Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Annelise Arth Date

22q11 Deletion Syndrome: B-cell Gene Expression Profiles and Potential Links to ASD and Schizophrenia

By

Annelise Arth
Master of Public Health

Epidemiology

Bradley Pearce, PhD

Committee Chair

22q11 Deletion Syndrome: B-cell Gene Expression Profiles and Potential Links to ASD and Schizophrenia

By

Annelise Arth
B.S.
Purdue University
2011

Thesis Committee Chair: Bradley Pearce, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2014

Abstract

22q11 Deletion Syndrome: B-cell Gene Expression Profiles
and Potential Links to ASD and Schizophrenia

By Annelise Arth

22q11.2 Deletion Syndrome (22q11DS) is associated with diverse neurobehavioral characteristics including mild intellectual impairment, reduced spatial reasoning, delayed speech and motor development, lowered executive functioning, autism spectrum disorder (ASD), and attention-deficit hyperactivity disorder. There is substantial-inter-individual variation in the neuropsychiatric challenges faced by these patients. Importantly, up to 30% of people with 22q11DS are diagnosed with schizophrenia by adulthood. The interaction of genetic and environmental factors and their impact on neurodevelopment is not yet understood. This study compared B-cell expression profiles of 12 individuals with 22q11DS and 12 controls in a RNA microarray to i) examine expression levels of genes in 22q11 deletion region ii) examine differentially expressed genes using Ingenuity Pathways Analysis (IPA) software to determine significant cellular and metabolic pathways iii) test for an association between expression of significant genes and externalizing and internalizing behavior. IPA analysis revealed that ubiquitination pathways played a central role in molecular networks predicted by our differential gene expression patterns. Disruptions in the ubiquitination pathway have previously been associated with idiopathic schizophrenia. A dysregulated gene in our study, UFD1L, may be a link to the etiology of this disease among 22q11 individuals. Other dysregulated genes include HGF and BASP1, associated with ASD and idiopathic schizophrenia respectively. No significant relationship between expression levels and behavior was found. This study uncovers potential mechanisms by which disruption of complex molecular pathways can lead to ASD or schizophrenia in this high-risk population.

22q11 Deletion Syndrome: B-cell Gene Expression Profiles and Potential Links to ASD and Schizophrenia

By

Annelise Arth
B.S.
Purdue University
2011

Thesis Committee Chair: Bradley Pearce, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2014

Acknowledgments

I would like to thank my mentor and thesis advisor, Dr. Brad Pearce, for his expertise, wisdom, advice, and support throughout this entire project.

I would also like to thank Dr. Myfanwy Hopkins for her help and assistance on this thesis. The laboratory analysis was supported by the essential funding from NIMH, grant 1 R21 MH083138-01A1. Additional data was collected by a grant from the Emory Neuroscience Initiative, and support was provided by the Emory NINDS core facilities and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000454.

Technical assistance was also supplied by Christie Harwell, Wyatt Wilson and Erica Smearman.

Table of Contents

Introduction

22q11.2 deletion syndrome (22q11DS) is the second most common chromosomal disorder and is estimated to affect 1 in 2000-5950 live births [1, 2]. Multisystem signs and symptoms of 22q11DS vary markedly between cases. Clinical presentations can include cardiac malformations, aplasia or hypoplasia of the thymus, hypoparathyroidism, immunodeficiency, dysmorphic facial features, and variable gastrointestinal, renal, and audiological problems [3].

Learning difficulties affect the majority of individuals with 22q11DS. Common manifestations include delayed onset of speech and motor skills. IQ is variably decreased in 22q11DS, and verbal IQ is generally 6-8 points higher than performance IQ [4]. In terms of behavior, children with 22q11DS can be affected by social difficulties, anxiety, obsessive compulsive disorder, or mood swings. Most remarkably, up to 50% of those with 22q1DS are diagnosed with an Autism Spectrum Disorder, including Autism, Asperger's, and PDD-NOS [3]. Later onset conditions include bipolar disorder and schizophrenia, which can affect up to 30% of cases [4].

Most cases of 22q11DS are caused by de novo deletions. Approximately 90% of cases are caused by a 3 Mb deletion while most of the remaining cases are caused by a 1.5 Mb deletion. The length of the deletion, however, does not clearly correlate with phenotype severity [5]. During development, it has been proposed that the diminished expression of subsets of genes at mesenchymal and epithelial sites leads to morphogenic problems in the brain, face, heart, and limbs [6]. However, while some features of the disorder are present at birth, the immune

and neurodevelopmental components evolve over a period of several years, or even decades.

While some genes have been revealed to cause certain phenotypes, such as TBX1 with cardiac and parathyroid manifestations, the cognitive and behavioral symptoms have not been linked with specific genes or biological mechanisms. Disruptions in higher-order neurological networks, such as the heteromodal association network and prefrontal striatal network, have been hypothesized to play a role in behaviors commonly seen in 22q11 individuals [4]. There is no evidence that the length of the deletion gives rise to specific cognitive and behavioral profiles [2].

A driving hypothesis of this thesis is that cellular and molecular pathways that utilize proteins encoded by 22q11 will be disrupted, and that this can be examined on a genome-wide basis by measuring gene expression. Furthermore, we hypothesize that individual differences in the gene expression profiles and molecular pathways affected in 22q11DS will correlate with a subset of neuropsychological deficits.

Using whole genome B-cell expression from children with 22q11 deletion syndrome, this study addresses the following research questions:

1. Using available expression data (collected in Dr. Pearce his lab) in order to compare expression profiles between 22q11 cases and healthy controls, which genes are differentially expressed and how are these genes connected in molecular pathways using systems biology software?

2. Based on molecular ensembles (closely interconnected networks of gene products) from data described above, is there an association between expression of key genes participating in these ensembles and neuropsychological outcomes?

Background

I. Phenotype Characterization

The manifestations of 22q11 deletion syndrome are highly variable. Individuals may have a few or many problems including cardiac defects, facial dysmorphia, cleft palate, immunodeficiency, mild intellectual impairment, hypoplasia or aplasia of the thyroid or parathyroid, and hypernasal speech, among others. Because of the range and irregularity of symptoms between patients, historically, the syndrome was classified as DiGeorge syndrome, velocardiofacial syndrome, or conotruncal anomaly face (CTAF) syndrome. In the early 1990's, studies showed that a majority of these classified individuals had a deletion on the long arm of chromosome 22 at band 11 [7]. Cases which are characterized by mild or atypical phenotypes may never be diagnosed. Thus, population estimates are most likely underestimates [8].

Congenital heart disease is found in a majority of patients, estimated between 75-80%. Oftentimes, a cardiac defect found in newborns is the first evidence of the syndrome. Typical cardiac problems include conotruncal, septal, and aortic arch abnormalities[9].

Facial dysmorphia is common in patients, and oftentimes is recognized by toddlerhood. A majority of patients experience velopharyngeal insufficiency, which accounts for hypernasal speech [9]. Speech difficulties—including problems with phonation and language acquisition and comprehension—are common. Up to 45% have conductive hearing loss [8].

Aplasia or hypoplasia of the thymus is thought to be the proximal cause of lifelong immune problems. Up to 80% of 22q11 DS patients have immune system abnormalities, mainly caused by lowered T-cell count [9]. There is also an increased prevalence of IgA deficiency, estimated between 2%-30% of cases [8]. Many individuals experience frequent and prolonged viral infections and respiratory bacterial infections. Aplasia or hypoplasia of the parathyroid causes low levels of calcium in the blood, which increases the chance of seizures.

II. Neuropsychological Summary

Learning disabilities are common in individuals with 22q11 deletion syndrome, albeit with a wide spectrum. Delays in motor development and speech are evident in many cases. Up to 50% of affected children show significant motor delays, while at least 50% have speech/language impairments [3]. Full-scale IQ scores are generally below average, with verbal scores higher than performance. Between 34%-53% of patients are diagnosed with intellectual disability [3]. Deficits in visuospatial ability and difficulty with mathematics are common. Other characterizing limitations include poor executive function, working memory, emotion identification, and ability to comprehend intentions of others [10].

Between 14%-50% of patients fall into an autistic spectrum disorder (ASD), including autistic disorder, Asperger's syndrome, and pervasive developmental disorder not otherwise specified [10]. There is an increased prevalence of attention deficit hyperactivity disorder, mood disorders, and anxiety disorders, including obsessive-compulsive disorder. Mood disorders

include major depressive disorder, dysthymia, and bipolar disorder. Up to 30% of these individuals develop schizophrenia by adulthood, and it is estimated that 22q11DS patients account for 1%-2% of all schizophrenia cases [10]. A sharp decline in attention and executive function can indicate the prodromal stage of later psychosis [3].

III. Genetics

A majority of people with 22q11DS have a 3 megabase (Mb) deletion, which contains approximately 75 genes. Less common is a 1.5 Mb deletion, accounting for 8-10% of cases, which contains 35 genes. It is thought that this subset of genes contains the basic set required for phenotypic characterization [10]. 94% of 22q11 deletions are de novo, presumably caused by an aberrant meiotic exchange event. The structure of the region is marked by several sets of large low copy number repeats, making it prone to misalignment during meiosis and thus rearrangement of the chromosome. Figure 1 shows the 3 MB typically deleted region, which is believed to arise from centromeric breakpoint, LCR22A, and telomeric breakpoint LCR22D. Breakpoints may vary between individuals. The typically deleted region in 22q11DS has been sequenced yet there is still some ambiguity as to the genes affected, in part because of individual differences in the precise breakpoints.

Genomic coordinates of LCR22A, based from build GRCh37/hg19, are 18,640,000–18,910,000, which contains 4 genes (centromeric to telomeric) — USP18, GGT3P, DGCR6, PRODH. Likewise, genomic coordinates of LCR22D are 21,470,000–21,920,00021, which contains 11 genes (centromeric to telomeric)—

BCRP2, FAM230B, POM121L8P, RIMBP3C, RIMBP3B, HIC2, TMEM191C, PI4KAP2, RIMBP3B, RIMBP3C, UBE2L3 [11]. A qPCR study of 44 adults with 22q11 showed the LCR22A breakpoint includes USP18 [12]. The LCR22D breakpoint rarely extends through HIC2 (5/110 cases) [13], but is believed to include RIMBP3B in most cases [14].

IV. Gene-phenotype relationships in 22q11DS

The mechanisms by which the 22q11 deletion produces the diversity of clinical phenotypes remain unclear. However, two possibilities have been ruled unlikely—an individual gene producing the entire phenotypic spectrum, or multiple genes in which each causes a distinct and independent phenotype [7].

There have been isolated incidences where patients, who were misdiagnosed with velocardiofacial syndrome, actually had a point mutation in TBX1. [8] During development, TBX1 is expressed in mesenchyme and endodermal pouches, which give rise to the face, parathyroid, and thymus.. As a consequence of reduced expression of this gene, these structures develop abnormally. Aberrant migration of neural crest cells account for many phenotypic characteristics which include the jaw, heart, nervous system, thymus, and parathyroid gland [3].

