 NMD3 NR6A1 NUP54 OR5C1 PAFAH1B2 PCDH9
NCAPH NEMF NME1 NRBP1 NUP62 OR5D16 PAFAH1B3 PCDHA13
NCAPH2 NEO1 NME1-NME2 NRCAM NUP93 OR5T1 PAFAH2 PCDHA3
NCBP1 NET1 NME2 NRF1 NUP98 OR6C75 PAG1 PCDHA6
NCBP2 NETO2 NME4 NRG3 NUPL1 OR7D2 PAH PCDHAC2
NCF4 NEU1 NME6 NRIP2 NXF1 OR9Q2 PAIP1 PCDHGA5
NCK2 NEU3 NME7 NRP1 NXF3 ORAI2 PAIP2B PCDHGA6
NCKAP1 NEURL4 NMI NRSN1 NXN ORAI3 PAK1 PCDHGB1
NCKAP1L NEUROD1 NMRAL1 NRXN1 NXPH1 ORC2 PAK2 PCDHGB4
NCKAP5L NEUROD6 NMT1 NRXN2 NXPH2 ORC5 PAK3 PCDHGB7
NCL NF1 NMT2 NRXN3 NXPH3 ORC6 PAK4 PCDHGC3
NCLN NF2 NNMT NSA2 NXPH4 ORMDL1 PAK7 PCDHGC4
NCOA1 NFAM1 NNT NSD1 NXT1 ORMDL2 PALM PCDHGC5
NCOA3 NFASC NOB1 NSDHL NYAP1 ORMDL3 PALM2 PCDP1
NCOA6 NFAT5 NOC4L NSFL1C O3FAR1 OS9 PAM PCED1B
NCOA7 NFATC1 NOD1 NSMAF OAS3 OSBP PAMR1 PCF11
NCOR1 NFATC2 NODAL NSMCE2 OASL OSBP2 PAN2 PCGF2
NCOR2 NFE2 NOL4 NSMCE4A OAT OSBPL10 PAN3 PCGF3
NCS1 NFE2L1 NOL6 NSUN3 ODSL1 OSBPL11 PANK2 PCGF5
NCSTN NFIA NOL9 NSUN6 OCIAD1 OSBPL1A PAPD4 PCGF6
NDC80 NFIB NOP56 NT5C1A OCRL OSBPL2 PAPD5 PCID2
NDE1 NFIC NOP58 NT5C2 ODC1 OSBPL3 PAPL PCIF1
NDFIP1 NFIX NOS2 NT5DC2 ODF2 OSBPL5 PAPOLA PCLO
NDNF NFKB1 NOSIP NT5DC3 ODF2L OSBPL6 PAPOLG PCM1
NDRG1 NFKB2 NOTCH1 NT5M ODZ1 OSBPL7 PAPPA PCMTD2
NDRG3 NFKBIA NOTCH2 NTF3 ODZ3 OSBPL8 PAPPA2 PCNA
NDRG4 NFKBIB NOTCH3 NTHL1 ODZ4 OSBPL9 PAPSS1 PCNX
NDST1 NFKBID NOTUM NTMT1 OFD1 OSGEV PAPSS2 PCNXL3
NDST2 NFKBIZ NOVA1 NTN1 OGDH OSGIN2 PAQR3 PCOLCE
NDUFA12 NFRKB NOVA2 NTN4 OGDHL OSM PAQR4 PCSK1
NDUFA4 NFS1 NOX3 NTNG1 OGFOD1 OSTalpha PAQR6 PCSK7
NDUFA4L2 NFX1 NOXRED1 NTNG2 OGFOD2 OSTF1 PAQR7 PCYOX1
NDUFA5 NFYA NPAS2 NTRK1 OGFRL1 OSTM1 PAQR8 PCYT1A
NDUFA8 NFYC NPAS3 NTRK2 OGN OTOA PAQR9 PCYT1B
NDUFAB1 NGDN NPAS4 NTRK3 OGT OTOF PARD6G PCYT2

PDAP1 PFKFB3 PIGN PLD5 POLK PPM1B PRKAR2A PSMA2
PDC PFKL PIGO PLEC POLL PPM1D PRKAR2B PSMA3
PDCD1 PFKM PIGR PLEK2 POLR1B PPM1E PRKCA PSMA4
PDCD11 PFKP PIGT PLEKHA1 POLR1C PPM1G PRKCB PSMA5
PDCD2L PFN1 PIGU PLEKHA3 POLR1D PPM1K PRKCG PSMA7
PDCD4 PGAM2 PIGV PLEKHA5 POLR2A PPME1 PRKCH PSMA8
PDCD6 PGAP1 PIK3C2A PLEKHA6 POLR2C PPOX PRKCI PSMB10
PDCD6IP PGAP2 PIK3C3 PLEKHA7 POLR2D PPP1CA PRKCQ PSMB4
PDCL PGBD2 PIK3CA PLEKHB1 POLR2E PPP1CB PRKCSH PSMB6
PDDC1 PGBD3 PIK3CB PLEKHB2 POLR2I PPP1R10 PRKD2 PSMC1
PDE12 PGBD4 PIK3CD PLEKHF1 POLR2L PPP1R12A PRKG1 PSMC2
PDE1A PGD PIK3CG PLEKHF2 POLR3A PPP1R12B PRKG2 PSMC3
PDE1B PGF PIK3R1 PLEKHG3 POLR3B PPP1R13B PRKRIR PSMC3IP
PDE4B PGK1 PIK3R2 PLEKHG4B POLR3D PPP1R16A PRKX PSMC4
PDE4D PGM1 PIK3R3 PLEKHG6 POLR3E PPP1R16B PRL PSMC5
PDE6A PGM3 PIK3R5 PLEKHH2 POLR3F PPP1R27 PRLH PSMC6
PDE6B PGM5 PIKFYVE PLEKHM1 POLR3G PPP1R7 PRM2 PSMD1
PDE7A PGPEP1 PILRB PLEKHM2 POLR3GL PPP1R8 PRMT1 PSMD10
PDE7B PGS1 PIM2 PLEKHO1 POLR3H PPP2CA PRMT10 PSMD11
PDE8A PHACTR3 PINK1 PLIN3 POLR3K PPP2CB PRMT3 PSMD12
PDE8B PHACTR4 PIP4K2C PLK1 POM121 PPP2R1A PRMT5 PSMD2
PDE9A PHAX PIP5K1A PLK2 POMGNT1 PPP2R2A PRMT6 PSMD3
PDGFA PHB2 PIP5K1C PLK4 POMP PPP2R2B PRMT7 PSMD4
PDGFD PHC2 PIPOX PLLP POMT1 PPP2R2C PRMT8 PSMD5
PDGFRA PHC3 PISD PLOD1 POP1 PPP2R3C PROC PSMD6
PDGFRB PHF1 PITPNNA PLRG1 POP5 PPP2R5A PROCA1 PSMD7
PDHB PHF10 PITPNM2 PLS1 POPDC2 PPP2R5B PROKR1 PSMD8
PDIA3 PHF11 PITPNM3 PLS3 POR PPP2R5D PROKR2 PSME3
PDIA4 PHF13 PITX1 PLSCR5 PORCN PPP2R5E PROL1 PSPC1
PDIA6 PHF14 PITX2 PLTP POU2AF1 PPP3CA PROM1 PTBP1
PDIK1L PHF15 PIWIL1 PLVAP POU2F1 PPP3CB PROM2 PTBP2
PDK1 PHF16 PIWIL2 PLXNA1 POU2F2 PPP4C PROS1 PTBP3
PDK2 PHF17 PJA2 PLXNA2 POU4F2 PPP5C PROSC PTCD2
PDLM1 PHF20 PKD2 PLXNA3 POU4F3 PPP6C PROSER1 PTCD3
PDLM3 PHF20L1 PKD2L2 PLXNA4 POU5F1 PPT1 PROX1 PTCH1
PDLM7 PHF21A PKDREJ PLXNB1 POU6F1 PPTC7 PROX2 PTCH2
PDP1 PHF23 PKM PLXNB2 POU6F2 PPWD1 PROZ PTCHD2
PDS5A PHF3 PKMYT1 PLXNB3 PPA1 PQLC1 PRPF19 PTCHD4
PDS5B PHGDH PKN3 PLXNC1 PPAN PQLC3 PRPF3 PTDSS1
PDSS2 PHIP PKNOX1 PLXND1 PPAN-P2RY11 PRC1 PRPF38A PTDSS2
PDXK PHKA1 PKNOX2 PMCH PPAP2A PRCC PRPF4 PTEN
PDZD4 PHKA2 PLA2G12B PMFBP1 PPAP2B PRDM15 PRPF40B PTGDS
PDZD7 PHKG1 PLA2G15 PMM1 PPAPDC1A PRDM16 PRPF4B PTGER2
PEAK1 PHKG2 PLA2G1B PMM2 PPAPDC3 PRDM2 PRPF6 PTGER4

PEF1 PHLDA2 PLA2G4A PMPCA PPARA PRDM4 PRPF8 PTGFRN
PEG10 PHLDDB1 PLA2G5 PMPCB PPARD PRDM5 PRPS2 PTGIS
PELI1 PHLPP2 PLA2G6 PNCK PPARG PRDX1 PRPSAP1 PTGR2
PELI2 PHRF1 PLAC1 PNKD PPARGC1A PRDX2 PRPSAP2 PTGS2
PELO PHTF1 PLAG1 PNKP PPBP PRDX6 PRR14 PTH
PEPD PHTF2 PLAGL1 PNLLIP PPCDC PREB PRRC2B PTH2R
PER1 PHYHIP PLAGL2 PNMA2 PPDPF PRELID1 PRRC2C PTK2
PER2 PHYHIPL PLAT PNN PPEF1 PREPL PRSS1 PTK2B
PES1 PI4K2B PLBD2 PNOC PPFIA1 PREX2 PRSS16 PTK7
PEX1 PI4KA PLCB1 PNPLA4 PPFIA4 PRF1 PRSS22 PTOV1
PEX10 PI4KB PLCB2 PNPLA8 PPFIBP1 PRG2 PRSS3 PTP4A1
PEX11A PIAS1 PLCB3 PNPO PPFIBP2 PRIC285 PRSS35 PTP4A3
PEX11B PIAS2 PLCB4 POC1B PPHLN1 PRICKLE1 PRSS53 PTPDC1
PEX14 PIAS3 PLCD1 POFUT1 PPIB PRICKLE3 PRSS57 PTPN1
PEX19 PIAS4 PLCG1 POFUT2 PPIE PRIMA1 PRSS8 PTPN11
PEX3 PICALM PLCG2 POGZ PPIH PRKAA1 PRTG PTPN12
PEX5 PICK1 PLCL1 POLA1 PPIL2 PRKAA2 PRX PTPN14
PFAS PID1 PLCL2 POLA2 PPIL4 PRKACA PSAT1 PTPN18
PFDN2 PIDD PLCXD2 POLB PPIL6 PRKACB PSD PTPN23
PFDN5 PIGH PLCXD3 POLD1 PPIP5K2 PRKACG PSEN1 PTPN5
PFKFB1 PIGK PLCZ1 POLE PPL PRKAG2 PSKH1 PTPN6
PFKFB2 PIGM PLD3 POLE3 PPM1A PRKAR1A PSMA1 PTPN9
PTPRD RAB40C RAPH1 RBMS1 RGS12 RNF150 RPLP2 RUSC1
PTPRE RAB41 RARA RBMS2 RGS13 RNF152 RPN2 RUSC2
PTPRF RAB4A RARB RBMS3 RGS17 RNF165 RPP25L RUVBL1
PTPRG RAB4B RARG RBP2 RGS19 RNF166 RPRD1A RWDD1
PTPRK RAB5A RARRES3 RBP4 RGS2 RNF167 RPRD1B RWDD2A
PTPRM RAB5B RARS RBPJ RGS20 RNF169 RPRD2 RXFP1
PTPRN RAB7A RARS2 RC3H1 RGS3 RNF182 RPS10 RXFP2
PTPRO RAB8A RASA1 RC3H2 RGS4 RNF19A RPS11 RXFP3
PTPRS RAB8B RASA2 RCAN1 RGS6 RNF19B RPS14 RXRA
PTPRU RAB9B RASA3 RCAN3 RGS7 RNF2 RPS16 RXRB
PTPRZ1 RABAC1 RASAL2 RCBTB1 RHBDD2 RNF20 RPS19 RXRG
PTRHD1 RABEP1 RASAL3 RCBTB2 RHBDD3 RNF208 RPS2 RYR1
PTS RABEP2 RASD1 RCC1 RHBDL1 RNF214 RPS26 RYR2
PTTG1 RABGAP1 RASD2 RCC2 RHBDL2 RNF216 RPS5 RYR3
PUF60 RABGEF1 RASGEF1B RCE1 RHBDL1 RNF219 RPS6 S100A1
PUM2 RABGGTA RASGEF1C RCHY1 RHBDL3 RNF220 RPS6KA4
S100A14
PURA RABGGTB RASGRF1 RCN2 RHCG RNF24 RPS6KA5 S100A16
PURG RABL3 RASGRF2 RCOR1 RHO RNF26 RPS6KA6 S100A4
PUS1 RABL5 RASGRP1 RCOR2 RHOBTB1 RNF31 RPS6KB1 S100P
PUS10 RABL6 RASGRP2 RCOR3 RHOBTB3 RNF34 RPS6KB2 S100PBP
PUS7 RAC1 RASL11A RDH12 RHOC RNF38 RPS7 S1PR2

PVRL1 RAC2 RASL11B RDH16 RHOH RNF40 RPS8 S1PR4
PVRL4 RAC3 RASL12 RDX RHOJ RNF41 RPS9 SACM1L
PWP2 RACGAP1 RASSF2 REC8 RHOQ RNF44 RPTOR SACS
PXDN RAD17 RASSF8 REEP2 RHOT1 RNF5 RPUSD1 SAE1
PXMP2 RAD21 RAVER1 REEP5 RHOU RNF8 RPUSD2 SAFB
PXN RAD23B RAVER2 REEP6 RHPN1 RNFT2 RQCD1 SALL1
PYCR1 RAD50 RB1 REG1A RIBC1 RNGTT RRAD SALL4
PYCR2 RAD51 RB1CC1 REG3A RIC8A RNH1 RRAGA SAMD13
PYGB RAD54L RBBP4 REL RIC8B RNPEP RRAGB SAMD14
PYGM RAD54L2 RBBP5 RELA RICTOR RNPS1 RRAGC SAMD3
PYGO2 RAD9A RBBP6 RELN RIF1 ROBO1 RRAGD SAMD4A
PYROXD1 RAE1 RBBP7 REM1 RILPL2 ROBO2 RREB1 SAMD8
QARS RAF1 RBBP8 REM2 RIMBP2 ROCK1 RRM1 SAMHD1
QDPR RAG1 RBBP9 REN RIMKLA ROGDI RRM2 SMM50
QKI RAI1 RBFOX1 REPS1 RIMS1 ROM1 RRM2B SAP130
QRICH1 RAI14 RBFOX2 REPS2 RIMS2 ROR1 RRNAD1 SAP18
QSER1 RAI2 RBFOX3 RER1 RIMS4 RORB RRP1B SAP30BP
QSOX1 RALB RBL1 RERE RING1 RPAP2 RRP9 SAR1B
QSOX2 RALBP1 RBL2 RERG RINT1 RPAP3 RSBN1 SARNP
QTRT1 RALGAPA1 RBM10 RET RIOK3 RPE RSBN1L SARS
R3HDM1 RALGAPA2 RBM12B REV3L RIPK1 RPE65 RSG1 SARS2
R3HDM2 RALGDS RBM14 REXO1 RIPK4 RPH3A RSL1D1 SASH1
RAB11A RALGPS1 RBM14-RBM4 REXO2 RIT1 RPH3AL RSL24D1 SASH3
RAB11B RALGPS2 RBM15B RFC2 RLIM RPIA RSPH9 SASS6
RAB11FIP4 RALY RBM17 RFC3 RLN1 RPL10A RSPO3 SATB1
RAB11FIP5 RALYL RBM18 RFC4 RMND5B RPL11 RSPRY1 SATB2
RAB12 RAMP1 RBM22 RFC5 RNASE12 RPL13A RSRC1 SAV1
RAB19 RAMP2 RBM23 RFFL RNASEH1 RPL15 RSRC2 SBF1
RAB1B RAN RBM25 RFK RNASEH2C RPL17 RSU1 SBF2
RAB20 RANBP1 RBM26 RFTN1 RNASEL RPL18 RTE1 SBNO1
RAB22A RANBP10 RBM27 RFWD2 RND2 RPL18A RTF1 SBNO2
RAB24 RANBP2 RBM3 RFWD3 RND3 RPL23A RTKN SC5DL
RAB26 RANBP3 RBM34 RFX1 RNF10 RPL24 RTKN2 SCAF11
RAB27A RANBP9 RBM39 RFX2 RNF103 RPL26L1 RTN2 SCAF4
RAB28 RANGAP1 RBM4 RFX3 RNF103-CHMP3 RPL27 RTN3 SCAI
RAB30 RANGRF RBM41 RFX4 RNF115 RPL29 RTN4IP1 SCAMP2
RAB31 RAP1A RBM42 RGL2 RNF121 RPL3 RUFY1 SCAMP5
RAB32 RAP1B RBM45 RGL3 RNF126 RPL34 RUFY2 SCAP
RAB34 RAP1GAP RBM46 RGMA RNF130 RPL6 RUFY3 SCARA3
RAB35 RAPGEF1 RBM47 RGMB RNF139 RPL7 RUNDC1 SCARB1
RAB3A RAPGEF2 RBM48 RGNEF RNF14 RPL7A RUNDC3A SCD
RAB3C RAPGEF3 RBM4B RGP1 RNF144A RPL7L1 RUNDC3B SCFD1
RAB3IL1 RAPGEF4 RBM6 RGR RNF145 RPL8 RUNX1 SCG2
RAB3IP RAPGEF6 RBM7 RGS1 RNF148 RPL9 RUNX2 SCG3

RAB40B RAPGEFL1 RBM8A RGS10 RNF149 RPLP1 RUNX3 SCG5
SCGB2A1 SEPT10 SGPP1 SLC12A7 SLC26A6 SLC4A3 SMAP2 SNX14
SCGB2A2 SEPT11 SGPP2 SLC13A2 SLC26A9 SLC4A4 SMARCA1 SNX18
SCHIP1 SEPT12 SGSM1 SLC13A3 SLC27A1 SLC4A5 SMARCA2 SNX2
SCIMP SEPT3 SGSM2 SLC13A4 SLC27A2 SLC4A8 SMARCA4 SNX24
SCLT1 SEPT4 SGSM3 SLC13A5 SLC27A3 SLC4A9 SMARCAL1 SNX25
SCMH1 SEPT5 SGTA SLC15A2 SLC27A6 SLC50A1 SMARCB1 SNX29
SCML1 SEPT6 SH2B1 SLC16A1 SLC29A2 SLC5A10 SMARCC1 SNX30
SCML2 SEPT8 SH2D3C SLC16A12 SLC2A13 SLC5A12 SMARCC2 SNX32
SCN10A SEPT9 SH2D5 SLC16A13 SLC2A14 SLC5A2 SMARCD1 SNX33
SCN11A SERBP1 SH3BP2 SLC16A14 SLC2A2 SLC5A3 SMARCD2 SNX4
SCN1A SERINC3 SH3BP4 SLC16A3 SLC2A4 SLC5A5 SMARCD3 SNX5
SCN2A SERINC5 SH3BP5L SLC16A5 SLC2A5 SLC5A6 SMC1A SNX6
SCN2B SERPINA3 SH3D19 SLC16A6 SLC2A6 SLC6A1 SMC2 SNX9
SCN3A SERPINA4 SH3GL2 SLC17A6 SLC30A1 SLC6A11 SMC3 SOAT1
SCN3B SERPINA6 SH3GLB1 SLC17A7 SLC30A10 SLC6A12 SMC5 SOCS4
SCN4A SERPINB13 SH3KBP1 SLC18A2 SLC30A3 SLC6A13 SMC6 SOD1
SCN4B SERPINB4 SH3PXD2A SLC18A3 SLC30A4 SLC6A14 SMCHD1
SOGA2
SCN5A SERPINB6 SH3PXD2B SLC18B1 SLC30A5 SLC6A16 SMCR7
SOGA3
SCN8A SERPINB7 SH3RF1 SLC19A1 SLC30A7 SLC6A17 SMCR8 SON
SCO1 SERPINB9 SH3RF2 SLC19A2 SLC31A2 SLC6A19 SMEK1 SORBS2
SCO2 SERPINE1 SHC1 SLC19A3 SLC34A1 SLC6A2 SMEK2 SORCS2
SCP2 SERPINE2 SHC4 SLC1A1 SLC35A2 SLC6A20 SMG1 SORL1
SCPEP1 SERPINF1 SHCBP1 SLC1A2 SLC35B1 SLC6A3 SMG6 SOWAHB
SCRIB SERPINF2 SHE SLC1A3 SLC35B2 SLC6A4 SMG7 SOX10
SCUBE1 SERPING1 SHISA3 SLC1A4 SLC35B3 SLC6A6 SMG8 SOX14
SCUBE2 SERPINH1 SHISA5 SLC1A6 SLC35B4 SLC6A7 SMG9 SOX2
SCUBE3 SERTAD2 SHMT2 SLC1A7 SLC35C2 SLC6A8 SMO SOX7
SDC4 SERTAD3 SHOC2 SLC20A1 SLC35D1 SLC6A9 SMOC2 SOX9
SDCBP SESN1 SHOX2 SLC20A2 SLC35D2 SLC7A1 SMPD2 SP2
SDCCAG8 SESN2 SHPK SLC22A13 SLC35E1 SLC7A10 SMPD3 SP3
SDF4 SESN3 SHROOM1 SLC22A17 SLC35E2B SLC7A3 SMPD4 SP4
SDHA SESTD1 SIAH2 SLC22A2 SLC35E3 SLC7A7 SMS SP7
SDK1 SETD1A SIAH3 SLC22A23 SLC35E4 SLC7A8 SMU1 SPACA1
SDK2 SETD4 SIGLEC1 SLC22A3 SLC35F3 SLC7A9 SMURF1 SPAG5
SDS SETD5 SIGLEC10 SLC22A4 SLC35G1 SLC9A1 SMURF2 SPAST
SEC13 SETD9 SIGLEC7 SLC22A6 SLC35G2 SLC9A2 SMYD1 SPATA12
SEC14L1 SETDB1 SIK1 SLC22A8 SLC35G3 SLC9A3 SMYD3 SPATA17
SEC14L5 SETDB2 SIK2 SLC23A3 SLC36A1 SLC9A5 SNAI2 SPATA19
SEC16A SETX SIKE1 SLC24A1 SLC37A1 SLC9A6 SNAP23 SPATA2
SEC22A SEZ6L SIM1 SLC24A2 SLC37A3 SLC9A7 SNAP91 SPATA20
SEC22C SF1 SIM2 SLC24A5 SLC38A1 SLC9A8 SNAPC3 SPATA5

