TOWARDS PERSONALIZED MEDICINE IN PSYCHIATRY: FOCUS ON SUICIDE

Daniel F. Levey

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Program of Medical Neuroscience,

Indiana University

April 2017

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Andrew J. Saykin, Psy. D. - Chair

____

Alan F. Breier, M.D.

Doctoral Committee

____

Gerry S. Oxford, Ph.D.

December 13, 2016

____

Anantha Shekhar, M.D., Ph.D.

____

Alexander B. Niculescu III, M.D., Ph.D.

Dedication

This work is dedicated to all those who suffer, whether their pain is physical or psychological.

Acknowledgements

The work I have done over the last several years would not have been possible without the contributions of many people. I first need to thank my terrific mentor and PI, Dr. Alexander Niculescu. He has continuously given me advice and opportunities over the years even as he has suffered through my many mistakes, and I greatly appreciate his patience. The incredible passion he brings to his work every single day has been inspirational. It has been an at times painful but often exhilarating 5 years.

I need to thank Helen Le-Niculescu for being a wonderful colleague and mentor. I learned a lot about organization and presentation working alongside her, and her tireless work ethic was an excellent example for a new graduate student.

I had the pleasure of working with a number of great people over the years. Mikias Ayalew showed me the ropes of the lab and began my understanding of the power of algorithms. Despite being a New England Patriots fan he's a pretty decent guy. Peter Phalen helped to teach me the ways of R coding and provided some clinical perspective (while also occasionally distracting me with talk of politics and philosophy.)

In no particular order, Evan Winiger, Rebecca Learman, Naga Venipenta, Farnoosh Khan, Vivek Venugopal, Helen Dainton, one of the greatest parts of my time in the Niculescu lab has been seeing you move on to take the next steps in

your lives. I'm glad to have known you, and thank you for allowing me to live vicariously through your successes.

I want to thank my parents and sister for their love and support over the years. I cannot adequately express in words my love and appreciation for you all. So instead I will thank my father for his marathon lectures, which helped prepare me for the rigors of graduate school.

I must thank my lifelong friend Brian Marler. He provided a hand in a time of great need and really helped me turn my life around. I am forever grateful. And he joins me as one of several hardcore IU football fans on the planet earth, so he's good for that as well.

Lastly, I'd like to thank my beautiful and patient wife and my amazing son for being the reason I wake up every morning.

Daniel F. Levey

Towards Personalized Medicine in Psychiatry: Focus on Suicide

Psychiatric disorders cost an estimated $273 billion annually. This cost comes largely in the form of lost income and the chronic disability that often strikes people when they are young and can last decades. While the monetary costs are quantifiable, the suffering of each individual patient is no less vital. As many as 1 in 5 persons diagnosed with mental illness will commit suicide, a contributing factor in suicide being the second leading cause of death of people age 15-34. There is a critical need to find better ways to identify and help those who are at risk.

Understanding mental illness and improving treatment has been difficult due to the heterogeneous and complex etiology of these illnesses. A significant challenge for the field is integrating findings from diverse laboratories all over the world contributing to the ever expanding literature and translating them into actionable treatment. Our lab employs a convergent functional genomics approach which incorporates multiple independent lines of evidence provided by genetic and functional genomic data published in the primary literature as a Bayesian strategy to prioritize experimental findings.

Heritability and genetics clearly play an important role in psychiatric disorders. We looked at schizophrenia and alcoholism in separate case-control analyses in order to identify and prioritize genes related to these disorders. We were able to reproduce these findings in additional independent cohorts using

polygenic risk scores. We found overlap in these disorders, and identified possible underlying biological processes.

Genetics play an important role in identifying clinical risk, particularly at the population level. At the level of the individual, gene expression may provide more proximal association to disease state, assimilating environmental, genetic, as well as epigenetic influence. We undertook N of 1 analyses in a longitudinally followed cohort of psychiatric participants, identifying genes which change in expression tracking an individual's change in suicidal ideation. These genes were able to predict suicidal behavior in independent cohorts. When combined with simple clinical instruments these predictions were improved. This work shows how multi-level integration of genetic, gene expression, and clinical data could be used to enable precision medicine in psychiatry.

Andrew J. Saykin, Psy. D. - Chair

Table of Contents

Introduction: Precision Medicine 1

Convergent Functional Genomics 7

Universal Materials and Methods 8

GWAS Studies 8

Gene Expression Studies 8

Convergent Functional Genomics 12

Workflow 15

Discovery 15

Prioritization 15

Validation 16

Prediction 16

Understanding 16

Chapter 1: Schizophrenia 17

Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction 21

Introduction 21

Materials and Methods 24

Figures 28

Results 28

Discussion 98

Conclusions and future directions 109

Chapter 2: Alcoholism 113
Genetic Risk Prediction and Neurobiological Understanding of Alcoholism 117
Introduction 117
Materials and Methods 120
Figures 129
Results 129
Discussion 150
Conclusion 168
Chapter 3: Focus on Suicide 170
Epidemiology 171
Clinical Risk Assessment Scales 172
Treatment 173
Etiology/genetics 179
Paths to Pathology 180
The Biology of Suicide 181
SKA2 184
SLC4A4 185
SAT1 and polyamines (SMOX) 185
Serotonin 186
Inflammatory markers 186
Identifying Biomarkers for Suicidal Behavior 187
Chapter 4: Discovery and Validation of Blood Biomarkers for Suicidality 188

Introduction 188
Methods 189
Human subjects 189
Medications 207
Human blood gene expression experiments and analyses 207
Analysis 207
Differential Gene Expression Analyses in the Discovery Cohort 208
Validation Analyses 209
Figures 210
Results 221
Discovery 221
Testing in suicide completers 225
Mechanistic understanding 230
Prospective Validation 239
Discussion 240
Apoptosis 243
Conclusions and future directions 246
Chapter 5: An INTRA n-of-1 approach to biomarker discovery 248
Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach 250
Introduction 250
Methods 251

Figures ... 270

Results ... 278

Discussion ... 304

Chapter 6: The Gender Paradox in Suicidal Behavior ... 318

Towards understanding and predicting suicidality in women: biomarkers and clinical risk assessment ... 320

Methods ... 322

Figures ... 338

Results ... 348

Discussion ... 384

Chapter 7: Conclusions ... 395

CFG ... 395

Within subject longitudinal design ... 397

Dimensionality ... 398

Limitations ... 399

Future Directions ... 401

Bibliography ... 403

Curriculum vitae

Introduction: Precision Medicine

Psychiatric disorders are the manifestation of a synthesis of underlying genetics, biology, and psychology interacting with environmental stressors. Some individuals may be more or less predisposed to illness. Some find themselves in privileged idyllic environments while others are deprived of nutrition, entertainment, education, stability, etc. It is in this complex interplay that an individual is forged. There is a growing understanding that none of these factors appear single-handedly sufficient for the development of psychopathology. It may therefore be necessary to take a personalized medicine approach, embracing the complexity of each individual patient and taking full advantage of developing tools from genetics, neuroscience, and psychology to improve nosology, diagnosis, and treatment.

The American Heart Association (AHA) defines heart failure as "a complex clinical syndrome that can result from any structural cardiac disorder that impairs the ability of the ventricle to fill or eject blood." This definition seems to embrace the pluropotentiality for distinctly different underlying structural dysfunction, in combination and interactions, arriving at a similar outcome. Diabetes mellitus, coronary heart disease, and hypertension are known risk factors of mechanistically dissimilar etiology. [1] If complex and interacting risk factors influence the pathology and predictability of heart failure, might brain disorders also be influenced by a confluence of interacting risk factors?

Not everyone responds the same way to treatments and interventions. Numerous studies indicate clozapine may be an effective alternative medication for schizophrenia patients who are 'treatment-resistant' to other typical and atypical anti-psychotics[2]. But the term itself, 'treatment-resistant', seems to imply that some patients with the same underlying disease pathology fail to see relief from the same drug. It is possible, however, that while patients appear to have similar symptoms the totality of underlying pathology that yields the symptoms may not be identical. Indeed, while it was thought that the benefits of clozapine in treatment resistant patients might be shared by other atypical antipsychotic medications, this may not be the case. [3]

The environment of an individual plays an important role in the development of psychopathology. Childhood trauma appears to be associated with psychopathology. This may be related to increased hypothalamic pituitary adrenal (HPA) axis activity and cortisol secretion in patients with a history of trauma. Rhesus monkeys exposed to social separation at 6 months of age they showed increased plasma cortisol. Those subjects who showed increased plasma cortisol also voluntarily consumed significantly more alcohol.[4] This system has been widely implicated as a risk factor in psychiatric disorders[5] and for suicide. The dexamethasone suppression test (DST) is a quantitative assessment of adrenal gland function, classically used to aid in the diagnosis of Cushing's syndrome[6] and still used today[7]. It consists of an IV injection of dexamethasone into the patient and measurement of cortisol levels in response on the following day. A patient

with normative HPA function should show reduced cortisol levels in response to dexamethasone through suppression of the adrenocorticotropic hormone (ACTH) secretion from the anterior pituitary. A failure to suppress cortisol levels suggests hyperactivity in the system. It has been used as a measure for diagnosis and for suicide risk assessment[8], though it has not proven to be sufficiently sensitive or specific to be clinically useful[9]. It does appear to have at least population level utility as a risk factor. Limitations at the level of individual patients could be due to heterogeneity in the clinical populations, and certain subtypes of individual may be better served with this kind of test.

Monozygotic twins of patients with psychiatric disorders tend to have a higher risk for psychopathology but the outcome is far from certain, though genetic predisposition as measured by singular markers or mutations are often neither necessary nor sufficient alone. Multiple common and de novo mutations probably interact to yield an observable disease phenotype, which might explain the reduced or incomplete penetrance often used to describe false positive or false negative findings from genetic risk factors. Additionally, development of disease may require additional environmental or acute pathological insult. This is why many modern theories posit that multiple 'hits' lead to illness. [10]

Some of the strongest common heritable findings in genetics, such as the APOE E4 allele as a risk factor for Alzheimer's disease (AD), represent significant increases (as much as 11x) of relative risk to develop the disorder. APOE E4 can be defined by two SNPs, the T allele of rs429358 and the C allele of rs7412, and

can be assessed in individuals by personal genomics companies such as 23andMe. But even those with the allele are not certain to develop AD, and disease risk is modified by factors such as gender. Additionally, ~40% of AD patients do not carry an E4 allele, showing that mutations in this gene have a powerful but perhaps not necessary role in the disease. [11] While the evidence is convincing that APOE E4 modifies risk in a dose dependent fashion and modifies population level average age of onset, it also displays incomplete penetrance and is likely influenced by genetic background and environmental context, and is therefore not sufficient in the development of AD.[12] Studies which embrace the complexity of interactions of a gene of large influence such as APOE with additional genetic risk modifiers such as CR1 and tie them to related amyloid pathology phenotype may point to future ways to better chart the disease risk and progression in individuals. [13]

Personalized medicine studies, sometimes called N-of-1 trials where a single individual is followed with significantly greater clinical and multi-omic depth, are beginning to gain traction. A landmark paper published in Cell by Michael Snyder and his lab at Stanford University follow a single individual over a period of 14 months. [14] The participant in this study was followed very closely with multiple high throughput methods and integrations of genomic, transcriptomic, proteomic, and metabolomic data. Various genetic markers, known to influence disease risk at the population level, were identified in the individual and used to prioritize biomarkers associated with diseases identified as highest risk were monitored, notably for elevated risk for diabetes. Over the course of the study the participants

blood glucose became elevated following an acute viral infection, and lifestyle changes were made along with beginning a low dose acetylsalicylic acid treatment. Risk factors were identified and This is an example of how personalized medicine can help inform choices made in treatment.

Personalized medicine offers the possibility of using established risk factors from epidemiological, molecular neuroscience, and psychological studies, which have population or sample level significance to disease, and integrate these factors at the level of the pathology of an individual patient. To that end we have worked to develop genetic and peripheral blood biomarkers. Initial work focused on identifying population level genetic risk factors using genome wide association studies (GWAS). These risk factors were identified in independent GWAS at a single nucleotide polymorphism level (SNP), converted into nearest associated genes, and then validated by multiple independent lines of evidence using a convergent functional genomics (CFG) approach, and top findings were finally integrated into a polygenic genetic risk prediction score (GRPS) and tested in multiple independent GWAS. Later work focused on longitudinal within subject designs for gene expression tracking the phenotype of individual study participants. Deep databases of quantitative phenotypes allowed for the integration of multiple different psychological and environmental modalities.

In this way we were able to begin to embrace the complexity of the individual and develop multi-dimensional genetic and phenomic biomarkers for

more personalized and accurate risk assessment of suicidal behavior in psychiatric patients.

Convergent Functional Genomics

Convergent functional genomics is an algorithmic approach that leverages large genomic datasets with the relevant peer reviewed primary literature to find reproducible disease relevant genes and biomarkers. Initially, high throughput technologies behind GWAS studies, microarray, RNA-Seq etc. produce a potentially large list of genes which are differentially associated with a target disease or phenotype. Nominally significant findings are then prioritized by integrating with an exhaustive database of all disease relevant primary literature to create a polyevidence CFG score. The primary literature is produced by a number of different labs utilizing a wide array of different approaches. Human studies of genetics and peripheral blood provide specificity to human disease, while human brain studies provide additional specificity to the likely target tissue, the brain. Animal studies provide sensitivity of in vivo analysis of implicated pathology and pathways.

It is essential to note that only those genes initially identified in the discovery analyses or experiments are made available for prioritization by CFG. This is crucial to the CFG process. What makes the CFG so effective is that it, by default, requires that prioritized genes from primary work done in the discovery analysis replicate peer reviewed research by multiple independent studies, often in multiple tissues and animal models. It prioritizes and cross-validates primary findings, which still means a great deal of importance and care must be invested in the quality and precision of the initial discovery analysis.

CFG has prior success in the identification of blood biomarkers for mood state in bipolar patients.[15] The resulting panel of 10 biomarkers showed sensitivity and specificity for predicting high and low mood states in 2 bipolar cohorts and 1 cohort with psychotic disorders (schizophrenia and schizoaffective). More importantly, this panel (which was previously shown to track elevated mood state) has been shown to increase following cognitive behavioral therapy treatment for depressed patients (an increased score indicates an elevated mood state) and to track changes in Hamilton Depression Rating Scale scores.[16] This highlights one of the key advantages of CFG: built in reproducibility. There is a wealth of excellent peer-reviewed literature available. Unlike a simple literature search, CFG cross-validates and enriches primary analyses by utilizing the literature in an a priori empirical and unbiased fashion to prioritize the most relevant findings.