The particular genes that are responsible for the neuropsychiatric phenotype of 22q11DS have proved difficult to discern, especially since autism spectrum disorders are typically diagnosed in these patients in early childhood, and schizophrenia usually becomes clinically apparent in the late teens or early adulthood. Anatomically, there is an overall decrease in brain volume and

decrease in cortical regions in the frontal, parietal, occipital and temporal lobes in 22q11DS. One hypothesis to explain psychosis is the reduction of white matter and excessive synaptic pruning thereby producing a combination of reduced cortical connectivity and synaptic plasticity [10].

A recent case-control gene expression study on lymphoblast samples correlated hemizygosity of the catechol-O-methyltransferase (COMT) gene with psychosis and lower IQ scores [3]. Several other genes, such as PRODH, ZDHHC8, and DGCR8 have been implicated indirectly in cognitive and psychiatric phenotypes. DGCR8 codes for an important component of miRNA processing, and also contains miRNAs (mir-185, mir-1306, mir1286). An individual miRNA has the capability of silencing multiple mRNAs, while mRNAs have the ability to be simultaneously controlled by multiple miRNA. By decreasing expression of DGCR8, production of a subset of mature miRNAs decreases which could potentially cause a widespread up-regulation of several downstream genes [14].

Because one copy of the 22q11 deleted region remains intact in 22q11DS syndrome, we hypothesized that some cellular and molecular pathways that utilize proteins encoded by 22q11.2 will be disrupted while others may remain relatively intact. The intact copy of the 22q11 genes also contains numerous SNPs that may determine inter-individual differences in gene expression, and hence relative compensation for the deletion. This concept is related to the question of why some individuals with 22q11DS will develop schizophrenia or autism, whereas others with the same deletion do not. Schizophrenia and ASD involve an interaction of genetic and environmental factors that disrupt inter- and intra-

cellular molecular pathways that in turn influence largely undefined neurodevelopmental and neurochemical processes.

The current study used microarray analysis to determine patterns of differential gene expression between 22q11DS individuals and controls. Rather than studying the impact of hemizygous deletion on phenotypic outcome, RNA analysis allows for a detailed view of expressed gene utilization in complex pathways. An advantage of using genome-wide mRNA profiles is the capability of simultaneous assessment of transcript expression.

This study used mRNA derived from B-cells, rather than peripheral blood mononuclear cells or T-cells, because B-cell numbers are unchanged in 22q11DS patients compared to age-matched controls [15]. In contrast, on average T-cell counts are lower in 22q11DS individuals than the general population despite considerable overlap and a tendency to normalize by early childhood [15, 16] . Moreover, studies analyzing PBMCs without first isolating lymphocyte subsets and allowing time in culture run the risk of capturing transient state variables related to the circulatory milieu. While there may be subtle defects in antibody repertoires in 22q11DS patients (as might be expected in cells with disrupted cell physiology), the culture conditions we used favor polyclonal (non-antibody-specific) B-cells.

Bioinformatics software was then used to identify influential signaling and metabolic pathways which may help to explain underlying intermolecular pathways that may underlie of the variety of phenotypes seen in 22q11DS.

Finally, in order to understand how neuropsychological outcomes might be influenced by gene expression, correlation of expression of individual genes

and neuropsychological outcomes was performed. To date, no other study on 22q11DS has used such an approach.

Methods

Research Participants

The study sample consisted of 12 patients with 22q11.2 deletion and 12 controls. To collect and assay-cells from fresh blood, Dr. Pearce collaborated with Drs. Opal Ousley, Lisa Kobrynski, Elaine Walker, Joseph Cubells, Erica Duncan and Karlene Coleman (who have various overlapping appointments in Autism Center, Emory Department of Psychiatry, Emory Department of Genetics, RSPH). Additional help was provided by The Atlanta Clinical and Translational Science Institute, and Emory Shared Resource Core (the ENNCF grant, NINDS) and Drs. Gregory H. Doho, and Robert Isett.

Drs. Joseph Cubells and Opal Ousley recruited children and young adults who have genetically-confirmed 22q11.2 deletion syndrome (22q11DS cases). To obtain a group of control participants who were age and sex appropriate comparisons to the 22q11DS group, we collaborated with Dr. Elaine Walker, who was recruiting controls for her study of prodromal schizophrenia symptoms, and Dr. Erica Duncan, who was recruiting a control group for her studies of schizophrenia. The inclusion criteria for control participants included that they did not have a current or past DSM IV Axis 1 diagnosis (by SCID) or known chromosomal disorder. The study was approved by Emory IRB (study #IRB00045086).

Behavioral Assessments

Each subject received the Child Behavior Checklist (CBCL) if the age was 4-18 or the Adult Behavior Checklist (ABCL) if the age was 18 or over; both were

completed by a parent or other care giver. Both measurement tools assess social, school, and total problems; anxiety; and attention deficit hyperactivity problems. Neuropsychological data of cases was provided by Dr. Opal Ousley (Emory Autism center and Marcus Autism Center), Dr. Elaine Walker (Emory Psychology Department), and Dr. Erica Duncan (Emory Psychiatry Department).

Blood and RNA Isolation

Peripheral blood mononuclear cells were isolated from heparinized peripheral blood by Histopaque density gradient centrifugation using Accuspin tubes (Sigma) [Colligan, 2007 #5703]. While there may be subtle defects in antibody repertoires in 22q11DS patients (as might be expected in cells with disrupted cell physiology), we performed cultures that favor polyclonal (non-antibody-specific) B-cells, which is a well-established technique [Colligan, 2007 #5703]. This techniques uses immunomagnetic positive selection of B-cells, which was performed by incubation of cells with CD19 MicroBeads, and purifying them on an LS column coupled to the SuperMACS separator according to manufactures instructions (Miltenyi Biotec). To stimulate proliferation, isolated B-cells were plated in a 96 well plate at $1x10^5$ cells/well in RPMI-1640 media supplemented with 10% FBS, 2mM L-glutamine, 100 U/ml Penicillin/100ug/ml streptomycin (Gibco), Staphylococcus aureus Cowan I (Calbiochem) 1:10,000, and 500 U/ml IL-2 (PeproTech Inc) and incubated at 37^oC in 5% CO2 (Colligan, 2007). After 3 days of proliferation in vitro, cells were expanded in a T25 flask and cultured an additional 7 days, pelleted by centrifugation, and frozen at -80 until used to isolate RNA (see below).

Total RNA was purified using organic extraction and solid-phase extraction with the mirVANA kit (Life Technologies, Grand Island, NY) per the manufacturer's protocol. RNA integrity was assessed for each sample on an Agilent Bioanalyzer (Agilent Technologies), which yields an RNA integrity number (RIN) from the electrophoretic trace. RIN values from 1 to 10 can be obtained, where 10 indicates the highest possible RNA quality.

Gene Expression Profiling

Hybridization was performed on the human HT12 (v4.0) chip (Illumina, San Diego, CA). RNA was amplified with the Ovation RNA Amplification System (Nugen). To determine fluorescence intensity, the beadchips were scanned on the Illumina BeadArray Reader. After initial scanning of array, the local background was subtracted from the signal intensity for each bead. Bead summary data was generated by averaging the intensities of all beads displaying a given probe. Background subtraction, of the average signal of the negative control probes for a given array from all other probe signals, was performed on bead summary data to adjust for between-array differences in non-specific signal intensities. In order to account for potential bias from systematic technical differences between chips, the data were quantile normalized using GenomeStudio software from Illumina. Quantile normalization is a common technique that forces the probe intensities of each chip into one standard distribution and subsequently matches every probe value to that distribution [17].

Microarray Data Analysis and Bioinformatics

After quantile normalization the data was reduced to 45,727 probes and was further examined in three ways (Figure A1). To summarize expression levels for each gene we considered techniques that give slightly different sets of significantly-expressed genes.

A. Probe ID:

All those probes detected (expressed at <0.05 using the Bead Array Algorithm) were kept. These 21,242 lines of probe-level, using the Illumina probe identifier (for example ILMN_1701731), were loaded into ArrayStar. Since there are often multiple probes for a given gene, this analysis only took probe-level data into account. To compare cases and controls, the geometric mean for each expressed probe was determined for all 22q11DS cases or all controls.

We also considered another alterative way to summarize probe data. The 21,242 lines of probe-level data, with corresponding accession number, were loaded into ArrayStar. ArrayStar subsequently sorted and grouped data by accession number and performed geometric mean on 19,247 lines of data by 22q11 or control class. Given that these gave similar results we did not use the accession number data in subsequent analysis.

B. Symbol:

To obtain a summarized gene-level expression value we created a list of detected probes such that only probes with a detection p-value of ≤ 0.05 (using 0.054444) were kept. These 21,892 lines of probe-level data, with corresponding symbol information, were loaded into ArrayStar. ArrayStar subsequently sorted

and grouped data by symbol and performed geometric mean on 16,410 lines of gene-symbol data by 22q11 or control class. However, 3,142 probe IDs lacked symbol information and were consequently not included in the analysis.

Student t-tests were used to determine significance of gene expression between cases and controls, using DNA ArrayStar.

Ingenuity Pathway Analysis (IPA) (Ingenuity Systems Redwood City, CA) was used to determine functional relationships and molecular pathways associated with a variety of cellular functions. Specifically, IPA software was used to identify genes and small molecules (e.g. steroid hormones) involved in cellular and molecular pathways represented by the observed expression pattern that differentiates 22q11DS cases versus controls. The biological networks in the Ingenuity database encompass millions of scientist-curated interactions that include genes, proteins, lipids, small molecules (i.e. neurotransmitters), cells, tissues, and diseases. The software uses an algorithm to map genes with differential expression between cases and controls into omnibus models that are constructed of interactions between this set of user-entered molecules (in this case genome wide expression levels for our B-cells), and the genome wide interaction data in the IPA knowledgebase. The modeled molecular pathways are ranked statistically and depicted visually to show network-based mechanistic associations with other uploaded biomarkers and partner molecules in the database [18].

We used advanced bioinformatics to model which molecular networks were most impacted by 22q11DS. By drawing on its knowledgebase, the IPA software creates a connectivity map that graphically models the optimal

interactome of cellular and molecular networks based on up and down regulated genes in our Illumina expression dataset.

Statistical Analysis

Differences between 22q11 and control groups were assessed using t-tests or chi-square or Fisher tests (as appropriate) for categorical variables. Pearson correlations were computed to evaluate associations between gene expression and neuropsychological data (CBCL or ABCL scores). Multivariate linear regression, controlling for age and sex was performed to determine the relationship of gene expression levels of certain genes to several candidate cognitive behaviors. SAS (Cary, NC) was used for statistical analysis.

Results

Demographics

The two groups differed in age, with means (±SD) of 23.3 ± 5.6 years for controls and 14.2 ± 7.3 years for 22q11 ($p=0.003$). Although the means differed between these groups, there was substantial overlap in the range (13-29 for controls and 7-27 for 22q11). The groups were comparable in sex distribution and RIN ($p>0.05$), but varied by race ($p<0.001$)

Analysis by Probe ID

Of the 21,241 expressed probes, 552 had differential expression with p-value ≤ 0.05, 61 had differential expression with p-value ≤ 0.01, and 2 had differential expression with p-value ≤ 0.001. Overall, there were fewer down-regulated genes than up-regulated genes—9,992 and 11,249 respectively. Of the probes with p-value ≤0.05, 177 were down-regulated and 375 were up-regulated in cases versus controls. Figure 2 shows a volcano plot to demonstrate the relationship between differential fold change and p-value.