SEC23A SF3A1 SIN3A SLC25A1 SLC38A5 SLC9A9 SNAPC5 SPATA5L1
SEC24B SF3A3 SIN3B SLC25A10 SLC38A6 SLCO1A2 SNCA SPATA6
SEC61A1 SF3B4 SIPA1 SLC25A11 SLC38A7 SLCO2B1 SNCB SPATS2
SEC61A2 SFMBT1 SIPA1L1 SLC25A12 SLC39A1 SLCO3A1 SNED1 SPDYC
SEC61B SFMBT2 SIPA1L2 SLC25A14 SLC39A10 SLCO4A1 SNF8 SPECC1
SEC63 SFN SIPA1L3 SLC25A16 SLC39A14 SLCO4C1 SNN SPECC1L
SECISBP2 SFR1 SIRPA SLC25A18 SLC39A5 SLIT1 SNPH SPEG
SEH1L SFRP1 SIRT4 SLC25A19 SLC39A6 SLIT2 SNRK SPEN
SEL1L3 SFRP2 SIRT7 SLC25A20 SLC3A1 SLIT3 SNRNP200 SPG11
SELENBP1 SFRP4 SIX1 SLC25A24 SLC3A2 SLITRK1 SNRNP25 SPG20
SELM SFT2D2 SKIL SLC25A25 SLC41A2 SLITRK2 SNRNP35 SPG21
SELO SFXN2 SKP2 SLC25A26 SLC43A1 SLITRK3 SNRNP40 SPHK1
SELT SFXN4 SLA SLC25A28 SLC43A3 SLITRK4 SNRNP48 SPHKAP
SEMA3C SFXN5 SLAIN1 SLC25A3 SLC44A1 SLITRK5 SNRPA SPI1
SEMA3D SGCA SLAIN2 SLC25A30 SLC44A5 SLITRK6 SNRPB SPIB
SEMA3E SGCB SLAMF6 SLC25A31 SLC45A1 SLK SNRPB2 SPIC
SEMA3F SGCD SLC10A1 SLC25A33 SLC45A3 SLMO1 SNRPD2 SPIN3
SEMA3G SGCE SLC10A2 SLC25A34 SLC45A4 SLPI SNRPN SPIRE2
SEMA4C SGCG SLC10A3 SLC25A35 SLC47A1 SLU7 SNTA1 SPN
SEMA5A SGcz SLC11A1 SLC25A36 SLC47A2 SMAD1 SNTB1 SPNS1
SEMA5B SGK1 SLC12A1 SLC25A42 SLC4A1 SMAD2 SNW1 SPNS3
SEMA6A SGK2 SLC12A2 SLC25A44 SLC4A10 SMAD3 SNX1 SPOCK1
SEMA6D SGK3 SLC12A3 SLC25A51 SLC4A11 SMAD4 SNX10 SPOCK2
SENP2 SGMS1 SLC12A4 SLC25A6 SLC4A1AP SMAD9 SNX11 SPPOP
SEPHS1 SGPL1 SLC12A5 SLC26A5 SLC4A2 SMAP1 SNX13 SPP2
SPPL3 SSU72 STRADB SYNJ2 TBC1D15 TESC TIE1 TMEM136
SPRED1 SSX2IP STRAP SYNPO TBC1D16 TESK1 TIFAB TMEM138
SPRED2 ST14 STRBP SYNPO2 TBC1D17 TET3 TIGD2 TMEM143
SPRTN ST18 STRN SYNPO2L TBC1D19 TEX10 TIGD3 TMEM145
SPRY1 ST3GAL2 STRN3 SYNPR TBC1D20 TEX11 TIGD7 TMEM147
SPRY3 ST3GAL3 STRN4 SYNRG TBC1D21 TEX14 TIGIT TMEM14B
SPRYD3 ST3GAL5 STS SYP TBC1D22B TEX261 TIMM17B TMEM150A
SPRYD7 ST3GAL6 STT3A SYPL1 TBC1D23 TEX29 TIMM21 TMEM150B
SPSB1 ST5 STT3B SYT1 TBC1D7 TEX30 TIMM8B TMEM161A
SPSB2 ST6GAL1 STUB1 SYT10 TBC1D9B TEX34 TIMP2 TMEM163
SPSB3 ST6GALNAC3 STX10 SYT11 TBCB TEX9 TIMP3 TMEM164
SPTAN1 ST6GALNAC4 STX12 SYT12 TBCD TF TIMP4 TMEM165
SPTB ST6GALNAC5 STX16 SYT13 TBCEL TFAP2A TINAGL1 TMEM168
SPTBN1 ST7 STX1A SYT14 TBK1 TFAP2B TIPARP TMEM169
SPTBN2 ST8SIA2 STX1B SYT17 TBL1X TFAP2C TJP1 TMEM170B
SPTBN4 ST8SIA3 STX4 SYT2 TBL1XR1 TFB2M TK1 TMEM179
SPTLC1 ST8SIA4 STX5 SYT3 TBL3 TFCP2 TK2 TMEM179B
SPTLC2 ST8SIA5 STX6 SYT6 TBP TFCP2L1 TKT TMEM180
SPTLC3 STAB1 STX7 SYT7 TBR1 TFDP2 TKTL1 TMEM182

SPTSSA STAB2 STXBP1 SYT9 TBRG1 TFEB TKTL2 TMEM184A
SQLE STAC STXBP3 SYTL1 TBRG4 TFG TLE1 TMEM184B
SQSTM1 STAC3 STXBP5 SYVN1 TBX10 TFIP11 TLE4 TMEM184C
SRC STAG2 STXBP5L SZRD1 TBX18 TFPT TLE6 TMEM186
SRCAP STAM STXBP6 TAAR1 TBX20 TFR2 TLK1 TMEM189
SRD5A1 STAM2 STYX TAB1 TBX21 TFRC TLK2 TMEM189-UBE2V1
SREBF1 STAP1 SUCLA2 TAB3 TBX22 TGDS TLN1 TMEM190
SREBF2 STAP2 SUCLG1 TACC1 TBX3 TGFA TLN2 TMEM192
SREK1 STAR SUCNR1 TACO1 TBX4 TGFB1 TLR2 TMEM2
SRF STARD10 SUCO TACR1 TBX5 TGFB1I1 TLR3 TMEM200A
SRGAP1 STARD13 SUDS3 TACR3 TBXA2R TGFB3 TLR7 TMEM203
SRGAP3 STARD3 SUFU TADA1 TC2N TGFBI TLR8 TMEM204
SRM STARD3NL SUGT1 TADA2B TCEA2 TGFBR1 TLR9 TMEM206
SRP19 STARD4 SULF1 TADA3 TCEA3 TGIF2 TLX3 TMEM208
SRP54 STAT1 SULF2 TAF1 TCEAL6 TGM1 TM4SF1 TMEM214
SRP72 STAT2 SULT1A1 TAF13 TCERG1 TGM2 TM4SF18 TMEM219
SRP9 STAT3 SULT1E1 TAF15 TCF12 TGM3 TM4SF4 TMEM222
SRPK1 STAT4 SULT4A1 TAF1L TCF20 TH1L TM4SF5 TMEM245
SRPK2 STAT5A SUMF1 TAF5 TCF3 THAP1 TM6SF1 TMEM246
SRPK3 STAT5B SUN5 TAF6 TCF4 THAP10 TM7SF2 TMEM248
SRPR STAT6 SUPT16H TAF6L TCF7L2 THAP11 TM7SF3 TMEM26
SRR STAU1 SUPT5H TAF7 TCHP THAP4 TM9SF2 TMEM27
SRRD STC1 SUPT6H TAF8 TCP1 THAP5 TM9SF3 TMEM30A
SRRM2 STC2 SUPT7L TAF9 TCTN2 THBS1 TM9SF4 TMEM33
SRRT STEAP1 SURF1 TAF9B TDP1 THBS2 TMBIM6 TMEM38A
SRSF11 STEAP3 SURF4 TAGLN2 TDRD1 THBS3 TMC1 TMEM39A
SRSF2 STIM1 SUSD1 TALDO1 TDRD3 THBS4 TMCC1 TMEM39B
SRSF5 STIM2 SUSD4 TAMM41 TDRD6 THEG TMCO3 TMEM40
SRSF6 STIP1 SUV39H1 TANC1 TDRD7 THEMIS TMED1 TMEM45A
SRSF7 STK11 SUV39H2 TANC2 TDRD9 THEMIS2 TMED2 TMEM48
SRSF9 STK11IP SUZ12 TAOK1 TEAD1 THG1L TMED4 TMEM51
SS18 STK16 SV2A TAOK2 TEAD3 THOC1 TMED8 TMEM55B
SS18L1 STK24 SV2B TAOK3 TEC THOC2 TMEFF1 TMEM56
SSB STK3 SV2C TARBP1 TECPR1 THOC7 TMEFF2 TMEM57
SSBP1 STK32A SVIL TARBP2 TECPR2 THOP1 TMEM100 TMEM59
SSBP3 STK32B SVOPL TARDBP TECR THPO TMEM101 TMEM62
SSBP4 STK32C SWAP70 TARS2 TECTA THRA TMEM108 TMEM63C
SSH1 STK33 SYAP1 TARSL2 TECTB THRAP3 TMEM109 TMEM64
SSH2 STK38 SYBU TAS1R3 TEF THRB TMEM11 TMEM66
SSH3 STK38L SYCP3 TAS2R10 TEFM THSD7A TMEM115 TMEM72
SSR2 STK39 SYK TASP1 TEK THTPA TMEM117 TMEM74
SSR3 STK4 SYMPK TAT TEKT2 THUMPD1 TMEM120B TMEM74B
SSRP1 STMN3 SYN3 TAX1BP1 TENC1 THUMPD2 TMEM127 TMEM87A
SSTR1 STOML1 SYNCRIPI TBC1D10A TEPP THYN1 TMEM129 TMEM87B

SSTR2 STOML2 SYNE1 TBC1D10C TERF1 TIAL1 TMEM132A TMEM8B
SSTR4 STON2 SYNGR1 TBC1D13 TERF2IP TIAM2 TMEM132B TMEM8C
SSTR5 STOX2 SYNGR2 TBC1D14 TES TICAM1 TMEM132D TMEM9
TMEM91 TOP2B TRIM17 TSN TUFM UBXN4 USP31 VPS37B
TMEM95 TOP3B TRIM21 TSNAZIP1 TULP4 UBXN6 USP32 VPS39
TMEM97 TOPBP1 TRIM23 TSPAN11 TUSC3 UBXN7 USP33 VPS41
TMEM9B TOR1A TRIM24 TSPAN12 TUT1 UCHL1 USP34 VPS45
TMF1 TOR1AIP2 TRIM25 TSPAN13 TWF2 UCHL3 USP38 VPS4A
TMIE TOR1B TRIM26 TSPAN14 TWISTNB UCK2 USP39 VPS51
TMIGD1 TOR3A TRIM27 TSPAN17 TWSG1 UCKL1 USP4 VPS52
TMLHE TOX TRIM29 TSPAN3 TXK UCN2 USP43 VPS53
TMOD1 TOX2 TRIM3 TSPAN33 TXLNA UEVLD USP44 VPS54
TMOD2 TOX4 TRIM32 TSPAN5 TXN2 UFC1 USP46 VPS8
TMOD3 TP53 TRIM33 TSPAN6 TXNDC15 UFL1 USP47 VRK1
TMPPE TP53BP1 TRIM35 TSPAN7 TXNL4A UGCG USP48 VRTN
TMPRSS11B TP53I11 TRIM37 TSPAN9 TXNRD1 UGDH USP5 VSIG1
TMPRSS11D TP53I13 TRIM41 TSPYL2 TYK2 UGT1A1 USP51 VSTM2A
TMPRSS3 TP53INP1 TRIM44 TSSK1B TYMS UGT1A10 USP6NL VSTM4
TMTc1 TP73 TRIM46 TSSK2 TYRO3 UGT1A6 USP7 VSX2
TMTc4 TPBG TRIM50 TSSK3 TYRP1 UGT1A7 USP9X VTI1A
TMUB2 TPCN2 TRIM52 TSSK6 TYW1 UGT1A8 UST VTI1B
TMX1 TPH1 TRIM54 TTBK1 TYW5 UGT1A9 UTRN VTN
TMX2 TPH2 TRIM55 TTBK2 U2AF1 UGT2B7 UVRAG VWCE
TNC TPI1 TRIM62 TTC12 U2AF1L4 UHMK1 UXS1 WAC
TNFAIP3 TPM1 TRIM63 TTC13 UACA UHRF2 UXT WAPAL
TNFAIP8 TPM2 TRIM8 TTC17 UAP1L1 ULK1 VAMP2 WARS
TNFAIP8L2 TPM4 TRIM9 TTC23 UBA1 ULK3 VAMP3 WASL
TNFRSF11B TPO TRIO TTC26 UBA2 UMOD VAMP5 WBP1
TNFRSF12A TPPP TRIP10 TTC27 UBA3 UMPS VANGL1 WBP11
TNFRSF6B TPPP3 TRIP12 TTC32 UBAC2 UNC119 VARS WBP4
TNFSF11 TPR TRIP13 TTC39A UBAP1 UNC13A VASH1 WBSCR17
TNFSF12 TPRG1L TRMT1 TTC39B UBASH3A UNC13B VASP WBSCR22
TNFSF12-TNFSF13 TPRKB TRMT11 TTC39C UBASH3B UNC13C VAT1
WDFY1
TNFSF14 TPRX1 TRMT112 TTC7B UBB UNC5A VAT1L WDFY3
TNFSF18 TPST1 TRMT1L TTC9C UBC UNC5B VAV1 WDR1
TNFSF4 TPST2 TRMT2B TTI1 UBE2C UNC5C VAV2 WDR12
TNFSF8 TPT1 TRMT6 TTK UBE2E2 UNC5D VAV3 WDR13
TNIK TPTE2 TRMT61A TTL UBE2F UNG VAX1 WDR16
TNIP2 TPX2 TRMU TTLL1 UBE2G2 UPB1 VCAN WDR20
TNK2 TRA2A TRNAU1AP TTLL11 UBE2I UPF1 VCL WDR24
TNKS TRA2B TRNT1 TTLL12 UBE2J1 UPF2 VCP WDR26
TNKS2 TRABD TROVE2 TTLL13 UBE2J2 UPP1 VCPIP1 WDR3
TNMD TRAF2 TRPC1 TTLL3 UBE2L6 UPP2 VDAC1 WDR33

TNNC1 TRAF3 TRPC3 TTLL6 UBE2O UPRT VDAC2 WDR34
TNNI2 TRAF4 TRPC4 TTLL7 UBE2Q1 UQCRC1 VDR WDR43
TNNI3K TRAF6 TRPC4AP TTPA UBE2R2 UQCRCFS1 VEZF1 WDR44
TNNT1 TRAF7 TRPC5 TTYH3 UBE2S URB2 VGLL1 WDR45
TNNT2 TRAFD1 TRPC6 TUB UBE2T URGCP VILL WDR45L
TNNT3 TRAIP TRPC7 TUBA1A UBE2Z URI1 VIM WDR46
TNPO1 TRAK1 TRPM5 TUBA1B UBE3A URM1 VIMP WDR47
TNPO3 TRAK2 TRPM6 TUBA1C UBE3B UROS VIP WDR48
TNR TRAM1 TRPM7 TUBA3C UBE4B USE1 VIPAS39 WDR5
TNRC6B TRAM1L1 TRPM8 TUBA3D UBIAD1 USF1 VIPR1 WDR53
TNRC6C TRAM2 TRPS1 TUBA4A UBL4A USH1C VIPR2 WDR59
TNS4 TRANK1 TRPV2 TUBA8 UBL7 USHBP1 VIT WDR6
TOB1 TRAP1 TRPV4 TUBAL3UBLCP1 USP1 VMP1 WDR61
TOB2 TRAPPC1 TRPV6 TUBB UBN1 USP10 VPS13A WDR63
TOE1 TRAPPC10 TRRAP TUBB2A UBN2 USP12 VPS13B WDR7
TOLLIP TRAPPC12 TSC2 TUBB3 UBOX5 USP14 VPS13C WDR70
TOM1 TRAPPC2L TSC22D1 TUBB4A UBP1 USP16 VPS13D WDR74
TOM1L1 TRAPPC4 TSC22D3 TUBB6 UBQLN1 USP19 VPS16 WDR75
TOMM22 TRAPPC6B TSEN2 TUBD1 UBQLN2 USP20 VPS18 WDR77
TOMM34 TRAPPC9 TSG101 TUBG1 UBR1 USP21 VPS26A WDR81
TOMM40 TREX1 TSGA10 TUBG2 UBR2 USP22 VPS26B WDR83
TOMM40L TRHR TSGA13 TUBGCP2 UBR4 USP24 VPS28 WDR89
TOMM70A TRIB2 TSHZ1 TUBGCP3 UBR5 USP25 VPS33B WDR91
TONSL TRIM11 TSHZ3 TUBGCP4 UBTD1 USP28 VPS35 WDR96
TOP1 TRIM13 TSKS TUBGCP5 UBXN10 USP3 VPS36 WDTC1
TOP1MT TRIM16L TSLP TUBGCP6 UBXN2B USP30 VPS37A WEE2
WFDC2 XPO5 ZBTB39 ZFP1 ZNF19 ZNF35 ZNF555 ZNF74
WFDC6 XPO6 ZBTB4 ZFP106 ZNF2 ZNF354A ZNF557 ZNF746
WFIKKN2 XPOT ZBTB40 ZFP161 ZNF202 ZNF383 ZNF560 ZNF75A
WFS1 XPR1 ZBTB41 ZFP3 ZNF205 ZNF384 ZNF563 ZNF75D
WHSC1 XRCC5 ZBTB43 ZFP36 ZNF207 ZNF385A ZNF564 ZNF766
WHSC1L1 XRCC6 ZBTB44 ZFP36L1 ZNF213 ZNF391 ZNF567 ZNF768
WIBG XRN1 ZBTB45 ZFP82 ZNF217 ZNF394 ZNF569 ZNF770
WIF1 XRN2 ZBTB46 ZFP90 ZNF22 ZNF395 ZNF570 ZNF773
WIPF1 XYLT2 ZBTB5 ZFR ZNF226 ZNF397 ZNF574 ZNF776
WIZ YAP1 ZBTB6 ZFX ZNF23 ZNF398 ZNF575 ZNF786
WLS YARS ZBTB7B ZFYVE1 ZNF232 ZNF410 ZNF581 ZNF791
WNK3 YARS2 ZBTB7C ZFYVE20 ZNF236 ZNF416 ZNF582 ZNF8
WNT10B YBX2 ZBTB8A ZFYVE26 ZNF238 ZNF423 ZNF583 ZNF814
WNT11 YEATS4 ZC3H10 ZFYVE28 ZNF239 ZNF425 ZNF584 ZNF816-
ZNF321P
WNT2 YES1 ZC3H11A ZGPAT ZNF248 ZNF426 ZNF589 ZNF821
WNT2B YIF1A ZC3H12B ZHX1 ZNF25 ZNF432 ZNF592 ZNF827
WNT3 YIF1B ZC3H12C ZHX1-C8ORF76 ZNF250 ZNF433 ZNF606 ZNF829

WNT3A YIPF1 ZC3H13 ZHX3 ZNF251 ZNF434 ZNF607 ZNF836
 WNT4 YIPF2 ZC3H14 ZIC1 ZNF253 ZNF436 ZNF608 ZNF862
 WNT5A YIPF3 ZC3H15 ZIK1 ZNF260 ZNF445 ZNF609 ZNFX1
 WNT5B YIPF6 ZC3H18 ZKSCAN1 ZNF263 ZNF449 ZNF618 ZNHIT1
 WNT7A YIPF7 ZC3H4 ZKSCAN4 ZNF266 ZNF451 ZNF619 ZNHIT2
 WNT8A YJEFN3 ZC3H6 ZKSCAN5 ZNF276 ZNF460 ZNF621 ZNHIT3
 WNT8B YKT6 ZC3H7A ZMAT3 ZNF280C ZNF461 ZNF622 ZNHIT6
 WNT9A YLPM1 ZC3H7B ZMIZ1 ZNF280D ZNF462 ZNF624 ZNRF3
 WRAP53 YME1L1 ZCCHC11 ZMIZ2 ZNF281 ZNF48 ZNF627 ZP2
 WRAP73 YOD1 ZCCHC12 ZMPSTE24 ZNF282 ZNF480 ZNF638 ZP4
 WSB1 YTHDC1 ZCCHC14 ZMYM2 ZNF287 ZNF483 ZNF644 ZPBP
 WSCD1 YTHDC2 ZCCHC6 ZMYM3 ZNF295 ZNF485 ZNF652 ZPLD1
 WSCD2 YTHDF1 ZCCHC7 ZMYM6 ZNF296 ZNF490 ZNF653 ZRANB1
 WT1 YWHAE ZDHHC1 ZMYND11 ZNF3 ZNF493 ZNF668 ZRANB2
 WTAP YWHAG ZDHHC15 ZMYND19 ZNF300 ZNF496 ZNF670
ZSCAN10
 WWC1 YWHAH ZDHHC17 ZMYND8 ZNF318 ZNF498 ZNF671 ZSCAN2
 WWP1 YWHAQ ZDHHC18 ZNF10 ZNF319 ZNF500 ZNF677 ZSCAN20
 WWP2 YY2 ZDHHC21 ZNF114 ZNF32 ZNF511 ZNF687 ZSCAN21
 XAB2 ZADH2 ZDHHC22 ZNF12 ZNF320 ZNF512 ZNF688 ZSCAN22
 XBP1 ZBED1 ZDHHC23 ZNF121 ZNF324 ZNF512B ZNF689 ZSCAN29
 XK ZBED4 ZDHHC5 ZNF124 ZNF324B ZNF513 ZNF691 ZSCAN4
 XKR4 ZBTB1 ZDHHC7 ZNF132 ZNF326 ZNF514 ZNF692 ZSCAN5A
 XKR6 ZBTB10 ZDHHC9 ZNF134 ZNF329 ZNF521 ZNF699 ZSWIM3
 XKR7 ZBTB11 ZEB2 ZNF136 ZNF331 ZNF526 ZNF70 ZSWIM4
 XKR8 ZBTB16 ZFAND1 ZNF142 ZNF33B ZNF527 ZNF704 ZSWIM5
 XPNPEP1 ZBTB17 ZFAND5 ZNF143 ZNF34 ZNF529 ZNF709 ZSWIM7
 XPNPEP2 ZBTB24 ZFAT ZNF148 ZNF345 ZNF532 ZNF710 ZXDC
 XPNPEP3 ZBTB25 ZFC3H1 ZNF17 ZNF346 ZNF536 ZNF711 ZZEF1
 XPO1 ZBTB26 ZFHX3 ZNF174 ZNF347 ZNF547 ZNF713 ZZZ3
 XPO4 ZBTB34 ZFHX4 ZNF18

9.11 Attachment 11

ABCC8 CARF EPS8L1 IGLON5 MIR124-2HG PLCB4 SASH1 SYT13
 ABCD2 CARMIL2 ERC1 INPP5F MIR22HG PLEKHA1 SCAMP5 SYT16
 ABCG4 CARMIL3 ERC2 IQSEC1 MIR3936HG PLK5 SCARA3 SYT3
 ABLIM2 CASKIN1 ETNK2 IRAK2 MMD PLPPR2 SCML1 SYT5
 ACKR1 CCDC149 FADS3 IRX4 MMP24 PLPPR4 SCN2B SYT7
 ACKR3 CCDC184 FAHD2B JAG2 MOAP1 PLXDC1 SCN3B TACR3
 ACOT7 CCDC186 FAM102B JAK1 MPP1 PLXNC1 SCN8A TATDN2P2
 ACTL6B CDH10 FAM131C JAKMIP1 MPP3 PNCK SCN9A TBC1D24

ACYP1 CDH12 FAM172A JAKMIP3 MTMR6 PNMA2 SELENBP1 TBC1D9
ADAM11 CDH9 FAM196A JPH3 MTND6P21 PNMA8B SEMA4F TCAF1
ADCY5 CDK5R2 FAM43B KCNA2 MTURN PPP1R13B SEMA5A TCAF1P1
ADD2 CDKN2D FAM57B KCNA3 MYT1L PPP1R16B SEMA7A TDRD7
ADD3 CELF5 FAXDC2 KCNA4 N4BP3 PPP1R1A SESTD1 TFPI
ADGRB1 CEP126 FBXL16 KCNAB2 NANOS1 PPP1R3B SEZ6L2 TFR2
ADGRL1 CEP170B FBXL2 KCNB1 NBEAL2 PRKACA SH3GL3 TIAF1
ADGRV1 CGREF1 FBXL4 KCNC1 NCAM2 PRKCE SH3KBP1 TMCC2
ADRA1B CHD5 FBXO2 KCNC3 NCOA3 PRKCZ SH3PXD2A TMEM107
ADRA2A CHGA FBXO27 KCNC4 NCS1 PRKN SH3RF1 TMEM130
AGAP1 CHRNA4 FBXO41 KCNH7 NDRG4 PRNP SHANK1 TMEM143
AGAP2 CKMT1A FBXO44 KCNJ11 NECAB2 PRRT2 SHANK3 TMEM145
AHNAK2 CKMT1B FGF12 KCNK3 NECTIN1 PRRT3 SIPA1 TMEM151A
ALDOC CLSTN3 FGF19 KCNN1 NEDD4L PRRT4 SLC12A5 TMEM169
AMER1 CNTN1 FHOD3 KCNQ3 NEK7 PSD SLC23A2 TMEM229B
AMER3 CORO6 FLRT3 KHK NENF PSD2 SLC24A3 TMEM246
ANK2 CPEB4 FMNL1 KIAA0513 NEURL1 PSD3 SLC25A27 TMEM44
ANKRD24 CPLX1 FNBP1P1 KIAA1024 NFASC PTGDS SLC29A1

TMEM54

AP3B2 CPNE5 FNDC10 KIAA1549L NHS PTGR2 SLC2A10 TMEM74B
AQP11 CPNE7 FOXC2 KIF16B NIPAL3 PTPN5 SLC2A12 TMOD2
ARHGAP20 CRB1 FRS3 KIF21B NMNAT3 PTPRN SLC36A1 TPPP
ARHGAP33 CREG2 FRY KLF3-AS1 NOVA1 PTPRN2 SLC38A3 TRIM3
ARHGDIG CTSH FSTL5 KLF9 NPC1 RAB11FIP2 SLC46A3 TRUB1
ARL2 CYGB FXYD7 KLHL15 NPIPA5 RAB11FIP4 SLC4A10 TSPOAP1
ATG16L2 DAB1 GABBR1 KLHL32 NPM2 RAB11FIP5 SLC6A17 TTBK1
ATP1A3 DACH2 GABRG2 KNDC1 NRG2 RAB26 SLC8A3 TTC14
ATP2B2 DACT3 GALNT14 L1CAM NRSN1 RAB39A SLC9A3R2 TTC9B
ATP2B3 DFNB59 GALNT9 L3MBTL1 NRXN2 RAB3A SLC9A5 TUBB2A
ATRNL1 DHRS13 GDAP1L1 LDB2 NTNG1 RAB3B SLC9A9 TUBB4A
B4GALNT1 DISP2 GDPD5 LGI3 NUTM2D RAB3IL1 SLTRK6 UBAP1L
BCO2 DLGAP3 GLS2 LHFPL4 NYAP2 RAB6B SMAP1 UBE2Q2P1
BEX5 DMTN GNAL LIN7B OGDHL RABGAP1L SMIM10L2A UNC13A
BICD1 DNAJC12 GNB5 LINC00265 OLFM1 RALGPS1 SMIM10L2B