Universal Materials and Methods

GWAS Studies

GWAS methodology is particular to the individual study, and will be discussed in greater detail in the appropriate chapter.

Gene Expression Studies

Human participants

Live psychiatric participants are part of a larger ongoing longitudinal cohort that is continuously collected. Participants are recruited from the patient population at the Richard L. Roudebush Veteran Affairs Medical Center (VAMC) and Indiana University School of Medicine (IUSM) in Indianapolis through referrals

from care providers, brochures left in plain sight in public areas, and mental health clinics through word of mouth. All participants understood and signed informed consent forms detailing research goals, procedure, caveats, and safeguards, per institutional review board approved protocol. Participants completed a Diagnostic Interview for Genetic Studies (DIGS) at the initial baseline visit, followed by up to 6 testing visits, generally 3-6 months apart or whenever a new psychiatric hospitalization occurred. At each testing visit they received a series of psychiatric rating scales and blood was drawn. Whole blood (10 ml) was collected in two RNA-stabilizing PAXgene tubes, labeled with deidentified ID number and stored at -80° C in a locked freezer until the time of future processing. Whole-blood RNA was extracted for microarray gene expression studies from the PAXgene tubes, as described below.

Postmortem subjects of suicide completers were used to validate findings from live participants, and were obtained through the Marion County coroner's office. We required a last observed alive postmortem interval of 24 hours or less, and cases selected had completed suicide by means other than overdose, which could affect gene expression. See the demographic tables for cause of death and age at time of death.

Medications

Live participants were all diagnosed with various psychiatric disorders. Their psychiatric medications were listed in their electronic medical records and documented at the time of each testing visit. Participants were on a variety of

different psychiatric medications. Medications can have a strong influence on gene expression. Because we focused our discovery of differentially expressed genes on within-participant analyses, and therefore relative within-participant gene expression changes, this influence is of negligible effect on our results. There was no consistent pattern in any particular type of medication or between type of medication and change in phenotype on the rare occasions where there was a medication change between visits.

Human blood gene expression experiments and analyses

Whole blood (2.5 – 5 ml) was collected into each PAXGene tube by routine venipuncture. PAXGene tubes contain proprietary reagents for the stabilization of RNA. The cells from whole blood were concentrated by centrifugation, the pellet washed, resuspended and incubated in buffers containing Proteinase K for protein digestion. A second centrifugation step was done to remove residual cell debris. After the addition of ethanol for an optimal binding condition, the lysate was applied to a silica-gel membrane/column. The RNA bound to the membrane as the column was centrifuged and contaminants were removed in three wash steps. The RNA was then eluted using diethlypyrocarbonate-treated water. The protocol for RNA extraction is carried out on a QIAgen QIAcube.

Sample labeling

Sample labeling was performed using the Ambion MessageAmp II-BiotinEnhanced antisense RNA (aRNA) amplification kit. The procedure is briefly outlined below:

1. Reverse transcription to synthesize first-strand cDNA was primed with T7 oligo(dT) primer to synthesize cDNA containing a T7 promoter sequence.
2. Second-strand cDNA synthesis converted the single-stranded cDNA into a double-stranded DNA template for transcription. The reaction employed DNA polymerase and RNase H to simultaneously degrade the RNA and synthesize the second-strand cDNA.
3. cDNA purification removed RNA, primers, enzymes, and salts that would have inhibited in vitro transcription.
4. In vitro transcription to synthesize aRNA with biotin-NTP Mix generated multiple copies of biotin-modified aRNA from the double-stranded cDNA templates; this is the amplification step.
5. aRNA purification removed unincorporated NTPs, salts, enzymes and inorganic phosphate to improve the stability of the biotin-modified aRNA.
6. aRNA fragmentation: the amplified RNA is fragmented in a reaction the employs a metal-induced hydrolysis to fragment the aRNA. The fragmented labeled aRNA is now ready for hybridization to the Affymetrix microarray chip (Affymetrix, Santa Clara, CA, USA).

Microarrays

Biotin-labeled aRNAs were hybridized to Affymetrix HG-U133 Plus 2.0 GeneChips (Affymetrix; with over 40,000 genes and expressed sequence tags), according to manufacturer's protocols http://media.affymetrix.com/support/downloads/manuals/expression_analysis_te

chnical_manual.pdf. Arrays were stained using standard Affymetrix protocols for antibody signal amplification and scanned on an Affymetrix GeneArray 2500 scanner with a target intensity set at 250. Quality-control measures, including 30/50 ratios for glyceraldehyde 3-phosphate dehydrogenase and β-actin, scale factors, background and Q-values, were within acceptable limits.

Convergent Functional Genomics

Databases

The Niculescu Lab of Neurophenomics has created a manually curated database of all of the primary literature for human and nonhuman gene expression (post-mortem brain, blood and cell cultures), genetic, (association, copy number variants, linkage, and transgenic) published to date on psychiatric disorders. We added only those findings which were described as significant by the authors, using their particular methodology. This database only contains primary literature and not reviews, meta-analyses, or other secondary analyses. This was done to avoid circularity with results already found in the primary literature. These databases are constantly updated as new research is published.

Human post-mortem brain, blood and other peripheral tissue gene expression

Literature search was performed in PubMed (http://www.ncbi.nlm.nih.gov/pubmed), searching the primary literature with various keyword combinations, human, brain, postmortem, lymphocytes, blood,

cells, gene expression, along with specific keywords for the disease or phenotype of interest.

Human genetic evidence (association, linkage)

We took special care to omit studies where subjects in our discovery cohorts overlapped with published studies scored for CFG. For linkage the location of each gene was obtained through GeneCards (http://www.genecards.org/). We input the chromosome location and start position into the Rutgers Map Interpolator[17] http://compgen.rutgers.edu/mapinterpolator to receive a sex-averaged centimorgan position. To be scored for linkage a gene had to map to within 5 centimorgans of a marker for linkage. Only published markers with a LOD score of >=2 were scored for convergence.

Animal model brain and blood gene expression

We used animal model evidence reported in the literature to score convergence. Where applicable we used data generated by our own lab, as described in the relevant chapters below.

Animal model genetic evidence

PubMed searches were performed, using project specific keywords, for relevant animal models for the disease or phenotype of interest. In addition we searched the Mouse Genome Informatics database (http://www.informatics.jax.org/) for transgenic studies using project specific categories.

CFG scoring

Internal score: Genes which were significantly associated with the disease of interest were first identified. The methods for identifying these genes are discussed in detail in the relevant chapters for schizophrenia, alcoholism, and suicide. Following the internal scoring of genes, external scoring was carried out on the same genes,

External score: There were a maximum possible of 6 lines of evidence which could be used to calculate a CFG score. Each line of evidence was capped in such a way that a positive hit within a line of evidence in the database would result in maximum points, no matter how many positive hits were present. This was done to avoid popularity biases. In the schizophrenia work, each line of evidence could contribute 1 single point, for a maximum possible CFG score of 6. All other work weighted the lines of evidence such that human evidence received twice as much as nonhuman evidence, and brain evidence received twice as many points as evidence from genetics or peripheral tissue. In this way, human brain evidence was given 4 points, human peripheral or genetic evidence would be given 2 points each, nonhuman brain evidence was given 2 points, and nonhuman peripheral, genetic, or transgenic evidence received 1 point each. In this weighting the maximum possible score was 12 (4 human brain, 2 human genetic, 2 human peripheral, 2 nonhuman brain, 1 nonhuman genetic, 1 nonhuman peripheral = 12).

Workflow

A vital part of our methodology is the work flow that we use. All of our projects have maintained an evolution of the basic framework of discovering markers in an initial independent cohort, prioritizing these findings with a polyevidence CFG score, and using top prioritized findings to predict in additional independent cohorts. Once we have demonstrated a consistency and reproducibility across cohorts we dig deeper into the biology and pathology of our findings, identifying overlap with other psychiatric disorders and treatments.

Discovery

Discovery is always performed in an independent cohort. In the case of GWAS, findings came from analyzing data in a standard case control design. All SNPs with a $p < 0.05$ were deemed to be nominally significant. For the initial suicide gene expression study a similar methodology was employed, with nominally significance required to proceed to the next step of analysis. Later gene expression studies applied a newer and different approach, and are defined in greater detail in the appropriate chapters.

Prioritization

Prioritization was applied to all genes deemed to be nominally significant in the first step. All nominally significant were prioritized with a CFG score as described above.

Validation

This step was not employed yet in the schizophrenia work. In all other studies validation was performed as an additional step to show reproducibility in other cohorts or approaches. This is discussed in greater detail in the following chapters, but briefly, an in house stress reactive alcohol consuming animal model was used to validate findings from the alcohol study. In the suicide study cohorts of gender and age matched suicide completers collected in collaboration with the Marion County Coroner's office were used to validate peripheral blood gene expression findings.

Prediction

In all studies outside independent cohorts were used to replicate findings.

Understanding

All work is completed naïve and blind to the actual nature of the biological processes and implications of findings. Only once findings have been independently discovered, prioritized, validated, and predicted do we dig deeper into the possible underlying biology. In this way work proceeds unbiased by favored genes or popularity effects. Once findings were verified we dug deeper into the biological implications, canonical pathways, overlap with other disorders, and treatment possibilities.

Chapter 1: Schizophrenia

Schizophrenia is a heterogeneous syndrome of serious and chronic cognitive dysfunction effecting as many as 51 million people worldwide. It is generally diagnosed by a co-occurrence of at least two positive and negative symptoms with the presence of delusions, hallucinations, or disorganized speech.

Disorganized speech could be thought of as an outwardly observable manifestation of fragmented thinking. This commonly may come in the form of tangential and hard to follow loose associations in conversation.

Hallucinations are the illusion of sensation in the absence of sensory stimuli. It commonly occurs in the form of hearing voices (auditory hallucination) or seeing things that do not exist (visual hallucination).

Firmly held but clearly false ideas make up the third major symptom of schizophrenia, delusions. Delusions may be particularly difficult to identify in screening as it requires identifying false ideas that a patient may not voluntarily express to a caregiver and may be difficult to distinguish from simply overvalued ideas without further probing.

Although environmental influences are important, genetics and heritability may play a significantly greater role. Schizophrenia has a high rate of heritability as evidenced by countless studies over the last 60 years (Nuechterlein and Dawson, 1984). Recent advances in the field of genetics have led to the hypothesis that schizophrenia heritability is probably not the result of the influence of any one gene but rather through the interaction of several different genes.

Genetics offer the potential of identifying individuals with a predisposition towards schizophrenia and intervening early in life to minimize additional environmental risks.

We sought to identify and prioritize a set of genes involved in schizophrenia. We began by using an independent GWAS study from the International Schizophrenia Consortium (ISC) for the discovery of genes, integrating all single nucleotide polymorphisms (SNPs) with a nominally significant p-value of <0.05 into the nearest gene. We then used CFG to prioritize these findings with a polyevidence score, resulting in a panel of 42 genes containing 542 SNPs.

We next validated these prioritized findings by creating a polygenic risk score derived entirely from the top SNPs identified in the discovery cohort and tested it in 4 independent cohorts provided by the Genetic Association Information Network (GAIN) and Molecular Genetics of Schizophrenia (MGS) non-GAIN studies. A polygenic panel of all 542 nominally significant SNPs was able to successfully separate schizophrenia cases from controls in all independent cohorts while a panel consisting of only the top SNP from each of the 42 genes failed to do so. This may be due to the small effect size of each individual SNP.

In post hoc analysis examining reproducibility across independent GWAS, we carried out discovery analysis in the GAIN cohorts, taking all nominally significant SNPs, integrating them at the nearest gene, and prioritizing with CFG. What we found was that there was very little overlap at the level of individual SNPs (0.4%), but we found increasing reproducibility between GWAS as we integrated

SNPs with the nearest gene (54.2%), with increasing CFG evidence (83.3%), and integrating genes at the level of canonical pathways (97.1%). This may reveal the importance of moving beyond simple statistical relevance of a SNP and shows the added value examining the impact of SNPs at the level of genes and the pathways in which they interact.

Along these lines we also sought to understand the biological roles of our top genes and how they interact in pathways by using our top findings from each independent GWAS with a CFG polyevidence score ≥3. As mentioned above, we found a striking consistency of findings across GWAS when results were integrated at the pathway level, with glutamate signaling in particular implicated as the top enriched pathway in 2 independent cohorts (ISC and GAIN EA) and the second highest enrichment in a third (GAIN AA). The danger in broadly implicating a neurotransmitter system such as glutamate is that they have been so widely study by so many labs and so many different approaches that the potential exists for popularity bias in the algorithm. With that said, many of the promising novel therapeutics in the literature are targeting this system. [18]

Given the heterogeneous nature of symptoms associated with schizophrenia, it has been an understandably difficult to understand and treat. Indeed, its definition has been evolving since early psychiatrists labeled the disease dementia praecox, through each iteration of the Diagnostic and Statistical Manual of Mental Disorders (DSM), currently on its fifth edition, through today. As method and technique continue to be refined, this work points to the possibility of

quantitative genetic risk assessment for predisposition to schizophrenia, and opens the door to early preventative interventions.

Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction

Introduction

Schizophrenia is a devastating disorder affecting ~1% of the population. While there is clear evidence for roles for both genes and environment, a comprehensive biological understanding of the disorder has been elusive so far. Most notably, there has been until recently a lack of concerted integration across functional and genetic studies, and across human and animal model studies, resulting in missed opportunities to see the whole picture.