The differential expression ratio and p-value of each gene in the 3MB deleted region is shown in Table 2. Six genes in this region are significantly down-regulated—DGCR2 ($p=0.002$), CDC45 ($p=0.003$), MED15 ($p=0.012$), ZNF74 ($p=0.012$), UFD1L ($p=0.02$), and RANBP1 ($p=0.04$). None of the genes in this region were significantly up-regulated, and all except three had fold change values of <1 (i.e. lower in 22q11DS then in control subjects).

45 probes (shown in red in Figure 2) had a p-value <0.05 and fold change ≥ |1.5| and were further examined using IPA. Two of the top 45 probes represent genes that are in the 22q11 3MB deletion region—CDC45 (fold change=0.53, p=0.0003) and RANBP1 (fold change=0.64, p=0.04). Of the 45 investigated Illumina probes, IPA was able to recognize gene annotations for 29 probes as shown in Table 3. IPA was unable to recognize 16 probes by Illumina Probe ID (Table 4).

The networks discovered by the IPA analysis were ranked based on their composite score that represents the log of the p-value for the likelihood these molecules would be found together by chance (Figure A2). The top ranked network (Score 35, i.e. $p=10^{-35}$) involves hematological system development and function, humoral immune response, and tissue morphology. As shown in Figure A3, this network is centered around B-cell function but also contains a number of molecules relevant to schizophrenia and autism, including HLA-DMB, CD74, and RFTN1 as well as the 22q11.2 gene, RANBP1 [19-21]. The second best ranked network (Score 25, i.e. $p=10^{-25}$) was examined further in network diagrams generated by the IPA algorithm (Figure 3). It is evident in this diagram this network involves a central role for ubiquitination, a signaling mechanism for protein degradation or promotion or prevention of protein interaction, which can affect cell cycle, DNA replication, recombination, and repair (Figure 3).

Analysis by Symbol

Of the 16,410 genes grouped by symbol, 473 had differential expression with p-value ≤ 0.05, 53 had differential expression with p-value ≤ 0.01, and 3 had

differential expression with p-value ≤ 0.001. Overall, there were more down-regulated genes than up-regulated genes—8,762 and 7,648 respectively. Of the genes with p-value ≤0.05, 273 of these genes were down-regulated and 200 were up-regulated.

Because the goal of the IPA analysis was to discover relevant molecular networks involving subtle alterations in gene expression, we set our cutoff for molecules differentially expressed at p≤0.01. 43 differentially expressed genes were further examined with IPA. The networks discovered by the IPA analysis were ranked based on their composite score that represents the log of the p-value for the likelihood these molecules would be found together by chance (Figure A4). The top ranked network (Score 28, i.e. $p=10^{-28}$) is shown in Figure 4. Note, the Gene Set Enrichment for this network is indicated as "Developmental Disorder, Hereditary Disorder, and Neurological Disease" even though the input was from B-cell gene expression. Despite the differences in gene input for this symbol method versus the probe ID method, the ubiquitin pathway again plays a central role.

IPA Upstream Regulator Analysis was used to identify upstream regulators that are likely to be responsible for the observed pattern of gene expression (Figure 5). The top ranked upstream regulator that had the expected pattern of downstream transcripts was HGF. The IPA transcription factor prediction algorithm indicated that 22q11DS was associated with altered transcription of several genes downstream of HGF. The p-value for overlap (p=0.003) suggested there is a statistically significant enrichment of genes regulated by HGF among differentially expressed genes (i.e. in cases versus controls). Moreover, the

algorithm predicted that the transcription pattern of these downstream genes is consistent with a loss of activity of HGF in cases versus controls (z-score of -2. 166, where a negative sign indicates inhibition or down regulation of HGF). While HGF mRNA itself was not substantially different between cases and controls (1.08 fold change), its downstream targets are of interest with respect to the role of reported deficient HGF levels in ASD (Figure 6 and Discussion).

Regression with Psychological Data

Multi-linear regression was performed on expression data of selected genes and behavioral data. CDC45 was selected because it is found in the 3 MB deleted region, of its significance and large fold change in our data, and presence in Network 2. BASP1 was selected because of its presence in Network 2 and its potential neurobiological role in schizophrenia (See discussion). The behavioral data was from either CBCL or ABCL (depending on age of subject) and included raw and t-scores of internalizing problems, externalizing problems, and summary score. There was no significant relationship found between expression levels of either gene and behavior (Tables 5 and 6).

Discussion

We hypothesized that the hemi-deletion in 22q11DS would diminish the expression of many of the genes in this region, but that others would have compensation, showing normal expression. Indeed, expression compensation appears to be operating to some extent because only six of the seventy five genes in 22q11.2 region are significantly down regulated (Table 2). However, none of the genes in the deleted region were significantly up-regulated despite our finding that across the genome there were overall more up-regulated than down regulated probe IDs for genes in the in the 22q11DS deleted region. We postulated that certain genes outside of the 22q11DS region would be dysregulated in 22q11DS patients based on their extensive interconnection with 22q11 genes. Thus we performed analysis of regulatory networks and gene set enrichment using our RNA expression data.

From the probe ID analysis, using fold-change |1.5| and $p<0.05$ a priori criteria, there were 2 highly significant networks i) a network that would be consistent with B-cell function and ii) a network that indicated a central role of ubiquitin. From the symbol analysis, using the $p<0.01$ a priori criteria, there were 4 significant networks, with the most significant involving UBC as a central gene and gene product.

In both the analyses by Probe ID and Symbol there were significant networks centered on Ubiquitin C (UBC), a polyubiquitin precursor involved in many cell functions such as protein degradation, DNA repair, cell cycle regulation, kinase modification, and other cell signaling [22]. While UBC, itself, was not dysregulated, two 22q11 genes, DGCR2 and CDC45, were significantly

down-regulated. A previous expression study of PBMCs showed the ubiquitin proteasome system (UPS) in BPD and psychosis to be a top canonical pathway [23] differentiating the psychiatric illnesses from controls. Another study, on postmortem superior temporal gyrus, found abnormal protein expression of UPS in idiopathic schizophrenia [24]. Gene expression of proteins involved in ubiquitin conjugation and deubiquitination have been correlated with positive symptoms of schizophrenia [25].

Both of these analyses point to the importance of ubiquitin. A 22q11 deleted gene involved in the ubiquitin-dependent proteolytic pathway, UFD1L, ubiquitin fusion degradation, was significantly down-regulated (Fold Change 0.73, p=0.020). Because the two methods of analyzing the data produced different gene sets uploaded into IPA, we view the fact that UBC emerged as a central character to provide complementary evidence of its importance rather than redundant information. Although not part of either ubiquitin-centered network, UFD1L binds to UBC and is involved in UPS. UFD1L polymorphisms have been correlated with presence of schizophrenia [26, 27], age of onset [28] and cognitive deficits [29]. The mouse homolog, UFD1, has been shown to influence congenital heart malformations similar to those seen in 22q11DS patients [30]. Moreover, UFD1 was present in mouse embryonic development of the first through fourth branchial arches, palatal precursors and frontonasal regions—all sites of 22q11DS phenotypic abnormalities [30].

From these associations, UFD1L may offer a key shared molecular link to idiopathic schizophrenia. We hypothesize that variation of polymorphisms between individuals play an important role in the heterogeneity of behavior-

related phenotypes in 22q11DS due to the presence of only one chromosomal copy. In other words, a recessive trait cannot be as easily masked.

The IPA algorithms revealed several important molecular connections. Since we expected that 22q11 deleted genes were upstream of many genes relevant to schizophrenia, we used IPA Upstream Regulator Analysis, on the symbol analyzed data, to identify the upstream regulators that are statistically most likely to be responsible for the observed pattern of gene expression. Most importantly, HGF, hepatocyte growth factor, was significantly down-regulated. HGF is a heterodimeric glycoprotein regulating many functions such as cell growth, cell motility, and morphogenesis by binding with c-MET receptor and activating a tyrosine kinase signaling cascade [22]. During development it is expressed in the cortex and hippocampus, and is believed to mediate interneuron development [31]. Decreased HGF has been previously implicated in the pathogenesis in high-functioning autism and OCD [32, 33].

Network 2 of the analysis by Probe ID showed a link to the up-regulation of BASP1, brain abundant membrane-attached signal 1. In a proteomic assessment of the prefrontal cortex, BASP1 was shown to be up-regulated in schizophrenia [34].

This study has some notable limitations. First, the number of subjects is small and the age range is large, from 7 to 29. Second, there is a possibility for type I error, by not correcting for multiple tests when calculating fold-change between cases and controls. Indeed, for some of the analysis, there were not substantially more genes over or under-expressed than expected by chance. In spite of this, the error could be limited when analyzed by IPA, because the

algorithms consider complex patterns of subtlety dysregulated genes and rank these patterns statistically.

Further study should be done to compare biological networks between idiopathic schizophrenia, 22qll associated schizophrenia, and controls in order to draw etiologic conclusions on shared pathways between the groups. Because polymorphisms of UFD1L have been shown to increase risk of schizophrenia, it would be beneficial to conduct genotype analysis on this sample and compare to previous publications.

References

1. Botto, L.D., et al., *A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population*. Pediatrics, 2003. **112**(1 Pt 1): p. 101-7.
2. Robin, N.H. and R.J. Shprintzen, *Defining the clinical spectrum of deletion 22q11.2*. J Pediatr, 2005. **147**(1): p. 90-6.
3. Ousley, O., et al., *A review of neurocognitive and behavioral profiles associated with 22q11 deletion syndrome: implications for clinical evaluation and treatment*. Curr Psychiatry Rep, 2007. **9**(2): p. 148-58.
4. Antshel, K.M., et al., *22q11.2 deletion syndrome: genetics, neuroanatomy and cognitive/behavioral features keywords*. Child Neuropsychol, 2005. **11**(1): p. 5-19.
5. Carlson, C., et al., *Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients*. Am J Hum Genet, 1997. **61**(3): p. 620-9.
6. Mega, M.S. and J.L. Cummings, *Frontal-subcortical circuits and neuropsychiatric disorders*. J Neuropsychiatry Clin Neurosci, 1994. **6**(4): p. 358-70.
7. Meechan, D.W., et al., *When half is not enough: gene expression and dosage in the 22q11 deletion syndrome*. Gene Expr, 2007. **13**(6): p. 299-310.
8. Kobrynski, L.J. and K.E. Sullivan, *Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes*. Lancet, 2007. **370**(9596): p. 1443-52.

9. Goldmuntz, E., *DiGeorge syndrome: new insights*. Clin Perinatol, 2005. **32**(4): p. 963-78, ix-x.

10. Schreiner, M.J., et al., *Converging levels of analysis on a genomic hotspot for psychosis: insights from 22q11.2 deletion syndrome*. Neuropharmacology, 2013. **68**: p. 157-73.

11. Mikhail, F.M., et al., *The recurrent distal 22q11.2 microdeletions are often de novo and do not represent a single clinical entity: a proposed categorization system*. Genet Med, 2014. **16**(1): p. 92-100.

12. Weksberg, R., et al., *Molecular characterization of deletion breakpoints in adults with 22q11 deletion syndrome*. Hum Genet, 2007. **120**(6): p. 837-45.