UNC5D

BICDL1 DNAJC6 GNG4 LINC02035 OPRL1 RAP1GAP SMKR1 UTRN
BMP2K DNM1 GNPTAB LINC02298 OR1F1 RAP1GAP2 SMPD3 VAMP1
BNIP3 DOC2A GOLGA7B LINGO1 PACS1 RAPH1 SNCA VEGFA
BOLA3-AS1 DTX1 GPRIN1 LMNTD2 PACSIN1 RASGRP2 SNCB VEGFB
C15orf59 DUSP23 GPRIN2 LMTK2 PAG1 RASL11A SNCG VSTM2B
C19orf66 DUSP5P1 GPRIN3 LRFN1 PAIP2B RASSF10 SNORA47 VWA5B2
C1QTNF4 ECE1 GREB1 LRRC3 PALM3 RASSF7 SNORD115-20 WASF3
C2orf27A EEF1A2 GRIK3 LRRTM3 PANX2 RBFOX3 SNPH WIPF3
C2orf72 EEF1AKMT2 GRIK5 LYNX1 PCDH1 RBM11 SNX10 WIPI1

C7orf31 EGFL7 H2AFJ LYSMD4 PCDHB13 RCAN2 SPACA9 WSCD2
C9orf172 EHHADH HCN2 MAGEE1 PCDHB15 REXO5 SPAG4 YBX2
CABLES2 EIF5A2 HCN3 MAOA PDE10A RGS2 SPHKAP ZBTB38
CABP1 ELAVL3 HIVEP2 MAP3K13 PDE1B RGS7BP SPRN ZBTB43
CACNA1B ELAVL4 HOXC5 MAP4K2 PDE2A RIC3 SPTBN4 ZCCHC12
CACNA1E ELFN2 HPCAL4 MAPK8IP2 PDIA2 RIMS4 SSPO ZDHHC4
CACNA2D3 ELMOD1 HPX MAPK9 PDXK RN7SL268P STAC2 ZFP28
CACNG2 EMID1 HRASLS MAPRE3 PDZD7 RNF180 STX1A ZFR2
CADM3 ENDOD1 HRH1 MAST1 PEBP1 RNF208 STX1B ZHX3
CADM3-AS1 ENO2 HRH3 MAST3 PEG10 RNU4-39P STX3 ZNF114
CALY ENPP5 HS6ST3 MAST4 PFKP RNY3P15 SWAP70 ZNF221
CAMK2B EPB41L2 HSPA12A MCF2L PID1 RP9P SYN1 ZNF382
CAMK2N1 EPB41L4B HTRA3 MCF2L-AS1 PIGZ RPL12P47 SYN2 ZNF467
CAMK2N2 EPHA10 IDS METTL7A PIK3CD RRN3P1 SYNGR1 ZNF540
CAMTA1 EPHA5 IFI44 MGAT3 PINK1 RSRC1 SYNGR3 ZNF547
CAPS2 EPHB6 IFNGR2 MGAT4A PIP5KL1 RUNDC3A-AS1 SYP ZNF699
ZNF774

9.12 Attachment 12

AAK1 CFH FAXC IDI1 MIR1273F PIK3CB RUSC1 SYT1
ABLIM3 CHRM2 FBN3 INA MMAB PJA2 SAMD12 SYT14
ACAP2-IT1 CISD1 FBXO25 ITGB1BP1 MRAS PKIA SCAI TAP2
ACSL3 CIT FCHSD2 JAZF1 MTMR7 PLCXD2 SCD TAS2R4
ACSL4 CLASP2 FDFT1 KANTR MTSS1 PLPPR5 SCN1A TBC1D30
ACSL6 CLCN4 FDPS KBTBD11 MVD POLE SCN2A TBC1D8
ADAM22 CLVS2 FGD4 KCNAB1 MVK POMK SCN3A TCP11L1
ADAM23 CMBL FGF9 KCNB2 MYCBP2 POU2F1 SCUBE1 TENM1
ADCY1 CNNM1 FHDC1 KCND2 MYO5A POU2F2 SEMA4A TF
ADGRL3 CNRIP1 FRRS1L KCNK1 MYO9A PPFIA2 SENP7 TLN2
AK5 CNTNAP2 FUT11 KCNK10 NAALAD2 PPP1R12B SGIP1 TM6SF1
AKAP9 COX6A1 GABBR2 KCNS2 NALCN PPP1R9A SGK1 TMEM135
ALDH5A1 CPB2-AS1 GABRA3 KIAA1107 NAP1L2 PPP2R2B SH3BP5
TMEM14A
AMPH CSMD3 GABRB2 KIAA1841 NAP1L3 PPP2R2C SHANK2
TMEM178B
ANK1 CSRNP3 GABRB3 KIF3A NAPB PPP2R3A SHISA7 TMEM196
ANK3 CTNNA2 GABRG3 KIFAP3 NAV3 PREPL SHROOM2 TMEM35A
ANKRD12 CTSC GALNT13 KLHL1 NBEA PRKAA2 SKIL TMEM55A
ANKRD18EP CTSO GFOD1 KSR2 NBEAL1 PRKAR1B SLC16A14
TMEM65

ANKRD36C CTTNBP2 GFRA2 LETMD1 NCOA2 PRKAR2B SLC1A1
TRAM1L1
ANKRD46 CYCSP34 GLRA1 LGI2 NCR3LG1 PRKCB SLC1A6 TRIB2
ANKRD55 CYP27A1 GLT1D1 LHFPL2 NELL1 PRUNE1 SLC25A12
TRIM67
APBA1 CYP46A1 GNAO1 LHFPL3 NETO1 PSPC1-AS2 SLC25A40 TRIQK
APC DCX GNG2 LIN28B NEXMIF PTGR1 SLC2A13 TRMO
APH1B DDHD1 GNG3 LIN7C NFS1 PTPRR SLC2A6 TRMT2B
ARG2 DDX25 GNG8 LINC00599 NGLY1 PUS7L SLC30A4 TRPM3
ARMCX5-GPRASP2 DGKE GPC5 LINC00632 NLGN1 PWAR6 SLC43A2
TSPAN13
ARRB1 DHCR24 GPHN LINC00652 NMNAT2 PYGO1 SLC6A11 TSPAN7
ASTN1 DHCR7 GPI LINC01128 NRG3 RAB27B SLC6A1-AS1 TTC9
ASXL3 DNER GPLD1 LINC01184 NRG4 RAB3C SLC7A11 TTLL7
ATL1 DNM3 GPR12 LINC02293 NRXN1 RAB9B SLC8A1 UBASH3B
ATP2B1-AS1 DOCK3 GPR21 LINC02486 NSDHL RALYL SLC9A6 UCHL5
ATP6V1G2 DPY19L2P2 GPR22 LINGO2 NUDT13 RAMP2-AS1 SLTRK1
UNC80
B4GALT6 DSCAM GPR85 LPGAT1 NXT2 RANBP17 SLTRK3 VEPH1
BNIP3L DSTNP1 GPRASP1 LPIN2 ODF2L RANBP3L SLTRK4 VN1R1
BRI3BP DYNC1I1 GPRASP2 LRBA OLFM3 RASA2 SMTNL2 VSNL1
C11orf87 EEF1AKMT1 GRB14 LRFN5 ONECUT2 RASAL2 SNAI3 WDR17
C14orf28 EFNA5 GRIA2 LRP2BP OXCT1 RBFA SNAI3-AS1 WDR7
C1R ELOVL4 GRIA3 LRRC28 P4HA1 REEP1 SNAP25 WEE2-AS1
C1S ELOVL6 GRIA4 LRRC3B PAK3 RFK SNAP91 XPR1
C9orf40 ENHO GRIN1 LRRC4C PAK5 RGS17 SOS1-IT1 ZBTB25
CA10 ENTPD1 GRIN2B LSAMP PALM2 RIMKLA SRD5A1 ZC2HC1A
CACNA1A EPB41L1 GRM5 LSM11 PANK3 RIMKLB SREBF2 ZFHX4
CACNA1D EXOC6B GRM8 LSS PARP8 RIMS1 SRGAP2B ZNF117
CADM2 EXTL1 HCG18 LYST PCDH11X RIMS2 STARD4 ZNF157
CALN1 FAAH HCN4 MAGI2 PCDH17 RNF144A STAU2 ZNF204P
CAMK2G FAIM2 HGNC:18790 MAGI3 PCDH7 RNF152 STK39 ZNF215
CAMK4 FAM110B HGNC:24955 MANEAL PCDHB3 RNF207 STMN2
ZNF334
CCDC71L FAM126B HK2 MAP1A PCLO RNU2-38P STOML1 ZNF441
CCDC88A FAM133A HMGCLL1 MAP3K21 PCYT2 RNU6-1016P STOX2
ZNF443
CD109 FAM155A HMGCR MAP3K9 PDE1A RNU6-130P STRIP2 ZNF514
CD8A FAM169A HMGCS1 MAP9 PDP1 RNU6-137P STXBP1 ZNF555
CDH8 FAM171B HMGXB4 MAPRE2 PGAM1 RNU6-30P SV2A ZNF571
CDKL2 FAM19A2 HOMER2 MAPT PGK1 RPARP-AS1 SV2C ZNF577
CDKL5 FAM216A HOOK1 MBLAC2 PGM2L1 RTN1 SVOP ZNF793
CDR1 FAM228B HPRT1 MDGA2 PGRMC1 RUFY2 SYNE1 ZNF804A
CEND1 FAM66C IAH1 MGST2 PI4K2B RUNDC3B SYNPR ZSWIM5

9.13 Attachment 13

ACTR8 CPNE8 HNRNPC NT5C3A PSMC1 RPL23 RPS25 ST13
AK2 CS HNRNPL NUDC PSMC6 RPL24 RPS6 STK3
ANAPC15 DNAJB6 HSP90AB1 OGFR1 PSME2 RPL26L1 RPS8 STRADB
ANP32B DNAJC8 HSPE1 OLA1 PTGES3 RPL3 RTN4 STX12
ARPC3 DYNC1I2 ISCA1 PABPC1 PXN RPL31 RWDD1 SUMO1
ATP1B3 EEF1B2 KCTD9 PDAP1 RAC1 RPL34 SDHC TANK
ATP5E EIF2S2 LPCAT2 PDCL3 RAN RPL35 SET TCP1
ATP5F1 EIF2S3 LYPLA2 PDE8A RAP1B RPL5 SH3BP4 TECR
ATP5G2 EIF4H MAP1LC3B PHC1 RBMX2 RPL6 SKP1 TPI1
BAK1 ELL2 MARK1 PIGH RCN1 RPLP0 SMS TTC27
BZW1 FAM45A MBD2 PIK3CB REEP5 RPLP1 SMU1 TUBG1
CACYBP FBXL3 MED26 PIN4 RHEB RPS12 SNRPA1 UBE2M
CD58 GALNT18 MFSD9 PKIB RPL13A RPS13 SNRNPB2 UBE2S
CDC20 GAPDH MRPS10 POLR2J RPL18A RPS15 SNRPF UQCR11
CDC42 GCSH MRPS36 POLRMT RPL19 RPS15A SNX5 WWP1
CDK7 GLRX3 MTRF1L PPP1R15A RPL21 RPS16 SPCS2 YBX1
COX17 GNB4 NDUFA2 PRKAG2 RPL22 RPS2 SSB ZNF706
COX7C HNRNPA1

9.14 Attachment 14

A2M C1orf198 CNN1 FBLN5 LRRC59 PARVA RAB23 STT3A
ACOT2 C20orf27 CNN2 FITM2 MATN2 PDLIM7 RASSF2 SULT1E1
ACTA2 C6orf106 COL11A1 GDI2 MGAT1 PFN1 RASSF3 SURF4
ACTG2 CALD1 COL14A1 GFPT2 MID1 PKP2 RCAN1 TAGLN
ACTN1 CAV1 COL4A1 GLIPR1 MSRB3 PLD3 RPL22L1 TES
ADAM19 CAVIN1 COL4A2 GNPNAT1 MYH9 PLEKHG3 RPS4X TFRC
ADPGK CCDC117 COTL1 GOLPH3L MYL12B PLK2 RSU1 TGFB1I1
AHCY CCND3 CPD IER3 MYL9 PMP22 SCRT2 TLN1
AMOTL2 CCZ1 CSRP1 IFITM3 MYOF PODXL SEC31A TMED3
ANXA2 CD24P4 DCBLD1 IL13RA1 NEXN POLR3D SFT2D2 TPM2
APEX2 CDKN2B DDX50 ITGB1 NFKB1 POU4F1 SIX4 TPM4
ARF4 CDR2 DEPDC7 KDELC1 NOMO2 PPM1F SLC38A2 TRAK2
ARID5B CDR2L DHX37 KIRREL1 NOP14 PRPS2 SLC52A2 TRAM2
ARSB CHPF2 DISP1 LAT52 NRTN PTDSS1 SNTB2 TTC9C
B3GALNT2 CHRAC1 EIF4EBP1 LIMS1 NXPH4 PUM3 SRM UGDH

BCAT1 CKAP4 EIF5A LMAN1 OGFOD1 PXDN SSR3 VASP
C12orf45 CLIC4 FAM46B LPP OPA3 PYCR1

9.15 Attachment 15

ADO CCT4 FLNA LARS MT-CYB PROX2 SAP30 TOMM34
AEN CHIC2 GAR1 LEO1 MTFR1 PRR7-AS1 SARS TOP3B
AGO1 COQ8B GIPC1 LIAS MTHFD1L PTOV1-AS1 SEC14L2 TRMT61A
AIMP2 CTDP1 GTF2F2 LIMA1 MYBBP1A PTTG1IP SETD7 TUBD1
AJUBA CTNNAL1 HAX1 LINC00461 NACC2 PUS3 SH3GL1 TWNK
ARHGEF2 CTPS2 HDGF LINC00969 NFATC3 PWP2 SHMT2 UBTD1
ARMCX6 DAPK3 HEG1 LINC02060 NFKBIL1 QARS SHROOM3 UBXN2A
ATAD3A DDX31 HOXA2 LRWD1 NOB1 RAB9A SMARCC1 UCK2
ATP5G2 DHX34 HS1BP3 M6PR NOL6 RANBP10 SMIM26 UMPS
AUTS2 DHX40 HSPBAP1 MAT2B NPM1 RASA4 SPSB4 UROD
BOD1 DKC1 HSPH1 MCM7 NUP188 RASL11B SS18 USP25
BVES DPF2 IER5 MED22 ORC5 RBM42 SSNA1 VARS
BYSL DPH5 IFRD1 MED27 PDCL3 RCC1 STARD7-AS1 VCL
C12orf65 EED IFRD2 METTL12 PHC2 RCC2 SZRD1 VEZF1
C17orf97 EFL1 IGF2BP1 METTL13 PHF23 RGS16 TAF12 WDR46
C19orf53 EIF2S2 ILDR2 MEX3A PIP5K1A RNF114 TCF12 WWC1
C6orf48 EIF3E INTS4 MFSD2A PLEKHA3 RNF182 TMEM104 WWC2
CA14 EIF3L INTS7 MIR181A2HG POC1B RNF41 TMEM161B-AS1 XPO5
CA2 ENAH IPO8 MKRN2 POLRMT RPA2 TMEM187 XRCC1
CASC10 EPB41L4A-AS1 ITGA4 MOCS3 POP1 RPL7 TMEM200C ZBTB12
CBFB FAM50A JADE3 MRI1 PPM1D RPP25L TMEM206 ZCCHC2
CCDC77 FBL KATNA1 MRPS6 PRELID1 RPS6KB1 TMEM243 ZNF664
CCDC85C FERMT2 KLHL25 MTA2 PROSER1 SAMD4B TNFRSF12A
ZNF79
CCNJL

9.16 Attachment 16

ACTB CUL3 FGF6 IL3 MAPKAPK5 PRKACA PSMD3 RASGRP4
ACTG1 DAB2IP FGF7 IL3RA MARK3 PRKACB PSMD4 RBX1
ACTN2 DLG4 FGF8 IL5 MOV10 PRKACG PSMD5 RET
AGO1 DNAJB1 FGF9 IL5RA MYC PSMA1 PSMD6 RPS27A
AGO2 DUSP1 FGFR1 IL6 NCAM1 PSMA2 PSMD7 SEPT7
AGO3 DUSP10 FGFR2 IL6R NCOA3 PSMA3 PSMD8 SHC1
AGO4 DUSP16 FGFR3 IL6ST NEFL PSMA4 PSMD9 SHC2

AKAP9 DUSP2 FGFR4 IQGAP1 NF1 PSMA5 PSME1 SHC3
ANGPT1 DUSP4 FGG IRS1 NRAS PSMA6 PSME2 SOS1
APBB1IP DUSP5 FN1 IRS2 NRG1 PSMA7 PSME3 SPRED1
ARAF DUSP6 FOXO1 ITGA2B NRG2 PSMA8 PSME4 SPRED2
ARRB1 DUSP7 FOXO3 ITGB3 NRG3 PSMB1 PSMF1 SPRED3
ARRB2 DUSP8 FRS2 JAK1 NRG4 PSMB10 PSPN SPTA1
ARTN DUSP9 FRS3 JAK2 NRTN PSMB11 PTK2 SPTAN1
BRAF EGF FYN JAK3 PAK1 PSMB2 PTPN11 SPTB
BRAP EGFR GDNF JUN PAK2 PSMB3 PTPRA SPTBN1
BTC ERBB2 GFRA1 KALRN PAK3 PSMB4 RAC1 SPTBN2
CALM1 ERBB3 GFRA2 KBTBD7 PAQR3 PSMB5 RAF1 SPTBN4
CAMK2A ERBB4 GFRA3 KIT PDGFA PSMB6 RAG1 SPTBN5
CAMK2B EREG GFRA4 KITLG PDGFB PSMB7 RAG2 SRC
CAMK2D ETV4 GRB2 KL PDGFRA PSMB8 RAP1A SYNGAP1
CAMK2G FGA GRIN1 KLB PDGFRB PSMB9 RAP1B TEK
CCND3 FGB GRIN2A KRAS PEA15 PSMC1 RAPGEF2 TLN1
CDC14A FGF1 GRIN2B KSR1 PEBP1 PSMC2 RASA1 TNRC6A
CDC14B FGF10 GRIN2C KSR2 PHB PSMC3 RASA2 TNRC6B
CDC42 FGF16 GRIN2D LAMTOR2 PPP2CA PSMC4 RASA3 TNRC6C
CDC42EP2 FGF17 HBEGF LAMTOR3 PPP2CB PSMC5 RASA4 TYK2
CDC42EP3 FGF18 HRAS LAT PPP2R1A PSMC6 RASAL1 UBA52
CDC42EP5 FGF19 HSPB1 MAP2K1 PPP2R1B PSMD1 RASAL2 UBB
CDK1 FGF2 IGF2BP1 MAP2K2 PPP2R5A PSMD10 RASAL3 UBC
CNKSR1 FGF20 IL17RD MAP3K11 PPP2R5B PSMD11 RASGEF1A VCL
CNKSR2 FGF22 IL2 MAPK1 PPP2R5C PSMD12 RASGRF1 VWF
CSF2 FGF23 IL2RA MAPK3 PPP2R5D PSMD13 RASGRF2 WDR83
CSF2RA FGF3 IL2RB MAPK4 PPP2R5E PSMD14 RASGRP1 XPO1
CSF2RB FGF4 IL2RG MAPK6 PPP5C PSMD2 RASGRP3 YWHAB
CSK FGF5

9.17 Attachment 17

ADNP ASH1L CUL3 GRIN2B KMT5B POGZ SCN2A SYNGAP1
ANK2 ASXL3 DSCAM KATNAL2 MYT1L PTEN SETD5 TBR1
ARID1B CHD8 DYRK1A KMT2A NAA15 RELN SHANK3 TRIP12

9.18 Attachment 18

ANKRD11 CIC ERBIN IRF2BPL MECP2 NRXN1 SLC6A1 TNRC6B
BCKDK CNTN4 FOXP1 KAT2B MED13 PHF3 SMARCC2 UBN2
BCL11A CNTNAP2 GABRB3 KDM5B MED13L PTCHD1 SPAST UPF3B
CACNA1D CTNND2 GIGYF2 KDM6A MET RANBP17 SRCAP USP7
CACNA1H DDX3X GRIP1 KMT2C MSNP1AS RIMS1 SRSF11 WAC
CACNA2D3 DEAF1 ILF2 MAGEL2 NCKAP1 SCN9A TBL1XR1 WDFY3
CHD2 DIP2C INTS6 MBOAT7 NLGN3 SHANK2 TCF20

9.19 Attachment 19

ACHE CC2D1A DAPP1 GPHN LMX1B PAH RAB2A STXBP5
ADA CCT4 DENR GRIK2 LZTR1 PARD3B RBFOX1 TAF6
ADCY3 CDC42BPB DIP2A GRIK5 MACROD2 PARK2 ROBO2 TCF7L2
AGAP2 CEP41 DISC1 GRIN1 MBD5 PAX5 SBF1 TERF2
AHDC1 CGNL1 DNMT3A HECW2 MFRP PER2 SCN1A TMLHE
AKAP9 CHRNA7 DOCK8 HIVEP3 MTOR PHB SEMA5A TRIO
AMPD1 CIB2 DPP10 HMGN1 MYO5A PHF2 SETBP1 TRPC6
AMT CLASP1 DYNC1H1 ITGB3 MYO9B PHRF1 SETD2 TRPM1
ANXA1 CNKSR2 EFR3A KAT6A NAV2 PLCB1 SHANK1 UBE3A
APH1A CNOT3 EHMT1 KCNJ10 NINL PLXNA4 SHOX UBE3C
ASTN2 CNR1 ELAVL3 KCNQ2 NLGN4X PLXNB1 SLC12A5 UBR5
ATP10A CNTN5 ELP4 KCNQ3 NR3C2 PON1 SLC35B1 USP15
ATP2B2 CNTN6 EP400 KDM5C NRXN3 PREX1 SLC38A10 USP45
AUTS2 CNTNAP4 ETFB KDM6B NUAK1 PRICKLE1 SLC6A3 VIL1
AVPR1A CTCF FAM92B KIAA1586 OPHN1 PRICKLE2 SLC7A3 WWOX
C11orf30 CTNNB1 FBN1 KIAA2022 OR52M1 PRKCB SLC7A5 ZBTB20
CACNA1E CTTNBP2 FOXP2 KIRREL3 OXTR PRODH SPARCL1 ZC3H4
CACNB2 CUL7 GIGYF1 KMT2E P2RX5 PTK7 STXBP1 ZMYND11
CAPRIN1 CUX1 GPC4 LAMB1 P4HA2 PYHIN1

9.20 Attachment 20

ADNP C12orf57 DHCR7 IQSEC2 NSD1 PRODH SIN3A TCF4
ADSL CACNA1C DMD KAT6A NTNG1 PSMD12 SLC1A2 TRIP12
AHDC1 CDKL5 DMPK KMT2A OCRL PTEN SLC45A1 TSC1
AHI1 CHD2 DYRK1A MAGEL2 PACS1 PTPN11 SLC9A6 TSC2
ALDH5A1 CHD7 EHMT1 MBD5 PAX6 RAI1 SMARCA2 UBE3A

ALG6 CHD8 EP300 MBOAT7 PCCA RERE SNX14 UNC13A
ANKRD11 CHKB FMR1 MECP2 PCCB RNF135 SOX5 UPF3B
ARID1B CIC GATM MEF2C PCDH19 RPS6KA3 STAG1 VPS13B
ARX CNKSR2 HCN1 NACC1 PIK3R2 SATB2 STXBP1 WAC
ASXL3 CNOT3 HDAC4 NF1 POGZ SCN1A SYNE1 YY1
ATP1A3 CNTNAP2 HEPACAM NFIX PPM1D SETD5 SYNGAP1 ZBTB20
BCL11A DDX3X HERC2 NIPBL PPP2R5D SGSH TAF1 ZNF462
BRAF DEAF1 HOXA1 NR2F1 PRKD1 SHANK3