As part of a translational convergent functional genomics (CFG) approach, developed by us over the last decade,[19],[20],[21],[22],[15] and expanding upon our earlier work on identifying genes for schizophrenia[23] and biomarkers for psychosis,[24] we set out to comprehensively identify candidate genes, pathways and mechanisms for schizophrenia, integrating the available evidence in the field to date. We have used data from published genome-wide association studies (GWAS) data sets for schizophrenia.[25],[26] We integrated those data with gene expression data---human postmortem brain gene expression data, human induced pluripotent stem cell-derived neuronal cells[27] and human blood gene expression data[24] published by others and us, as well as with relevant animal model brain and blood gene expression data generated by our group[28] and others. In addition, we have integrated as part of this comprehensive approach other genetic data---human genetic data (linkage, copy number variant (CNV) or association) for

schizophrenia, as well as relevant mouse model genetic evidence (Figure 1-1, Table 1-1 and Figure 1-2). Animal model data provide sensitivity of detection, and human data provide specificity for the illness. Together, they help to identify and prioritize candidate genes for the illness, using a polyevidence CFG score, resulting in essence in a de facto field-wide integration putting together the best available evidence to date. Once that is done, biological pathway analyses can be conducted and mechanistic models can be constructed (Figure 1-3).

Convergent Functional Genomics
Multiple Independent Lines of Evidence
For Identification and Prioritization

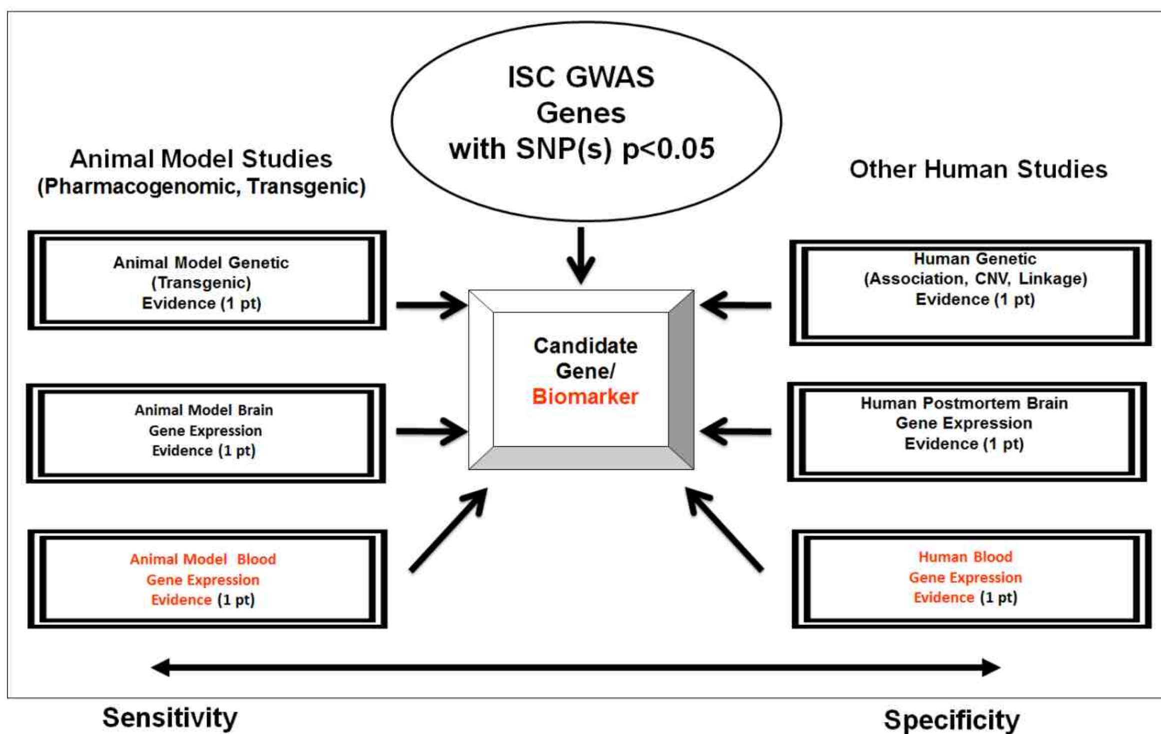


Figure 1-1. Convergent Functional Genomics.

An obvious next step is developing a way of applying that knowledge to genetic testing of individuals to determine risk for the disorder. On the basis of our comprehensive identification of top candidate genes described in this paper, we have chosen the nominally significant single-nucleotide polymorphisms (SNPs) inside those genes in the GWAS data set used for discovery (International Schizophrenia Consortium, ISC), and assembled a genetic risk prediction (GRP) panel out of those SNPs. We then developed a genetic risk prediction score (GRPS) for schizophrenia based on the presence or absence of the alleles of the SNPs associated with the illness in ISC, and tested the GRPS in independent cohorts (GAIN European Americans (EA), GAIN African Americans (AA), nonGAIN EA, nonGAIN AA) [26] for which we had both genotypic and clinical data available, comparing the schizophrenia subjects to normal controls. Our results show that a panel of SNPs in top genes identified and prioritized by CFG analysis can differentiate between schizophrenia subjects and controls at a population level, although at an individual level the margin is minimal. The latter point suggests that, like for bipolar disorder,[29] the contextual cumulative combinatorics of common variants and environment[30] plays a major role in risk for illness. Moreover, the genetic risk component identified by us seems to be stronger for classic age at onset schizophrenia than for early onset and late-onset schizophrenia, suggesting that those subtypes may be different, either in having a larger environmental component or having a different genetic component.

We have also looked at genetic heterogeneity, overlap and reproducibility between independent GWAS for schizophrenia. We show that the overlap is minimal at a nominal P-value SNP level, but increases dramatically at a gene level, then at a CFG-prioritized gene level and finally at a pathway level. CFG provides a fit-to disease prioritization of genes that leads to generalizability in independent cohorts, and counterbalances the fit-to-cohort prioritization inherent in classic SNP level genetic-only approaches, which have been plagued by poor reproducibility across cohorts. Finally, we have looked at overlap with candidate genes for other psychiatric disorders (bipolar disorder, anxiety disorders), as well as with other disorders affecting cognition (autism, Alzheimer disease (AD)), and provide evidence for shared genes.

Overall, this work sheds comprehensive light on the genetic architecture and pathophysiology of schizophrenia, provides mechanistic targets for therapeutic intervention and has implications for genetic testing to assess risk for illness before the illness manifests itself clinically.

Materials and Methods

Genome-wide association studies data for schizophrenia

The GWAS data from the ISC was used for the discovery CFG work. [25] This cohort consists of EA subjects (3322 schizophrenics and 3587 controls). SNPs with a nominal allelic P-value <0.05 were selected for our analysis. No Bonferroni correction was performed.

Four independent cohorts, [26] two EA (GAIN EA 1170 schizophrenics and 1378 controls; nonGAIN EA 1149 schizophrenics and 1347 controls) and two AA (GAIN AA 915 schizophrenics and 949 controls; nonGAIN AA 78 schizophrenics and 20 controls), were used for testing the results of the discovery analyses. The GWAS GAIN and nonGAIN data used for analyses described in this paper were obtained from the database of Genotype and Phenotype (dbGaP) found at www.ncbi.nlm.nih.gov.

The software package PLINK (http://pngu.mgh.harvard.edu/~purcell) was used to extract individual genotype information for each subject from the GAIN GWAS data files. We analyzed EA, and separately, AA, schizophrenia subjects and controls.

Gene identification

To identify the genes that correspond to the selected SNPs, the lists of SNPs from the GWAS were uploaded to NetAFFX (Affymetrix, Santa Clara, CA, USA; http://www.affymetrix.com/analysis/index.affx). We used the Netaffx na32 Genotyping Annotation build. In the cases where a SNP mapped to multiple genes, we selected all the genes. SNPs for which no gene was identified were not included in our subsequent analyses.

Keywords for Convergent Functional Genomics

Methods were as described above in the introduction, except that keywords used for scoring convergence were schizophrenia and psychosis. Additionally, in

this work all converging lines of evidence were weighted equally, with a maximum possible score of 6.

Pathway analyses

IPA 9.0. (Ingenuity Systems, Redwood City, CA, USA) was used to analyze the biological roles, including top canonical pathways, of the candidate genes resulting from our work (Table 1-2 and Supplementary Table S1-5), as well as used to identify genes in our data sets that are the target of existing drugs (Supplementary Table S1-2).

Intra-pathway epistasis testing

As an example, [29] the ISC GWAS data were used to test for epistatic interactions among the best P-value SNPs in genes from our data set present in a top canonical biological pathway identified by Ingenuity pathway analysis (Supplementary Table S1-4). SNPxSNP allelic epistasis was tested for each distinct pair of SNPs between genes, using the PLINK software package.

Genetic risk prediction panel and scoring

As we had previously done for bipolar disorder, [29] we developed a polygenic GRPS for schizophrenia based on the presence or absence of the alleles of the SNPs associated with illness, and tested the GRPS in independent cohorts for which we had both genotypic and clinical data available, comparing the schizophrenia subjects to normal controls. We tested two panels: a smaller one (GRPS-42) containing the single best P-value SNP in ISC in each of the top CFG prioritized genes (n = 42), and a larger one (GRPS-542), containing all the nominally

significant SNPs (n = 542) in ISC in the top CFG prioritized genes (n = 42; Tables 1-3, 1-4, Supplementary Table S1-3, and Figure 1-4).

Of note, our SNP panels and choice of affected alleles were based solely on analysis of the ISC GWAS, which is our discovery cohort, completely independently from the test cohorts. Each SNP has two alleles (represented by base letters at that position). One of them is associated with the illness (affected), the other not (non-affected), based on the odds ratios from the discovery ISC GWAS. We assigned the affected allele a score of 1 and the non-affected allele a score of 0. A two-dimensional matrix of subjects by GRP panel alleles is generated, with the cells populated by 0 or 1. A SNP in a particular individual subject can have any permutation of 1 and 0 (1 and 1, 0 and 1, 0 and 0). By adding these numbers, the minimum score for a SNP in an individual subject is 0, and the maximum score is 2. By adding the scores for all the alleles in the panel, averaging that, and multiplying by 100, we generate for each subject an average score corresponding to a genetic loading for disease, which we call Genetic Risk Predictive Score (GRPS).

The software package PLINK (http://pngu.mgh.harvard.edu/~purcell) was used to extract individual genotype information for each subject from the GAIN and nonGAIN GWAS data files. We analyzed separately EA and AA schizophrenia subjects and controls, to examine any potential ethnicity variability (Tables 1-3 and 1-4, and Supplementary Table S1-3). To test for significance, a one-tailed t-

test was performed between the schizophrenia subjects and the control subjects, looking at differences in GRPS.

Figures

Each figure in this chapter was completed by Mikias Ayalew, Helen Le-Niculescu, and Daniel Levey. This work has been published. [31]

Results

Top candidate genes

To minimize false negatives, we initially cast a wide net, using as a filter a minimal requirement for a gene to have both some GWAS evidence and some additional independent evidence. We thus generated an initial list of 3194 unique genes with at least a SNP at P<0.05 in the discovery GWAS analyzed (ISC), [25] that also had some additional evidence (human or animal model data), implicating them in schizophrenia (CFG score >1; Table 1-5). This suggests, using these minimal thresholds and requirements, that the repertoire of genes potentially involved directly or indirectly in cognitive processes and schizophrenia may be quite large, similar to what we have previously seen for bipolar disorder. [29]

To minimize false positives, we then used the CFG analysis integrating multiple lines of evidence to further prioritize this list of genes, and focused our subsequent analyses on only the top CFG scoring candidate genes. Overall, 186 genes had a CFG score of 3 and above (≥50% of maximum possible score of 6), and 42 had a CFG score of 4 and above (Tables 1-1 and 1-5, and Figure 1-2). Our top findings from ISC (Table 1-1) were over-represented in two independent

schizophrenia GWAS cohorts, the GAIN EA and GAIN AA. In total, 37 of the top 42 genes identified by our approach (88.1%) had at least a SNP with a P-value of <0.05 in those independent cohorts, an estimated twofold enrichment over what would be expected by chance alone at a genetic level (as there were 9002 genes at P<0.05 in the GAIN-EA GWAS, and the number of genes in the human genome is estimated at 20,500,[32] the enrichment factor provided by our approach is (37/42)/(9002/20 500) ≈ 2). Of note, there was no correlation between CFG prioritization and gene size, thus excluding a gene-size effect for the observed enrichment (Supplementary Figure S1-1).

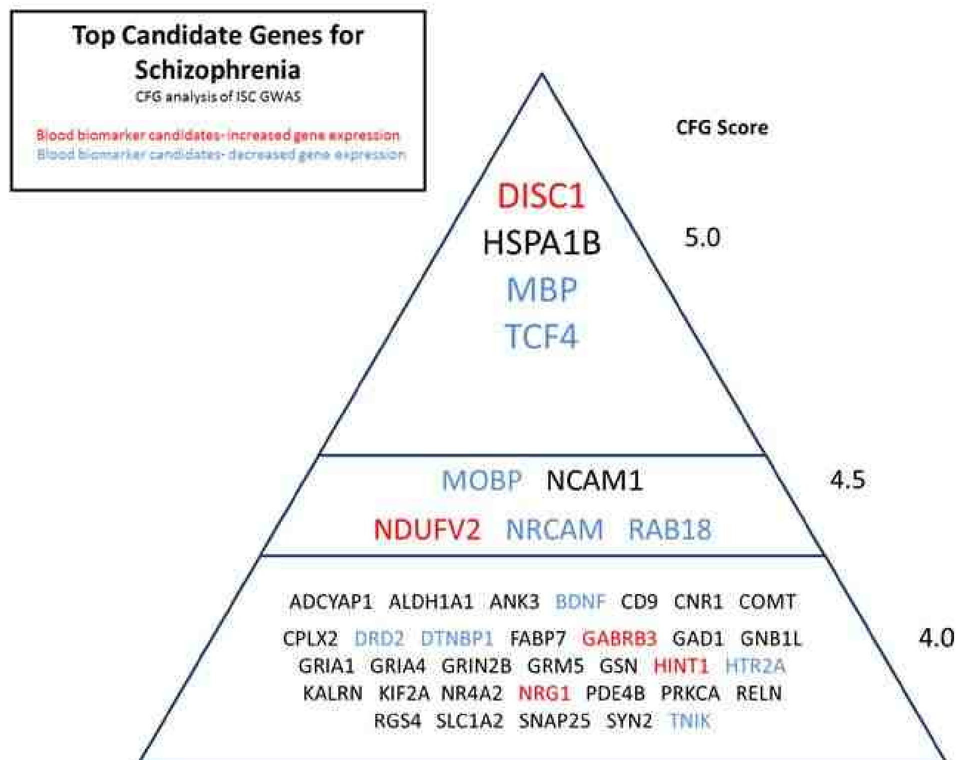


Figure 1-2. Top Candidate Genes for Schizophrenia

Table 1-1. Top candidate genes for schizophrenia - CFG analysis of ISC GWAS data. Top genes with a CFG score of 4 and above (n= 42) are shown. A more complete list of genes with CFG score of 3 and above (n= 186) is available in the Supplementary Information section

Gene Symbol/Name	ISC GWAS Best p-value SNP	Animal Model Genetic Evidence (TG)	Animal Model Brain Gene Expression Evidence	Animal Model Blood Gene Expression Evidence	Human Genetic Evidence (Association CNV or Linkage)	Human Postmortem Brain Gene Expression Evidence	Human Blood/ Other Peripheral Tissue Gene Expression	CFG Score	GAIN-EA GWAS Best p-value SNP	GAIN-AA GWAS Best p-value SNP

				Balaraman et al. 2007)			**Evidence**			
DISC1 disrupted in schizophrenia 1	0.002934 rs10910616	SZ (Koike, Arguello et al. 2006, Clapcote, Lipina et al. 2007, Duan, Chang et al. 2007, Hikida, Jaaro-Peled	(I) Antipsychotic treatment (Chiba, Hashimoto et al. 2006) Mouse neur		1q42.2 (Assoc) SZ (Callicott, Straub et al. 2005) (Liu, Fann et al. 2006, Palo, Antila et al.	(I) SZ HIP(Nakata, Lipska et al. 2009)	(I) SZ Lymphoblasts (Nakata, Lipska et al. 2009) PMBCs(Olincy, House et al. 2011)	5.0	0.001562 rs12754490	**0.0001308 rs11122318**

		et al. 2007, Kvajo, McKellar et al. 2008, Pletnikov, Ayhan et al. 2008, Pletnikov, Ayhan et al. 2008, Shen, Lang et al. 2008, Ayhan,	odevelopment (Schurov, Handford et al. 2004)		2007, Hennah, Thomson et al. 2009, Schumacher, Laje et al. 2009, Glessner, Reilly et al. 2010, Potkin, Macciardi et al. 2010, Green, Grozeva et al.					