13. Michaelovsky, E., et al., *Genotype-phenotype correlation in 22q11.2 deletion syndrome*. BMC Med Genet, 2012. **13**: p. 122.

14. Drew, L.J., et al., *The 22q11.2 microdeletion: fifteen years of insights into the genetic and neural complexity of psychiatric disorders*. Int J Dev Neurosci, 2011. **29**(3): p. 259-81.

15. Kanaya, Y., et al., *Maturational alterations of peripheral T cell subsets and cytokine gene expression in 22q11.2 deletion syndrome*. Clin Exp Immunol, 2006. **144**(1): p. 85-93.

16. Chinen, J., et al., *Long-term assessment of T-cell populations in DiGeorge syndrome*. J Allergy Clin Immunol, 2003. **111**(3): p. 573-9.

17. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias*. Bioinformatics, 2003. **19**(2): p. 185-93.

18. Calvano, S.E., et al., *A network-based analysis of systemic inflammation in humans*. Nature, 2005. **437**(7061): p. 1032-7.

19. Havik, B., et al., *The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia*. Biol Psychiatry, 2011. **70**(1): p. 35-42.

20. Garbett, K., et al., *Transcriptome alterations in the prefrontal cortex of subjects with schizophrenia who committed suicide*. Neuropsychopharmacol Hung, 2008. **10**(1): p. 9-14.

21. Mistry, M., J. Gillis, and P. Pavlidis, *Genome-wide expression profiling of schizophrenia using a large combined cohort*. Mol Psychiatry, 2013. **18**(2): p. 215-225.

22. Pruitt, K.D., et al., *RefSeq: an update on mammalian reference sequences*. Nucleic Acids Res, 2014. **42**(Database issue): p. D756-63.

23. Bousman, C.A., et al., *Preliminary evidence of ubiquitin proteasome system dysregulation in schizophrenia and bipolar disorder: convergent pathway analysis findings from two independent samples*. Am J Med Genet B Neuropsychiatr Genet, 2010. **153B**(2): p. 494-502.

24. Rubio, M.D., et al., *Dysfunction of the ubiquitin proteasome and ubiquitin-like systems in schizophrenia*. Neuropsychopharmacology, 2013. **38**(10): p. 1910-20.

25. Bousman, C.A., et al., *Positive symptoms of psychosis correlate with expression of ubiquitin proteasome genes in peripheral blood*. Am J Med Genet B Neuropsychiatr Genet, 2010. **153B**(7): p. 1336-41.

26. Xie, L., et al., *A family- and population-based study of the UFD1L gene for schizophrenia*. Am J Med Genet B Neuropsychiatr Genet, 2008. **147B**(7): p. 1076-9.

27. De Luca, A., et al., *Association study of a promoter polymorphism of UFD1L gene with schizophrenia*. Am J Med Genet, 2001. **105**(6): p. 529-33.

28. Ota, V.K., et al., *The UFD1L rs5992403 polymorphism is associated with age at onset of schizophrenia*. J Psychiatr Res, 2010. **44**(15): p. 1113-5.

29. Ota, V.K., et al., *Polymorphisms in schizophrenia candidate gene UFD1L may contribute to cognitive deficits*. Psychiatry Res, 2013. **209**(1): p. 110-3.

30. Yamagishi, H., et al., *A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects*. Science, 1999. **283**(5405): p. 1158-61.

31. Levitt, P., K.L. Eagleson, and E.M. Powell, *Regulation of neocortical interneuron development and the implications for neurodevelopmental disorders*. Trends Neurosci, 2004. **27**(7): p. 400-6.

32. Russo, A.J. and S.C. Pietsch, *Decreased Hepatocyte Growth Factor (HGF) and Gamma Aminobutyric Acid (GABA) in Individuals with Obsessive-Compulsive Disorder (OCD)*. Biomark Insights, 2013. **8**: p. 107-14.

33. Sugihara, G., et al., *Decreased serum levels of hepatocyte growth factor in male adults with high-functioning autism*. Prog Neuropsychopharmacol Biol Psychiatry, 2007. **31**(2): p. 412-5.

34. Behan, A.T., et al., *Proteomic analysis of membrane microdomain-associated proteins in the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder reveals alterations in LAMP, STXBP1 and BASP1 protein expression.* Mol Psychiatry, 2009. **14**(6): p. 601-13.

Tables

Table 1. Characteristics of Study Participants

Characteristic	Control (n=12)	22q11 (n=12)	P-value
Age, mean (SD, range)	23.3 (5.6, 13-29)	14.2 (7.3, 7-27)	0.003
RIN, mean (SD, range)	8.4 (0.9, 6.7-9.7)	8.0 (1.1, 6.1-9.3)	0.33
Sex, n (% Males)	6 (50.0)	7 (58.3)	0.68
Race, n (%)			
Caucasian	8 (66.7)	4 (33.3)	<0.001
Black	3 (25.0)	1 (8.3)	
Hispanic	0 (0)	2 (16.7)	
Asian	1 (8.3)	0 (0)	
Unknown	0 (0)	5 (41.7)	

Table 2. Expression of Genes in 3 MB Deleted Region

Gene Symbol	Alternate Name(s)	Illumina ID	Official Full Name	P-value [a]	Fold Change [b]
USP18	ISG43; UBP43	ILMN_3240420	ubiquitin specific peptidase 18	0.203	0.93
GGT3P [c]	GGT3		gamma-glutamyltransferase 3 pseudogene		
DGCR6		ILMN_1791396	DiGeorge syndrome critical region gene 6	0.888	0.98
DGCR6		ILMN_1663685	DiGeorge syndrome critical region gene 6	0.729	0.95
PRODH [c]	POX; PIG6; HSPOX2; PRODH1; PRODH2; TP53I6		proline dehydrogenase (oxidase) 1		
DGCR5	LINC00037; NCRNA00037	ILMN_2221808	DiGeorge syndrome critical region gene 5	0.204	1.11
DGCR9	DGS-A	ILMN_3241987	DiGeorge syndrome critical region gene 9	0.425	1.11
DGCR10 [c]	DGS-B		DiGeorge syndrome critical region gene 10		
DGCR2 [d]	IDD; LAN; DGS-C; SEZ-12	ILMN_1713301	DiGeorge syndrome critical region gene 2	0.002	0.87
DGCR11	DGS-D	ILMN_3236211	DiGeorge syndrome critical region gene 11	0.792	0.96
DGCR14	ES2; DGSH; DGSI; DGS-H; DGS-I; Es2el; DGCR13	ILMN_1719232	DiGeorge syndrome critical region gene 14	0.130	0.76
TSSK2 [c]	TSK2; DGS-G; SPOGA2; STK22B		testis-specific serine kinase 2		
GSC2 [c]	GSCL		goosecoid homeobox 2		
LOC100652736 [c]			uncharacterized LOC100652736		

Table 2 cont. Expression of Genes in 3 MB Deleted Region

Gene Symbol	Alternate Name(s)	Illumina ID	Official Full Name	P-value [a]	Fold Change [b]
SLC25A1 [c]	CTP; SEA; D2L2AD; SLC20A3		solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1		
CLTCL1 [c]	CLTD; CHC22; CLH22; CLTCL		clathrin, heavy chain-like 1		
HIRA	TUP1; DGCR1; TUPLE1	ILMN_1813664	histone cell cycle regulator	0.860	0.97
MRPL40	URIM; NLVCF; MRPL22; MRP-L22	ILMN_1687403	mitochondrial ribosomal protein L40	0.745	0.95
C22orf39 [c]			chromosome 22 open reading frame 39		
UFD1L [d]	UFD1	ILMN_1665737	ubiquitin fusion degradation 1 like (yeast)	0.020	0.74
CDC45 [d]	CDC45L; CDC45L2; PORC-PI-1	ILMN_1670238	cell division cycle 45	0.003	0.53
CLDN5 [c]	AWAL; BEC1; TMVCF; CPETRL1		claudin 5		
LINC00895 [c]			long intergenic non-protein coding RNA 895		
SEPT5 [c]	H5; CDCREL; PNUTL1; CDCREL1; CDCREL-1; HCDCREL-1		septin 5		
GP1BB [c]	BS; CD42C; GPIBB; BDPLT1		glycoprotein Ib (platelet), beta polypeptide		
TBX1	DGS; TGA; VCF; CAFS; CTHM; DGCR; DORV; VCFS; TBX1C; CATCH22	ILMN_2248112	T-box 1	0.863	0.98

Table 2 cont. Expression of Genes in 3 MB Deleted Region

Gene Symbol	Alternate Name(s)	Illumina ID	Official Full Name	P-value [a]	Fold Change [b]
GNB1L	GY2; FKSG1; WDR14; WDVCF; DGCRK3	ILMN_1761044	guanine nucleotide binding protein (G protein), beta polypeptide 1-like	0.076	0.90
C22orf29	BOP	ILMN_1759595	chromosome 22 open reading frame 29	0.348	0.85
TXNRD2	TR; TR3; SELZ; TRXR2; TR-BETA	ILMN_1657893	thioredoxin reductase 2	0.430	0.93
COMT	HEL-S-98n	ILMN_1730084	catechol-O-methyltransferase	0.908	0.98
MIR4761 [c]			microRNA 4761		
ARVCF [c]			armadillo repeat gene deleted in velocardiofacial syndrome		
TANGO2	C22orf25	ILMN_1789405	transport and golgi organization 2 homolog (Drosophila)	0.295	0.80
MIR185 [c]	MIRN185; miR-185		microRNA 185		
DGCR8 [c]	Gy1; pasha; DGCRK6; C22orf12		DGCR8 microprocessor complex subunit		
MIR3618 [c]			microRNA 3618		
MIR1306 [c]	MIRN1306; hsa-mir-1306		microRNA 1306		
TRMT2A	HTF9C	ILMN_2310075	tRNA methyltransferase 2 homolog A (S. cerevisiae)	0.059	0.86
MIR6816 [c]	hsa-mir-6816		microRNA 6816		
RANBP1	HTF9A	ILMN_1721457	RAN binding protein 1	0.538	0.93
RANBP1 [d]	HTF9A	ILMN_2109156	RAN binding protein 1	0.039	0.64
ZDHHC8	DHHC8; ZNF378; ZDHHCL1	ILMN_1789492	zinc finger, DHHC-type containing 8	0.340	0.93
LOC388849		ILMN_1676043	uncharacterized LOC388849	0.730	0.95
LOC284865 [c]			uncharacterized LOC284865		