9.21 Attachment 21

ACHE CACNA1H DEAF1 GRIN2B MAGEL2 PARD3B RIMS1 STXBP5
ADA CACNA2D3 DENR GRIP1 MBD5 PARK2 ROBO2 SYNGAP1
ADCY3 CACNB2 DHCR7 HCN1 MBOAT7 PAX5 SBF1 TAF1
ADNP CAPRIN1 DIP2A HDAC4 MECP2 PAX6 SCN1A TAF6
ADSL CC2D1A DIP2C HECW2 MED13 PCCA SCN2A TBL1XR1
AGAP2 CCT4 DISC1 HEPACAM MED13L PCCB SCN9A TBR1
AHDC1 CDC42BPB DMD HERC2 MET PCDH19 SEMA5A TCF20
AHI1 CDKL5 DMPK HIVEP3 MFRP PER2 SETBP1 TCF4
AKAP9 CEP41 DNMT3A HMGN1 MSNP1AS PHB SETD2 TCF7L2
ALDH5A1 CGNL1 DOCK8 HOXA1 MTOR PHF2 SETD5 TERF2
ALG6 CHD2 DPP10 ILF2 MYO5A PHF3 SGSH TMLHE
AMPD1 CHD7 DSCAM INTS6 MYO9B PHRF1 SHANK1 TNRC6B
AMT CHD8 DYNC1H1 IRF2BPL MYT1L PIK3R2 SHANK2 TRIO
ANK2 CHKB DYRK1A ITGB3 NAA15 PLCB1 SHANK3 TRIP12
ANKRD11 CHRNA7 EFR3A KAT2B NAV2 PLXNA4 SHOX TRPC6
ANXA1 CIB2 EHMT1 KAT6A NCKAP1 PLXNB1 SLC12A5 TRPM1
APH1A CIC ELAVL3 KATNAL2 NF1 POGZ SLC1A2 TSC1
ARID1B CLASP1 ELP4 KCNJ10 NFIX PON1 SLC35B1 TSC2
ARX CNKSR2 EP400 KCNQ2 NINL PPM1D SLC38A10 UBE3A
ASH1L CNOT3 ERBIN KCNQ3 NIPBL PREX1 SLC45A1 UBE3C
ASTN2 CNR1 ETFB KDM5B NLGN3 PRICKLE1 SLC6A1 UBN2
ASXL3 CNTN4 FAM92B KDM5C NLGN4X PRICKLE2 SLC6A3 UBR5
ATP10A CNTN5 FBN1 KDM6A NR3C2 PRKCB SLC7A3 UPF3B
ATP1A3 CNTN6 FMR1 KDM6B NRXN1 PRKD1 SLC7A5 USP15
ATP2B2 CNTNAP2 FOXP1 KIAA1586 NRXN3 PRODH SLC9A6 USP45
AUTS2 CNTNAP4 FOXP2 KIAA2022 NSD1 PSMD12 SMARCA2 USP7
AVPR1A CTCF GABRB3 KIRREL3 NUAK1 PTCHD1 SMARCC2 VIL1
BCKDK CTNNB1 GATM KMT2A OCRL PTEN SNX14 VPS13B
BCL11A CTNND2 GIGYF1 KMT2C OPHN1 PTK7 SOX5 WAC
BRAF CTTNBP2 GIGYF2 KMT2E OR52M1 PYHIN1 SPARCL1 WDFY3
C11orf30 CUL3 GPC4 KMT5B OXTR RAB2A SPAST WWOX

C12orf57 CUL7 GPHN LAMB1 P2RX5 RAI1 SRCAP YY1
CACNA1C CUX1 GRIK2 LMX1B P4HA2 RANBP17 SRSF11 ZBTB20
CACNA1D DAPP1 GRIK5 LZTR1 PACS1 RBFOX1 STAG1 ZC3H4
CACNA1E DDX3X GRIN1 MACROD2 PAH RELN STXBP1 ZMYND11

9.22 Attachment 22

"zen_code" "ID_pac" "iniciais" "nascimento" "diagnostico" "sexo" "status"
"alvo" "exoma" "fra_fr1" "fra_fr2" "fra_fr3" "fra_fr4" "spl_fr1" "spl_fr2" "spl_fr3"
"spl_fr4" "sto_fr1" "sto_fr2" "sto_fr3" "sto_fr4" "nfr_fr1" "nfr_fr2" "nfr_fr3"
"nfr_fr4" "nsy_fr1" "nsy_fr2" "nsy_fr3" "nsy_fr4" "syn_fr1" "syn_fr2" "syn_fr3"
"syn_fr4" "utr5_fr1" "utr5_fr2" "utr5_fr3" "utr5_fr4" "utr3_fr1" "utr3_fr2"
"utr3_fr3" "utr3_fr4" "ncRNA_fr1" "ncRNA_fr2" "ncRNA_fr3" "ncRNA_fr4"
"miRNA_fr1" "miRNA_fr2" "miRNA_fr3" "miRNA_fr4" "bnd_fr1" "bnd_fr2"
"bnd_fr3" "bnd_fr4" "reg_fr1" "reg_fr2" "reg_fr3" "reg_fr4" "lof_fr3" "lof_fr4"
"1" "C11045" 156 "MVMB" 16807 "ELA" "F" "probando" 0 20 81 12 40 3 78 3
16 4 80 5 16 3 183 17 44 0 9881 441 479 77 11620 431 321 40 1681 64 57 18 2495
97 105 15 1847 84 113 9 5 1 1 1 14071 627 786 82 324 15 13 2 72 10
"2" "C22147" 157 "JCAO" 21915 "ELA" "M" "probando" 0 20 89 5 36 4 73 3 10
0 53 3 12 1 145 13 40 2 9013 449 495 73 10917 395 296 35 1398 59 56 4 1958 89
103 8 1444 71 89 11 3 0 0 0 11683 552 748 66 292 15 6 1 58 5
"3" "C25973" 158 "GJO" 23485 "ELA" "M" "probando" 0 20 81 10 25 3 61 4 5 1
65 3 14 3 161 8 40 1 8928 498 511 73 10915 494 320 27 1466 77 71 10 1885 112 88
11 1566 70 113 7 3 0 0 0 11973 644 741 64 314 16 11 1 44 7
"4" "C26551" 159 "AMS" 25732 "ELA" "F" "probando" 0 20 76 7 23 0 66 3 4 0
66 3 16 2 144 9 44 4 8766 528 547 73 10629 457 404 52 1343 83 78 8 1833 85 98 10
1495 81 129 7 4 0 0 0 11486 677 765 80 265 13 10 1 43 2
"5" "CC1139" 1 "MJR" NA "cfc" "F" "probando" 0 6 88 4 33 3 100 3 7 1 71 12 4
1 174 11 41 3 8956 599 526 95 10865 601 321 24 3514 235 168 27 21390 1098 1301
155 4348 198 225 31 23 1 3 0 20557 1305 1340 165 526 49 31 6 44 5
"6" "CC1155" 2 "NSR" NA "noonan" "M" "probando" 0 6 88 4 27 4 98 5 5 3 70
4 8 1 200 10 35 1 9210 416 440 79 11077 397 254 38 3902 149 138 20 22177 733
1074 155 4343 126 215 26 20 2 1 0 22850 927 1257 158 558 20 19 0 40 8
"7" "CC1158" 3 "NAO" NA "noonan" "M" "probando" 0 6 78 4 24 4 95 4 6 1
61 7 9 0 189 11 35 2 8806 503 460 86 10978 494 311 49 3788 205 172 24 21353
1080 1178 153 4372 171 216 28 19 3 0 0 21225 1211 1318 157 522 24 30 5 39 5
"8" "CC1162" 4 "LBL" NA "noonan" "M" "probando" 0 6 77 10 30 4 101 9 11 5
65 8 13 3 180 13 44 2 9656 646 558 99 11753 689 382 39 4327 278 247 26 22567
1317 1262 159 4714 247 257 24 24 0 1 0 24152 1651 1659 175 583 47 23 2 54 12
"9" "CC1169" 5 "LBS" NA "noonan" "F" "probando" 0 6 90 11 31 2 100 8 12 1
70 7 10 1 183 10 40 4 9076 682 492 60 11060 642 353 41 3790 255 161 26 22064
1309 1289 138 4164 251 226 22 23 2 0 0 21807 1516 1388 151 544 35 21 0 53 4

"10" "CC1216" 6 "LAS" NA "noonan" "M" "probando" 0 4 65 6 12 5 96 2 8 1
58 5 5 1 151 8 35 0 8387 509 448 52 10153 536 342 46 3519 211 170 25 18541 1025
1051 157 3963 139 213 27 15 2 0 0 19047 1164 1248 136 484 30 29 3 25 7

"11" "CC1264" 7 "JMLS" NA "noonan" "M" "probando" 0 4 88 10 43 2 93 6 7 2
66 10 5 1 156 13 36 2 8898 581 515 69 10878 601 362 42 3603 232 197 27 20612
1290 1193 164 4311 207 245 32 22 1 3 0 20349 1333 1358 138 558 39 20 2 55 5

"12" "CC1371" 8 "RHGP" NA "noonan" "F" "probando" 0 6 77 8 20 1 97 5 5 2
62 2 7 4 172 13 39 2 8893 469 421 94 10763 420 274 44 3634 162 173 31 21349 866
1098 159 4240 167 210 23 21 2 2 0 21395 989 1171 158 528 35 27 1 32 7

"13" "CC1413" 9 "GFL" NA "noonan" "M" "probando" 0 6 91 11 22 2 106 2 11
2 82 5 8 2 192 7 33 2 9684 519 453 74 11634 514 353 45 4196 213 191 28 23064
1092 1301 149 4694 192 266 22 18 0 0 0 24058 1200 1556 164 610 37 26 2 41 6

"14" "CC1471" 10 "MCS" NA "noonan" "M" "probando" 0 6 91 10 26 4 107 9 9
0 62 9 8 1 189 10 34 3 9296 686 605 80 11542 715 443 38 4207 348 260 43 22275
1473 1402 170 4559 294 256 22 16 1 2 1 24216 1760 1789 159 570 53 43 5 43 5

"15" "CC1586" 11 "MAOF" NA "noonan" "M" "probando" 0 4 67 8 23 3 86 7 6
0 64 8 10 3 133 14 28 0 7820 618 455 77 9331 662 356 46 2543 191 165 31 20546
1197 1234 161 3960 244 247 20 15 1 1 1 14790 1160 1068 141 391 29 14 0 39 6

"16" "CC1599" 12 "PVIS" NA "noonan" "M" "probando" 0 5 66 8 32 4 83 10 11
6 78 5 8 2 156 12 29 1 8344 745 627 85 10247 799 441 38 3037 271 197 32 21458
1585 1424 180 3895 259 259 19 25 0 1 0 17606 1631 1339 148 443 46 32 4 51 12

"17" "CC1980" 13 "DB" NA "noonan" "M" "probando" 0 6 69 7 23 5 120 10 15
3 76 4 21 6 178 12 48 2 9359 668 539 80 11298 656 379 46 3928 247 238 42 22315
1317 1424 187 4624 280 249 23 18 1 2 0 22740 1556 1641 181 538 36 26 4 59 14

"18" "CC2016" 14 "GPZ" NA "noonan" "M" "probando" 0 6 84 4 28 5 101 2 9
2 67 5 9 1 166 8 29 0 8872 389 433 102 10613 338 282 42 3847 151 146 21 20710
730 1033 174 4270 168 220 24 21 2 1 0 21237 828 1177 156 556 27 30 5 46 8

"19" "CC245" 15 "MSGJ" NA "noonan" "M" "probando" 0 4 76 4 22 2 73 7 8 2
60 7 10 3 140 9 37 0 7895 617 539 89 9549 660 378 39 2626 238 148 26 20577 1412
1309 164 3952 239 251 21 14 1 3 0 14736 1207 1086 123 353 33 23 3 40 7

"20" "CC2492" 16 "WSS" NA "cfc" "M" "probando" 0 4 69 7 33 3 81 3 9 0 75 7
5 1 154 18 33 2 8347 624 480 74 10201 635 383 48 3357 223 195 29 18643 1215
1091 128 3944 231 223 29 18 1 0 0 18670 1350 1261 166 465 45 28 4 47 4

"21" "CC3533" 17 "LFG" NA "noonan" "M" "probando" 0 6 95 9 29 5 121 3 9 2
77 4 8 1 196 11 47 2 9659 436 401 70 11719 430 260 54 4335 178 200 36 22730 803
1169 176 4524 179 218 25 29 0 4 1 24890 1003 1358 161 597 27 31 1 46 8

"22" "CC3546" 18 "ASF" NA "noonan" "M" "probando" 0 6 63 3 22 1 75 6 6 3
52 9 10 1 134 8 37 1 7658 528 381 60 9378 511 298 37 2475 128 135 29 19648 982
1081 143 3624 182 183 18 26 1 3 0 13542 835 837 97 324 17 15 4 38 5

"23" "CC3680" 19 "PVACM" NA "noonan" "M" "probando" 0 6 94 5 17 4 107
6 9 2 61 5 10 2 183 5 35 3 9203 498 477 84 11455 482 331 34 4025 214 171 29
21672 902 1236 159 4343 174 178 22 23 2 2 1 22556 1130 1360 157 511 27 27 8 36
8

"24" "CC3682" 20 "LMS" NA "noonan" "F" "probando" 0 4 91 4 16 1 118 7 12
 3 68 3 15 4 184 6 44 1 9314 533 472 92 11438 533 310 40 4023 226 199 35 21323
 994 1130 154 4406 207 212 27 23 5 2 0 22741 1236 1398 153 554 37 26 5 43 8
 "25" "CC3766" 21 "A" NA "noonan" "F" "probando" 0 2 78 6 21 3 91 4 9 0 60 7
 12 2 146 15 32 2 8535 565 454 91 10285 555 340 49 3030 196 150 27 20302 1074
 1181 169 3935 184 203 26 22 1 1 0 16556 1017 1078 144 448 25 14 3 42 5
 "26" "CC3770" 22 "DFSD" NA "noonan" "F" "probando" 0 4 82 10 28 2 98 3
 11 1 69 5 8 1 167 8 24 0 8549 546 486 67 10541 572 332 26 3453 224 161 19 20189
 1066 1081 114 4161 215 213 16 18 1 2 0 19335 1220 1194 103 512 38 22 1 47 4
 "27" "CC4388" 23 "LVA" NA "noonan" "M" "probando" 0 6 97 9 26 2 109 3 8
 1 61 8 10 2 178 11 40 2 9509 749 546 83 11693 733 457 49 4144 297 256 39 22576
 1497 1467 163 4629 293 253 25 19 2 4 0 24173 1780 1682 190 580 47 29 2 44 5
 "28" "CC4565" 24 "DNAC" NA "noonan" "M" "probando" 0 6 91 11 27 6 105
 9 7 0 65 4 10 2 186 6 45 0 9100 694 569 118 11117 771 373 37 3782 256 222 32
 21259 1459 1324 202 4124 272 280 33 19 2 2 0 21680 1683 1556 167 536 46 45 6 44
 8
 "29" "CC4602" 25 "SCSS" NA "cfc" "F" "probando" 0 6 79 7 32 8 121 7 7 1 67 5
 11 3 156 16 43 2 9119 676 583 76 11129 793 387 37 3999 278 232 38 21124 1501
 1415 179 4281 293 226 28 21 5 2 0 22445 1688 1578 184 528 34 20 4 50 12
 "30" "CC4640" 26 "RPS" NA "noonan" "M" "probando" 0 4 75 4 17 3 111 3 10
 1 66 6 13 2 172 8 33 1 9338 592 460 77 11598 543 334 45 3925 199 161 35 21423
 1076 1155 158 4410 164 224 29 14 3 4 0 22582 1234 1356 173 578 38 19 1 40 6
 "31" "CC4653" 27 "KSG" NA "noonan" "M" "probando" 0 4 70 7 30 4 104 4 12
 5 74 3 8 1 162 12 38 3 8622 507 446 85 10497 556 338 54 3624 222 185 31 19533
 932 1050 150 3996 171 214 27 18 1 2 0 19833 1125 1289 167 482 34 24 2 50 10
 "32" "CC4902" 28 "JMMS" NA "noonan" "M" "probando" 0 5 73 9 19 0 102 7
 7 1 67 7 6 1 158 14 36 1 8432 716 529 59 10517 722 418 32 4060 307 258 47 17463
 1292 1147 153 4046 261 196 29 17 2 0 0 20972 1580 1463 174 528 38 44 5 32 2
 "33" "CC4965" 29 "MMEF" NA "noonan" "M" "probando" 0 5 73 5 26 4 94 8
 12 1 69 7 10 1 144 10 30 1 8676 581 554 90 10402 673 405 41 3295 228 165 25
 20957 1211 1328 188 4188 225 231 33 16 1 0 0 19082 1411 1312 127 473 34 30 6 48
 6
 "34" "CC4996" 30 "ASP" NA "noonan" "M" "probando" 0 6 87 6 23 4 111 5 11
 1 63 11 8 1 170 12 37 2 9173 681 614 91 11333 717 419 55 4020 264 288 34 21659
 1405 1393 169 4414 321 279 33 18 1 2 0 22544 1713 1637 169 527 38 29 1 42 6
 "35" "CC5035" 31 "BMB" NA "noonan" "M" "probando" 0 6 92 6 26 5 105 6 5
 0 68 5 9 2 163 7 42 4 8949 424 429 110 10888 371 238 39 3791 159 173 35 21276
 726 1038 159 4342 114 173 32 18 3 2 0 21091 889 1179 189 519 22 18 6 40 7
 "36" "CC5054" 32 "GHOS" NA "noonan" "M" "probando" 0 6 75 12 37 4 104 9
 7 0 61 9 13 2 156 18 44 1 9016 922 770 84 11003 1046 610 39 3668 412 275 34
 21619 1924 1808 214 4350 357 272 31 26 1 0 0 20982 2255 1987 179 481 58 36 4 57
 6

"37" "CC5059" 33 "OAFE" NA "cfc" "F" "probando" 0 5 65 4 22 2 92 0 5 2 54 6
10 1 158 9 34 5 8519 392 402 96 10093 322 252 60 3120 127 119 38 20436 709 945
176 4193 125 150 21 23 3 3 1 17799 694 986 160 467 23 24 5 37 5

"38" "CC5064" 34 "SWSS" NA "noonan" "M" "probando" 0 5 57 6 27 5 91 10 7
1 72 14 12 5 148 11 31 3 8246 791 666 78 10024 786 478 37 3079 275 202 17 20359
1691 1563 160 4019 280 288 30 20 0 0 0 17621 1819 1558 140 413 53 27 3 46 11

"39" "CC5071" 35 "VSS" NA "noonan" "M" "probando" 0 6 72 12 29 3 109 11
10 3 65 8 11 2 182 16 44 2 9024 802 588 83 10863 810 411 47 3746 307 235 39
21950 1531 1458 158 4360 304 278 26 30 3 0 0 21157 1823 1569 155 506 57 34 4 50
8

"40" "CC5072" 36 "DJOS" NA "noonan" "M" "probando" 0 6 85 5 32 5 107 7 9
2 50 9 6 2 168 5 42 4 8373 583 548 87 10448 645 375 45 3471 280 193 27 20508
1217 1215 153 4129 202 199 25 27 0 4 1 19683 1388 1376 158 475 23 15 4 47 9

"41" "CC5088" 37 "RSAM" NA "nf1" "M" "probando" 0 6 74 12 26 6 108 3 10
3 50 5 15 0 180 13 38 3 8843 635 502 91 10735 577 362 47 3601 218 195 28 21365
1182 1186 154 4377 158 189 19 17 5 2 1 20494 1236 1364 158 517 26 20 3 51 9

"42" "CC5097" 38 "RLA" NA "noonan" "M" "probando" 0 5 79 7 27 2 90 6 5 0
47 3 10 1 154 16 39 2 8960 743 616 81 10823 845 460 43 3419 249 192 25 21523
1548 1459 184 4194 311 255 22 17 4 1 0 19730 1682 1512 153 520 44 42 5 42 3

"43" "CC5098" 39 "ALFC" NA "cfc" "M" "probando" 0 5 94 10 42 1 116 17 15
3 72 11 10 1 153 19 45 2 9218 1059 815 65 11214 1125 648 46 3620 416 291 40
22219 2238 2005 210 4358 437 357 37 16 3 1 0 20967 2472 1983 197 550 63 35 3 67
5

"44" "CC5099" 40 "MVS" NA "noonan" "M" "probando" 0 4 85 16 38 5 112 9 7
1 69 2 8 2 175 17 35 1 9235 585 500 73 11410 641 379 49 3827 251 181 31 21551
1212 1235 165 4439 217 211 24 31 1 1 0 20975 1370 1404 167 561 33 30 7 53 8

"45" "CC5105" 41 "KMV" NA "nf1" "M" "probando" 0 4 67 5 27 3 98 4 10 1 63
7 8 3 147 15 29 2 8674 572 433 83 10508 531 329 49 3559 156 150 23 20034 989
1132 170 4219 190 207 27 22 2 1 0 19653 1094 1142 138 505 35 25 9 45 7

"46" "CC5114" 42 "HBF" NA "noonan" "M" "probando" 0 5 76 7 25 5 89 4 6 2
57 10 8 1 148 7 35 2 8680 588 537 61 10306 587 359 42 3105 187 159 23 21154
1191 1277 170 4063 206 271 27 17 2 4 0 17867 1165 1260 142 445 30 19 3 39 8

"47" "CC5119" 43 "JPMS" NA "noonan" "M" "probando" 0 5 80 5 25 5 95 3 10
1 68 4 6 1 172 11 39 2 8855 498 476 83 10871 488 316 41 3402 174 168 28 22375
1008 1295 174 4220 172 252 41 16 2 3 0 20536 1066 1265 163 564 33 29 2 41 7

"48" "CC5120" 44 "ALN" NA "noonan" "M" "probando" 0 6 74 5 28 4 99 5 13
3 61 4 7 2 157 9 44 1 8841 624 516 73 10797 652 381 31 3473 267 186 25 21864
1379 1356 145 4428 224 236 26 15 1 2 0 20396 1456 1446 142 510 36 23 4 48 9

"49" "CC5133" 45 "APS" NA "noonan" "M" "probando" 0 6 69 6 30 3 95 5 5 2
72 7 9 1 166 12 38 0 9033 580 459 73 10949 540 363 30 3521 172 169 25 21882
1014 1279 165 4365 176 242 22 24 1 0 0 20289 1137 1412 153 476 38 27 5 44 6

"50" "CC5188" 46 "BHMS" NA "noonan" "M" "probando" 0 6 70 7 31 5 98 7
11 2 64 8 8 3 171 7 44 3 9064 521 470 88 10875 505 344 32 3731 198 216 39 21795
1044 1195 169 4400 178 241 23 14 0 2 0 21426 1148 1388 169 521 18 27 5 50 10

"51" "CC5198" 47 "JLC" NA "noonan" "F" "probando" 0 5 72 1 31 2 93 1 7 1
61 2 7 2 163 15 38 2 8999 536 451 102 10760 474 285 43 3162 163 135 26 22002
892 1155 174 4276 197 217 32 24 1 1 0 19786 1034 1136 160 479 27 20 3 45 5
"52" "CC5200" 48 "JCS" NA "noonan" "M" "probando" 0 4 68 12 31 4 101 14
12 3 56 15 17 3 162 30 43 3 9072 1166 908 74 11154 1369 741 46 3692 458 376 39
21354 2566 2062 205 4454 464 409 31 23 2 0 20897 2714 2258 216 542 78 59 7 60
10
"53" "CC5217" 50 "MPS" NA "noonan" "M" "probando" 0 6 76 11 31 4 89 4 11
1 72 2 13 3 178 4 41 3 8998 456 402 79 10915 463 285 44 3573 163 121 15 21567
913 1126 161 4166 160 214 25 18 1 1 0 20937 1021 1230 176 526 16 20 3 55 8
"54" "CC5219" 51 "AAD" NA "noonan" "M" "probando" 0 4 80 9 19 2 88 2 9 1
59 2 11 2 171 11 27 0 8675 412 391 83 10709 388 283 45 3473 157 174 35 20491
804 1041 181 4013 165 191 28 18 3 3 1 19417 881 1125 152 453 23 20 5 39 5
"55" "CC5284" 52 "LFSS" NA "noonan" "M" "probando" 0 5 71 4 25 8 103 1 9
3 55 9 16 6 160 11 40 1 8722 586 498 87 10562 585 337 42 3298 184 178 25 21658
1163 1364 158 4221 206 228 23 21 1 3 0 19754 1254 1285 163 481 36 30 4 50 17
"56" "CC5299" 54 "ERBS" NA "noonan" "M" "probando" 0 5 75 5 26 2 98 7 5 1
70 4 17 4 165 14 38 0 8733 594 492 81 10516 593 357 36 3285 224 194 30 21948
1294 1310 152 4337 228 234 26 22 0 0 0 19287 1230 1304 141 490 35 30 4 48 7
"57" "CC5305" 56 "AHS" NA "cfc" "M" "probando" 0 6 62 13 28 3 98 7 10 2 72
8 12 1 156 19 49 0 9004 802 657 93 10684 847 526 43 3472 321 234 27 22217 1687
1567 186 4204 272 277 34 20 2 1 0 20164 1934 1712 157 529 56 31 6 50 6
"58" "CC5308" 57 "WSS" NA "cfc" "M" "probando" 0 6 83 9 29 4 89 9 7 1 73 7
7 1 131 10 37 4 8479 715 558 74 10215 706 468 36 3259 223 217 35 20711 1447
1295 158 4111 253 238 19 15 1 2 0 18772 1526 1401 145 475 39 17 3 43 6
"59" "CC5399" 58 "MBLF" NA "noonan" "M" "probando" 0 4 74 6 25 4 99 7 9
2 60 6 15 3 147 6 29 3 8588 630 561 88 10457 671 416 39 3511 264 193 19 19105
1255 1274 163 4150 278 218 26 15 2 1 0 19376 1427 1414 135 477 46 27 5 49 9
"60" "CC5407" 59 "GAC" NA "noonan" "M" "probando" 0 6 70 3 25 4 99 6 6 1
56 1 11 3 166 7 32 6 8765 358 452 101 10401 330 292 52 3349 116 148 35 20122
625 998 177 4108 118 193 23 24 0 3 0 19271 715 1117 152 492 18 21 4 42 8
"61" "CC722" 61 "MAM" NA "noonan" "M" "probando" 0 4 79 11 28 3 92 3 5
2 64 7 13 2 160 9 32 0 8863 405 386 93 10700 382 315 53 3520 146 147 30 19633
788 954 163 3975 150 184 23 14 0 0 0 19342 845 1084 163 460 23 22 1 46 7
"62" "CC726" 62 "MCS" NA "noonan" "F" "probando" 0 4 76 9 31 2 98 10 9 2
55 12 15 5 160 14 28 3 8174 732 576 81 10156 800 438 56 3487 288 219 31 18956
1397 1307 184 3776 215 255 36 19 2 1 0 18872 1567 1419 156 438 46 29 1 55 9
"63" "CC736" 63 "SMFC" NA "noonan" "F" "probando" 0 4 82 8 27 0 97 5 8 0
77 4 9 0 161 12 38 1 9203 596 457 39 11013 520 299 16 3206 193 154 9 23250 1093
1266 64 4525 191 219 14 23 3 4 0 19686 1154 1219 59 498 34 34 2 44 0
"64" "CC804" 64 "JFCS" NA "noonan" "M" "probando" 0 6 56 4 16 0 65 4 8 4
51 6 4 0 135 5 23 1 7158 410 419 81 8784 349 261 41 2817 128 140 30 15838 685
871 137 3408 154 140 24 16 2 1 0 14637 742 945 148 352 12 20 3 28 4