		Abazyan et al. 2011, Ishizuka, Kamiya et al. 2011)			2011, Greenwood, Lazzeroni et al. 2011)					
HSPA1B heat shock 70kDa protein 1B	**0.00090 03 rs27639 79**		(D) PCP HIP; (D) CLZ VT (Le-Niculescu, Balaraman et al. 2007)	(D) CLZ	6p21.33 (Assoc) SZ(Pae, Drago et al. 2009)	(I) SZ(Arion, Unger et al. 2007)	(I) SZ IPS derived neurons (Brennand, Simone et al. 2011)	5.0		

MBP myelin basic protein	0.01002 rs12959006		(I) PCP and CLZ AMY; (D) PCP PFC[Le-Niculescu, Balaraman et al. 2007] (I) SZ[Ouchi, Kubota et al. 2005]	(I) PCP and CLZ	18q23 (Assoc) SZ[Baruch, Silberberg et al. 2009]	(D) SZ[Tkachev, Mimmack et al. 2003]	(D) SZ Peripheral blood[Glatt, Stone et al. 2011]	5.0	0.01345 rs9967028	0.03257 rs1629089

			(D) SZ (Fatemi, Pearce et al. 2005)							
TCF4 transcription factor 4	**0.0002902 rs17594665**	SZ(Brzozka, Radyushkin et al. 2010)	(I) PCP NAC (Le-Niculescu, Balaraman et al. 2007)		18q21.2 (Assoc) SZ(Stefansson, Ophoff et al. 2009, Li, Li et al. 2010, Lennertz,	(I) SZ (Mudge, Miller et al. 2008)	(I) SZ IPS derived neurons(Brennand, Simone et al. 2011	5.0	0.01039 rs17594665	0.00126 rs1539951

					Rujescu et al. 2011, Quednow, Ettinger et al. 2011)		); (D) Delusions SZ (Kurian, Le-Niculescu et al. 2011)			
MOBP myelin-associated oligodendrocyte basic protein	0.003529 rs170804 4		(I) PCP and CLZ AMY; (D) CLZ CP; (I)	(I) CLZ	3p22.1 (Linkage) SZ (Lewis, Levinson et al. 2003)	(D) SZ(Tkachev, Mimmack et al. 2003) (I) SZ and substance	(D) SZ Lymphocytes (Bowden, Weidenhofer et al. 2006)	4.5	0.02583 rs14057 98	0.004474 rs1538783

			CLZ NAC; (D) PCP PFC; (D) CLZ VT (Le-Niculescu, Balaraman et al. 2007) (I) Psychosis			abuse (Mitkus, Hyde et al. 2008)				

			(Lowe, Lu et al. 2007) (I) Response to antipsychotics (Takahashi, Kumanishi et al. 2004) (D) SZ(Matsuok							

			a, Tsunoda et al. 2008)							
NCAM1 neural cell adhesion molecule 1	0.003917 rs11214441	SZ (Wood, Tomasiewicz et al. 1998)	(I) SZ(Ouchi, Kubota et al. 2005)		11q23.2 (Assoc) SZ (Atz, Rollins et al. 2007)	(I) SZ (Atz, Rollins et al. 2007, Chan, Tsang et al. 2010)	(D) SZ IPS derived neurons (Brennand, Simone et al. 2011)	4.5	0.002043 rs1245133	0.001454 rs600964
<u>NDUFV2</u> NADH dehydrogena	0.003243 rs8084822		Response	(D) PCP	18p11.22 (Assoc) SZ(Wa	(D) SZ Striatum (Ben-	(I) SZ lymphocytes	4.5	0.004361 rs18	**0.0009294 rs10468792**

se (ubiquinon) flavoprotein 2, 24kDa			to antipsychotics (Ji, La et al. 2009)	and CLZ	shizuka, Kametani et al. 2006)	Shachar and Karry 2007) (I) SZ Parito-occipital cortex (Ben-Shachar and Karry 2007) (I) SZ(Prabakaran, Swatton et al. 2004)	(Ben-Shachar and Karry 2007)		931 44	

NRCAM neuronal cell adhesion molecule	0.006234 rs102500 83	SZ(Moy, Nonneman et al. 2009)	(I) CLZ VT (Le-Niculescu, Balaraman et al. 2007)		7q31.1 (Linkage) SZ(Ekelund, Lichtermann et al. 2000)	(D) SZ(Roussos, Katsel et al. 2011)	(D) SZ Serum(Schwarz, Guest et al. 2011)	4.5	0.004773 rs404287	0.002343 rs409797
RAB18 RAB18, member RAS oncogene family	0.03817 rs122616 90		(I) PCP AMY; (D) PCP PFC; (D)	(I) PCP	10p12.1 (Linkage) SZ (Faraone, Matise et	(D) SZ (Chu, Liu et al. 2009)	(D) SZ Whole Blood (Kuzman, Medved	4.5	0.01716 rs7476899	0.01231 rs11015796

			CLZ VT (Le-Niculescu, Balaraman et al. 2007)		al. 1998, Freedman, Leonard et al. 2001, Straub, MacLean et al. 2002, Maziade, Roy et al.		et al. 2009)			

					2005, Takahashi, Faraone et al. 2005)					
ADCYAP1 adenylate cyclase activating polypeptide 1 (pituitary)	0.002876 rs995457 4	SZ (Hashimoto, Hashimoto et al. 2007)	(I) CLZ NAC; (D) CLZ VT (Le-Niculescu, Balara		18p11.32 (Assoc) SZ (Hashimoto, Hashimoto et al. 2007);(Koga,	(D) SZ (Chu, Liu et al. 2009)		4.0	0.02558 rs1394890	0.005448 rs16953183

			man et al. 2007)		Ishiguro et al. 2010)					
ALDH1A1 aldehyde dehydrogenase 1 family, member A1	0.02526 rs111434 38		(I) PCP and CLZ AMY; (D) PCP and CLZ NAC (Le-Niculescu, Balaraman et al.	(I) PCP		(D) SZ(Galter, Buerven ich et al. 2003, Prabakaran, Swatton et al. 2004) (I) SZ(Barley, Dracheva et al. 2009)	(I) SZ Fibroblasts(Wang, Lockstone et al. 2010)	4.0	0.01389 rs7028573	0.01285 rs11999628

			2007) (I) Psychosis (Lowe, Lu et al. 2007)							
ANK3 ankyrin 3, node of Ranvier (ankyrin G)	0.001727 rs494825 6		(D) PCP AMY; (I) PCP CP; (D) PCP NAC; (I)		10q21.2 (Assoc) SZ (Athanasiu, Mattingsdal et al. 2010, Roussos,	(D) SZ (Roussos, Katsel et al. 2011)	(I) SZ IPS derived neurons (Brennand, Simone et al. 2011)	4.0	0.006456 rs10509133	0.005837 rs7906905

			CLZ VT (Le-Niculescu, Balaraman et al. 2007)		Katsel et al. 2011)					
BDNF brain-derived neurotrophic factor	0.001666 rs10742178		(I) SZ (Grotticк, Bagnol et al. 2005, Pillai and Mahadik 2008, Somm		11p14.1 (Assoc) SZ (Gratacos, Gonzalez et al. 2007, Chao, Kao et al.	(D) SZ(Torrey, Barci et al. 2005, Pillai 2008) (Weickert, Hyde et al. 2003)	(D) SZ Serum (Grillo, Ottoni et al. 2007, Miodownik, Maayan et al. 2011,	4.0		0.005507 rs2131060

			er, Schmitt et al. 2010) (I) PCP in rats (Kaiser, Foltz et al. 2004) (I) MK-801 in rats(Guo, Yang et al. 2010)		2008) (Decoster, van Os et al. 2011, Suchanek, Owczarek et al. 2011)		Rizos, Papathanasiou et al. 2011, Schwarz, Guest et al. 2011) (D) Leukocytes Psychosis (Mondelli, Cattaneo et al. 2011)			

CD9 CD9 molecule	0.0455 rs318129 1		(I) PCP AMY; (D) PCP PFC; (I) PCP VT (Le-Niculescu, Balaraman et al. 2007)	(I) PCP and CLZ		(D) SZ(Katsel, Davis et al. 2005)	(D) SZ IPS derived neurons (Brennand, Simone et al. 2011)	4.0	0.01167 rs2268014	0.04739 rs7342306
CNR1 cannabinoid recept	0.002567 rs13	SZ(Haller, Szirmai	(D) PCP and CLZ		<u>6q15</u> (Assoc) SZ	(D) SZ(Eggan, Hashimo		4.0	0.001542 rs94	0.002128 rs873413

or 1 (brain)	24073	et al. 2005)	VT (Le-Niculescu, Balaraman et al. 2007) (D) response to antipsychotics (Takahashi, Kumanishi et al.		(Ujike, Takaki et al. 2002, Xu, Woodroffe et al. 2009)	to et al. 2008)			51023	

			2004)							
COMT catechol-O-methyltransferase	0.04098 rs1544325	SZ(Huotari, Garcia-Horsman et al. 2004, Babovic, O'Tuathaigh et al. 2008)	(I) CLZ VT (Le-Niculescu, Balaraman et al. 2007) (D) SZ(Jurata, Gallagher et al. 2006)		22q11.21 (Assoc) SZ(Shifman, Bronstein et al. 2002)	(I) SZ(Abdolmaleky, Cheng et al. 2006)		4.0		0.01457 rs1544325
CPLX2 complexin 2	0.04338 rs10	SZ(Yamauchi, Qin et	(D) PCP and		5q35.2	(D)		4.0	0.01658 rs68	

	213927	al. 2005)	CLZ VT (Le-Niculescu, Balaraman et al. 2007) (D) SZ (Sommer, Schmitt et al. 2010)		(Assoc) SZ (Lee, Song et al. 2005)	SZ(Eastwood and Harrison 2005, Torrey, Barci et al. 2005, Weidenhofer, Bowden et al. 2006)			87620	
DRD2 dopamine receptor D2	0.01151 rs12791990		(D) PCP and CLZ PFC (Le-		11q23.2 (Assoc) SZ (Liu,	(D) SZ (Dean, Pavey et	(D) SZ Delusions (Kurian, Le-	4.0	0.007265 rs4938021	0.01096 rs17529477

			Niculescu, Balaraman et al. 2007) (D) SZ(Ortega-Alvaro, Aracil-Fernandez et al. 2011)		Downing et al. 2010, McClay, Adkins et al. 2011)	al. 2004, Torrey, Barci et al. 2005, Zhan, Kerr et al. 2011)	Niculescu et al. 2011)			
DTNBP1 dystrobrevin binding	0.002634 rs12209943	SZ (Chen, Feng et al. 2008, Feng,			6p22.3 (Assoc) SZ(Schwab,	(D) SZ(Weickert, Straub et al. 2004) (Talbot,	(D) Lymphocytes(Chagnon, Roy	4.0	0.009146 rs9477021	**0.0001501 rs16876575**

protein 1		Zhou et al. 2008, Hattori, Murotani et al. 2008, Takao, Toyama et al. 2008)			Knapp et al. 2003, Kirov, Ivanov et al. 2004, Fallin, Lasseter et al. 2005, Rethelyi, Bakker et al. 2010, Voisey, Swagell et al. 2010, Fatjo-Vilas, Papiol	Eidem et al. 2004, Weickert, Rothmond et al. 2008, Talbot, Louneva et al. 2011)	et al. 2008)			

					et al. 2011)					
FABP7 fatty acid binding protein 7, brain	0.01053 rs9490546	SZ (Watanabe, Toyota et al. 2007)	(I) CLZ NAC; (I) PCP and CLZ PFC (Le-Niculescu, Balaraman et al. 2007)		6q22.31 (Assoc) SZ (Watanabe, Toyota et al. 2007)	(I) SZ (Watanabe, Toyota et al. 2007)		4.0		
GABRB3 gamma-amino	0.004635 rs8037461		(I) PCP AMY; (D)		15q12 (Assoc) SZ (Bergen	(I) SZ (Bullock, Cardon	(I) SZ Serum (Silver,	4.0	0.01579 rs12904865	**0.0009769 rs4906835**

butyric acid (GABA) A receptor, beta 3			PCP HIP; (D) PCP PFC (Le-Niculescu, Balaraman et al. 2007)		, Fanous et al. 2009)	et al. 2008) (D) SZ(Hashimoto, Arion et al. 2008)	Susser et al. 2011)			
GAD1 glutamate decarboxylase 1 (brain, 67kDa)	0.03907 rs16859026		(I) PCP AMY (Le-Niculescu, Balaraman et al.	(I) CLZ	2q31.1 (Assoc) SZ (Straub, Lipska et al.	(D) SZ (Impagnatiello, Guidotti et al. 1998,		4.0	0.008447 rs10191129	0.01776 rs2883888

			2007) (D) SZ(Sommer, Schmitt et al. 2010); (Pillai and Mahadik 2008) (I) SZ(Bullock, Cardon et al. 2008)		2007)	Guidotti, Auta et al. 2000, Mirnics, Middleton et al. 2000, Costa, Davis et al. 2001, Heckers, Stone et al. 2002, Hashimoto, Volk et al. 2003, Veldic, Caruncho et al. 2004, Woo,				

						Walsh et al. 2004, Torrey, Barci et al. 2005, Veldic, Guidotti et al. 2005, Akbarian and Huang 2006, Straub, Lipska et al. 2007, Benes, Lim et al. 2008, Bullock, Cardon et al.				