Table 2 cont. Expression of Genes in 3 MB Deleted Region

Gene Symbol	Alternate Name(s)	Illumina ID	Official Full Name	P-value [a]	Fold Change [b]
LINC00896 [c]			long intergenic non-protein coding RNA 896		
RTN4R [c]	NGR; NOGOR		reticulon 4 receptor		
MIR1286 [c]	MIRN1286; hsa-mir-1286		microRNA 1286		
DGCR6L		ILMN_3245688	DiGeorge syndrome critical region gene 6-like	0.509	0.91
LOC729444 [e]					
TMEM191B		ILMN_1794989	transmembrane protein 191B	0.320	1.13
PI4KAP1 [c]			phosphatidylinositol 4-kinase, catalytic, alpha pseudogene 1		
RIMBP3	RIM-BP3; RIMBP3A; RIMBP3.1; RIM-BP3.1; RIM-BP3.A	ILMN_3243124	RIMS binding protein 3	0.505	0.92
ZNF74 [d]	COS52; hZNF7; ZFP520; ZNF520	ILMN_2383871	zinc finger protein 74	0.012	0.79
ZNF74	COS52; hZNF7; ZFP520; ZNF520	ILMN_2281404	zinc finger protein 74	0.596	0.93
SCARF2	NSR1; SREC2; VDEGS; SREC-II; SRECRP-1		scavenger receptor class F, member 2		
KLHL22	KELCHL	ILMN_1705390	kelch-like family member 22	0.475	0.81
MED15 [d]	TIG1; CAG7A; CTG7A; PCQAP; TIG-1; TNRC7; ARC105	ILMN_1706200	mediator complex subunit 15	0.012	0.74
MED15	TIG1; CAG7A; CTG7A; PCQAP; TIG-1; TNRC7; ARC105	ILMN_1668748	mediator complex subunit 15	0.190	0.84

Table 2. Expression of Genes in 3 MB Deleted Region

Gene Symbol	Alternate Name(s)	Illumina ID	Official Full Name	P-value [a]	Fold Change [b]
POM121L4P		ILMN_3249444	POM121 transmembrane nucleoporin-like 4 pseudogene	0.214	0.76
TMEM191A	TMEM191AP	ILMN_3228529	transmembrane protein 191A (pseudogene)	0.784	0.94
PI4KA	PIK4CA; pi4K230; PI4K-ALPHA	ILMN_1655177	phosphatidylinositol 4-kinase, catalytic, alpha	0.306	0.90
SERPIND1 [c]	HC2; LS2; HCF2; HCII; HLS2; THPH10; D22S673		serpin peptidase inhibitor, clade D (heparin cofactor), member 1		
SNAP29	CEDNIK; SNAP-29	ILMN_1659857	synaptosomal-associated protein, 29kDa	0.340	0.91
CRKL		ILMN_1690122	v-crk avian sarcoma virus CT10 oncogene homolog-like	0.237	0.82
LOC101928891 [c]			uncharacterized LOC101928891		
AIFM3 [c]	AIFL		apoptosis-inducing factor, mitochondrion-associated, 3		
LZTR1	BTBD29; LZTR-1; SWNTS2	ILMN_1805161	leucine-zipper-like transcription regulator 1	0.621	0.95
THAP7 [c]			THAP domain containing 7		
THAP7-AS1 [c]			THAP7 antisense RNA 1		
TUBA3FP [c]			tubulin, alpha 3f, pseudogene		
P2RX6	P2X6; P2XM; P2RXL1	ILMN_1716335	purinergic receptor P2X, ligand-gated ion channel, 6	0.563	0.90
SLC7A4	VH; CAT4; CAT-4; HCAT3	ILMN_1774229	solute carrier family 7, member 4	0.768	0.94

Table 2 cont. Expression of Genes in 3 MB Deleted Region

Gene Symbol	Alternate Name(s)	Illumina ID	Official Full Name	P-value [a]	Fold Change [b]
P2RX6P [c]			purinergic receptor P2X, ligand-gated ion channel, 6 pseudogene		
LOC400891 [c]			leucine-rich repeat-containing protein LOC400891		
BCRP2 [c]	BCR2; BCR-2; BCRL2		breakpoint cluster region pseudogene 2		
FAM230B [c]			family with sequence similarity 230, member B (non-protein coding)		
POM121L8P	DKFZp434K191	ILMN_1652533	POM121 transmembrane nucleoporin-like 8 pseudogene	0.840	0.96
RIMBP3C [c]	RIMBP3.3; RIM-BP3.3		RIMS binding protein 3C		
RIMBP3B	RIMBP3.2; RIM-BP3.2; RIM-BP3.B	ILMN_3247152	RIMS binding protein 3B	0.146	1.39

[a] Student t-test

[b] Ratio of 22q11:control

[c] Gene not expressed in study

[d] Significant ($p<0.05$)

[e] RefSeq status: withdrawn

Table 3. IPA Mapped Differentially Expressed Genes

Gene Symbol	Illumina ID	Fold Change	P-value [a]	Name	Chromosome Region	Location	Type
FCRLA	ILMN_1691071	1.54	0.001	Fc receptor-like A	1q23.3	Plasma Membrane	other
CDC45	ILMN_1670238	0.53	0.003	cell division cycle 45	22q11.21	Nucleus	other
PIK3AP1	ILMN_1652787	1.75	0.005	phosphoinositide-3-kinase adaptor protein 1	10q24.1	Cytoplasm	other
CD19	ILMN_1782704	1.53	0.006	CD19 molecule	16p11.2	Plasma Membrane	transmembrane receptor
CCR6 [b]	ILMN_1690907	1.78	0.007	chemokine (C-C motif) receptor 6	6q27	Plasma Membrane	G-protein coupled receptor
SIGLEC6 [b]	ILMN_2299095	1.52	0.007	sialic acid binding Ig-like lectin 6	19q13.33	Extracellular Space	other
ATRIP [b]	ILMN_1653896	0.66	0.011	ATR interacting protein	3p21.31	Nucleus	kinase
GSTM4 [b]	ILMN_1651800	1.55	0.013	glutathione S-transferase mu 4	1p13.3	Cytoplasm	enzyme
CD55	ILMN_1800540	1.52	0.015	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	1q32.2	Plasma Membrane	other
LOC152586	ILMN_1772092	1.58	0.015	glycosyltransferase 54 domain-containing protein	4q31.1	Other	other
EGR1	ILMN_1762899	1.75	0.015	early growth response 1	5q31.2	Nucleus	transcription regulator
RFTN1	ILMN_1800787	2.00	0.016	raftlin, lipid raft linker 1	3p24.3	Plasma Membrane	other
PTRF	ILMN_1757552	1.82	0.021	polymerase I and transcript release factor	17q21.31	Nucleus	transcription regulator

Table 3 cont. IPA Mapped Differentially Expressed Genes

Gene Symbol	Illumina ID	Fold Change	P-value	Name	Chromosome Region	Location	Type
CCNB1IP1 [b]	ILMN_1714386	0.65	0.021	cyclin B1 interacting protein 1, E3 ubiquitin protein ligase	14q11.2	Nucleus	enzyme
TRIM4 [b]	ILMN_2323385	1.56	0.023	tripartite motif containing 4	7q22.1	Cytoplasm	other
TNFRSF13B	ILMN_1759075	1.96	0.024	tumor necrosis factor receptor superfamily, member 13B	17p11.2	Plasma Membrane	transmembrane receptor
PSAP [b]	ILMN_1749109	2.13	0.027	prosaposin	10q22.1	Extracellular Space	other
BASP1	ILMN_1651826	1.89	0.029	brain abundant, membrane attached signal protein 1	5p15.1	Nucleus	transcription regulator
CD74 [b]	ILMN_2379644	1.96	0.029	CD74 molecule, major histocompatibility complex, class II invariant chain	5q33.1	Plasma Membrane	transmembrane receptor
LBH [b]	ILMN_2315979	1.76	0.034	limb bud and heart development	2p23.1	Nucleus	transcription regulator
LGR6 [b]	ILMN_1713550	1.55	0.034	leucine-rich repeat containing G protein-coupled receptor 6	1q32.1	Plasma Membrane	G-protein coupled receptor
ZNF91	ILMN_1802053	1.51	0.035	zinc finger protein 91	19p12	Nucleus	other
MCM10	ILMN_2413898	0.65	0.036	minichromosome maintenance complex component 10	10p13	Nucleus	other
RANBP1 [b]	ILMN_2109156	0.64	0.039	RAN binding protein 1	22q11.21	Nucleus	other
FEM1C	ILMN_1703330	0.66	0.042	fem-1 homolog c (C. elegans)	5q22.3	Nucleus	transcription regulator
RBAKDN	ILMN_2054938	1.53	0.044	RBAK downstream neighbor (non-protein coding)	7p22.1	Other	other

Table 3 cont. IPA Mapped Differentially Expressed Genes

Gene Symbol	Illumina ID	Fold Change	P-value	Name	Chromosome Region	Location	Type
HLA-DMB	ILMN_1761733	1.61	0.044	major histocompatibility complex, class II, DM beta	6p21.32	Plasma Membrane	transmembrane receptor
PLCG2	ILMN_1815719	1.69	0.044	phospholipase C, gamma 2 (phosphatidylinositol-specific)	16q23.2	Cytoplasm	enzyme
HNRNPA1L2 [b]	ILMN_2387742	1.57	0.049	heterogeneous nuclear ribonucleoprotein A1-like 2	13q14.3	Nucleus	other

Table 4. Differentially Expressed Unmapped IPA genes

ID (LOC)	New ID	Gene Symbol	Gene Type	Fold Change	p-value	Name	Chromosome Region	Notes
LOC100129742		RPS3AP13	Pseudo	1.63	0.036	ribosomal protein S3A pseudogene	2q24.2	
LOC100132521			Protein coding	1.82	0.007		14q13.2	RefSeq Withdrawn
LOC152586		LOC152586	Protein coding	1.58	0.015	glycosyltransferase 54 domain-containing protein	4q31.1	
LOC285741		TPT1P4	Pseudo	1.51	0.045	tumor protein, translationally-controlled 1 pseudogene 4	6q24.2	
LOC388474		RPL7AP66	Pseudo	2.24	0.032	ribosomal protein L7a pseudogene 66	18q12.2	
LOC389458		RBAKDN	ncRNA	1.53	0.044	RBAK downstream neighbor (non-protein coding)	7p22.1	
LOC440396		HNRNPA1P16	Pseudo	1.68	0.042	heterogeneous nuclear ribonucleoprotein A1 pseudogene 16	17p13.3	
LOC441034		RPL7AP30	Pseudo	1.74	0.026	ribosomal protein L7a pseudogene 30	4q25	
LOC442727			Pseudo	1.87	0.048	prothymosin, alpha pseudogene	7q33	
LOC644701			Protein coding	1.51	0.029	hypothetical LOC644701	15q14	RefSeq Withdrawn
LOC644928		RPS26P15	Pseudo	1.51	0.047	ribosomal protein S26 pseudogene 15	1p32.2	
LOC647841			Protein coding	1.56	0.019	hypothetical protein LOC647841	X	RefSeq Withdrawn
LOC648169	642801		Protein coding	1.52	0.029	hypothetical protein LOC642801	17p13.3	RefSeq Withdrawn
LOC653458		BMS1P9	Pseudo	0.64	0.045	BMS1 pseudogene 9	9q13	

Table 4 cont. Differentially Expressed Unmapped IPA genes

ID (LOC)	New ID	Gene Symbol	Gene Type	Fold Change	p-value	Name	Chromosome Region	Notes
LOC653773		RPL31P17	Pseudo	1.56	0.019	ribosomal protein L31 pseudogene 17	2q36.1	
LOC728973	440732	RPS7P3	Pseudo	1.88	0.020	ribosomal protein S7 pseudogene 3	1q42.2	
LOC729903		RPS16P10	pseudo	1.96	0.017	ribosomal protein S16 pseudogene 10	19p12	

Table 5. Multivariate Linear Regression Between Gene Expression in 22Q11 Individuals and Raw Scores From Behavioral Checklist (n=12), controlling for age and sex