"65" "CC815" 65 "VGS" NA "noonan" "M" "probando" 0 6 62 7 16 1 50 5 8 2
44 3 9 2 120 9 23 1 6602 469 415 70 8054 492 297 37 2597 179 137 28 14227 831
847 147 3331 193 160 24 9 3 2 1 13135 882 859 106 313 34 18 9 33 5
"66" "CC883" 67 "VCAB" NA "cfc" "M" "probando" 0 4 70 6 25 2 87 3 4 1 65 5
6 3 171 8 31 1 8368 411 378 89 10109 351 242 32 3496 123 124 22 18965 681 930
190 3979 132 192 22 20 3 1 0 19122 838 1055 138 441 22 25 3 35 6
"67" "CC884" 68 "AGN" NA "noonan" "F" "probando" 0 6 70 8 15 2 62 2 3 1
43 6 9 3 131 7 30 1 6685 380 360 77 8143 348 238 39 2461 116 113 27 13959 683
761 142 2935 113 155 29 10 0 0 0 12995 650 712 124 334 15 15 3 27 6
"68" "CC893" 69 "AGVC" NA "cfc" "F" "probando" 0 5 83 10 19 2 70 4 6 3 73 4
9 3 161 6 28 1 8450 462 386 71 10229 450 289 44 3060 121 111 26 20765 823 1039
156 3940 140 190 22 21 0 2 0 18283 911 1029 123 449 25 19 5 34 8
"69" "CC941" 70 "VMS" NA "noonan" "F" "probando" 0 4 79 9 28 4 81 6 6 2
60 4 13 5 170 8 35 1 8643 664 532 72 10742 736 446 35 3602 253 219 37 19758
1346 1296 167 4241 265 249 20 17 1 3 0 20258 1445 1465 161 498 32 28 2 47 11
"70" "CC962" 71 "LTSN" NA "noonan" "F" "probando" 0 5 73 16 34 5 97 4 7 0
54 6 10 3 135 17 31 3 8502 796 643 87 10381 901 521 43 3200 265 225 28 21361
1724 1584 207 4086 322 280 21 24 1 2 1 18992 1821 1562 164 495 52 34 1 51 8
"71" "CC963" 72 "KFFV" NA "cfc" NA "probando" 0 5 85 14 23 5 102 4 6 4 63
11 13 6 159 8 33 0 9001 529 476 88 10806 486 334 45 3384 206 163 28 21876 1017
1148 168 4303 171 232 31 24 0 0 0 19993 1165 1248 165 493 18 21 1 42 15
"72" "F-3078-1" 177 "HF" 36741 "TEA" "M" "probando" 1 18 41 2 30 3 37 2 9 5
49 2 10 3 93 2 25 2 5985 317 506 80 7563 311 292 41 701 44 44 5 977 60 54 12 1041
53 102 8 3 0 0 0 6880 358 542 55 207 12 6 1 49 11
"73" "F010644-1" 149 "MRC" 40964 "TEA" "M" "probando" 1 21 97 10 28 2 81
6 13 3 65 11 12 2 199 17 60 7 9560 786 731 83 11623 813 567 43 2147 172 126 22
2873 260 188 19 1861 145 160 10 5 0 0 0 17206 1483 1513 138 419 43 21 2 53 7
"74" "F010992-1" 150 "ALVT" 40267 "TEA" "M" "probando" 1 21 87 13 50 4
96 5 11 0 74 7 20 5 229 14 45 0 9827 752 653 79 12045 737 453 55 2084 158 140 22
3137 223 182 23 1942 110 135 7 6 0 0 0 17971 1244 1224 144 396 34 16 3 81 9
"75" "F10291-1" 151 "LIG" 40057 "TEA" "M" "probando" 1 21 83 5 42 2 89 1
14 4 83 4 20 7 198 10 49 2 10223 407 555 111 12181 411 334 49 2041 91 89 18 3210
128 159 29 2133 69 177 12 6 0 0 0 17726 790 1192 127 444 13 24 3 76 13
"76" "F10658-1-II" 168 "JPPA" 42573 "zika" "M" "probando" 0 19 62 13 39 3 59
2 9 1 55 5 12 2 134 9 33 2 8118 570 591 63 9825 572 401 38 1143 65 58 7 1768 129
101 13 1458 96 131 8 2 0 0 0 10426 726 831 59 278 25 26 3 60 6
"77" "F10661-1-I" 160 "LGS" 42234 "zika" "M" "probando" 0 20 71 9 39 3 80 2
10 0 63 12 26 2 163 10 35 1 9416 644 692 89 11531 650 489 40 1294 94 66 15 2460
159 149 18 1687 104 129 10 7 0 0 0 12658 961 1104 97 366 25 22 6 75 5
"78" "F10691-1" 178 "BSS" 42360 "zika" "M" "probando" 0 18 56 11 25 0 43 4
11 3 42 10 11 2 118 28 30 3 6705 874 763 79 8011 951 536 43 793 100 62 4 1154
163 115 11 1053 108 120 10 1 0 0 0 7488 987 859 68 210 30 15 1 47 5

"79" "F10692-1" 179 "MAC" 42163 "zika" "F" "probando" 0 18 48 10 35 4 45 2
7 0 51 5 19 2 94 13 27 2 6523 722 702 72 8004 735 525 36 792 92 78 9 1127 139 104
10 1130 101 111 7 2 1 0 0 7480 856 766 67 215 27 15 0 61 6

"80" "F10693-1-II" 161 "MAC" 42302 "zika" "F" "probando" 0 20 86 12 40 2 78
7 7 0 65 7 8 2 153 9 46 5 9533 684 721 95 11397 669 505 31 1245 106 53 4 2617 189
156 23 1793 113 159 7 9 0 0 0 12775 962 1065 87 382 27 16 4 55 4

"81" "F10701-1-I" 162 "MOSL" 42373 "zika" "F" "probando" 0 20 79 9 25 2 84
3 16 5 53 8 14 4 131 13 36 2 8962 649 667 88 10733 654 463 43 1071 84 57 5 2450
186 133 13 1530 108 123 9 5 0 0 0 11695 830 908 77 338 23 20 1 55 11

"82" "F10741-1-I" 163 "PGS" 42330 "zika" "M" "probando" 0 20 87 12 36 4 68 4
10 3 72 11 20 3 148 8 35 2 9308 552 579 83 10945 606 432 26 1216 72 82 10 2427
146 134 15 1616 96 123 12 5 0 0 0 12282 858 1007 88 315 18 15 2 66 10

"83" "F10747-1-I" 164 "RNLCCB" 42371 "zika" "F" "probando" 0 20 74 8 48 5
94 5 10 0 63 4 12 0 163 9 36 1 9389 646 736 86 11231 674 493 49 1178 86 76 12
2614 202 171 18 1815 130 146 6 5 1 1 0 12767 1016 1096 86 373 25 15 1 70 5

"84" "F10760-1" 152 "SEO" 41269 "TEA" "M" "probando" 1 21 80 7 37 4 91 7
16 4 72 7 14 4 213 19 43 2 9855 758 743 93 11710 813 549 33 2018 190 127 19 2907
267 212 24 1930 137 162 12 5 0 0 0 16946 1443 1451 113 420 34 24 1 67 12

"85" "F10763-1-I" 169 "ITE" 42007 "zika" "F" "probando" 0 19 100 3 39 3 78 5 7
0 78 7 15 4 212 14 38 7 10216 596 650 107 12088 578 458 57 1859 114 94 10 3346
194 183 22 2137 144 184 9 8 0 1 0 17439 1066 1284 125 451 35 25 3 61 7

"86" "F10777-1" 170 "GSI" 41519 "TEA" "M" "probando" 1 19 87 8 51 2 90 6 8
2 65 6 23 3 207 14 47 3 10141 656 643 94 12240 678 429 49 1888 130 100 10 2862
208 150 19 1950 129 152 6 7 0 0 0 16390 1129 1198 105 417 33 27 1 82 7

"87" "F10788-1-I" 171 "JLSA" 42236 "zika" "M" "probando" 0 19 76 6 36 3 71 4
10 0 66 12 17 3 158 13 29 4 8679 646 703 97 10554 639 492 41 1294 84 76 9 1881
147 135 22 1569 104 138 4 6 0 0 0 11390 866 962 89 312 26 25 3 63 6

"88" "F10789-1-I" 172 "HBG" 42474 "zika" "F" "probando" 0 19 65 7 35 5 60 3
10 3 59 8 13 1 135 16 32 1 8135 658 648 85 9974 734 484 55 1083 95 62 12 1594
141 99 12 1336 117 149 5 4 0 0 0 10072 786 853 79 265 19 18 5 58 9

"89" "F10798-1" 173 "NPC" 41657 "TEA" "F" "probando" 1 19 88 4 45 1 72 1 7
1 80 5 12 4 186 12 46 4 9973 387 545 93 11863 368 327 40 1752 61 74 17 2725 107
116 11 1871 73 125 11 2 0 0 0 15150 588 914 102 392 12 22 1 64 6

"90" "F10865-1" 165 "IVSN" 39455 "desconhecido" "F" "probando" 0 20 89 9
41 3 62 3 13 3 63 5 18 3 156 14 37 7 9037 589 633 84 10709 563 388 43 1181 81 60
10 2197 121 115 15 1550 106 103 6 3 0 2 1 11579 757 898 87 300 18 22 3 72 9

"91" "F10885-1" 154 "GBI" 38803 "TEA" "M" "probando" 1 21 81 9 40 6 78 2
10 0 63 8 15 4 211 6 36 3 9485 416 517 84 11573 397 324 43 1921 85 66 15 2693 87
148 21 1825 57 103 6 9 0 0 0 16361 699 971 111 389 22 15 5 65 10

"92" "F10980-1" 155 "DMAS" 40402 "TEA" "M" "probando" 1 21 73 5 38 3 81
8 10 1 77 5 17 1 200 18 38 4 9250 583 570 68 11241 563 437 46 1807 92 82 13 2546
147 156 23 1695 97 125 8 3 0 0 0 15205 814 999 121 356 27 14 2 65 5

"93" "F1385-1" 73 "MBA" NA "TEA" "M" "probando" 1 7 83 5 31 3 94 7 9 3 75
6 12 7 163 9 42 1 8879 598 588 77 10785 560 388 43 3581 197 180 23 20392 1072
1139 162 4346 240 296 31 19 3 2 0 20301 1202 1360 129 550 37 20 2 52 13
"94" "F1389-1" 74 "MBG" NA "TEA" "F" "probando" 1 7 77 1 33 4 108 0 12 2
63 3 13 5 167 12 31 1 8479 370 422 80 10196 373 222 40 3442 137 160 38 19415
615 918 125 4186 124 268 23 24 1 2 0 19013 764 1098 148 447 22 21 6 58 11
"95" "F1665-1" 76 "IRS" 34761 "TEA" "M" "probando" 1 7 90 3 33 3 128 2 11 1
78 9 7 1 199 12 45 2 9761 607 511 69 11875 557 366 42 4457 236 266 46 22086
1229 1408 157 4490 228 302 28 26 2 3 0 22579 1272 1481 154 526 21 27 2 51 5
"96" "F1697-1" 77 "LMF" 36390 "TEA" "M" "probando" 1 7 70 5 34 3 91 7 13 6
62 1 11 2 158 4 30 3 8639 391 391 76 10422 328 247 44 3494 131 123 31 19699 647
853 127 4172 115 255 31 24 1 2 0 19940 803 967 149 512 9 23 4 58 11
"97" "F1722-1" 78 "AR" 32549 "TEA" "M" "probando" 1 7 93 3 34 2 98 6 8 3 81
6 12 5 174 7 25 1 9078 423 423 81 10989 348 271 39 3690 150 157 42 20968 722
1014 164 4320 136 253 28 27 1 2 0 21479 871 1175 183 537 14 22 3 54 10
"98" "F1729-1" 79 "GSS" 36989 "TEA" "F" "probando" 1 7 65 10 25 2 106 2 9 5
57 5 11 2 159 11 33 3 8619 749 628 95 10482 783 499 44 3512 289 220 27 19993
1569 1410 161 4613 334 333 26 24 2 1 0 20210 1769 1630 151 461 48 24 6 45 9
"99" "F1903-1" 80 "LMC" 38023 "craniosinostose" "F" "probando" 0 6 60 6 21
3 66 6 8 3 45 6 10 2 120 7 33 2 7349 399 383 84 8827 369 278 48 2746 140 133 23
16455 813 865 142 3531 131 146 18 14 0 0 0 14981 739 935 136 386 18 23 4 39 8
"100" "F2452-1" 81 "IM" NA "TEA" "F" "probando" 1 7 76 9 31 5 72 4 12 6 66
7 11 3 153 11 40 3 8106 445 442 68 9794 424 291 32 3333 174 143 27 18498 792
928 125 3969 198 285 19 21 0 2 0 17739 894 1017 129 475 24 14 2 54 14
"101" "F2570-1" 82 "MMCS" 27933 "CLP" "F" "probando" 0 3 73 5 20 1 109 7
17 4 66 5 6 3 185 12 48 3 9423 718 604 102 11572 739 431 39 4109 318 257 33
22597 1542 1489 194 4608 262 248 25 24 2 0 0 24376 1727 1735 188 580 49 28 6 43
8
"102" "F2613-1" 83 "IPCC" 36240 "TEA" "M" "probando" 1 3 74 7 40 6 101 5 5
3 72 3 12 1 145 8 38 1 8664 406 450 64 10481 411 278 35 3414 158 154 27 19602
746 917 145 4291 158 262 15 24 1 2 2 19198 831 1083 120 474 17 20 6 57 10
"103" "F2685-1" 84 "EHK" 37519 "TEA" "M" "probando" 1 11 81 4 15 6 106 3
10 2 69 4 7 4 161 8 26 3 9034 339 542 108 10975 383 354 62 4007 155 204 50 18830
666 965 157 3975 157 217 25 20 0 2 1 16084 643 1015 157 364 15 19 1 32 12
"104" "F2688-1" 85 "CAMC" 37482 "TEA" "M" "probando" 1 7 67 9 23 2 71 6 4
2 48 7 4 1 159 8 38 1 7898 521 431 77 9620 487 276 38 2725 186 144 24 19918 1099
1126 177 3872 183 191 23 16 1 2 0 15920 1061 1058 149 412 20 21 2 31 5
"105" "F2735-1" 86 "LTC" 35529 "TEA" "M" "probando" 1 7 72 4 25 4 82 4 6 2
60 9 7 1 155 6 34 4 7846 414 406 96 9414 381 261 47 2679 129 112 24 19770 769
988 192 3651 146 132 24 14 1 2 0 15311 767 967 156 374 24 19 5 38 7
"106" "F2848-1" 87 "HEN" 39284 "CLP" "M" "probando" 0 11 59 8 31 0 69 3 4
1 58 6 9 1 121 9 25 1 7195 467 478 78 8855 515 298 37 2980 174 148 24 14306 907
870 119 3253 146 220 22 9 1 0 0 11817 765 772 96 258 19 12 2 44 2

"107" "F3104-1" 88 "LPM" 38964 "FG" "M" "probando" 0 14 85 6 37 5 81 6 6 3
55 8 16 2 155 7 28 0 8769 543 630 73 10613 550 416 48 4194 195 204 38 17064 914
907 150 3620 181 250 23 16 0 1 1 15355 841 1107 139 322 23 17 2 59 10

"108" "F3196-1" 89 "JCS" NA "CLP" "M" "probando" 0 4 72 10 28 4 93 4 10 2
60 3 14 2 165 12 38 0 8300 564 454 108 10244 580 306 43 3523 210 194 34 19002
1032 1138 168 4046 226 242 22 14 1 1 0 19681 1248 1347 161 531 23 20 2 52 8

"109" "F4442-1" 90 "JVM" 39209 "CLPSy" "F" "probando" 0 1 65 1 34 5 94 4 6
1 62 5 8 1 155 13 29 3 8689 550 514 78 10564 564 338 34 3515 200 171 30 19800
1105 1176 144 4054 190 180 19 20 1 0 0 17985 1133 1201 158 449 26 26 5 48 7

"110" "F4482-1" 91 "VSS" 38249 "TEA" "M" "probando" 1 3 83 8 37 5 103 5 4 1
67 7 8 2 171 14 49 2 9037 634 606 77 10958 634 447 51 3757 220 181 31 21003
1312 1143 126 4493 238 348 26 21 2 2 0 20705 1422 1420 160 516 44 35 6 49 8

"111" "F4483-1" 92 "BOJ" 39807 "Saethre-Chotzen" "F" "probando" 0 2 72 4 21
2 79 5 8 0 56 9 12 0 135 7 29 4 7414 623 490 87 9091 666 377 36 2709 223 152 26
17835 1307 1297 139 3670 243 225 18 14 0 6 3 14643 1248 1081 128 406 37 22 2 41
2

"112" "F4485-1" 93 "DNS" 31131 "acrofrontofacionasal" "F" "probando" 0 1
60 7 22 0 51 4 6 1 52 3 7 2 81 10 26 2 5740 471 431 57 6948 470 332 30 1754 148
121 11 12703 931 938 142 2914 189 163 13 15 1 0 9429 720 748 72 276 23 22 3 35
3

"113" "F5389-1" 94 "KDS" 36860 "frontonasal" "F" "probando" 0 6 68 16 28 0
103 9 7 2 62 4 6 0 170 14 43 3 9027 745 619 75 10975 778 467 41 3844 312 222 32
20967 1579 1389 178 4256 272 250 29 12 1 1 0 21128 1722 1569 172 511 43 30 7 41
2

"114" "F5414-1" 95 "VEFA" 39792 "OAV" "M" "probando" 0 6 68 7 24 3 84 7 9
4 40 6 13 2 135 14 48 3 7000 491 418 70 8969 451 294 29 3996 235 236 34 13262
792 759 117 3693 181 241 24 8 0 0 0 22345 1291 1557 160 597 36 24 0 46 9

"115" "F5502-1" 96 "JPP" 38373 "TEA" "M" "probando" 1 11 82 6 29 6 97 2 4 0
64 6 8 1 141 8 29 2 9054 600 562 89 11070 565 365 41 3983 243 209 40 19492 1069
1151 145 3962 187 216 17 18 2 4 2 16509 1027 1028 138 388 22 15 1 41 7

"116" "F5547-1" 97 "BCP" 39389 "TEA" "F" "probando" 1 11 84 6 29 2 98 5 11
4 61 6 18 7 138 9 33 2 9022 421 439 85 10962 448 254 32 4034 155 156 36 19276
694 902 151 3909 151 184 28 18 1 3 0 16688 720 880 151 377 14 15 3 58 13

"117" "F5576-1" 98 "LCSS" 38549 "TEA" "F" "probando" 1 3 79 7 26 1 99 6 10
1 64 7 17 4 181 16 38 4 9198 497 527 88 11139 447 353 49 3723 167 168 22 21084
914 1016 147 4391 211 321 18 20 1 3 0 20648 1109 1247 132 495 33 16 4 53 6

"118" "F5577-3" 99 "MSP" 25809 "auriculo_condilar" "M" "probando" 0 13 65
7 32 3 69 1 5 1 49 7 11 4 109 4 18 1 7438 453 417 67 9292 461 335 28 3290 173 143
22 13801 714 687 106 2992 162 169 18 12 2 2 0 11237 600 649 86 211 13 8 1 48 8

"119" "F5616-1" 100 "VCB" 38964 "TEA" "M" "probando" 1 7 84 5 28 5 90 5 11
1 62 2 7 0 159 11 32 1 8528 448 480 77 10323 381 306 36 3569 185 170 24 18999
755 984 148 4063 159 297 26 20 1 3 1 18522 879 1138 134 448 28 18 2 46 6

"120" "F5674-1" 101 "VBG" 36352 "TEA" "M" "probando" 1 7 76 5 24 6 90 4 14
1 67 5 11 1 167 9 33 3 8803 404 455 108 10659 334 255 46 3633 141 167 43 20229
720 958 163 4320 133 243 21 23 2 3 0 19471 827 1104 153 474 17 18 3 49 8
"121" "F5979-1" 102 "LDO" 38552 "TEA" "M" "probando" 1 7 77 8 25 3 99 2 11
1 62 8 14 4 186 9 45 5 8737 490 537 93 10901 458 350 41 3665 208 214 41 20351
952 1123 134 4279 187 260 18 22 2 0 20242 1065 1327 172 499 17 29 2 50 8
"122" "F5980-1" 103 "JNS" 38137 "TEA" "M" "probando" 1 9 65 12 23 1 109 5
15 3 52 5 11 1 169 14 35 4 8347 691 536 64 10181 745 422 42 4484 352 283 37
19971 1346 1448 160 4041 264 301 36 15 0 1 0 23095 1827 1796 187 527 56 23 2 49
5
"123" "F5996-1" 104 "TSS" 38480 "TEA" "F" "probando" 1 7 71 5 42 5 72 7 8 1
65 10 15 2 153 9 38 2 8272 543 546 94 10056 551 374 42 3268 208 195 42 18789
1093 1151 161 4046 195 258 30 18 2 2 0 17912 1175 1266 149 421 31 24 4 65 8
"124" "F6021-1" 105 "CBA" 37713 "TEA" "M" "probando" 1 9 57 6 28 4 74 4 7
1 41 3 10 2 132 11 32 2 6265 506 419 82 7889 464 280 34 3448 231 172 18 13371
831 852 130 2867 161 203 15 14 1 2 0 16586 1080 1134 132 372 22 14 2 45 7
"125" "F6025-1" 106 "JGNG" 27939 "DE" "M" "probando" 0 7 73 9 28 7 84 13 8
3 59 7 14 5 165 6 49 3 8903 531 416 104 10922 484 308 62 3890 209 174 42 20550
935 1056 175 4244 152 189 22 14 3 3 0 20852 1133 1251 224 481 17 28 7 50 15
"126" "F6034-1" 107 "BMK" 37859 "TEA" "M" "probando" 1 9 62 8 15 2 70 3 4
3 34 3 6 2 134 16 27 2 6077 379 368 41 7558 362 255 31 3383 196 154 32 12769 757
811 106 2862 131 181 21 9 1 1 0 16509 985 1066 127 376 23 17 3 25 7
"127" "F6083-1" 108 "ASO" 38495 "oculoauriculofrontonasal" "F" "probando"
0 6 64 2 35 5 89 6 7 1 72 3 7 0 159 9 29 1 8488 510 515 87 10344 507 313 42 3483
205 191 38 20064 1017 1011 135 4099 189 227 41 20 1 2 0 19841 1201 1268 153
495 31 26 5 49 6
"128" "F6123-1" 109 "VCB" 38857 "TEA" "F" "probando" 1 9 59 6 16 7 67 7 6 1
45 2 6 2 129 7 25 1 6198 413 379 58 7642 410 258 25 3442 193 188 43 13155 777
854 133 2926 135 195 19 16 2 0 0 16444 952 1129 137 346 16 16 1 28 10
"129" "F6131-1" 110 "GSC" 38501 "TEA" "M" "probando" 1 9 58 3 21 4 72 4 6
1 32 6 11 3 131 7 31 0 6239 554 473 74 7828 590 349 35 3395 268 234 36 13474
1094 932 141 3000 192 237 29 21 2 0 0 16117 1270 1251 160 365 29 20 6 38 8
"131" "F6163-1" 112 "GTSM" 39185 "TEA" "F" "probando" 1 9 50 3 17 2 68 3
15 2 43 6 11 3 131 9 26 2 6121 477 514 89 7619 502 377 37 3383 220 216 39 13133
950 1033 166 3100 162 271 29 14 0 2 0 15583 1029 1348 184 339 19 28 5 43 7
"132" "F6167-1" 113 "IAN" 39353 "TEA" "F" "probando" 1 9 63 10 21 2 71 3 9
3 32 3 12 1 123 7 31 6 6088 454 457 64 7614 488 318 44 3227 227 187 44 12908 902
897 123 2929 163 212 21 13 2 0 0 15466 1034 1118 128 339 21 15 3 42 6
"133" "F617-19" 114 NA NA "CLP" NA "probando" 0 2 57 6 17 1 76 7 4 1 54 3
8 2 112 12 31 1 7573 518 411 73 9041 427 294 49 2666 153 138 29 17812 925 1016
168 3530 193 125 22 16 1 4 2 13634 849 827 104 317 25 8 0 29 4
"134" "F6281-1" 115 "RGM" 36802 "TEA" "M" "probando" 1 7 82 5 24 8 95 2 9
0 69 4 6 1 171 9 31 1 8966 546 475 79 10715 527 334 41 3836 220 153 24 20838 990
1145 149 4348 184 206 23 22 2 0 21355 1184 1357 147 503 33 27 5 39 9