						2008, Hashimoto, Arion et al. 2008, Hashimoto, Bazmi et al. 2008) (I) SZ(Hakak, Walker et al. 2001)				
GNB1L guanine nucleotide binding	0.03659 rs17745302	Impaired PPI (Paylor, Glaser	(I) Chronic haloperidol (Ishig		22q11.21 (Assoc) SZ (Williams,	(D) SZ (Ishiguro, Koga et al. 2010)		4.0		

protein (G protein), beta polypeptide 1-like		et al. 2006)	uro, Koga et al. 2010) (D) SZ(Jurata, Gallagher et al. 2006);		Glaser et al. 2008)					
GRIA1 glutamate receptor, ionotropic, AMPA 1	**0.0008031 rs2962816**	SZ (Wiedholz, Owens et al. 2008)	(D) PCP and CLZ AMY (Le-Niculescu, Balaraman et		5q33.2 (Assoc) SZ (Magri, Gardella et al. 2006, Leon,	(D) SZ(Sokolov 1998, Ibrahim, Hogg et al. 2000) (I) SZ(Dracheva,		4.0	0.00659 rs10044974	0.006037 rs498660

			al. 2007) (I) Response to antipsychotics (Takahashi, Kumanishi et al. 2004) (D) Response to		Schumacher et al. 2011)	McGurk et al. 2005, Stadler, Kolb et al. 2005, Benes, Lim et al. 2008)				

			PCP (Toyooka, Usui et al. 2002)							
GRIA4 glutamate receptor, ionotrophic, AMPA 4	0.02792 rs649098	Cognition(Impaired PPI)(Sagata, Iwaki et al. 2010)	(I) PCP and CLZ AMY (Le-Niculescu, Balaraman et al. 2007) (D) SZ(Chong, Young		<u>11q22.3</u> (Assoc) SZ(Makino, Fujii et al. 2003)	(D) SZ (Beneyto and Meador-Woodruff 2008, Chu, Liu et al. 2009) (I) SZ(Dracheva, McGurk et al. 2005)		4.0	0.001526 rs7116118	0.00343 rs2277280

			et al. 2002)							
GRIN2B glutamate receptor, ionotropic, N-methyl D-aspartate 2B	0.001569 rs436370 3	SZ (Kochlamazashvili, Senkov et al. 2010)	(D) CLZ AMY; (I) CLZ VT (Le-Niculescu, Balaraman et al. 2007) MK-801 treated		12p13.1 (Assoc) SZ(Fallin, Lasseter et al. 2005, Liu, Downing et al. 2010)	(D) SZ(Ibrahim, Hogg et al. 2000) (I) SZ(Bowden, Scott et al. 2008)		4.0	0.001427 rs1072388	0.003102 rs12826365

			rats(Marvanova, Lakso et al. 2004) (D) Response to antipsychotics (Takahashi, Kumanishi et al. 2004)							

			(D) SZ (Bullock, Cardon et al. 2008)							
GRM5 glutamate receptor, metabotropic 5	0.002559 rs992259	SZ (Kinney, Burno et al. 2003)	(D) PCP AMY; (D) PCP VT (Le-Niculescu, Balaraman et al. 2007) (D)		<u>11q14.3</u> (Assoc) SZ (Devon, Anderson et al. 2001)	(D) SZ(Benes, Lim et al. 2008)		4.0	0.01842 rs170110	0.001263 rs1846475

			SZ(Grottick, Bagnol et al. 2005)							
GSN gelsolin	0.04739 rs12376078		(I) PCP AMY; (D) CLZ CP (Le-Niculescu, Balaraman et al. 2007)	(I) PCP and CLZ	33.2 (Assoc) SZ(Xi, Qin et al. 2004)	(D) SZ(Hakak, Walker et al. 2001, Prabakaran, Swatton et al. 2004) (Davis and Haroutunian 2003, Katsel,		4.0	0.002313 rs767770	**0.0001564 rs4837835**

						Davis et al. 2005)				
HINT1 histidine triad nucleotide binding protein 1	**0.00008672 rs11242025**	SZ (Barbier, Zapata et al. 2007)			5q23.3 (Assoc) SZ (Chen, Wang et al. 2008, Kurotaki, Tasaki et al. 2011)	(D) SZ (Vawter, Crook et al. 2002)	(I) SZ Whole Blood (Kuzman, Medved et al. 2009)	4.0	0.008637 rs7734177	
HTR2A 5-hydroxytryptamine (serotonin)	0.02014 rs7985155		(I) Response to antipsyc		13q14.2 (Assoc) SZ (Liu, Downin	(D) SZ (Polesskaya and Sokolov 2002, Torrey,	(D) SZ Lymphocytes (Fukuda, Koga	4.0	0.003461 rs17070879	0.002529 rs1886439

receptor 2A			hotics (Takahashi, Kumanishi et al. 2004) (D) SZ(Steward, Kennedy et al. 2004) (I) SZ(Ortega-Alvaro, Aracil-		g et al. 2010, McClay, Adkins et al. 2011)	Barci et al. 2005, Garbett, Gal-Chis et al. 2008)	et al. 2006)			

			Fernandez et al. 2011)							
KALRN kalirin, RhoGEF kinase	0.006285 rs377275 6		(I) CLZ VT (Le-Niculescu, Balaraman et al. 2007)		3q21.1 (Assoc) SZ(Kushima, Nakamura et al. 2010)	(D) SZ(Hill, Hashimoto et al. 2006)	(D) SZ IPS derived neurons (Brennand, Simone et al. 2011)	4.0	0.01015 rs9832419	0.009074 rs1822791
KIF2A kinesin heavy chain member 2A	0.005374 rs686479 3		(D) CLZ VT (Le-Nicule	(I) PCP and	5q12.1 (Assoc) SZ(Li, Zheng	(D) SZ (Chu, Liu et al. 2009)		4.0	0.003396 rs153864	0.005207 rs10069830

			scu, Balaraman et al. 2007)	CLZ	et al. 2006)					
NR4A2 nuclear receptor subfamily 4, group A, member 2	**0.0006887 rs12465886**	SZ (Rojas, Joodmardi et al. 2007)	(I) PCP and CLZ HIP; (I) PCP and CLZ NAC; (D) PCP PFC; (I)		2q24.1 (Assoc) SZ(Huang, Perlis et al. 2010)		(I) SZ IPS derived neurons (Brennand, Simone et al. 2011)	4.0	0.001624 rs6743834	0.004081 rs16840214

			PCP and CLZ VT (Le-Niculescu, Balaraman et al. 2007)							
<u>NRG1</u> neuregulin 1	0.001731 rs1158001	SZ (Stefansson, Sigurdsson et al. 2002, O'Tuathaigh, Babovic et al.			<u>8p12</u> (Assoc) SZ (van Schijndel, van Loo et al. 2009, Walker,	(D) SZ (Tkachev, Mimmack et al. 2003, Law, Shannon Weickert et al.	(I) SZ IPS derived neurons (Brennand, Simone	4.0	0.00104 rs2716960	**0.000006564 rs6989777**

		2007, Chen, Johnson et al. 2008)			Christoforou et al. 2010, Greenwood, Lazzeroni et al. 2011, McClay, Adkins et al. 2011) (Stefansson, Sarginson et al. 2003, Mozhui 2011)	2004, Colantuoni, Hyde et al. 2008, Harris, Lockstone et al. 2009) (I) SZ(Hashimoto, Straub et al. 2004, Hahn, Wang et al. 2006, Law, Lipska et al. 2006)	et al. 2011) (I) SZ Leucocytes(Petryshen, Middleton et al. 2005) (I) SZ Lymphocyte (Middleton, Pato et al. 2005) (I)			

							SZ Delusions (Kurian, Le-Niculescu et al. 2011)			
PDE4B phosphodiesterase 4B, cAMP-specific	0.003042 rs658819 3	SZ (Siuciak, McCarthy et al. 2008)	(D) PCP (Le-Niculescu, Balaraman et al. 2007) (I) Psychosis (Dlaboga,		1p31.3 (Assoc) SZ (Pickard, Thomson et al. 2007, Fatemi, King et al.	(D) SZ (Fatemi, King et al. 2008)		4.0	0.02102 rs11805090	**0.000103 rs17417507**

			Hajjhussein et al. 2008) (D) Response to antipsychotics(Fatemi, Folsom et al. 2010)		2008, Tomppo, Hennah et al. 2009, Kahler, Otnaess et al. 2010, Mozhui 2011)					
PRKCA protein kinase	0.007991 rs65		(I) CLZ PFC		17q24.2 (Assoc)	(D) SZ(Torrey, Barci	(D) SZ IPS deriv	4.0	0.01166 rs99	**0.0004001 rs16959057**

C, alpha	0442 8		; (D) CLZ VT (Le- Nicule scu, Balara man et al. 2007) (I) Res pon se to anti psyc hoti cs (Chen and		SZ(Car roll, William s et al. 2010, Jablens ky, Morar et al. 2011)	et al. 2005)	ed neur ons (Brenna nd, Simone et al. 2011)		088 14	

			Chen 2005)							
RELN reelin	0.01368 rs27 11865	SZ (Ballmaier, Zoli et al. 2002)	(I) CLZ PFC ; (D) PCP and CLZ VT (Le-Niculescu, Balaraman et al. 2007) ; (D) SZ(Pillai		7q22.1 (Assoc) SZ (Kahler, Djurovic et al. 2008, Shifman, Johannesson et al. 2008, van Schijndel, van Loo et al.	(D) SZ(Impagnatiello, Guidotti et al. 1998, Guidotti, Auta et al. 2000, Knable, Barci et al. 2004, Veldic, Caruncho et al. 2004, Grayson, Jia et al.		4.0	0.007165 rs10227303	0.004365 rs7794418

			and Mahadik 2008, Fatemi, Reutiman et al. 2009)		2009, Liu, Downing et al. 2010, Alkelai, Lupoli et al. 2011, Greenwood, Lazzeroni et al. 2011, Ovadia and Shifman 2011)	2005, Torrey, Barci et al. 2005)				
RGS4 regulator of	0.004835 rs46	SZ (Grillet,	(I) PCP and		1q23.3	(D) SZ(Mirn		4.0	0.007928	0.007516

G-protein signaling 4	57235	Pattyn et al. 2005)	CLZ AMY; (I) CLZ HIP; (D) CLZ PFC; (I) PCP and CLZ VT (Le-Niculescu, Balaraman et al.		(Assoc) SZ(Fallin, Lassetter et al. 2005, So, Chen et al. 2008, Prasad, Almasy et al. 2010, McClay, Adkins et al. 2011)	ics, Middleton et al. 2001, Prasad, Chowdari et al. 2005, Erdely, Tamminga et al. 2006, Lipska, Mitkus et al. 2006, Arion, Unger et al. 2007, Bowden, Scott et al. 2007, Bowden,			rs12403644	rs10917637

			2007) (D) SZ(MacDonald, Eaton et al. 2005, McOmish, Burrows et al. 2008)			Scott et al. 2008) (I) SZ(Volk, Eggan et al. 2010)				
SLC1A2 solute carrier family 1 (glial high affinity	0.02565 rs379408 6		(D) CLZ AMY; (D) CLZ VT	(I) CLZ	11p13 (Assoc) SZ(Deng, Shibata et al.	(D) SZ (Weidenhofer, Bowden et al. 2006) (I)		4.0	0.03109 rs3829280	0.002563 rs12270460

glutamate transporter), member 2			(Le-Niculescu, Balaraman et al. 2007)		2004)	SZ(Matute, Melone et al. 2005, Shao and Vawter 2008)					
SNAP25 synaptosomal-associated protein, 25kDa	0.01815 rs6032783	SZ (Jeans, Oliver et al. 2007)	(I) SZ (MacDonald, Eaton et al. 2005, Sommer, Schmitt et al. 2010)		<u>20p12.2</u> (Assoc) SZ(Fanous, Zhao et al. 2010)	(D) SZ (Karson, Mrak et al. 1999, Hemby, Ginsberg et al. 2002, Thompson, Egbufoama et al.		4.0	0.005819 rs362616	0.01192 rs362560	

						2003, Glatt, Everall et al. 2005, Benes, Lim et al. 2008) (I) SZ(Gabriel, Haroutunian et al. 1997, Clark, Dedova et al. 2006, Benes, Lim et al. 2008)				

SYN2 synapsin II	0.003144 rs2960421	SZ (Dyck, Skoblenick et al. 2007); (Dyck, Skoblenick et al. 2009)	(D) PCP and CLZ AMY; (D) PCP CP (D) PCP and CLZ VT (Le-Niculescu, Balaraman et al. 2007) (I)		3p25.2 (Assoc) SZ SZ (Chen, He et al. 2004, Lee, Song et al. 2005, Saviouk, Moreau et al. 2007)	(D) SZ(Vawter, Thatcher et al. 2002, Arion, Unger et al. 2007) (Mirnics, Middleton et al. 2000) (I) SZ antipsychotic treatment(Chan,		4.0	0.042 rs2618406	0.02541 rs17671592

			SZ(Chong, Young et al. 2002) (D) SZ (MacDonald, Eaton et al. 2005) ((Sommer, Schmitt et al. 2010)			Tsang et al. 2010)				
<u>TNIK</u> TRAF2 and NCK interac	0.00 1377 rs26 0769		(I) CLZ VT (Le-Nicule		<u>3q26 .31</u> (Ass oc) SZ	(I) SZ (Glatt, Everall et	(D) SZ Lym phob lastoi	4.0	0.00 698 7 rs12	**0.0000 7179 rs1306 5441**

ting kinase			scu, Balaraman et al. 2007)		(Potkin, Turner et al. 2009, Shi, Levinson et al. 2009, Potkin, Macciardi et al. 2010, Mozhui 2011)	al. 2005)	d Cell Lines (Martin, Rollins et al. 2009)		639373	

Candidate blood biomarkers

Of the top candidate genes from Table 1-1 (see also Figure 1-2), 15 out of 42 have prior human blood evidence for change in schizophrenia, implicating them as potential blood biomarkers. The additional evidence provided by GWAS data suggests a genetic rather than purely environmental (medications, stress) basis for their alteration in disease, and their potential utility as trait rather than purely state markers.