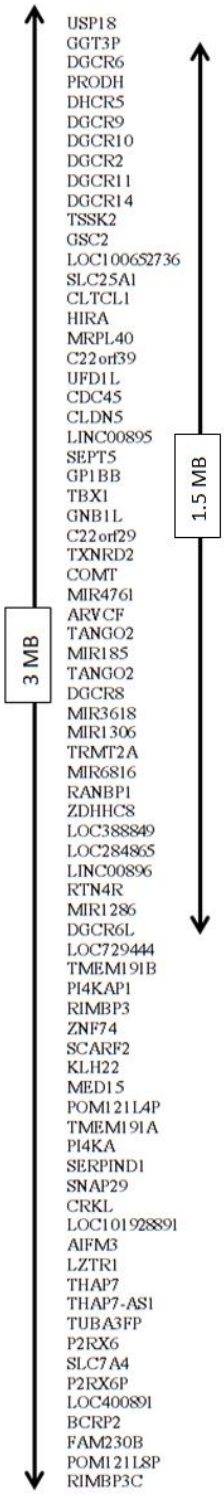
	CDC45			BASP1		
Behavior	**B(SE)**	**P-value**	**R^2**	**B(SE)**	**P-value**	**R^2**
Internalizing	-0.014 (0.095)	0.89	0.40	0.002 (0.009)	0.86	0.41
Externalizing	0.047 (0.095)	0.64	0.07	-0.003 (0.009)	0.73	0.05
Summary	0.066 (0.840)	0.84	0.14	-0.014 (0.031)	0.66	0.16

Table 6. Multivariate Linear Regression Between Gene Expression in 22Q11 Individuals and T-Scores From Behavioral Checklist (n=12), controlling for age and sex

	CDC45			BASP1		
Behavior	**B(SE)**	**P-value**	**R^2**	**B(SE)**	**P-value**	**R^2**
Internalizing	0.003 (0.107)	0.98	0.16	0.003 (0.010)	0.76	0.17
Externalizing	0.039 (0.108)	0.73	0.13	-0.004 (0.010)	0.69	0.14
Summary	0.076 (0.097)	0.46	0.10	-0.003 (0.010)	0.81	0.04

Figures

Figure 1. Genes in 3 MB Deleted Region



Genes listed according to the RefSeq genes, provided by UCSC Browser, Assembly Feb 2009 (GRCh37/hg19) https://genome.ucsc.edu/cgi-bin/hgTables.

Figure 2. Volcano Plot of Differentially Expressed Genes

A scatter-plot of the –log10 p-value against the log2 fold change in expression in B-cells (22q11/control). Genes with statistically significant differential expression (p <0.05 and fold change ≥ |1.5|) are shown in red.

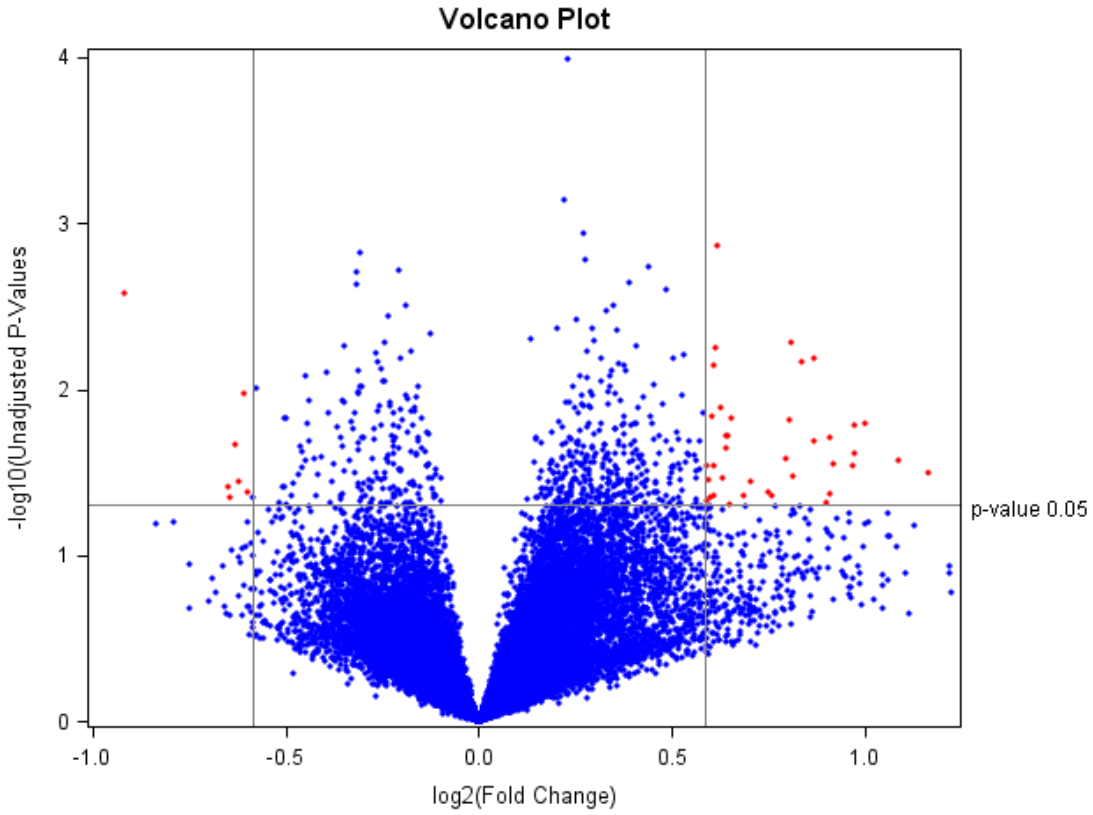


Figure 3. Network 2 in Probe ID Analysis

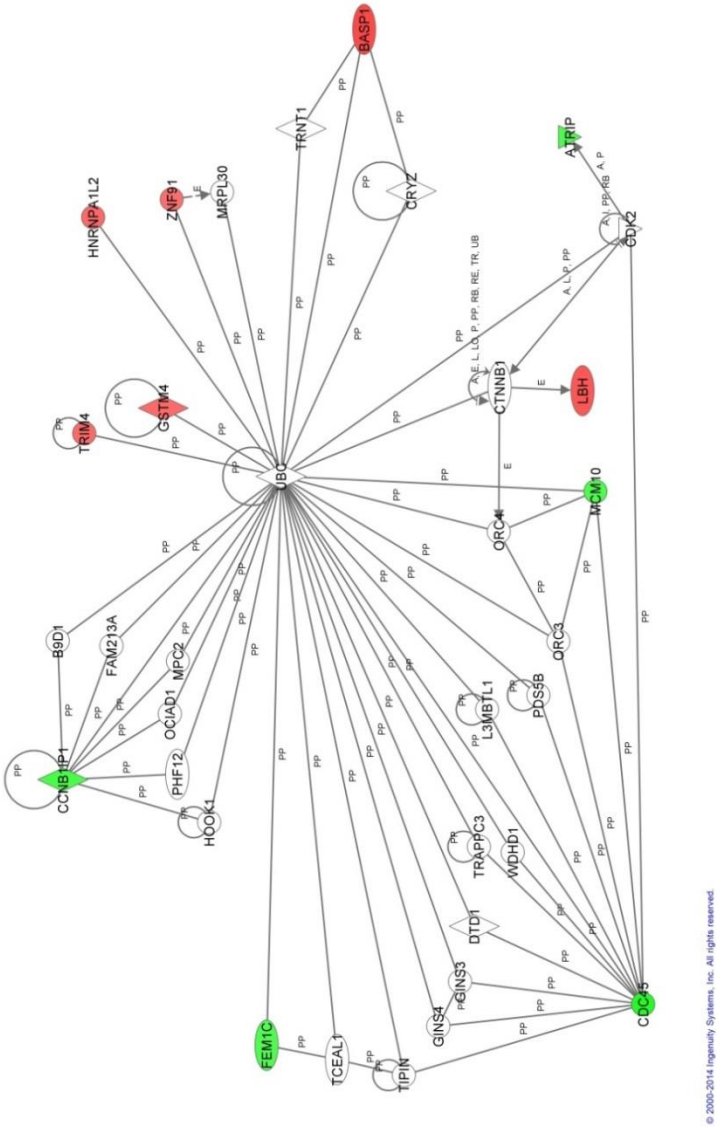


Figure 3 cont. Network 2 in Probe ID Analysis

Interconnected 22q.11 genes network determined by Ingenuity analysis.

IPA Legend: → acts on; — binding only; Molecules: red-upregulated transcripts in 22q11DS versus controls participants; Green, downregulated; no color – additional interconnected genes, transcripts or proteins that are most interconnected with our uploaded molecules based on the IPA connection algorithm. Solid lines show direct interaction (binding/physical contact). A dashed line, indirect interaction supported by the literature but possibly involving one or more intermediate molecules that have not been investigated definitively.