"135" "F6325-1" 116 "MRNS" 39099 "TEA" "M" "probando" 1 11 92 11 18 0
105 6 9 1 71 9 16 5 148 14 36 4 9311 667 589 74 11311 746 420 47 4194 257 213 27
19767 1252 1082 126 4043 214 254 24 21 5 2 0 17247 1188 1106 124 415 29 25 3 43
6

"136" "F6331-1" 117 "LAF" 38395 "TEA" "F" "probando" 1 4 84 4 33 4 84 6 9 2
51 7 8 4 171 9 29 3 8831 641 502 71 10787 660 362 36 3881 251 193 36 20567 1195
1270 185 4432 245 255 30 19 0 2 0 21317 1474 1455 173 531 44 37 7 50 10

"137" "F6366-1" 118 "PCO" 39333 "TEA" "M" "probando" 1 9 60 6 22 2 85 4 9
2 45 9 6 0 128 11 23 1 6450 444 466 59 7930 467 329 35 3436 218 208 35 14033 855
903 123 3048 154 226 22 11 1 4 0 16837 1052 1213 148 357 24 27 6 37 4

"138" "F6594-1" 119 "IBB" 35182 "quebra_Ina" "F" "probando" 0 3 87 5 30 3 85
10 6 0 55 4 9 1 157 10 37 3 8320 636 540 79 10194 706 362 30 3449 261 185 28
20180 1415 1227 127 4196 272 259 23 16 3 1 0 19115 1429 1385 128 450 43 30 3 45
4

"139" "F6642-1" 120 "PRV" 39545 "TEA" "M" "probando" 1 9 60 5 20 3 70 7 5
0 47 4 4 0 114 7 29 3 6187 391 386 54 7560 383 244 44 3267 191 164 36 13037 755
808 126 2850 124 226 27 14 1 3 1 16171 854 1040 141 361 18 24 3 29 3

"140" "F6645-1" 121 "VMA" 37145 "TEA" "M" "probando" 1 8 73 10 21 0 91 3
11 5 51 6 8 1 165 5 37 3 8381 480 501 89 10158 407 323 32 3413 158 189 20 19752
844 1045 148 4116 184 284 20 23 2 1 0 19733 965 1205 140 472 24 27 5 40 6

"141" "F7511-1" 122 "VHFPS" 38343 "TEA" "M" "probando" 1 7 59 11 35 2 89
3 6 1 54 3 12 2 143 15 36 3 7752 545 561 87 9320 569 343 43 3085 198 180 28
18766 1051 1116 157 3605 217 319 32 15 0 2 0 16866 1175 1260 163 406 36 21 5 53
5

"142" "F7614-1" 123 "AOLB" 40471 "CLP" "M" "probando" 0 3 72 7 15 4 96 3
9 1 57 5 16 3 161 8 32 2 8333 504 428 91 10131 485 300 40 3484 179 177 27 19073
1004 968 153 4026 236 223 21 18 1 1 0 18603 1119 1010 149 439 30 23 1 40 8

"143" "F7935-1" 124 "JGCE" 40060 "TEA" "F" "probando" 1 3 77 10 34 7 109
12 17 4 60 11 18 4 155 24 32 3 8931 1020 794 79 10924 1062 613 46 3835 383 294
36 21381 2040 1695 169 4399 416 390 36 20 0 2 0 21088 2290 1932 181 553 72 41 3
69 15

"144" "F7936-1" 125 "CESS" 37677 "TEA" "M" "probando" 1 4 79 8 22 3 103 7
10 1 66 7 7 1 173 13 37 5 8853 807 616 72 10801 773 463 49 3652 322 237 36 20528
1570 1434 186 4277 258 221 19 14 1 3 0 21207 1775 1547 163 521 46 26 1 39 5

"145" "F8069-1" 126 "DHF" 34976 "SMD Cone Rod Dystrophy" "M"
"probando" 0 1 77 10 22 4 95 9 6 0 69 4 11 2 158 8 35 6 9035 703 601 82 10962 744
479 40 3232 288 212 44 21854 1637 1504 171 4114 291 233 24 22 2 1 0 17774 1508
1352 155 432 39 30 3 39 6

"146" "F8118-1" 127 "HOGBS" 34173 "TEA" "M" "probando" 1 4 77 7 25 6 88
2 5 0 60 3 8 1 160 4 31 1 8281 426 372 71 10227 375 241 30 3521 148 148 29 19678
726 960 164 4140 152 190 29 19 1 1 0 20322 875 1047 141 503 17 23 3 38 7

"147" "F8123-1" 128 "LSS" NA "SMD Cone Rod Dystrophy" "F" "probando"
0 1 88 9 29 2 111 8 6 0 66 5 7 1 175 18 27 1 9439 713 580 64 11295 767 450 32 3618

274 225 39 21651 1488 1442 154 4489 294 229 18 13 0 1 0 19003 1534 1428 158
475 40 20 4 42 3
 "148" "F8124-1" 129 "TN" 37765 "SMD Cone Rod Dystrophy" "F" "probando"
0 1 79 10 22 5 112 9 6 1 75 8 7 1 153 18 38 0 9342 753 604 65 11340 791 474 31
3722 301 218 24 22373 1552 1553 137 4442 283 260 14 22 4 2 0 19718 1589 1554
135 484 33 32 2 35 7
 "149" "F8128-1" 130 "IRS" 41141 "condrodisplasia_punctata" "F" "probando"
0 7 69 10 33 5 98 8 7 5 70 4 15 1 164 9 41 3 8412 547 480 78 9970 575 342 40 3302
209 169 33 19949 1130 1172 160 4159 244 211 22 18 4 2 0 17819 1154 1159 146
411 27 20 5 55 11
 "150" "F8368-1" 133 "MHR" 41014 "craniossinostose" "M" "probando" 0 12 82
6 24 2 79 5 8 3 65 8 16 4 203 14 33 4 9322 508 481 100 11616 489 327 54 1736 102
74 11 2860 135 157 33 1383 80 90 12 6 0 0 0 14675 738 888 112 361 17 19 2 48 9
 "151" "F8382-1" 134 "AJA" 41102 "craniossinostose" "F" "probando" 0 12 101
15 30 2 85 3 9 2 75 3 15 2 191 18 44 3 9706 647 565 71 11779 681 383 48 2048 151
121 13 3458 220 187 10 1606 103 101 12 8 0 1 0 16918 1169 1215 112 418 28 20 2
54 6
 "152" "F8385-1" 135 "TGS" 36189 "Sener" "F" "probando" 0 6 79 9 23 3 94 5 6
1 63 6 9 1 158 12 38 3 8421 879 653 79 10388 913 545 38 3641 355 236 32 19809
1831 1532 173 4171 311 276 19 19 2 3 0 20150 2008 1792 156 498 59 39 2 38 5
 "153" "F8402-1" 136 "IZSS" 39910 "TEA" "M" "probando" 1 3 73 9 25 3 109 8 9
0 69 1 10 3 203 9 42 7 9482 513 595 96 11404 550 358 40 4055 179 202 35 22176
1042 1204 182 4897 201 377 34 20 2 2 0 23143 1308 1359 191 546 45 30 2 44 6
 "154" "F8442-1" 137 "VFP" 33870 "TEA" "M" "probando" 1 4 84 4 26 4 99 18 9
0 60 9 14 3 153 15 35 4 8619 957 757 102 10458 972 596 49 3661 366 282 33 20173
1916 1588 168 4151 348 271 28 21 2 3 0 20650 2159 1831 167 506 61 46 2 49 7
 "155" "F8474-1" 138 "GAO" 34001 "DE" "M" "probando" 0 6 60 4 30 5 91 5 12
3 64 7 10 2 141 5 38 2 8402 468 389 68 10100 442 309 48 3281 148 149 35 19267
813 999 152 4040 177 200 30 18 0 3 1 18678 874 1144 164 531 24 16 3 52 10
 "156" "F8552-1" 139 "TPSP" 41096 "craniossinostose" "M" "probando" 0 14 64
5 33 5 82 3 9 3 52 3 17 6 139 12 37 3 8425 518 636 99 10067 517 383 45 3923 197
205 35 16942 858 956 146 3380 196 268 34 19 2 2 1 14253 804 1067 122 330 15 23
2 59 14
 "157" "F8613-1" 140 "MBG" 31811 "OFD" "F" "probando" 0 9 73 13 28 4 103 3
5 2 71 4 16 5 186 9 33 1 8245 489 463 72 10118 448 292 44 4436 233 243 45 19813
891 1138 140 3904 188 274 32 14 0 1 0 22886 1190 1523 183 506 26 24 1 49 11
 "158" "F8807-1" 142 "MAUN" 40471 "Rolland-Desbuquois" "F" "probando" 0
9 65 7 25 5 100 5 7 3 71 8 14 3 173 10 30 4 8543 488 485 87 10105 459 371 53 4482
174 244 47 19363 969 1228 178 3806 164 245 34 18 2 2 1 22244 1060 1630 219 479
27 38 4 46 11
 "159" "F8823-1" 175 "COPJ" 39564 "TEA" "M" "probando" 1 19 74 2 35 5 60 1
9 1 58 3 15 3 142 9 31 2 8170 446 507 68 10164 404 301 34 1154 75 62 19 1748 96
64 13 1341 56 104 9 4 0 0 0 10413 554 694 97 271 17 10 7 59 9

"160" "F8864-1" 180 "RN2SVA" NA "espondilocostal" "M" "probando" 0 18
53 13 28 5 46 2 1 0 45 7 11 1 110 9 18 1 6397 549 548 66 7991 500 435 34 908 60 61
9 1189 84 69 12 1098 84 127 2 2 0 0 0 8088 658 735 57 204 14 10 3 40 6

"161" "F8959-1" 143 "KMTF" 39712 "noonan" "F" "probando" 0 9 85 11 22 3
106 9 9 0 49 7 10 1 179 16 53 4 8140 720 663 54 9716 776 481 19 4331 338 309 33
19372 1506 1406 105 4169 308 329 20 22 0 1 0 23355 1972 1995 148 512 53 44 5 41
4

"162" "F8978-1" 144 "STFS" 40514 "displasia espondilo-epifisária" "F"
"probando" 0 15 71 5 37 2 76 3 9 1 51 10 19 2 119 11 37 4 7971 613 671 60 9803
522 477 46 3583 198 204 39 15275 1016 1010 138 3416 184 327 26 15 0 2 0 12831
888 1160 106 259 20 18 1 65 5

"163" "F8982-1" 145 "DMFC" 36259 "noonan" "M" "probando" 0 9 90 15 25 3
100 5 10 2 51 2 10 0 161 12 36 0 8114 479 442 34 9922 442 285 14 4341 200 242 19
19180 858 1149 58 3819 161 241 7 26 3 3 0 22466 1125 1550 92 450 30 18 0 45 5

"164" "F9187-1" 166 "AYHC" 40982 "craniossinostose" "M" "probando" 0 20
91 10 52 9 85 6 15 1 71 5 19 3 178 8 45 3 9813 494 658 121 11584 461 449 59 1385
63 69 11 2674 117 146 23 1888 79 144 9 7 0 0 0 13598 651 998 123 356 17 15 1 86
13

"165" "F9320-1" 146 "DACL" 35042 "craniossinostose" "F" "probando" 0 14
56 6 21 6 75 5 11 2 47 4 15 4 101 8 24 2 7550 385 532 79 9561 344 304 33 3685 161
167 34 14112 574 641 110 3157 134 207 17 15 0 1 0 12572 589 765 99 261 11 13 4
47 12

"166" "F9470-1" 147 "MNS" 39025 "craniossinostose" "F" "probando" 0 15 77
9 57 5 91 5 13 1 64 2 22 2 149 9 35 1 8976 667 762 84 10913 604 486 36 3910 236
210 28 18234 1111 1227 168 3894 202 361 25 20 1 3 0 15858 1070 1468 157 348 28
13 0 92 8

"167" "F9485-4" 176 "JFS" 24898 "mao_torta" "M" "probando" 0 19 101 6 29 2
93 5 12 1 77 9 14 0 238 18 44 3 10127 878 749 84 12215 815 567 38 1826 178 112
19 3070 275 198 14 2160 145 198 7 6 0 0 0 17083 1470 1412 106 452 29 23 3 55 3

"168" "F9893-1" 148 "CECS" 41879 "DE" "M" "probando" 0 15 75 7 47 3 80 6 9
2 52 10 23 3 118 14 32 4 8077 600 739 62 10020 576 482 32 3622 197 197 29 16293
997 1032 121 3468 263 354 32 16 3 1 0 13072 898 1197 107 273 24 15 1 79 8

9.22.1 Prefilter

```
#!/bin/bash
```

```
# este script irá filtrar o vcf a partir de um arquivo bed e irá preparar o  
arquivo para ir no R
```

```
# Versão 1: esta versão é útil apenas para 179 indivíduos, fora isso tem que  
mudar o script
```

```
MENSAGEM_HELP=""
```

```
Uso: $(basename "$0") [OPÇÕES]
```

OPÇÕES:

- a vcf original (obrigatório)
- b arquivo bed dos genes
- h Mostra esta tela de ajuda e sai
- V Mostra a versão do programa e sai
- "

```
while test -n "$1"  
do  
    case "$1" in  
        -a)  
            shift  
            Vcf="$1"  
            ;;  
        -b)  
            shift  
            bed="$1"  
            ;;  
        -h)  
            echo "$MENSAGEM_HELP"  
            exit 0  
            ;;  
        -V)  
            echo $(basename "$0")  
            grep "^\# Versão" "$0" | tail -1 | cut -d : -f1 | tr -d '\#'  
            exit 0  
            ;;  
    esac  
    shift  
done
```

```

bgzip $Vcf
tabix -p vcf $Vcf.gz
vcftools --gzvcf $Vcf.gz --bed $bed --out $bed.output --recode --keep-
INFO-all

perl /usr/local/annovar/convert2annovar.pl -format vcf4old
$bed.output.recode.vcf -includeinfo -outfile $bed.avinput

perl /usr/local/annovar/table_annovar.pl $bed.avinput
/usr/local/annovar/humandb/ -protocol
refGene,exac03,kaviar_20150923,hrcr1,esp6500siv2_all,1000g2015aug_all,gn0
mad_exome,60mais609_AC,60mais609_AN,snp135,avsnp147,clinvar_2016030
2,dbnsfp30a,dbnsfp31a_interpro,dbscsnv11 -operation g,f,f,f,f,f,f,f,f,f,f,f,f,f
-argument '-exonicsplicing -splicing 2' --buildver hg19 --otherinfo --
remove -nastring NA -outfile $bed.2017

perl /usr/local/bin/IncluiColunasInfo2017mar.pl $bed.2017

bash /home/saturno/ngs/alunos/tfalmeida/scripts/prep.sh -a
$bed.2017.hg19_multiannoFINALREG2017.txt -b 60mais609_AC -c
60mais609_AN

zgrep ^#CHROM $Vcf.gz > headpac.txt

head -1 $bed.2017.hg19_multiannoFINALREG2017.txt.filtro.txt >
headann.txt

cut -f1-99 headann.txt > header1.txt

cut -f100-102 headann.txt > header2.txt

cut -f1-7 headpac.txt > header3.txt

cut -f8-188 headpac.txt > header4.txt

paste header1.txt header3.txt header2.txt header4.txt > header5.txt

sed -e 's/ACMG\ /Exclusion/XACMG/g' header5.txt | sed -e
's/esp6500siv2_all/Freq1/g' | sed -e 's/1000g2015aug_all/Freq2/g' | sed -e
's/60mais609_AF/Freq3/g' > headp.txt

cat headp.txt $bed.2017.hg19_multiannoFINALREG2017.txt.filtro.txt | sed
'2d' > $bed.anotado.txt

```

```
cut -f1-7,9,19,42,43,55,107-110,112-290 $bed.anotado.txt >  
$bed.anotado.cut.txt
```

```
rm $bed.avinput $bed.2017.* head*
```

9.22.2 Spring

Classificação das variantes

```
#!/bin/bash  
#  
# spring.sh  
# Objetivo: este script pega de uma lista de pacientes e um vcf preparado  
as variantes não-sinônimas para colocar no SPRING  
#  
#  
# Autor: Tatiana Almeida  
#  
# Versão 1: a partir de um vcf preparado este script filtra as variantes não  
sinônimas de qualidade e frequencia descritas para utilizar no SPRING para  
cada paciente  
  
MENSAGEM_HELP=""  
  
Uso: $(basename "$0") [OPÇÕES]  
  
OPÇÕES:  
  
-a vcf anotado (obrigatório)  
-b lista das identificações dos pacientes em uma coluna, sem header  
-c última coluna antes de começar os pacientes (obrigatório, se colocado os  
pacientes)  
-d dá um valor para filtrar a probabilidade genotípica, se não quiser filtrar  
nenhum coloque 0 (obrigatório)  
-e filtro do VAF, se quiser manter as variantes acima de 30% de proporção  
o VAF a ser colocado é 0,6. Se não quiser filtrar colocar 0.  
-f Nome da coluna com 6500exomas  
-g Nome da coluna com 1000Genomas  
-i Nome da coluna de frequencia alélica da amostras (^AF)  
-j Nome da coluna com frequencia alélica dos velhinhos  
-k frequencia de variante benigna
```

- l frequencia de polimorfismo
- m Nome da coluna com contagem alélica da amostra (^AC)
- n Número de variantes privativas para o modelo (indivíduo = 1)
- o escolher a frequencia em que vai calcular as variantes (benigna, polirara, privativa)
- p arquivo ou OMIM para o seed do SPRING
- h Mostra esta tela de ajuda e sai
- V Mostra a versão do programa e sai
- "

```

while test -n "$1"
do
    case "$1" in
        -a)
            shift
            anotVcf="$1"
            if test -z "$anotVcf"
            then
                echo "Faltou vcf anotado"
                exit 1
            fi
            ;;
        -b)
            shift
            listapac="$1"
            if test -z "$listapac"
            then
                echo "Faltou coluna do paciente"
                exit 1
            fi
            ;;
        -c)
            shift
            ultimo="$1"
            ;;
        -d)
            shift
            filter="$1"
            ;;
        -e)
            shift
            vaf="$1"
            ;;
        -f)
            ;;
    esac
done

```

```

shift
evs6500="$1"
;;
-g)
shift
MilG="$1"
;;
-i)
shift
AF="$1"
;;
-j)
shift
AF609="$1"
;;
-k)
shift
benigna="$1"
;;
-l)
shift
poli="$1"
;;
-m)
shift
AC="$1"
;;
-n)
shift
sing="$1"
;;
-o)
shift
frequencia="$1"
;;
-p)
shift
seed="$1"
;;
-h)
echo "$MENSAGEM_HELP"
exit 0
;;

```

```

-v)
echo $(basename "$0")
grep "^\# Versão" "$0" | tail -1 | cut -d : -f1 | tr -d \#
exit 0
;;
esac
shift
done

head -1 "$anotVcf" | sed 's/\./_/g' > header

numultimo=$(awk -F\      '{for(i=1;i<=NF;i++){if($i ~ /'$ultimo') {print i}}}' header)
num6500=$(awk -F\    '{for(i=1;i<=NF;i++){if($i ~ /'$evs6500') {print i}}}' header)
numMilG=$(awk -F\   '{for(i=1;i<=NF;i++){if($i ~ /'$MilG') {print i}}}' header)
numAF=$(awk -F\    '{for(i=1;i<=NF;i++){if($i ~ /'$AF') {print i}}}' header)
num609=$(awk -F\   '{for(i=1;i<=NF;i++){if($i ~ /'$AF609') {print i}}}' header)
numAC=$(awk -F\    '{for(i=1;i<=NF;i++){if($i ~ /'$AC') {print i}}}' header)

echo "O número da última coluna antes dos pacientes é: $numultimo"
echo "O número da coluna do 6500 é: $num6500"
echo "O número da coluna do 1000Genomas é: $numMilG"
echo "O número da coluna do AF é: $numAF"
echo "O número da coluna do AF609 é: $num609"
echo "O número da coluna do AC é: $numAC"
echo "O valor do benigno do benigno é: $benigna"
echo "O valor do benigno do polimorfismo é: $poli"
echo "O valor da variante privativa é: $sing"
echo "O valor do filtro de GL é: $filter"
echo "O valor do filtro de VAF é: $vaf"
echo "As frequencias feitas são: $frequencia #ainda não funciona"

# Retirando apenas as variantes não-sinônimas

grep -w nonsynonymous $anotVcf > $anotVcf.nsy

# Retirando apenas as variantes benignas

```

```

awk -F\" '$$num6500' >= '$benigna' || $$numMilG' >= '$benigna' || 
$$num609' >= '$benigna' {print $0}' $anotVcf.nsy > $anotVcf.nsy.benigna

# Retirando as variantes entre benigna e polimorfismo

awk -F\" 'BEGIN{OFS=""};}{$$num6500' < '$benigna' && $$numMilG' <
'$benigna' && $$num609' < '$benigna' {print $0}' $anotVcf.nsy | awk -F\" 
$$num6500' >= '$poli' || $$numMilG' >= '$poli' || $$num609' >= '$poli' 
{print $0}' > $anotVcf.nsy.poli

# Retirando as variantes raras

awk -F\" '$$num6500' < '$poli' && $$numMilG' < '$poli' && $$num609' 
< '$poli' {print $0}' $anotVcf.nsy > $anotVcf.nsy.rara

# Retirando as variantes privativas
# preciso melhorar essa conta porque pode ter algumas frequencias que
não trazem valor, ou é NA. Checar isso.

awk -F\" '$$numAC' == '$sing' {print $0}' $anotVcf.nsy >
$anotVcf.nsy.privativa

# Criando arquivos SPRING por indivíduo e frequencia

for colpac in $(cat $listapac)
do
    echo $colpac
    numpac=$(awk -F\" '{for(i=1;i<=NF;i++){if($i ~ /'$colpac'/) {print i}}}' 
header)
    echo "$numpac"
    awk -F\" '$$numpac' !~ "0/0|\.|\./\." {print $0}' 
$anotVcf.nsy.$frequencia | sort | uniq > $colpac.tmp.txt
    awk -F\" 'BEGIN{OFS=""};}{split($$numpac,a,":"); if (a[4] > '$filter') 
print $0}' $colpac.tmp.txt > $colpac.filter.txt
    rm $colpac.tmp.txt
    awk -F\" 'BEGIN{OFS=""};}{split($$numpac,a,":"); split(a[2],b,",");
gsub("\.\.", "0", b[1]); if (b[1] != 0) n = b[2]/b[1]; else n = 0; if (n > '$vaf') print 
$0}' $colpac.filter.txt > $colpac.nsy.$frequencia.txt
    rm $colpac.filter.txt
done

# Cortando o arquivo para ser compatível com o SPRING

# Cromossomo Posição Ref Alt

```

```
for colpac in $(cat $listapac)
do
    cut -f1,2,4,5 $colpac.nsy.$frequencia.txt | sed -e 's/^chr/ /g' >
$colpac.$frequencia.spring.txt
done
```

Rodando o SPRING

```
for colpac in $(cat $listapac)
do
    echo $colpac
    /home/saturno/ngs/alunos/tfalmeida/scripts/spring -p -o
$colpac.result -s $seed $colpac.$frequencia.spring.txt
done
```

Choice of rank 1

```
lista="$1"
for pac in $(cat $lista); do cut -f5-9 $pac.result | awk -F\
";" $3 == 1 {print "$pac",$0}' >> rank1_$lista; done
```