Biological pathways

Pathway analyses were carried out on the top genes (Table 1-2), and on all the candidate genes (Supplementary Table S1-5). Notably, glutamate receptor signaling, G-protein--coupled receptor signaling and cAMP-mediated signaling were the top canonical pathways over-represented in schizophrenia, which may be informative for new drug discovery efforts by pharmaceutical companies.

Genetic risk prediction

Once the genes involved in a disorder are identified, and prioritized for likelihood of involvement, then an obvious next step is developing a way of applying that knowledge to genetic testing of individuals to determine risk for the disorder. Based on our identification of top candidate genes described above using CFG, we pursued a polygenic panel approach, with digitized binary scoring for presence or absence, similar to the one we have devised and used in the past for biomarkers testing[15] and for genetic testing in bipolar disorder. [29] Somewhat similar approaches but without CFG prioritization, attempted by other groups, have been either unsuccessful[33] or have required very large panels of markers. [9],[34]

We first chose the single best P-value SNPs in each of our top CFG prioritized genes (n = 42) in the ISC GWAS data set used for discovery, and assembled a GRP

Table 1-2. Ingenuity Pathway analyses of top candidate genes.

Discovery in ISC and reproducibility in two independent cohorts, GAIN EA and GAIN AA.

Top Canonical Pathways CFG ≥ 3	**P-Value**	**Ratio**
ISC (n=186 genes)		
Glutamate Receptor Signaling	9.25E-13	12/69 (0.174)
G-Protein Coupled Receptor Signaling	9.33E-13	27/530 (0.051)
CREB Signaling in Neurons	1.76E-12	17/202 (0.084)
cAMP-mediated signaling	3.55E-11	17/219 (0.078)
Neuropathic Pain Signaling In Dorsal Horn Neurons	3.64E-11	13/112 (0.116)
---	--------------	----------------------
GAIN EA (n=173 genes)		
Glutamate Receptor Signaling	4.57E-16	14/69 (0.203)
CREB Signaling in Neurons	4.72E-14	18/202 (0.089)
G-Protein Coupled Receptor Signaling	2E-13	27/530(0.051)
cAMP-mediated signaling	1.2E-12	18/219 (0.082)

Synaptic Long Term Potentiation	1.58E-12	14/114 (0.123)
GAIN AA (n= 201 genes)		
cAMP-mediated signaling	7.6E-17	23/219 (0.105)
Glutamate Receptor Signaling	1.09E-16	15/69 (0.217)
Synaptic Long Term Potentiation	2.24E-15	17/114(0.149)
G-Protein Coupled Receptor Signaling	2.43E-14	30/530 (0.057)
CREB Signaling in Neurons	4.52E-14	19/202 (0.094)

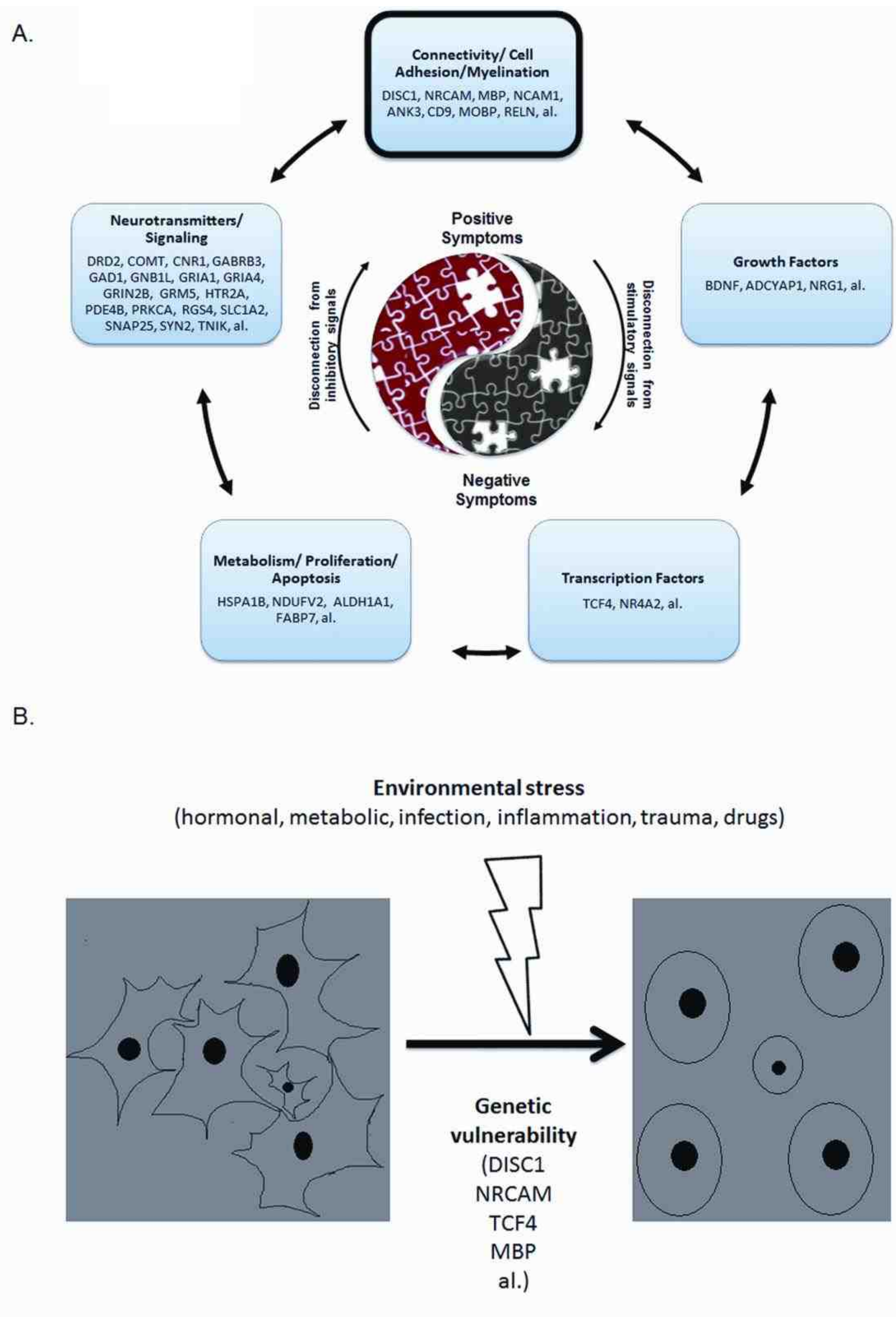


Figure 1-3. Schizophrenia as a Disease of Disconnection.

A. Biology of Schizophrenia B. Gene-Environment Interplay

panel out of those SNPs (Table 1-3). We then developed a GRPS for schizophrenia based on the presence or absence of the alleles of the SNPs associated with the illness, and tested the GRPS in independent cohorts (GAIN EA and GAIN AA), comparing the schizophrenia subjects to normal controls (Table 1-3). The results were not significant. We concluded that genetic heterogeneity at a SNP level is a possible explanation for these negative results. We then sought to see if we get better separation with a larger panel, containing all the nominally significant SNPs (n = 542) in the top CFG prioritized genes in ISC (n = 42), on the premise that a larger panel may reduce the heterogeneity effects, as different SNPs might be more strongly associated with illness in different cohorts. We found that our larger panel of SNPs was indeed able to significantly distinguish schizophrenics from controls in both GAIN EA and GAIN AA, two independent cohorts of different ethnicities. To verify this unexpectedly strong result, we further tested our panel in two other independent cohorts, nonGAIN EA and nonGAIN AA, and obtained similarly significant results (Table 1-4 and Figure 1-4).

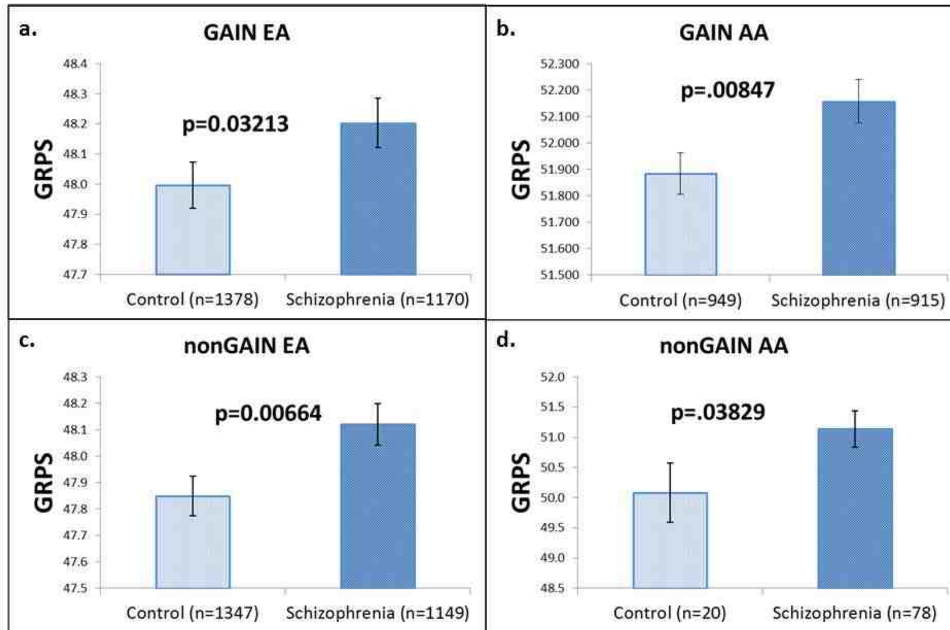


Figure 1-4. Genetic Risk Prediction of Schizophrenia in four independent cohorts

We also looked at whether our GRPS score distinguishes classic age of onset schizophrenia (defined by us as ages 15 to 30 years) from early onset (before 15 years) and late-onset (after 30 years) illness. Our results show that classic age of onset schizophrenia has a significantly higher GRPS than early or late-onset schizophrenia, in three out of the four independent cohorts of two different ethnicities (Figure 1-5).

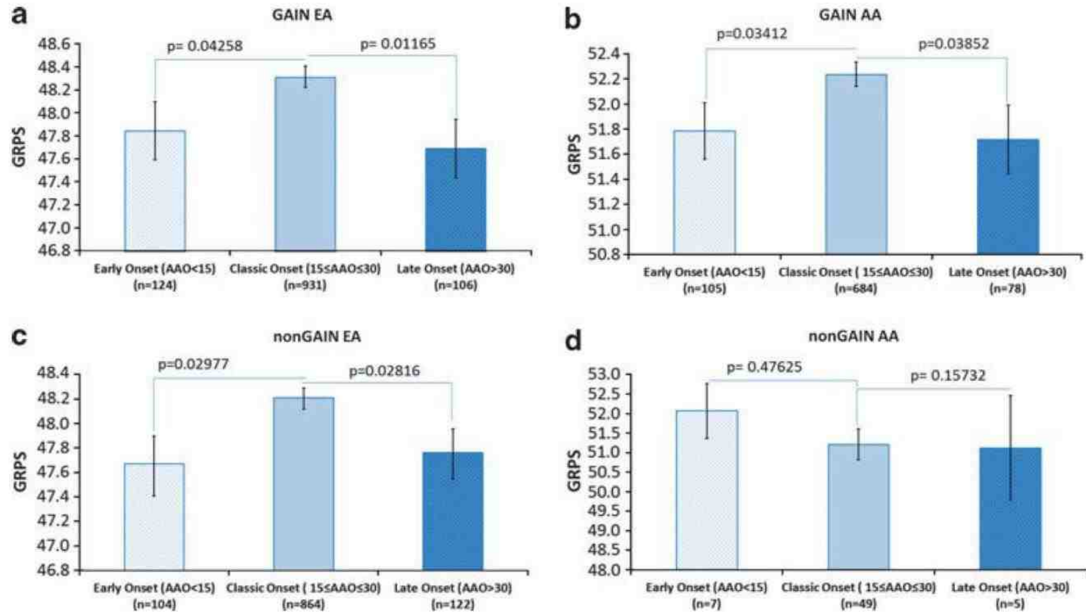


Figure 1-5. Genetic Risk Score and Age at Onset of Schizophrenia. AA, African American; AAO, age at onset; EA, European American; GRPS, genetic risk prediction score.