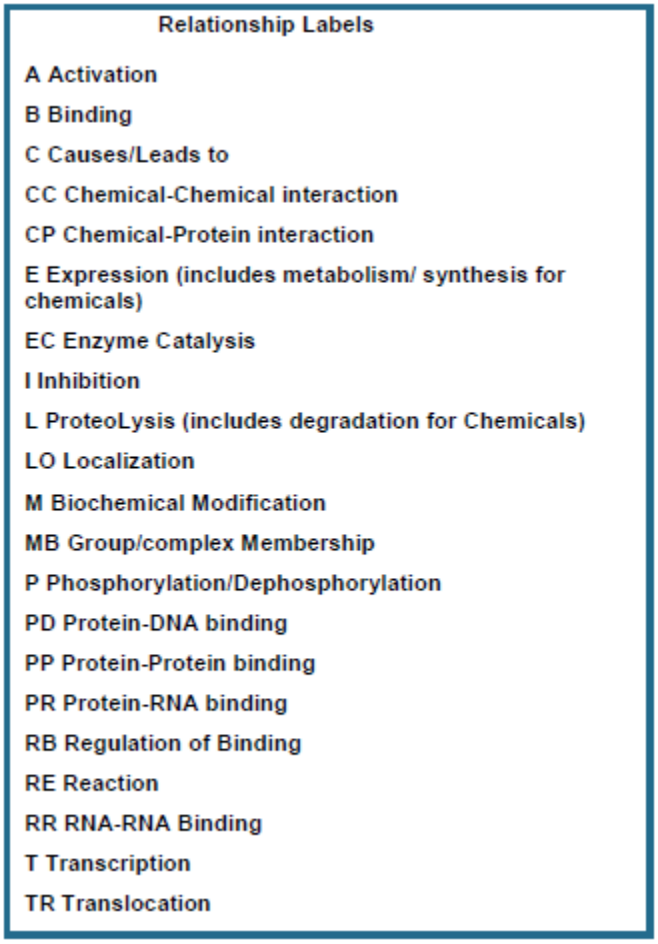
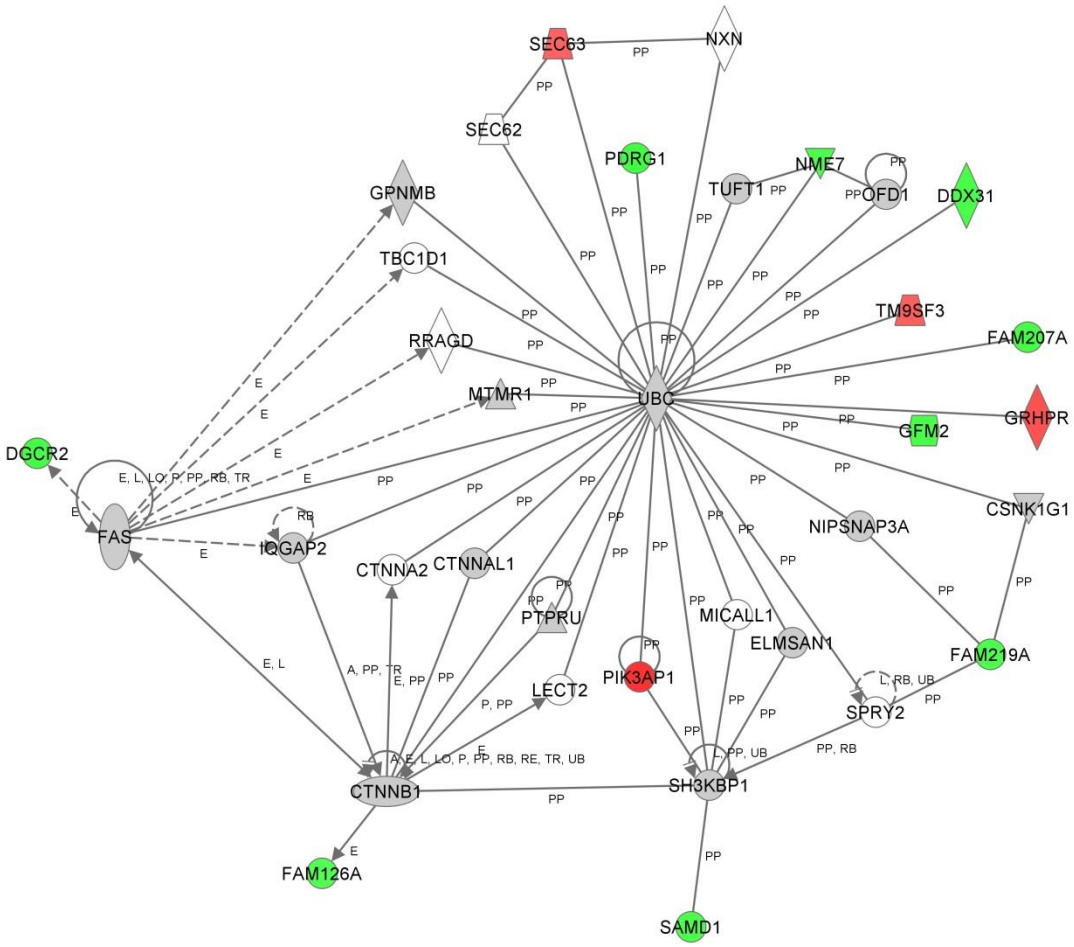


Figure 4. Network 1 in Symbol Analysis



See IPA legend above.

Figure 5. Upstream Regulators in Symbol Analysis

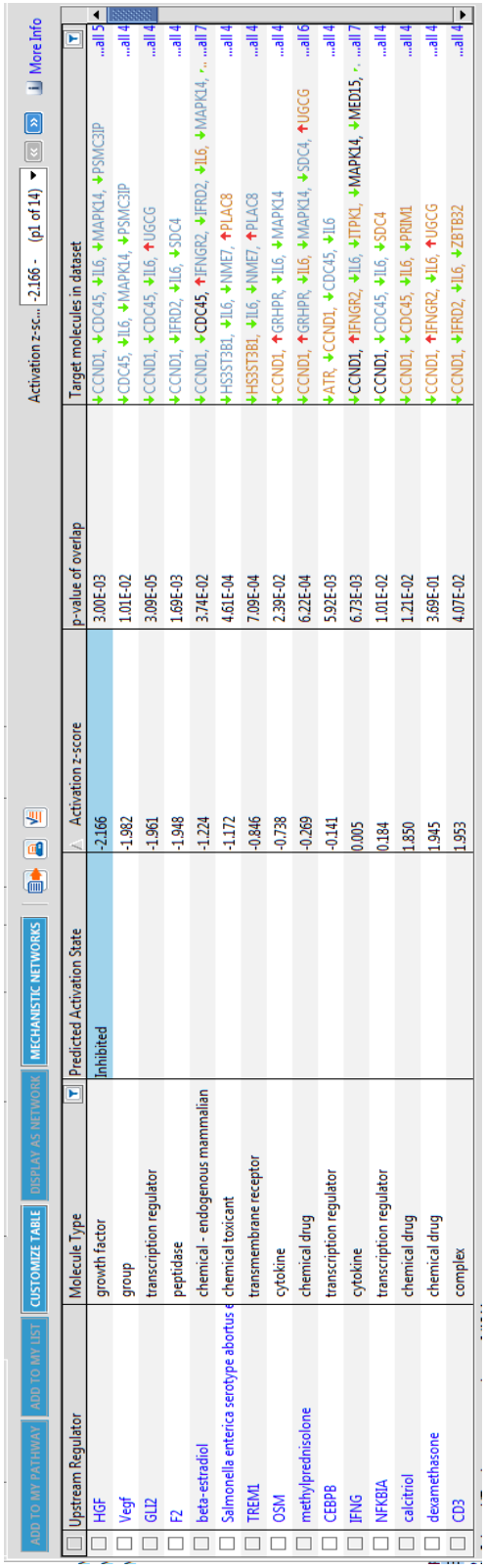
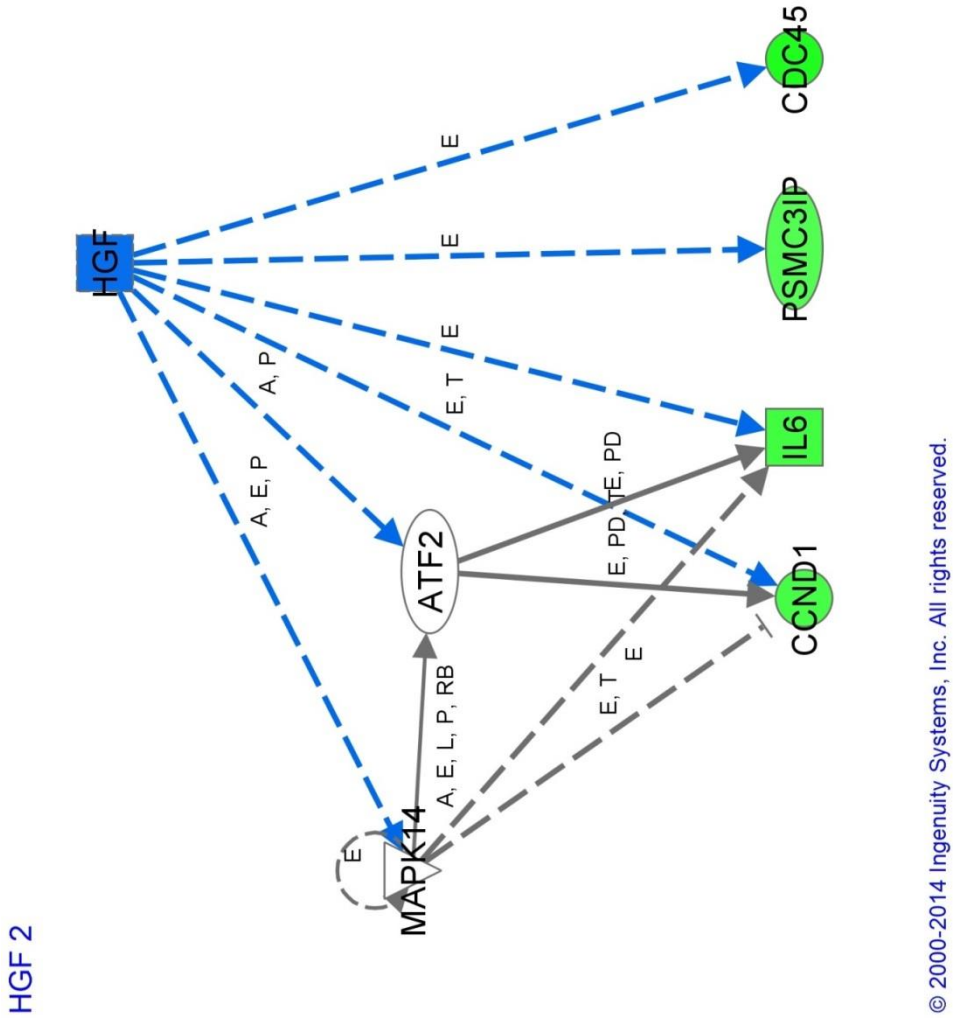


Figure 6. Downstream Targets of HGF



Legend: HGF in IPA upstream regulator analysis, which is based how many known targets of each transcription regulator are present in our uploaded dataset, and also compares their direction of change (i.e. expression in the 22q11DS relative to control) to what is expected from the literature in order to predict likely relevant transcriptional regulators. HGF (transcriptional regulator) is shown in blue. The downstream targets include genes that are down regulated in our 22q11DS versus control samples (green). HGF is also upstream of other transcription regulators for these genes (no color).