9.22.3 Counting

```
library(dplyr)
library(magrittr)
library(tidyr)

### funções
count <- function(df,fr1,fr2,pac){
  soma <- matrix(ncol = 5,nrow = nrow(pac))
  for ( j in 1:nrow(pac)){
    as.character(pac[j,1]) -> n
    df[is.na(df)] <- 0
    tidyr::separate_(df, n, into = c("GT", "AD","DP","GQ","PL"), sep = ":") -> a
    a <- dplyr::filter(a, (GT == "0/1" | GT == "1/1") & as.numeric(DP) > 10 &
      as.numeric(GQ) == 99)
    tidyr::separate(a, AD, into = c("AD1", "AD2"), sep = ",") -> a
    a <- dplyr::filter(a, as.numeric(AD2)/as.numeric(DP) > 0.3)
    i = 2
    frq1 <- dplyr::filter(a, as.numeric(Freq3) >= fr1 | as.numeric(Freq1) >=
      fr1 | as.numeric(Freq2) >= fr1)
    frq2 <- dplyr::filter(a, (as.numeric(Freq3) < fr1 & as.numeric(Freq1) < fr1 |
      & as.numeric(Freq2) < fr1) & (as.numeric(Freq3) >= fr2 | as.numeric(Freq1) |
      >= fr2 | as.numeric(Freq2) >= fr2))
    frq3 <- dplyr::filter(a, as.numeric(Freq3) < fr2 & as.numeric(Freq1) < fr2 |
      & as.numeric(Freq2) < fr2)
    frq4 <- dplyr::filter(a, grepl("^AC=1;", INFO) & as.numeric(Freq3) <= 0 |
      & as.numeric(Freq2) <= 0 & as.numeric(Freq1) <= 0)
    n[i] <- dplyr::filter(frq1, GT == "0/1") %>% nrow()
    i = i+1
    n[i] <- dplyr::filter(frq1, GT == "1/1") %>% nrow()*2
    i = i+1
    n[i] <- dplyr::filter(frq2, GT == "0/1") %>% nrow()
    i = i+1
    n[i] <- dplyr::filter(frq2, GT == "1/1") %>% nrow()*2
    i = i+1
    n[i] <- dplyr::filter(frq3, GT == "0/1") %>% nrow()
    i = i+1
    n[i] <- dplyr::filter(frq3, GT == "1/1") %>% nrow()*2
    i = i+1
    n[i] <- dplyr::filter(frq4, GT == "0/1") %>% nrow()
    i = i+1
    soma[j,] <- cbind(n[1],(as.numeric(n[2]) +
      as.numeric(n[3])),(as.numeric(n[4]) + as.numeric(n[5])),(as.numeric(n[6]) +
      as.numeric(n[7])),as.numeric(n[8]))
  }
}
```

```

}

return(soma)
}

count_total <- function(daf,frq1,frq2,pact){
  fra <- dplyr::filter(daf, grepl("^\$frameshift", ExonicFunc_refGene))
  count(fra,frq1,frq2,pact) -> frat
  as.data.frame(frat) -> frat
  colnames(frat) <- c("ID", "fra_fr1", "fra_fr2", "fra_fr3", "fra_fr4")
  spl <- dplyr::filter(daf, grepl("^\$splicing", Func_refGene))
  count(spl,frq1,frq2,pact) -> splt
  as.data.frame(splt) -> splt
  colnames(splt) <- c("ID", "spl_fr1", "spl_fr2", "spl_fr3", "spl_fr4")
  sto <- dplyr::filter(daf, grepl("^\$stop", ExonicFunc_refGene))
  count(sto,frq1,frq2,pact) -> stot
  as.data.frame(stot) -> stot
  colnames(stot) <- c("ID", "sto_fr1", "sto_fr2", "sto_fr3", "sto_fr4")
  nfr <- dplyr::filter(daf, grepl("^\$nonframeshift", ExonicFunc_refGene))
  count(nfr,frq1,frq2,pact) -> nfrt
  as.data.frame(nfrt) -> nfrt
  colnames(nfrt) <- c("ID", "nfr_fr1", "nfr_fr2", "nfr_fr3", "nfr_fr4")
  nsy <- dplyr::filter(daf, grepl("^\$nonsynonymous", ExonicFunc_refGene))
  count(nsy,frq1,frq2,pact) -> nsyt
  as.data.frame(nsyt) -> nsyt
  colnames(nsyt) <- c("ID", "nsy_fr1", "nsy_fr2", "nsy_fr3", "nsy_fr4")
  syn <- dplyr::filter(daf, grepl("^\$synonymous", ExonicFunc_refGene))
  count(syn,frq1,frq2,pact) -> synt
  as.data.frame(synt) -> synt
  colnames(synt) <- c("ID", "syn_fr1", "syn_fr2", "syn_fr3", "syn_fr4")
  utr5 <- dplyr::filter(vcf, Func_refGene == "UTR5")
  count(utr5,frq1,frq2,pact) -> utr5t
  as.data.frame(utr5t) -> utr5t
  colnames(utr5t) <- c("ID", "utr5_fr1", "utr5_fr2", "utr5_fr3", "utr5_fr4")
  utr3 <- dplyr::filter(vcf, Func_refGene == "UTR3")
  count(utr3,frq1,frq2,pact) -> utr3t
  as.data.frame(utr3t) -> utr3t
  colnames(utr3t) <- c("ID", "utr3_fr1", "utr3_fr2", "utr3_fr3", "utr3_fr4")
  ncRNA <- dplyr::filter(vcf, grepl("^\$ncRNA_exonic", Func_refGene))
  count(ncRNA,frq1,frq2,pact) -> ncRNAt
  as.data.frame(ncRNAt) -> ncRNAt
  colnames(ncRNAt) <- c("ID", "ncRNA_fr1", "ncRNA_fr2", "ncRNA_fr3",
  "ncRNA_fr4")
  miRNA <- dplyr::filter(vcf, grepl("miRNA", Oreganno_Dez15))
  count(miRNA,frq1,frq2,pact) -> miRNAt
}

```



```

aracter","character","character","character","character","character","
character","character","character","character","character","character",
"character","character","character","character","character","character",
"character","character")  

read.delim("bedcromatinaall.txt.anotado.cut.txt", sep = "\t", dec = ".",
header = T, stringsAsFactors = F, check.names = F, colClasses = classvcf) ->
vcf  

read.table("dados_2016_06_15.txt") -> dados  

dplyr::filter(dados, zen_code %in% names(vcf)) -> dadosval
dplyr::filter(dadosval, diagnostico != "NA") -> dadosval  

dadosval$zen_code -> paciente  

as.data.frame(paciente, col.names = "paciente") -> paciente  

paciente -> pact  

count_total(vcf,0.05,0.01,pact) -> cont
merge(dadosval,cont, by.x = "zen_code", by.y = "ID") -> final
final$llof_fr3 <- final$fra_fr3+final$sto_fr3+final$spl_fr3
final$llof_fr4 <- final$fra_fr4+final$sto_fr4+final$spl_fr4  

write.table(final, "final.txt")

```

9.23 Attachment 23

```
---
```

```
title: "Target X Control Analysis"
author: "Tatiana Almeida"
date: "`r format(Sys.time(), '%d %B, %Y')`"
output:
  html_document:
    fig_caption: yes
    keep_md: yes
    toc: yes
  editor_options:
    chunk_output_type: console
---
```

```
```{r bibliotecas, echo=FALSE, message=FALSE, warning=FALSE}
library
library(dplyr)
library(magrittr)
library(ggplot2)
library(tidyr)
library(lazyeval)
library(openxlsx)
library(SNPRelate)
library(gridExtra)
library(devtools)
library(ggbiplot)
library(knitr)
library(caret)
library(rpart)
library(rpart.plot)
library(gmodels)
library(arules)
library(hmeasure)
library(pROC)
library(neuralnet)
library(GGally)
library(corrplot)
library(broom)
library(psych)
library(ggpubr)
library(e1071)
library(kableExtra)
```

```

vcf.fn<-“(raw VCF file)”
```
```

```{r função all_num, echo=FALSE,message=FALSE, warning=FALSE}
all_num <- function(df,status,id) {
  nums <- sapply(df, is.numeric)
  df_nu <- df[,nums]
  df_nu[,sapply(df_nu, function(v) var(v, na.rm=TRUE)!=0)] -> df_nu
  cbind(df[,id],df_nu) -> df_num
  colnames(df_num)[1] <- "paciente"
  #df_num[,sapply(df_num[-1], function(v) var(v, na.rm=TRUE)!=0)]
  return(df_num)
}
```

```{r função multiplot, echo=FALSE,message=FALSE, warning=FALSE}
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots / cols)),
                    ncol = cols, nrow = ceiling(numPlots / cols))
  }

  if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout),
      ncol(layout))))
    # Make each plot, in the correct location
    for (i in 1:numPlots) {

```

```

# Get the i,j matrix positions of the regions that contain this subplot
matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                layout.pos.col = matchidx$col))
}

}
}

```
``{r função aed, fig.height=10, fig.width=15,echo=FALSE,message=FALSE,
warning=FALSE}
aed <- function(df,nc){
 all_num(df,alvo,"zen_code") -> df_num
 df_num[,nc] <- as.factor(df_num[,nc])
 dplyr::rename_(df_num, target = nc) -> rdf
 ggpairs(rdf[,which(names(rdf) %in%
c("target","fra_fr1","spl_fr1","sto_fr1","nfr_fr1","nsy_fr1","syn_fr1","utr5_fr1","u
tr3_fr1","ncRNA_fr1","miRNA_fr1","bnd_fr1","reg_fr1"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p1
 ggpairs(rdf[,which(names(rdf) %in%
c("target","fra_fr2","spl_fr2","sto_fr2","nfr_fr2","nsy_fr2","syn_fr2","utr5_fr2","u
tr3_fr2","ncRNA_fr2","miRNA_fr2","bnd_fr2","reg_fr2"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p2
 ggpairs(rdf[,which(names(rdf) %in%
c("target","fra_fr3","spl_fr3","sto_fr3","nfr_fr3","nsy_fr3","syn_fr3","utr5_fr3","u
tr3_fr3","ncRNA_fr3","miRNA_fr3","bnd_fr3","reg_fr3"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p3
 ggpairs(rdf[,which(names(rdf) %in%
c("target","fra_fr4","spl_fr4","sto_fr4","nfr_fr4","nsy_fr4","syn_fr4","utr5_fr4","u
tr3_fr4","ncRNA_fr4","miRNA_fr4","bnd_fr4","reg_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p4
 ggpairs(rdf[,which(names(rdf) %in%
c("target","fra_fr1","fra_fr2","fra_fr3","fra_fr4","lof_fr3","lof_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p5
 ggpairs(rdf[,which(names(rdf) %in%
c("target","sto_fr1","sto_fr2","sto_fr3","sto_fr4","lof_fr3","lof_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p6
 ggpairs(rdf[,which(names(rdf) %in%
c("target","spl_fr1","spl_fr2","spl_fr3","spl_fr4","lof_fr3","lof_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p7
 ggpairs(rdf[,which(names(rdf) %in%
c("target","nfr_fr1","nfr_fr2","nfr_fr3","nfr_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p8

```

```

ggpairs(rdf[,which(names(rdf) %in%
c("target","nsy_fr1","nsy_fr2","nsy_fr3","nsy_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p9
ggpairs(rdf[,which(names(rdf) %in%
c("target","syn_fr1","syn_fr2","syn_fr3","syn_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p10
ggpairs(rdf[,which(names(rdf) %in%
c("target","utr5_fr1","utr5_fr2","utr5_fr3","utr5_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p11
ggpairs(rdf[,which(names(rdf) %in%
c("target","utr3_fr1","utr3_fr2","utr3_fr3","utr3_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p12
ggpairs(rdf[,which(names(rdf) %in%
c("target","ncRNA_fr1","ncRNA_fr2","ncRNA_fr3","ncRNA_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p13
ggpairs(rdf[,which(names(rdf) %in%
c("target","miRNA_fr1","miRNA_fr2","miRNA_fr3","miRNA_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p14
ggpairs(rdf[,which(names(rdf) %in%
c("target","bnd_fr1","bnd_fr2","bnd_fr3","bnd_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p15
ggpairs(rdf[,which(names(rdf) %in%
c("target","reg_fr1","reg_fr2","reg_fr3","reg_fr4"))],
mapping=ggplot2::aes(colour = target, alpha=.25)) -> p16
return(list(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12,p13,p14,p15,p16))
}

plotpc <- function(df){
 all_num(df,alvo,"zen_code") -> df_num
 select(df_num, -exoma, -paciente, -ID_pac, -nascimento) -> df_num2
 corr.test(df_num2) -> c1
 plot1 <- function() {
 corrplot(c1$r, method = "circle", order = "original", type = "upper",
 sig.level = 0.05, tl.col = "black", tl.srt = 45, tl.cex = 0.7)
 }
 plot2 <- function() {
 corrplot(c1$p, method = "circle", order = "original", type = "full", sig.level
 = 0.05, tl.col = "black", tl.srt = 45, tl.cex = 0.7, cl.lim = c(0,1))
 }
 return(list(plot1=plot1,plot2=plot2))
}

aed2 <- function(df){
 all_num(df,alvo,"zen_code") -> df_num

```

```

df_num[, "alvo"] <- as.factor(df_num[, "alvo"])
dplyr::rename_(df_num, target = "alvo") -> rdf
i = 1
for (nm in names(rdf[6:length(rdf)])){
 ggboxplot(rdf, x = "target", y = nm, color = "target") +
 stat_compare_means(method = "wilcox.test",label.x = 1.5,label.y =
max(df_num[,nm])+1) -> l
 paste("p",i,sep="") -> q
 assign(q,l)
 i = i+1
}
multiplot(p1,p2,p3,p4,p5,p6,p7,p8,p9,cols=3)
multiplot(p10,p11,p12,p13,p14,p15,p16,p17,p18,cols=3)
multiplot(p19,p20,p21,p22,p23,p24,p25,p26,p27,cols=3)
multiplot(p28,p29,p30,p31,p32,p33,p34,p35,p36,cols=3)
multiplot(p37,p38,p39,p40,p41,p42,p43,p44,p45,cols=3)
multiplot(p46,p47,p48, p49,p50, cols=3)
}

aed3 <- function(df){
 all_num(df,kmeans,"zen_code") -> df_num
 df_num[, "kmeans"] <- as.factor(df_num[, "kmeans"])
 i = 1
 for (nm in names(df_num[6:(length(df_num)-1)])){
 my_comparisons <- list(c("1", "2"))
 ggboxplot(df_num, x = "kmeans", y = nm, color = "kmeans") +
 stat_compare_means(comparisons = my_comparisons) -> l
 paste("p",i,sep="") -> q
 assign(q,l)
 i = i+1
 }
 multiplot(p1,p2,p3,p4,p5,p6,p7,p8,p9,cols=3)
 multiplot(p10,p11,p12,p13,p14,p15,p16,p17,p18,cols=3)
 multiplot(p19,p20,p21,p22,p23,p24,p25,p26,p27,cols=3)
 multiplot(p28,p29,p30,p31,p32,p33,p34,p35,p36,cols=3)
 multiplot(p37,p38,p39,p40,p41,p42,p43,p44,p45,cols=3)
 multiplot(p46,p47,p48,p49,p50, cols=3)
}
```

```

```

```{r função hp_lof, echo=FALSE,message=FALSE, warning=FALSE}
hp_lof <- function(df,ct){
 ifelse(df$alvo==1,"affected","control") -> df$alvobi
 if (var(df$lof_fr3) != 0 & var(df$lof_fr4) != 0){

```

```

df$lof_fr3d <- discretize(df$lof_fr3, method = "frequency", categories = ct)
table(dflof_fr3d, dfalvobi) -> m3
prop.table(m3, 2) -> m3p
round(m3p, digits = 3) -> m3p
df$lof_fr4d <- discretize(df$lof_fr4, method = "frequency", categories = ct)
table(dflof_fr4d, dfalvobi) -> m4
prop.table(m4, 2) -> m4p
round(m4p, digits = 3) -> m4p
ifelse(dim(m3)[1] + dim(m3)[2] > 3, chisq.test(m3)[[3]], "sem cálculo de Chi-square") -> f1
round(as.numeric(f1), digits = 3) -> f1
ifelse(dim(m4)[1] + dim(m4)[2] > 3, chisq.test(m4)[[3]], "sem cálculo de Chi-square") -> f2
round(as.numeric(f2), digits = 3) -> f2

df$lof_fr3d <- discretize(df$lof_fr3, method = "fixed", categories =
c(0, 1, max(df$lof_fr3)))
table(dflof_fr3d, dfalvobi) -> m5
prop.table(m5, 2) -> m5p
round(m5p, digits = 3) -> m5p
df$lof_fr4d <- discretize(df$lof_fr4, method = "fixed", categories =
c(0, 1, max(df$lof_fr3)))
table(dflof_fr4d, dfalvobi) -> m6
prop.table(m6, 2) -> m6p
round(m6p, digits = 3) -> m6p
ifelse(dim(m5)[1] + dim(m5)[2] > 3, chisq.test(m5)[[3]], "sem cálculo de Chi-square") -> f3
round(as.numeric(f3), digits = 3) -> f3
ifelse(dim(m6)[1] + dim(m6)[2] > 3, chisq.test(m6)[[3]], "sem cálculo de Chi-square") -> f4
round(as.numeric(f4), digits = 3) -> f4

df$lof_fr3d <- discretize(df$lof_fr3, method = "cluster", categories = ct)
table(dflof_fr3d, dfalvobi) -> m7
prop.table(m7, 2) -> m7p
round(m7p, digits = 3) -> m7p
df$lof_fr4d <- discretize(df$lof_fr4, method = "cluster", categories = ct)
table(dflof_fr4d, dfalvobi) -> m8
prop.table(m8, 2) -> m8p
round(m8p, digits = 3) -> m8p
ifelse(dim(m7)[1] + dim(m7)[2] > 3, chisq.test(m7)[[3]], "sem cálculo de Chi-square") -> f5
round(as.numeric(f5), digits = 3) -> f5

```

```

ifelse(dim(m8)[1]+dim(m8)[2] > 3, chisq.test(m8)[[3]], "sem cálculo de Chi-square") -> f6
round(as.numeric(f6), digits = 3) -> f6

tabela <- rbind(m3, m3p, m4, m4p, m5, m5p, m6, m6p, m7, m7p, m8, m8p)
tabela <- cbind(tabela, ChiSquare =
c(f1, f1, f1, f1, f2, f2, f2, f2, f3, f3, f3, f4, f4, f4, f4, f5, f5, f5, f6, f6, f6))
as.data.frame(tabela) -> tabela
names <- rownames(tabela)
rownames(tabela) <- NULL
tabela <- cbind(names, tabela)
ifelse(tabela$ChiSquare == "NaN", 99, tabela$ChiSquare) ->
tabela$ChiSquare
colnames(tabela) <- c("Number of LoF", "Affected", "Control", "ChiSquare")

options(knitr.table.format = "html")
tabela %>%
mutate(
 ChiSquare = cell_spec(ChiSquare, color = ifelse(ChiSquare < .05, "red",
"black")))
%>%
select(`Number of LoF`, `Affected`, `Control`, `ChiSquare`) %>%
kable(escape = F) %>%
kable_styling(bootstrap_options = c("striped", "hover", "condensed",
"responsive", position = "left")) %>%
collapse_rows(columns = 4) %>%
group_rows("Grouping Method: Frequency", 1, 8) %>%
group_rows("Rare variants", 1, 4) %>%
group_rows("Singletons", 5, 8) %>%
group_rows("Grouping Method: Presence versus Absence", 9, 16) %>%
group_rows("Rare variants", 9, 12) %>%
group_rows("Singletons", 13, 16) %>%
group_rows("Grouping Method: Kmeans", 17, 24) %>%
group_rows("Rare variants", 17, 20) %>%
group_rows("Singletons", 21, 24)
}
}

lof_roc <- function(df){
 cbind(dfzen_code, dfalvo, dflof_fr3, dflof_fr4) %>% as.data.frame() ->
pac_lof
 colnames(pac_lof) <- c("zen_code", "alvo", "lof_fr3", "lof_fr4")
 min(df$lof_fr3) -> mn
 max(df$lof_fr3) -> mx
 matrix(data = NA, nrow = (mx-mn)+1, ncol = 2) -> roc_1
}

```

```

colnames(roc_1) <- c("TVP","TFP")
rownames(roc_1) <- mn:mx
j <- 1
for (i in mn:mx){
 filter(pac_lof, alvo == 1 & lof_fr3 >= i) %>% nrow()/nrow(filter(pac_lof,
alvo == 1)) -> roc_1[j,1]
 filter(pac_lof, alvo == 0 & lof_fr3 >= i) %>% nrow()/nrow(filter(pac_lof,
alvo == 0)) -> roc_1[j,2]
 j <- j +1
}
min(df$lof_fr4) -> mn
max(df$lof_fr4) -> mx
matrix(data = NA, nrow = (mx-mn)+1, ncol = 2) -> roc_2
colnames(roc_2) <- c("TVP","TFP")
rownames(roc_2) <- mn:mx
j <- 1
for (i in mn:mx){
 filter(pac_lof, alvo == 1 & lof_fr4 >= i) %>% nrow()/nrow(filter(pac_lof,
alvo == 1)) -> roc_2[j,1]
 filter(pac_lof, alvo == 0 & lof_fr4 >= i) %>% nrow()/nrow(filter(pac_lof,
alvo == 0)) -> roc_2[j,2]
 j <- j +1
}
plot(roc_1[,2],roc_1[,1], ylab = "True Positive Rate", xlab = "False Positive
Rate",main = "Rare variants",abline(a = 0, b = 1)) -> pl1
plot(roc_2[,2],roc_2[,1], ylab = "True Positive Rate", xlab = "False Positive
Rate",main = "Singleton variants",abline(a = 0, b = 1)) -> pl2
#return(list(pl1,pl2))
}
```

```

```

```{r função spring2, echo=FALSE,message=FALSE, warning=FALSE}
spring2 <- function(df,dados){
 colnames(df) <- c("paciente","gene","line","rank","pvalue","qvalue")
 merge(dados,df, by.x = "zen_code", by.y = "paciente") -> spr
 spr$logqv <- round(-log(as.numeric(spr$qvalue)), digits = 0)
 ggplot(spr) +
 geom_violin(aes(x=as.factor(alvo), y = logqv)) +
 labs(x="", y="log(q-value)", title="Q-value distribution") -> g1
 table(sprzen_code, sprlogqv) %>% as.data.frame()-> a
 colnames(a) <- c("zen_code","logqv","cont")
 merge(dados,a, by= "zen_code") -> b
 filter(b, alvo == 0) -> ntea
 filter(b, alvo == 1) -> tea

```

```

aggregate(data = ntea, cont ~ logqv, mean) -> c
aggregate(data = tea, cont ~ logqv, mean) -> d
merge(c,d, by = "logqv") -> e
colnames(e)<- c("logqv", "Control", "Target")
f <- e[,c(1,2)] %>% cbind("Conrtol")
g <- e[,c(1,3)] %>% cbind("Target")
colnames(f) <- c("log(qvalue)","average","status")
colnames(g) <- c("log(qvalue)","average","status")
rbind(f,g) -> h
ggplot(h) +
 geom_point(aes(x=`log(qvalue)`,y=average,color=status))+

 labs(title="Average of Mutations per individual") -> g2
multiplot(g1,g2, cols =1) -> m1
}
```

```

```

```{r função rl, echo=FALSE,message=FALSE, warning=FALSE}
rl <- function(df,pc,ac){
 df$kmeans <- as.factor(df$kmeans)
 all_num(df,alvo,"zen_code") -> df_num
 dplyr::select(df_num, -exoma, -ID_pac, -nascimento) -> df_num
 fml <- as.formula(paste("alvo ~", paste(colnames(df_num[c(-1,-2)]),
 collapse = "+"))))
 set.seed(1981)
 flag <- createDataPartition(df_num$alvo,p=ac, list=F)
 lrn=df_num[flag,]
 tst=df_num[-flag,]
 if(var(tst$alvo != 0)){
 glm(fml, data = lrn[,-1], family = binomial(link = "logit"), trace = 0) -> lr
 summary(lr) -> s1
 lrf <- step(lr, trace = 0)
 summary(lrf) -> s2
 tidy(lrf) -> t1
 round(t1[,-1], digits = 3) -> t2
 t1[,1] -> t2$term
 tst$pred <- predict(lrf, newdata = tst, type = "response")
 tst$ktea <- discretize(tst$pred, method = "frequency", categories = 5)
 } else {print("Regression not possible")}
 return(list(t2=t2,tst=tst))
}

tbl <- function(t2){
 t2 %>%
 mutate(p.value = cell_spec(p.value, color = ifelse(p.value < .05, "red",
 "black")))) %>%

```