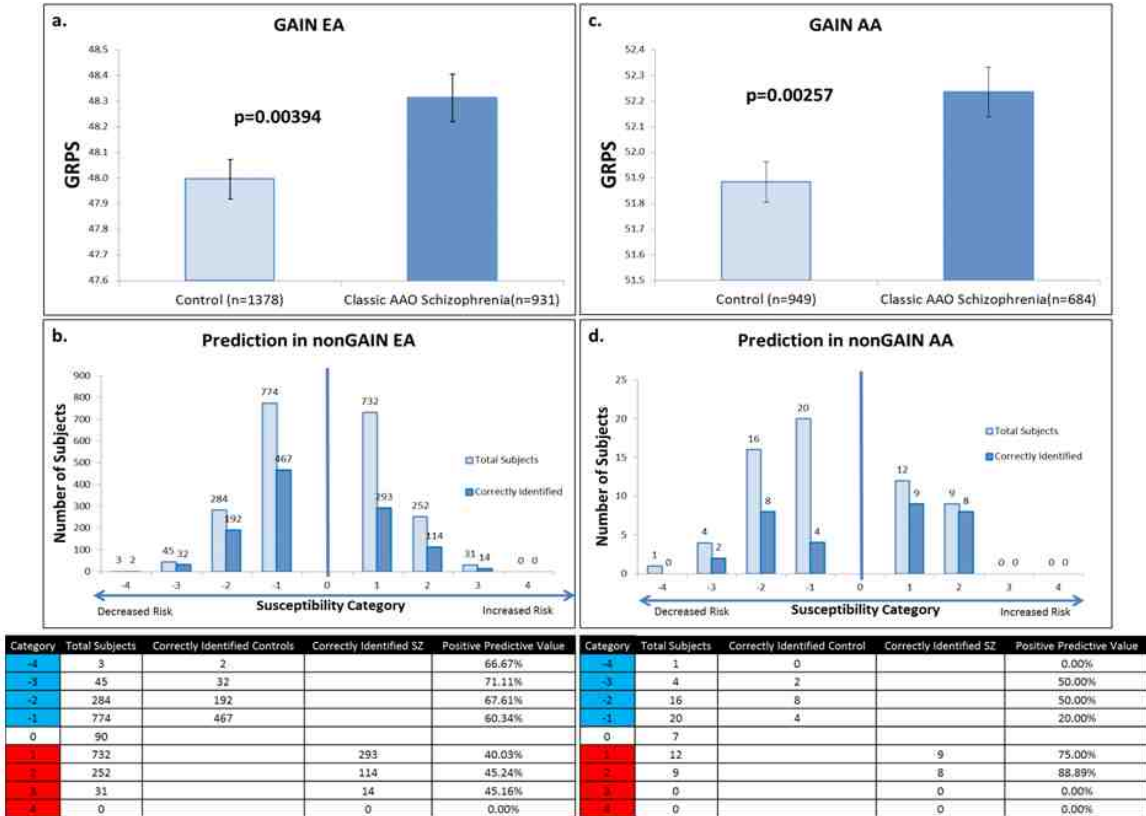
Finally, as we had done previously for bipolar disorder, [29] we developed a prototype of how the GRPS score could be used in testing individuals to establish their category of risk for schizophrenia (Figure 1-6). The current iteration of the test, using the panel of 542 SNPs, seems to be able to distinguish in independent cohorts who is at lower risk for classic age of onset schizophrenia in two out of three EA subjects, and who is at higher risk for classic age of onset schizophrenia in three out of four AA subjects.

Table 1-3. GRPS-42: non- differentiation between schizophrenics and controls in independent cohorts using a panel composed of the single best SNP from ISC in each of the top candidate genes (42 SNPs, in 42 genes).

Description of panel	GAIN-EA	GAIN-AA
Single Best p-value SNPS in each of the top 42 candidate genes from ISC GWAS n=42	p= 0.10308 39 out of the 42 ISC SNPs were present in GAIN-EA	p= 0.13567 37 out of the 42 ISC SNPs were present in GAIN-AA

Table 1-4. GRPS-542: differentiation between schizophrenics and controls in four independent cohorts using a panel composed of all the nominally significant SNPs from ISC in the top candidate genes (542 SNPs in 42 genes).

GAIN EA	GAIN AA
p= 0.03213 527 SNPs in 41 genes were present in GAIN-EA	**p= 0.00847** 516 SNPs in 42 genes were present in GAIN-AA
nonGAIN EA	**nonGAIN AA**
p= 0.00664 537 SNPs in 42 genes were present in nonGAIN EA	**p= 0.03829** 537 SNPs in 42 genes were present in nonGAIN AA



Category	Total Subjects	Correctly Identified Controls	Correctly Identified SZ	Positive Predictive Value
-4	3	2		66.67%
-3	45	32		71.11%
-2	284	192		67.61%
-1	774	467		60.34%
0	90			
1	732		293	40.03%
2	252		114	45.24%
3	31		14	45.16%
4	0		0	0.00%

Category	Total Subjects	Correctly Identified Control	Correctly Identified SZ	Positive Predictive Value
-4	1	0		0.00%
-3	4	2		50.00%
-2	16	8		50.00%
-1	20	4		20.00%
0	7			
1	12		9	75.00%
2	9		8	88.89%
3	0		0	0.00%
4	0		0	0.00%

Figure 1-6. Prototype of how GRPS testing could be used at an individual rather than population level, to aid diagnostic and personalized medicine approaches. We used the average values and standard deviation values for GRPS from the GAIN samples from each ethnicity (EA and AA) as thresholds for predictive testing in the independent nonGAIN EA and nonGAIN AA cohorts. The average GRPS score for schizophrenics in the GAIN cohort is used as a cut-off for schizophrenics in the test cohort (i.e., being above that threshold), and the average GRPS score for controls in the GAIN cohort is used as a cutoff for controls in the test nonGAIN cohort (i.e., being below that threshold). The subjects who are in between these two thresholds are called undetermined. Furthermore, to stratify risk, we categorized subjects into risk categories (in red increased risk, in blue decreased risk):

Category 1 if they fall within one standard deviation above the schizophrenics' threshold, and Category -1 if they fall within one standard deviation below the controls threshold. Category 2 and -2 subjects are between one and two standard deviations from the thresholds, Category 3 and -3 between two and three standard deviations, and Category 4 and -4 are those who fall beyond three standard deviations of the thresholds. The positive predictive value (PPV) of the tests increases in the higher categories, and the test is somewhat better at distinguishing controls in EA (i.e., in a practical application, individuals that are lower risk of developing the illness), and schizophrenics in AA (i.e., in a practical application, individuals that are higher risk of developing the illness).

Overlap among studies

We examined the overlap at a nominally significant (P<0.05) SNP level between ISC, GAIN EA and GAIN AA, and found that a minority of these SNPs (0.4%) overlap (Table 1-5 and Figure 1-7). We then examined the overlap at a gene level, then CFG prioritized genes level and finally biological pathways level, and found increasing evidence of commonality and reproducibility of findings across studies.

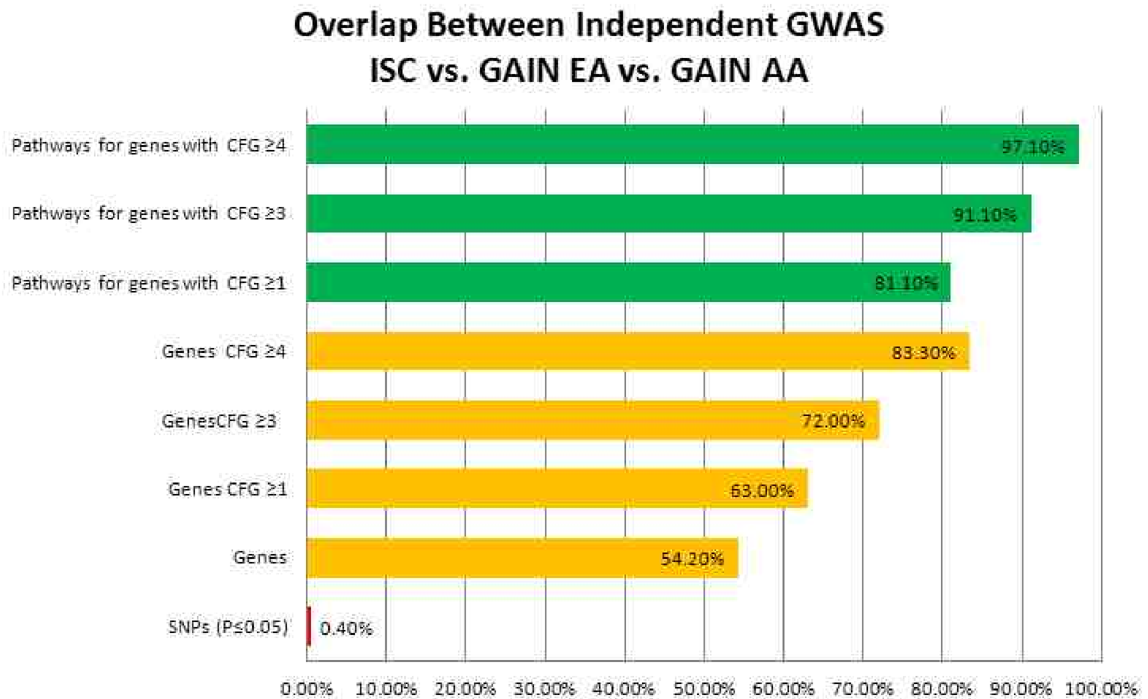


Figure 1-7. Overlap Between Independent GWAS. AA, African American; EA, European American; CFG, convergent functional genomics; ISC, International Schizophrenia Consortium; SNP, single-nucleotide polymorphism.

Table 1-5. Reproducibility between independent GWAS.

Increasing consistency and overlap observed from nominally significant SNPs (0.4%) to genes, then to CFG prioritized genes , and finally to pathways of CFG prioritized genes (97.1%).

Numbers and overlap across studies	ISC	GAIN-EA	GAIN-AA	ISC vs. GAIN-EA	ISC vs. GAIN-AA	GAIN-EA vs. GAIN-AA	ISC vs. GAIN-EA vs. GAIN-AA (% of ISC)
SNPs P≤0.05	45,972	42,336	57,118	2,649	2,986	2,839	163 (0.4%)
Genes	10,180	9,002	11,260	6,470	7,583	6,807	5,518 (54.2%)
Genes CFG ≥1	3,194	2,913	3,524	2,243	2,564	2,384	2,012 (63.0%)
Genes CFG ≥3	186	173	201	147	160	153	134 (72.0%)
Genes CFG ≥4	42	41	45	37	37	38	35 (83.3%)

Pathways for genes with CFG ≥1	217	210	205	194	188	180	176 (81.1%)
Pathways for genes with CFG ≥3	79	85	108	72	76	81	72 (91.1%)
Pathways for genes with CFG ≥4	34	50	75	33	34	48	33 (97.1%)

Discussion

Our CFG approach helped prioritize genes, such as DISC1 and MBP, with weaker evidence in the GWAS data but with strong independent evidence in terms of gene expression studies and other prior human or animal genetic work. Conversely, some of the top findings from GWAS, such as ZNF804A, have fewer different independent lines of evidence, and thus received a lower CFG prioritization score in our analysis (Supplementary Information Table S1-1),

although ZNF804A is clearly involved in schizophrenia-related cognitive processes.[35] While we cannot exclude that more recently discovered genes have had less hypothesis driven work done and thus might score lower on CFG, it is to be noted that the CFG approach integrates predominantly non-hypothesis driven, discovery-type data sets, such as gene expression, GWAS, CNV, linkage and quantitative traits loci. We also cap each line of evidence from an experimental approach (Figure 1-1) at a maximum score of 1, to minimize any 'popularity' bias, whereas multiple studies of the same kind are conducted on better-established genes. In the end, it is gene-level reproducibility across multiple approaches and platforms that is built into the approach and gets prioritized most by CFG scoring during the discovery process. Our top results subsequently show good reproducibility and predictive ability in independent cohort testing, the litmus test for any such work.

At the very top of our list of candidate genes for schizophrenia, with a CFG score of 5, we have four genes: DISC1, TCF4, MBP and HSPA1B. An additional five genes have a CFG score of 4.5: MOBP, NRCAM, NCAM1, NDUFV2 and RAB18.

DISC1 (Disrupted-in Schizophrenia 1), encodes a scaffold protein that has an impact on neuronal development and function,[36], [37], [38] including neuronal connectivity.[39] DISC1 has been identified as a susceptibility gene for major mental disorders by multiple studies.[40],[41],[42] DISC1 isoforms are upregulated in expression in blood cells in schizophrenia, thus serving as a potential peripheral biomarker as well.[43],[44] Developmental stress interacts with DISC1 expression to produce

neuropsychiatric phenotypes in mice.[45] Notably, its interacting partners PDE4B,[46] TNIK,[47] FEZ1[48] and DIXDC1[49] are also present on our list of prioritized candidate genes, with CFG scores of 4, 4, 3.5 and 2.5, respectively (Table 1-1 and Supplementary Table S1-1).

TCF4 (transcription factor 4) encodes a basic helix-turn-helix transcription factor, expressed in immune system as well as neuronal cells. It is required for the differentiation of subsets of neurons in the developing brain. There are multiple alternatively spliced transcripts that encode different proteins, providing for biological diversity and heterogeneity. Defects in this gene are a cause of Pitt-Hopkins syndrome, characterized by mental retardation with or without associated facial dimorphisms and intermittent hyperventilation. TCF4 has additional genetic evidence for association with schizophrenia-relevant phenotypes.[50],[51],[52],[53] It is changed in expression in postmortem brain,[54] induced pluripotent stem cell-derived neurons[27] and blood from schizophrenia patients. [24] Notably, it is a candidate blood biomarker for level of delusional symptoms (decreased in high delusional states) based on our previous work. [24]

MBP (myelin basic protein) is a major constituent of the myelin sheath of oligodendrocytes and Schwann cells in the nervous system. MBP-related transcripts are also present in the bone marrow and the immune system. MBP has additional genetic evidence for association with schizophrenia.[55] It is decreased in expression in postmortem brain[56] and blood[57] from schizophrenia patients. MBP is also changed in expression in the brain and blood of a pharmacogenomics mouse

model of schizophrenia, based on our previous work. [23] It was also decreased in expression in a stress-reactive genetic mouse model of bipolar disorder, [58] and treatment with the omega-3 fatty acid docosahexaenoic acid led to an increase in expression. Notably, MBP is a candidate blood biomarker for level of mood symptoms (increased in high mood states in bipolar subjects), based on our previous work. [15] Overall, the data indicate that MBP and other myelin-related genes[59],[60] may be involved in the effects of stress on psychosis and mood. Demyelinating disorders such as multiple sclerosis tend to be precipitated and exacerbated by stress, and have co-morbid psychiatric symptoms.[61] Of note, other myelin-related genes are also present on our list of prioritized candidate genes: MOBP and MOG, with CFG scores of 4.5 and 3, respectively (Table 1-1 and Supplementary Table S1-1).