Appendix

Figure A1. Data Flow Chart

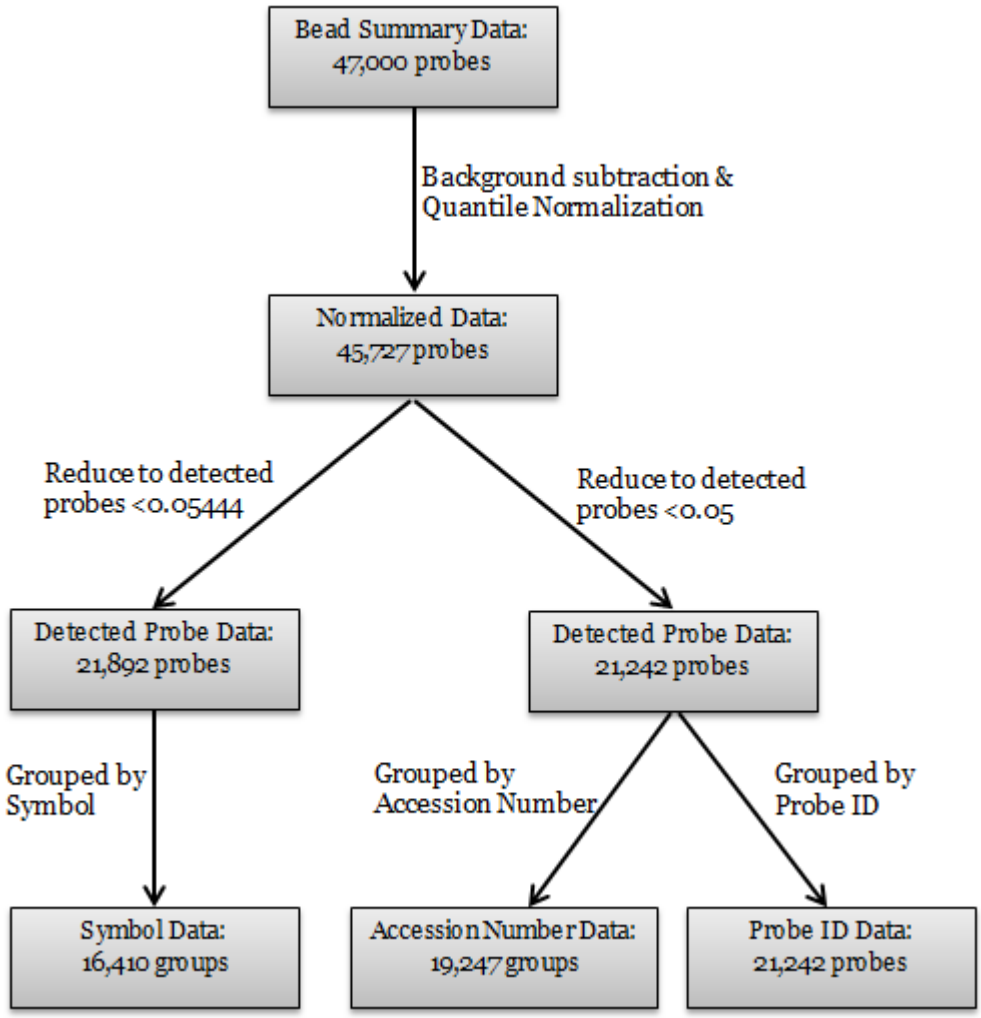


Table A1. Non-Significant Probes With Shared Gene Symbol of Significant Probes

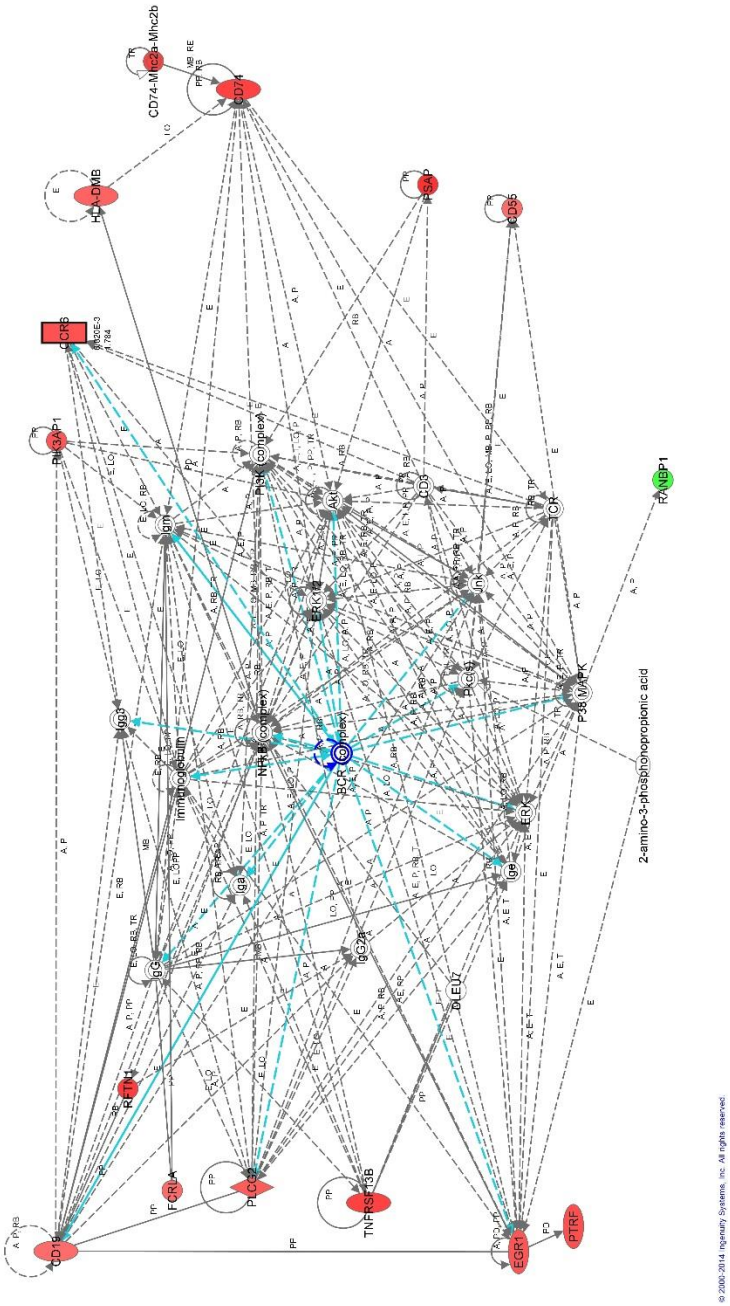
Gene Symbol	Illumina ID	Fold Change	P-value
CCR6	ILMN_2387696	1.44	0.301
SIGLEC6	ILMN_1685630	0.93	0.725
ATRIP	ILMN_2370464	0.89	0.473
GSTM4	ILMN_1716979	0.95	0.433
CCNB1IP1	ILMN_1752394	1.14	0.349
CCNB1IP1	ILMN_2260833	1.19	0.557
CCNB1IP1	ILMN_2347349	0.98	0.855
TRIM4	ILMN_1792265	1.03	0.902
PSAP	ILMN_2355559	2.00	0.065
CD74	ILMN_1736567	1.79	0.138
CD74	ILMN_1761464	1.00	0.984
LBH	ILMN_1660794	0.88	0.512
LBH	ILMN_1670369	1.08	0.578
LBH	ILMN_1684834	0.84	0.303
LBH	ILMN_1811507	0.85	0.185
LBH	ILMN_2244009	0.84	0.341
LGR6	ILMN_1677464	1.24	0.072
LGR6	ILMN_1703081	0.85	0.248
RANBP1	ILMN_1721457	0.93	0.538
HNRNPA1L2	ILMN_1736894	0.85	0.260

Figure A2. IPA Significant Networks by Probe ID Analysis

The analysis is composed of 4 networks. To view a network, select the appropriate network(s) and click View Networks. To merge selected networks, click Merge Networks.

ID	Molecules in Network	Score	Focus Molecul	Top Diseases and Functions
1	2-amino-3-phosphonopropionic acid, Akt, BCR (complex), ↑**CCR6**, CD3, ↑**CD19**, ↑**CD55**, ↑**CD74**, CD74-Mhc2a-Mhc2b, DLEU7, ↑**EGR1**, ERK, ERK1/2, ↑**FCRLA**, ↑**HLA-DMB**, Iga, Ige, IgG, Igg3, IgG2a, Igm, Immunoglobulin, Jnk, NFkB (complex), P38 MAPK, PI3K (complex), ↑**PIK3AP1**, Pkc(s), ↑**PLCG2**, ↑**PSAP**, ↑**PTRF**, ↓**RANBP1**, ↑**RFTN1**, TCR, ↑**TNFRSF13B**	35	14	Hematological System Development and Function, Humoral Immune Response, Tissue Morphology
2	↓**ATRIP**, B9D1, ↑**BASP1**, ↓**CCNB1IP1**, ↓**CDC45**, CDK2, CRYZ, CTNNB1, DTD1, FAM213A, ↓**FEM1C**, GINS3, GINS4, ↑**GSTM4**, ↑**HNRNPA1L2**, HOOK1, L3MBTL1, ↑**LBH**, ↓**MCM10**, MPC2, MRPL30, OGIAD1, ORC3, ORC4, PDS5B, PHF12, TCEAL1, TIPIN, TRAPPC3, ↑**TRIM4**, TRNT1, UBC, WDHD1, ↑**ZNF91**	25	11	Drug Metabolism, Cell Cycle, DNA Replication, Recombination, and Repair
3	↑**LGR6**, LNX1	3	1	Cardiovascular Disease, Inflammatory Disease, Protein Synthesis
4	PTPN6, PTPN11, ↑**SIGLEC6**, SOCS3	2	1	Cell Death and Survival, Hematological System Development and Function, Cell Cycle

Figure A3. Network 1 in Probe ID Analysis



See IPA legend above.

Figure A4. IPA Significant Networks by Symbol Analysis

ID	Molecules in Network	Score	Focus Mol	Top Diseases and Functions
1	↓CSNK1G1, CTNNA2, ↑CTNNAL1, ↓CTNNB1, ↓**DDX31**, ↓**DGCR2**, ↑ELMSAN1, ↓**FAM126A**, ↓**FAM207A**, ↓**FAM219A**, ↓FAS, ↓**GFM2**, ↑GPNMB, ↑**GRHPR**, ↓IQGAP2, LECT2, MICALL1, ↓MTMR1, ↓NIPSNAP3A, ↓**NME7**, NXN, ↓OFD1, ↓**PDRG1**, ↑**PIK3AP1**, ↑PTPRU, RRAGD, ↓**SAMD1**, SEC62, ↑**SEC63**, ↑SH3KBP1, SPRY2, TBC1D1, ↑**TM9SF3**, ↓TUFT1, ↑UBC	28	13	Developmental Disorder, Hereditary Disorder, Neurological Disease
2	AMPK, Ap1, ↓**ATR**, BCR (complex), CD3, ↑**CD19**, Cdc2, Creb, ERK1/2, ↑**FCRLA**, Hsp27, ↑**IFNGR2**, IgG, Igg3, Igm, IL1, IL12 (complex), IL12 (family), Immunoglobulin, LDL, ↓**MAPK14**, Mek, P38 MAPK, p85 (pik3r), PDGF BB, PI3K (complex), ↑**PIK3AP1**, ↑**PLAC8**, Rock, ↓**RPTOR**, Rxr, ↓**SDC4**, Tgf beta, ↓**UFD1L**, ↑**UGCG**	24	11	Cell Cycle, Lipid Metabolism, Nervous System Development and Function
3	Akt, Androgen-AR, ↓**CCND1**, ↓**CDC45**, ↓**CHD8**, Collagen type I, ERK, estrogen receptor, GP130 dimer, Histone h3, Histone h4, Hsp70, Hsp90, ↓**IL6**, IL1/IL6/TNF, Insulin, Jnk, Mapk, ↓**MED15**, NFkB (complex), Pka, Pkc(s), ↓**PSMC3IP**, Rac, Ras, Ras homolog, ↑**RHOQ**, RNA polymerase II, ↓S1PR4, ↓**SLC39A14**, TCR, ↑**UBE2E2**, Vegf, VTCN1, ↓**ZBTB32**	21	10	Cellular Development, Hematological System Development and Function, Hematopoiesis
4	↓ADAM19, ↑APBB1IP, ↓ARID5B, CDC42EP3, CDC42EP4, ↑EIF4H, ↓**HS3ST3B1**, ↓**IFRD2**, ↑ILKAP, ↓**ITPK1**, ↓KDELR2, MICALL1, ↑MRPL38, ↓**NME7**, NNMT, ↓P2RY6, ↓**PACSIN3**, ↓PI4K2B, ↑**PLAC8**, ↓**PRIM1**, ↑PRIM2, ↑PTTG1IP, RNF220, ↑ROBO3, ↑ROR1, SCG5, ↑SIGIRR, ↓**SLC39A14**, ↑**SNX3**, ↓STARD4, TBC1D1, ↑TGFB1, TREM1, ↑UBC, WNT5B	19	9	DNA Replication, Recombination, and Repair, Gene Expression, Cell-To-Cell Signaling and Interaction
5	↓CEP350, ↓**SAPCD2**	3	1	Cellular Assembly and Organization, Cellular Function and Maintenance, Gene Expression