```

dplyr::select(term, estimate, p.value) %>%
kable(escape = F) %>%
kable_styling(bootstrap_options = c("striped", "hover", "condensed",
font_size = 6))}

```
```{r função tree, echo=FALSE,message=FALSE, warning=FALSE}
tree <- function(df,pc){
 df$kmeans <- as.factor(df$kmeans)
 all_num(df,alvo,"zen_code") -> df_num
 dplyr::select(df_num, -exoma, -ID_pac, -nascimento) -> df_num
 ifelse(df_num$alvo == 1,"Target","Control") -> df_num$tea_bi
 fml <- as.formula(paste("tea_bi ~", paste(colnames(df_num)[,-c(1,2,length(df_num))]), collapse = "+")))
 set.seed(1981)
 flag <- createDataPartition(df_num$alvo,p=pc, list=F)
 lrn=df_num[flag,]
 if (var(lrn$alvo) != 0){
 ad1 <- rpart(data = lrn, fml)
 prp(ad1, type=2, extra=100, nn=T, fallen.leaves = T, branch.col = "red",
branch.lty = 5, box.col = c("white","green")) -> a1
 printcp(ad1) -> e1
 } else {
 a1 <- NULL
 e1 <- NULL
 ad1 <- NULL
 }
 return(list(a1,e1,ad1))
}

prune <- function(ad1,cut){
 ad2<-prune(ad1,cp=cut)
 prp(ad2, type = 2, extra=104, nn=T, fallen.leaves = T, branch.col
="red",branch.lty = 5, box.col = c("white",'green')) -> a2
 return(list(ad2,a2))
}

treetest <- function(df,ad,pc){
 df$kmeans <- as.factor(df$kmeans)
 all_num(df,alvo,"zen_code") -> df_num
 dplyr::select(df_num, -exoma, -ID_pac, -paciente, -nascimento) -> df_num
 ifelse(df_num$alvo == 1,"Target","Control") -> df_num$tea_bi
 set.seed(1981)
 flag <- createDataPartition(df_num$alvo,p=pc, list=F)

```

```

tst=df_num[-flag,]
if (var(tst$alvo) != 0){
 prob <- predict(ad,newdata = tst,type = "prob")
 prob[,1] -> tst$prob
 return(prob=tst$prob)}}
```

```{r função rede neural, echo=FALSE,message=FALSE, warning=FALSE}
rede <- function(df,pc){
 df$kmeans <- as.factor(df$kmeans)
 all_num(df,alvo,"zen_code") -> df_num
 dplyr::select(df_num, -exoma, -ID_pac, -nascimento, -paciente) -> df_num
 fml <- as.formula(paste("alvo ~", paste(colnames(df_num[,-1]), collapse =
 "+"))))
 maxs <- apply(df_num[,-1],2,max)
 mins <- apply(df_num[,-1],2,min)
 scaled <- as.data.frame(scale(df_num[,-1], center = mins, scale = maxs -
 mins))
 df_num <- data.frame(alvo=df_num[,1],scaled)
 set.seed(1981)
 flag <- createDataPartition(df_num$alvo,p=pc, list=F)
 lrn=df_num[flag,]
 tst=df_num[-flag,]
 if (var(tst$alvo) != 0){
 set.seed(1234)
 rn1=neuralnet(data = lrn,fml,hidden = c(15,5),lifesign="none",
 linear.output = F, rep = 5)
 pp=tst[,-1]
 xx=compute(rn1, pp)
 out2=xx$net.result
 plot(rn1, rep = "best")
 } else {print("Neural Network Not Possible to compute")}
 return(out2=out2)
}
```

```{r função svm, echo=FALSE,message=FALSE, warning=FALSE}
spvm <- function(df,ac,pc){
 df$kmeans <- as.factor(df$kmeans)
 all_num(df,alvo,"zen_code") -> df_num
 dplyr::select(df_num, -exoma, -ID_pac, -nascimento, -paciente) -> df_num
 fml <- as.formula(paste("alvobi ~", paste(colnames(df_num[,-1]), collapse =
 "+"))))
 df_num$alvobi <- ifelse(df_num$alvo == 1,"target","control")
}

```

```

set.seed(1981)
flag <- createDataPartition(df_num$alvo,p=ac, list=F)
lrn=df_num[flag,]
tst=df_num[-flag,]
if (var(tst$alvo) != 0){
 sv <- svm(fml,data = lrn, kernel = "linear", type = "C-classification",
probability = T)
 pp=tst[,-1]
 pred <- predict(sv, pp, probability = T)
 attr(pred, 'probabilities')[,2] -> tst$pred
 p1 <- ggplot(tst) +
 geom_point(aes(x=as.factor(alvo),y=pred, color=as.factor(alvo)),
position = "jitter") +
 labs(x="Target", y="SVM prediction", color = "Target")
} else{
 print("Impossible to calculate the Support Vector Machine")
}
return(list(tst=tst,p1=p1))
}
```

```

```

```{r função hms, echo=FALSE,message=FALSE, warning=FALSE}
hms <- function(tst,hdf){

```

```

 HMeasure(tst, hdf, severity.ratio = 0.75)$metrics -> met
 round(met, digits = 3) -> met
 met %>%
 mutate(
 AUC = cell_spec(AUC, color = ifelse(AUC > .8, "blue", "black")),
 KS = cell_spec(KS, color = ifelse(KS > .4, "blue", "black")),
 H = cell_spec(H, color = ifelse(H > .4, "blue", "black"))
) %>%
 dplyr::select(H,AUC,KS,TP,FP,TN,FN) %>%
 kable(escape = F) %>%
 kable_styling(bootstrap_options = c("hover", "condensed")))

```

```

plotrc <- function(tst,hdf){
 HMeasure(tst$alvo, hdf, severity.ratio = 0.75) -> hmsp
 plotROC(hmsp,which = 1)
}
```

```

```

```{r função kmeans, echo=FALSE,message=FALSE, warning=FALSE}
km <- function(df){

```

```

all_num(df,alvo,"zen_code") -> df_num
dplyr::select(df_num, -exoma, -ID_pac, -nascimento) -> df_num
set.seed(1981)
k.max <- 15
wss <- sapply(1:k.max,
 function(k){kmeans(df_num[,-1], k, nstart=50,iter.max = 15
)${tot.withinss}})
plot(1:k.max, wss,
 type="b", pch = 19, frame = FALSE,
 xlab="Number of clusters K",
 ylab="Total within-clusters sum of squares")
}

kb <- function(df,kn){
all_num(df,alvo,"zen_code") -> df_num
dplyr::select(df_num, -exoma, -ID_pac, -nascimento) -> df_num
set.seed(1981)
wss <- kmeans(df_num[,-1], kn, nstart=50,iter.max = 15)$cluster
return(wss)
}
```
```
`{r função pcaf, echo=FALSE,message=FALSE, warning=FALSE}
pcaf <- function(df,nome){
 df$kmeans <- as.factor(df$kmeans)
 all_num(df,alvo,"zen_code") -> df_num
 select(df_num, -exoma, -ID_pac, -nascimento, -paciente) -> df_num
 prcomp(df_num[,-1], center = TRUE, scale = TRUE) -> pca_df
 k <- ggbiplot(pca_df, obs.scale = 1, var.scale = 1, groups =
 as.factor(df$alvo), ellipse = TRUE, circle = TRUE)
 k <- k + scale_color_discrete(name = "")
 k <- k + theme(legend.direction = 'horizontal', legend.position = 'top')
 l <- ggbiplot(pca_df, obs.scale = 1, var.scale = 1, groups =
 as.factor(df$alvo), ellipse = TRUE, circle = TRUE, choices = 2:3)
 l <- l + scale_color_discrete(name = "")
 l <- l + theme(legend.direction = 'horizontal', legend.position = 'top')
 grid.arrange(k, l, ncol=2)}
```
```
`{r função pca_vcf, echo=FALSE,message=FALSE, warning=FALSE}
pca_vcf <- function(genofile, out, dados){
 tea_pca<-snpgdsPCA(genofile)
 as.data.frame(cbind(tea_pca$sample.id,tea_pca$eigenvect[,1],tea_pca$eige
 nvect[,2],tea_pca$eigenvect[,3])) -> a
}

```

```

left_join(a,dados, by = c("V1" = "zen_code")) -> data_pca

dplyr::filter(data_pca, !V1 %in% out) -> data2_pca

k <- ggplot(data = data2_pca,
aes(x=as.numeric(as.character(V2)),y=as.numeric(as.character(V3)))) +
 geom_point() +
 geom_text(aes(label = data2_pca$V1)) +
 geom_point(data = data2_pca[data2_pca$alvo == 0], col =3) + # verde
 geom_point(data = data2_pca[data2_pca$alvo == 1], col =2) + # vermelho
 labs(x="PC1", y="PC2")

l <- ggplot(data = data2_pca,
aes(x=as.numeric(as.character(V3)),y=as.numeric(as.character(V4)))) +
 geom_point() +
 geom_text(aes(label = data2_pca$V1)) +
 geom_point(data = data2_pca[data2_pca$alvo == 0], col =3) + # verde
 geom_point(data = data2_pca[data2_pca$alvo == 1], col =2) + # vermelho
 labs(x="PC2", y="PC3")

grid.arrange(k, l, ncol=2)
}

```
```{r função kmcomp, echo=FALSE,message=FALSE, warning=FALSE}
kmcomp <- function(df,nome,up,pc,ct){
 data1 <- NULL
 data2 <- NULL
 data3 <- NULL
 data4 <- NULL
 data5 <- NULL
 data6 <- NULL
 set.seed(1981)
 for (i in 1:up){
 filter(df, kmeans == i) -> df1
 if (var(df1$alvo) != 0){
 if (var(df1$lof_fr3) != 0 & var(df1$lof_fr4) != 0){
 hp_lof(df1,ct) -> lofr
 lof_roc(df1) -> lofroc
 } else {print("No loss-of-fcunction mutation") -> lofr; lofroc <- NULL}
 }
 }
}

```

```

rl(df1,.5,pc) -> rlr
tree(df1,pc) -> astay1
if (!is.null(astay1[[1]])){
treetest(df1,astay1[[3]],pc)$metrics -> trr
} else { ttr <- NULL}
rede(df1,pc) -> rdr
spvm(df1,.5,.1) -> sv
pcaf(df1, nome) -> pcar
plotpc(df1) -> pl
paste("data",i, sep = "") -> a
list(nome,i,print("Perda de função"),lofr,lofroc, print("regressão logística"),
rlr, print("Árvore de decisão"), trr,print("rede neural"), rdr,print("Support
Vector Machine"),sv,print("Componente principal"), pcar, pl$plot1()) -> b
assign(a,b)
}
}
return(list(data1,data2,data3,data4,data5,data6))
}
```

```

```

``{r dados, message=FALSE, warning=FALSE, echo=FALSE}
read.xlsx("../2017_06_15_dados.xlsx",sheet = 1) -> dados
read.table("final.txt") -> final
```

```

# Exploratory Analysis of the variants

## Variants distribution

```

``{r frequências, echo=FALSE,message=FALSE, warning=FALSE}

syn <- aggregate(syn_fr1+syn_fr2+syn_fr3+syn_fr4 ~ zen_code, data =
final, sum)
fra <- aggregate(fra_fr1+fra_fr2+fra_fr3+fra_fr4 ~ zen_code, data = final,
sum)
sto <- aggregate(sto_fr1+sto_fr2+sto_fr3+sto_fr4 ~ zen_code, data = final,
sum)
spl <- aggregate(spl_fr1+spl_fr2+spl_fr3+spl_fr4 ~ zen_code, data = final,
sum)
nsy <- aggregate(nsy_fr1+nsy_fr2+nsy_fr3+nsy_fr4 ~ zen_code, data =
final, sum)
nfr <- aggregate(nfr_fr1+nfr_fr2+nfr_fr3+nfr_fr4 ~ zen_code, data = final,
sum)

```

```

utr5 <- aggregate(utr5_fr1+utr5_fr2+utr5_fr3+utr5_fr4 ~ zen_code, data =
final, sum)
utr3 <- aggregate(utr3_fr1+utr3_fr2+utr3_fr3+utr3_fr4 ~ zen_code, data =
final, sum)
ncRNA <- aggregate(ncRNA_fr1+ncRNA_fr2+ncRNA_fr3+ncRNA_fr4 ~
zen_code, data = final, sum)
miRNA <- aggregate(miRNA_fr1+miRNA_fr2+miRNA_fr3+miRNA_fr4 ~
zen_code, data = final, sum)
bnd <- aggregate(bnd_fr1+bnd_fr2+bnd_fr3+bnd_fr4 ~ zen_code, data =
final, sum)
reg <- aggregate(reg_fr1+reg_fr2+reg_fr3+reg_fr4 ~ zen_code, data = final,
sum)
expl <-
join_all(list(syn,fra,sto,spl,nsy,nfr,utr5,utr3,ncRNA,miRNA,bnd,reg))
colnames(expl) <-
c("zen_code","Synonymous","Frameshift","Stopgainloss","Splicing","Nonsynon
ymous","Nonframeshift","UTR5","UTR3","ncRNA","miRNA","Binding","Regul
atory")
p1 <- ggplot(data = expl, aes(x=Frameshift)) +
 geom_histogram()
p2 <- ggplot(data = expl, aes(x=Stopgainloss)) +
 geom_histogram()
p3 <- ggplot(data = expl, aes(x=Splicing)) +
 geom_histogram()
p4 <- ggplot(data = expl, aes(x=Nonframeshift)) +
 geom_histogram()
p5 <- ggplot(data = expl, aes(x=Nonsynonymous)) +
 geom_histogram()
p6 <- ggplot(data = expl, aes(x=Synonymous)) +
 geom_histogram()
p7 <- ggplot(data = expl, aes(x=UTR5)) +
 geom_histogram()
p8 <- ggplot(data = expl, aes(x=UTR3)) +
 geom_histogram()
p9 <- ggplot(data = expl, aes(x=ncRNA)) +
 geom_histogram()
p10 <- ggplot(data = expl, aes(x=miRNA)) +
 geom_histogram()
p11 <- ggplot(data = expl, aes(x=Binding)) +
 geom_histogram()
p12 <- ggplot(data = expl, aes(x=Regulatory)) +
 geom_histogram()
multiplot(p1,p2,p3,p4,p5,p6, cols =3)
multiplot(p7,p8,p9,p10,p11,p12,cols=3)

```

```
```
```

```
## Kmeans
```

```
```{r kmeans, echo=FALSE,message=FALSE, warning=FALSE}
km(final)
kb(final,2) -> final$kmeans
```

```
#read.table("../ all/kmeans.txt") -> a
#colnames(a) <- "kmeans"
#cbind(final,a) -> final
```

```
```
```

```
## Correlation matrices
```

```
### Correlation matrix for all individuals
```

```
```{r cma, echo=FALSE,message=FALSE, warning=FALSE}
plotpc(final) -> pl
pl$plot1()
```
```

```
### P-value matrix for all individuals
```

```
```{r pma, echo=FALSE,message=FALSE, warning=FALSE}
pl$plot2()
```
```

```
### Correlation matrix for target individuals
```

```
```{r cmt, echo=FALSE,message=FALSE, warning=FALSE}
plotpc(final[final$alvo == 1,]) -> pla
pla$plot1()
```
```

```
### Correlation matrix for control individuals
```

```
```{r cmc, echo=FALSE,message=FALSE, warning=FALSE}
plotpc(final[final$alvo == 0,]) -> plc
plc$plot1()
```
```

```
## Comparison for each variant between affected and control
```

```

```{r ea, echo=FALSE,message=FALSE, warning=FALSE, fig.height=10,
fig.width=15,}
aed2(final)
```

## Comparison for each variant for each kmeans cluster

```{r eac, echo=FALSE,message=FALSE, warning=FALSE, fig.height=10,
fig.width=15,}
aed3(final)
```

# Comparison analysis

## Loss-of-function grouping

```{r hp_lof, echo=FALSE,message=FALSE, warning=FALSE}
hp_lof(final,2)
```

## Loss-of-function ROC curve

```{r lof_roc, echo=FALSE,message=FALSE, warning=FALSE}
lof_roc(final)
```

## Comparison of Missense Mutations

#### Rare mutations

```{r spring_rara, echo=FALSE,message=FALSE, warning=FALSE}
read.table("rank1_rara.txt", sep = "\t", dec = ".",fill = T) -> springr
spring2(springr,dados)
```

#### Singleton mutations

```{r spring_privativa, echo=FALSE,message=FALSE, warning=FALSE}
read.table("rank1_privativa.txt", sep = "\t", dec = ".",fill = T) -> springp
spring2(springp,dados)
```

## Logistic Regression

```

```
```{r regressão, echo=FALSE,message=FALSE, warning=FALSE}
rl(final,.5,.5) -> rl1
kbl(rl1$t2)
```
```

Decicion Tree

```
```{r arvore, echo=FALSE,message=FALSE, warning=FALSE}
tree(final,.5) -> afinal
treetest(final,afinal[[3]],.5) -> tr
```
```

Neural Network

```
```{r rede neural, echo=FALSE,message=FALSE, warning=FALSE}
rede(final,.5) -> nn
```
```

Support Vector Machine

```
```{r svm, echo=FALSE,message=FALSE, warning=FALSE}
spvm(final,.5,.1) -> sp1
sp1$p1
```
```

Classification Performance

```
```{r clp, echo=FALSE,message=FALSE, warning=FALSE}
hdf <- data.frame(LR=rl1$tst['pred'],DT=tr,NN=nn,SVM=sp1$tst['pred'])
colnames(hdf) <- c("LR","DT","NN","SVM")
rl1$tst -> tst
hms(tst$alvo,hdf)
plotrc(rl1$tst,hdf)
```
```

PCA with counting variants

```
```{r pca tabela, echo=FALSE,message=FALSE, warning=FALSE}
pcraf(final, "Principal Component Analysis")
```
```

PCA with raw vcf

```
```{r pca dados brutos, echo=FALSE,message=FALSE, warning=FALSE}
snpgdsVCF2GDS(vcf.fn, "final.gds", method="biallelic.only")
genofile <- openfn.gds("final.gds")
out <- NULL
pca_vcf(genofile,out,dados)
```
```

```
```{r cluster, echo=FALSE,message=FALSE, warning=FALSE}
filter(final, kmeans == 1) -> final1
filter(final, kmeans == 2) -> final2
```
```

Cluster 1

LoF

Loss-of-function grouping

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
hp_lof(final1,2)
```
```

Loss-of-function ROC curve

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
lof_roc(final1)
```
```

Logistic Regression

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
rl(final1,.5,.5) -> rl11
tbl(rl11$t2)
```
```

Decicion Tree

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
tree(final1,.5) -> afinal1
treetest(final1,afinal1[[3]],.5) -> tr1
```
```

Neural Network

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
rede(final1,.5) -> nn1
````
```

```
## Support Vector Machine
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
spvm(final1,.5,.1) -> sp11
sp11$p1
````
```

```
## Classification Performance
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
hdf <-
data.frame(LR=rl11$tst['pred'],DT=tr1,NN=nn1,SVM=sp11$tst['pred'])
colnames(hdf) <- c("LR","DT","NN","SVM")
rl11$tst -> tst
hms(tst$alvo,hdf)
plotrc(rl11$tst,hdf)
````
```

```
## Principal Component Analysis
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
pcaf(final1, "Principal Component Analysis")
````
```

```
# Cluster 2
```

```
## LoF
```

```
### Loss-of-function grouping
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
hp_lof(final2,2)
````
```

```
### Loss-of-function ROC curve
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
lof_roc(final2)
````
```

```
## Logistic Regression
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
rl(final2,.5,.5) -> rl12
kbl(rl12$t2)
```
```

```
## Decicion Tree
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
tree(final2,.5) -> afinal2
treetest(final2,afinal2[[3]],.5) -> tr2
```
```

```
## Neural Network
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
rede(final2,.5) -> nn2
```
```

```
## Support Vector Machine
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
spvm(final2,.5,.1) -> sp12
sp12$p1
```
```

```
## Classification Performance
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
hdf <-
data.frame(LR=rl12$tst['pred'],DT=tr2,NN=nn2,SVM=sp12$tst['pred'])
colnames(hdf) <- c("LR","DT","NN","SVM")
rl12$tst -> tst
hms(tst$alvo,hdf)
plotrc(rl12$tst,hdf)
```
```

```
## Principal Componente Analysis
```

```
```{r, echo=FALSE,message=FALSE, warning=FALSE}
pcaf(final2, "Principal Component Analysis")
```
```

```

# Only Exonic Regions

```{r exonic,, echo=FALSE,message=FALSE, warning=FALSE}
exonic <- c("fra_fr1","fra_fr2", "fra_fr3","fra_fr4","spl_fr1","spl_fr2",
"spl_fr3", "spl_fr4", "sto_fr1", "sto_fr2", "sto_fr3", "sto_fr4", "nfr_fr1", "nfr_fr2",
"nfr_fr3", "nfr_fr4", "nsy_fr1", "nsy_fr2", "nsy_fr3", "nsy_fr4", "syn_fr1",
"syn_fr2", "syn_fr3", "syn_fr4","lof_fr3","lof_fr4")
dd <- c("zen_code", "ID_pac", "iniciais", "nascimento", "diagnosticos", "sexo",
"status", "alvo", "exoma","kmeans")

fe <- c(dd,exonic)

finalex <- final[,which(names(final) %in% fe)]
```
## LoF

#### Loss-of-function grouping

```{r, echo=FALSE,message=FALSE, warning=FALSE}
hp_lof(finalex,2)
```

#### Loss-of-function ROC curve

```{r, echo=FALSE,message=FALSE, warning=FALSE}
lof_roc(finalex)
```

## Logistic Regression

```{r, echo=FALSE,message=FALSE, warning=FALSE}
rl(finalex,.5,.5) -> rlex
tbl(rlex$t2)
```

## Decision Tree

```{r, echo=FALSE,message=FALSE, warning=FALSE}
tree(finalex,.5) -> afinalex
treetest(finalex,afinalex[[3]],.5) -> trex
```

```

```

## Neural Network

```{r, echo=FALSE,message=FALSE, warning=FALSE}
rede(finalex,.5) -> nnex
```

## Support Vector Machine

```{r, echo=FALSE,message=FALSE, warning=FALSE}
spvm(finalex,.5,.1) -> spex
spex$p1
```

## Classification Performance

```{r, echo=FALSE,message=FALSE, warning=FALSE}
hdf <-
data.frame(LR=rlex$tst['pred'],DT=trex,NN=nnex,SVM=spex$tst['pred'])
colnames(hdf) <- c("LR","DT","NN","SVM")
rlex$tst -> tst
hms(tst$alvo,hdf)
plotrc(rlex$tst,hdf)
```

## PCA with exonic variants

```{r, echo=FALSE,message=FALSE, warning=FALSE}
pcaf(finalex, "Principal Component Analysis")
```

# Only Regulatory Regions

```{r regulatory, echo=FALSE,message=FALSE, warning=FALSE}
finalrg <- final[,-which(names(final) %in% exonic)]
finalrg$lof_fr3 <-
finalrg$ncRNA_fr3+finalrg$miRNA_fr3+finalrg$bnd_fr3+finalrg$reg_fr3
finalrg$lof_fr4 <-
finalrg$ncRNA_fr4+finalrg$miRNA_fr4+finalrg$bnd_fr4+finalrg$reg_fr4
```

## All regulatory Regions
```

```
### All regulatory grouping  
```{r, echo=FALSE,message=FALSE, warning=FALSE}  
hp_lof(finalrg,2)
```
```

```
### All regulatory ROC curve  
```{r, echo=FALSE,message=FALSE, warning=FALSE}  
lof_roc(finalrg)
```
```

```
## Logistic Regression  
```{r, echo=FALSE,message=FALSE, warning=FALSE}  
rl(finalrg,.5,.5) -> rlg
kbl(rlg$t2)
```
```

```
## Decicion Tree  
```{r, echo=FALSE,message=FALSE, warning=FALSE}  
tree(finalrg,.5) -> afinalrg
treetest(finalrg,afinalrg[[3]],.5) -> trrg
```
```

```
## Neural Network  
```{r, echo=FALSE,message=FALSE, warning=FALSE}  
rede(finalrg,.5) -> nnrg
```
```

```
## Support Vector Machine  
```{r, echo=FALSE,message=FALSE, warning=FALSE}  
spvm(finalrg,.5,.1) -> sprg
sprg$p1
```
```

```
## Classification Performance  
```{r, echo=FALSE,message=FALSE, warning=FALSE}  
hdf <-
data.frame(LR=rlg$tst['pred'],DT=trrg,NN=nnrg,SVM=sprg$tst['pred'])
```

```
colnames(hdf) <- c("LR","DT","NN","SVM")
rlrg$tst -> tst
hms(tst$alvo,hdf)
plotrc(rlrg$tst,hdf)
```
## PCA with regulatory variants
```{r, echo=FALSE,message=FALSE, warning=FALSE}
pcaf(finalrg, "Principal Component Analysis")
```

```

9.24 Attachment 24

Table for the variants abbreviations

| Type of variants | |
|------------------|-----------------------------------|
| fra | Frameshift |
| sto | Stopgain/loss |
| spl | Canonical splicing (+1,+2,-1,-2) |
| nfr | Nonframeshift |
| nsy | Nonsynonymous |
| syn | Synonymous |
| utr3 | 3'-UTR |
| utr5 | 5'-UTR |
| ncRNA | Noncoding RNA |
| miRNA | Micro RNA |
| bnd | Transcription binding sites |
| reg | Regulatory sites |
| lof | Frameshift+Splicing+Stopgain/loss |
| Frequencies | |
| fr1 | Above 0.05 |
| fr2 | Between 0.05 and 0.01 |
| fr3 | Below 0.01 |
| fr4 | Singletons |