HSPA1B (heat-shock 70-kDa protein 1B), a chaperone involved in stress response, stabilizes existing proteins against aggregation and mediates the folding of newly translated proteins. HSPA1B has additional genetic evidence for association with schizophrenia.[62] It is changed in expression in postmortem brain[63] and induced pluripotent stem cell-derived neurons[27] from schizophrenia patients. HSPA1B is also increased in expression in the brain and blood of a pharmacogenomics mouse model of schizophrenia, based on our previous work. [23] It was also codirectionally changed in the brain and blood in a pharmacogenomics mouse model of anxiety disorders, we have recently described,[64] as well as in a stress-reactive genetic mouse model.40 Treatment with

the omega-3 fatty acid docosahexaenoic acid reversed the increase in expression of HSPA1B in this stress reactive genetic mouse model.[65] Another closely related molecule, HSPA1A (heat-shock 70-kDa protein 1A), is also present on our list of prioritized candidate genes, with a CFG score of 3.5 (Supplementary Table S1-1). Heat-shock proteins may be involved in the biological and clinical overlap and interdependence between response to stress, anxiety and psychosis.

NRCAM (neuronal cell adhesion molecule) encodes a neuronal cell adhesion molecule. This ankyrin-binding protein is involved in neuron--neuron adhesion and promotes directional signaling during axonal cone growth. NRCAM is also expressed in non-neural tissues and may have a general role in cell--cell communication via signaling from its intracellular domain to the actin cytoskeleton during directional cell migration. It is decreased in expression in postmortem brain[66] and peripherally in serum[67] from schizophrenia patients. NRCAM is also changed in expression in the brain of a pharmacogenomics mouse model of schizophrenia, based on our previous work. [23] It was also increased in the amygdala in a stress-reactive genetic mouse model studied by our group.40 Another closely related molecule, NCAM1 (neural cell adhesion molecule 1), is among our top candidate genes as well. These data support a central role for cell connectivity and cell adhesion in schizophrenia.

Another top candidate gene is CNR1 (cannabinoid receptor 1, brain). CNR1 is a member of the guanine-nucleotide-binding protein (G-protein) coupled receptor family, which inhibits adenylate cyclase activity in a dose-dependent

manner. CNR1 has additional genetic evidence for association with schizophrenia.[68],[69] It is decreased in expression in postmortem brain from schizophrenics.[70] The other main cannabinoid receptor, CNR2 (cannabinoid receptor 2), is among our top candidate genes too (Supplementary Table S1-1), and is decreased in expression in postmortem brain from schizophrenics as well. These data support a role for the cannabinoid system in schizophrenia, perhaps through a deficiency of the endogenous cannabinoid signaling that leads to vulnerability to psychotogenic stress,[71] and is accompanied by increased compensatory exogenous cannabinoid consumption that may have additional deleterious consequences.[72]

A number of glutamate receptor genes are present among our top candidate genes for schizophrenia (GRIA1, GRIA4, GRIN2B and GRM5), as well as GAD1, an enzyme involved in glutamate metabolism, and SLC1A2, a glutamate transporter (Table 1-1). Other genes involved in glutamate signaling present in our data, with a lower scores, are GRIN2A, SLC1A3, GRIA3, GRIK4, GRM1, GRM4 and GRM7 (Supplementary Table S1-1). Glutamate receptor signaling is one of the top canonical pathways over-represented in our analyses (Table 1-2), and that finding is reproduced in independent GWA data sets (Table 1-2). One has to be circumspect with interpreting such results, as glutamate signaling is quasi-ubiquitous in the brain, and a lot of prior hypothesis-driven work has focused on this area, potentially biasing the available evidence. Nevertheless, our results are striking, and contribute to the growing body of evidence that has emerged over

the last few years implicating glutamate signaling as a point of convergence for findings in schizophrenia,[73] as well as for autism[74] and AD.[75] Glutamate signaling is the target of active drug development efforts,[76] which may be informed and encouraged by our current findings.

Our analysis also provides evidence for other genes that have long been of interest in schizophrenia, but have had previous variable evidence from genetic-only studies: BDNF, COMT, DRD2, DTNBP1 (dystrobrevin binding protein1/dysbindin; Table 1-1). In addition, our analysis provides evidence for genes that had previously not been widely implicated in schizophrenia, but do have relevant biological roles, demonstrating the value of empirical discovery-based approaches such as CFG (Table 1-1): ANK3, [66] ALDH1A1 and ADCYAP1, which is a ligand for schizophrenia candidate gene VIPR2,[77],[78] also present in our data set, albeit with a lower CFG score of 2. Other genes of interest in our full data set (Supplementary Table S1-1) include ADRBK2 (GRK3), first described by us as a candidate gene for psychosis, [19] CHRNA7,[79] and PDE10A,[80] which are targets for drug development efforts.

Pathways and mechanisms

Our pathway analyses results are consistent with the accumulating evidence about the role of synaptic connections and glutamate signaling in schizophrenia, most recently from CNV studies[81] (Table 1-2, Supplementary Table S1-5, Figure 1-3). Very importantly, the same top pathways were consistent across independent GWA studies we analyzed (Tables 1-2, 1-5, and Supplementary Table S1-5). We

also did a manual curation of the top candidate genes and their grouping into biological roles examining them one by one using PubMed and GeneCards, to come up with a heuristic model of schizophrenia (Figure 1-3). Overall, while multiple mechanistic entry points may contribute to schizophrenia pathogenesis (Figure 3a-1), it is likely at its core a disease of decreased cellular connectivity precipitated by environmental stress during brain development, on a background of genetic vulnerability (Figure 1-3b).

Genetic risk prediction

Of note, our SNP panels and choice of affected alleles were based solely on analysis of the discovery ISC GWAS, completely independently from the test GAIN EA, GAIN AA, nonGAIN EA and nonGAIN AA GWAS. Our results show that a relatively limited and well-defined panel of SNPs identified based on our CFG analysis could differentiate between schizophrenia subjects and controls in four independent cohorts of two different ethnicities, EA and AA. Moreover, the genetic risk component identified by us seems to be stronger for classic age of onset schizophrenia than for early or late-onset illness, suggesting that the latter two may be more environmentally driven or have a somewhat different genetic architecture. It is likely that such genetic testing will have to be optimized for different cohorts if done at a SNP level. Interestingly, at a gene and pathway level, the differences between studies seem much less pronounced than at a SNP level, if at all present (Table 1-5), suggesting that gene-level and pathway-level tests may have more universal applicability. In the end, such genetic data, combined

with family history and other clinical information (phenomics),[82] as well as with blood biomarker testing, [15] may provide a comprehensive picture of risk of illness.[83],[84]

Reproducibility among studies

Our work provides striking evidence for the advantages, reproducibility and consistency of gene-level analyses of data, as opposed to SNP level analyses, pointing to the fundamental issue of genetic heterogeneity at a SNP level (Table 1-5 and Figure 1-7). In fact, it may be that the more biologically important a gene is for higher mental functions, the more heterogeneity it has at a SNP level[85] and the more evolutionary divergence,[86] for adaptive reasons. On top of that, CFG provides a way to prioritize genes based on disease relevance, not study-specific effects (that is, fit-to-disease as opposed to fit-to-cohort). Reproducibility of findings across different studies, experimental paradigms and technical platforms is deemed more important (and scored as such by CFG) than the strength of finding in an individual study (for example, P-value in a GWAS). The CFG prioritized genes show even more reproducibility among independent GWAS cohorts (ISC, GAIN EA, GAIN AA) than the full list of unprioritized genes with nominal significant SNPs. The increasing overlap and reproducibility between studies of genes with a higher average CFG score points out to their biological relevance to disease architecture. Finally, at a pathway level, there is even more consistency across studies. Again, the pathways derived from the top CFG scoring genes show more consistency than the pathways derived from the lower CFG scoring genes. Overall,

using our approach, we go from reproducibility between independent studies of 0.4% at the level of nominally significant SNPs to a reproducibility of 97.1% at the level of pathways derived from top CFG scoring genes.

Overlap with other psychiatric disorders

Despite using lines of evidence for our CFG approach that have to do only with schizophrenia, the list of genes identified has a notable overlap with other psychiatric disorders (Figure 1-8, Supplementary Table S1-1). This is a topic of major interest and debate in the field.[30],[87] We demonstrate an overlap between top candidate genes for schizophrenia and candidate genes for anxiety and bipolar disorder, previously identified by us through CFG (Figure 1-8), thus providing a possible molecular basis for the frequently observed clinical co-morbidity and interdependence between schizophrenia and those other major psychiatric disorders, as well as cross-utility of pharmacological agents. In particular, PDE10A is at the overlap of all three major psychiatric domains, and may be of major interest for drug development.[80] The overlap between schizophrenia and bipolar may have to do primarily with neurotrophicity and brain infrastructure (underlined by genes such as DISC1, NRG1, BDNF, MBP, NCAM1, NRCAM, PTPRM). The overlap between schizophrenia and anxiety may have to do primarily to do with reactivity and stress response (underlined by genes such as NR4A2, QKI, RGS4, HSPA1B, SNCA, STMN1, LPL). Notably, the overlap between schizophrenia and anxiety is of the same magnitude as the previously better appreciated overlap between schizophrenia and bipolar disorder, [23],[88] supporting the consideration of

a nosological domain of schizoanxiety disorder,[64] by analogy to schizoaffective disorder. Clinically, while there are some reports of co-morbidity between schizophrenia and anxiety,[89] it is an area that has possibly been under-appreciated and understudied. 'Schizoanxiety disorder' may have heuristic value and pragmatic clinical utility.

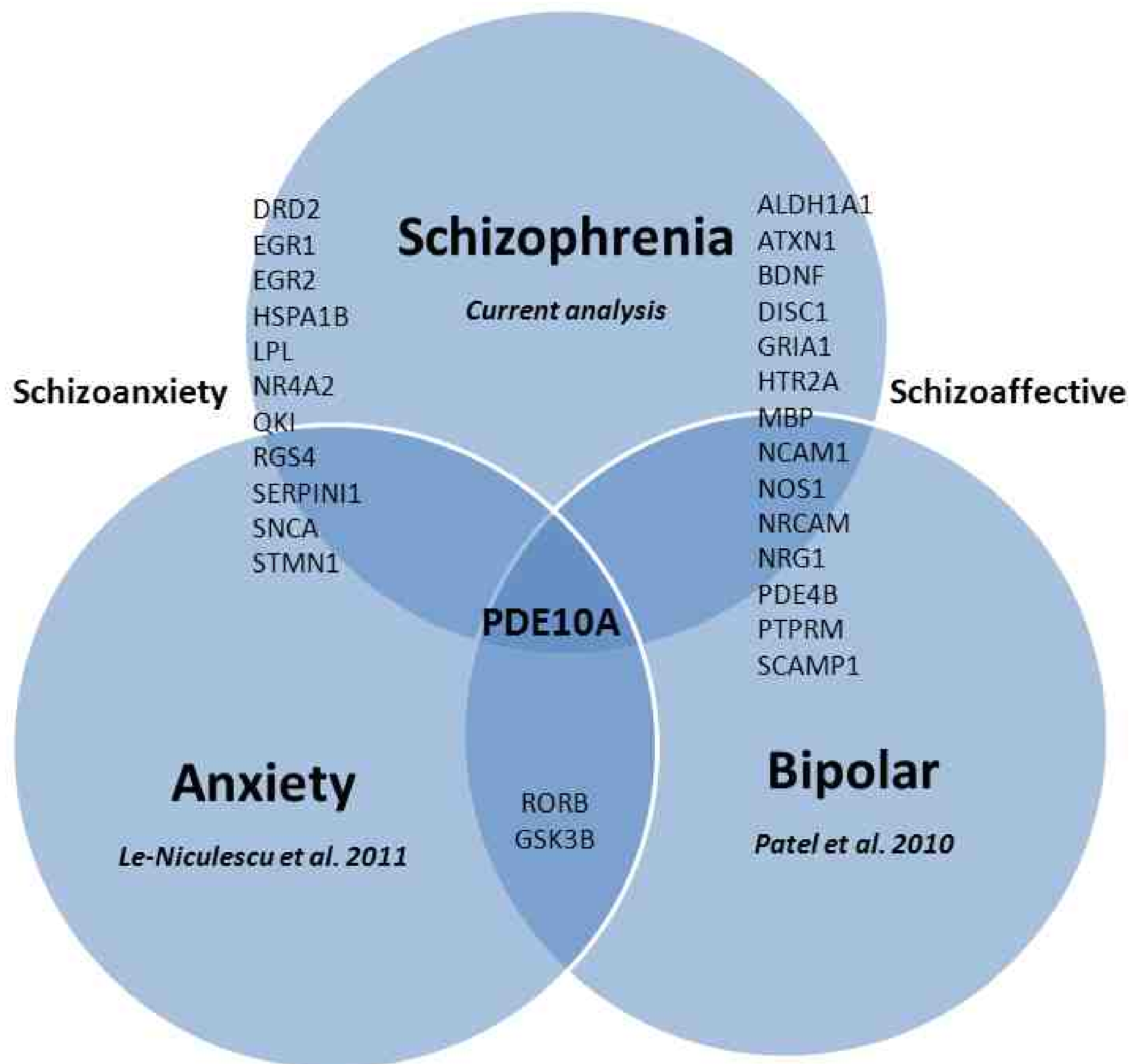


Figure 1-8. Genetic Overlap Among Psychiatric Disorders.

We also looked at the overlap with candidate genes for autism and AD from the literature (Supplementary Table S1-1), to elucidate whether schizophrenia, autism and AD might be on a spectrum, that is, whether autism might be a form of 'schizophrenia praecox', similar to schizophrenia being referred to as 'dementia praecox' (Kraepelin). We see significant overlap between the three disorders among the top genes with a CFG score of 4: a third of the genes overlap between schizophrenia and autism, and a quarter between schizophrenia and AD. Additional key genes of interest are lower on the list as well, with a CFG score of 3: CNTNAP2 for autism, MAPT and SNCA for AD (Supplementary Table S1-1).

Conclusions and future directions

First, in spite of its limitations, our analysis is arguably the most comprehensive integration of genetics and functional genomics to date in the field of schizophrenia, yielding a comprehensive view of genes, blood biomarkers, pathways and mechanisms that may underlie the disorder. From a pragmatic standpoint, we would like to suggest that our work provides new and/or more comprehensive insights on genes and biological pathways to target for new drug development by pharmaceutical companies, as well as potential new uses in schizophrenia for existing drugs, including omega-3 fatty acids (Supplementary Table S1-2).

Second, our current work and body of work over the years provides proof how a combined approach, integrating functional and genotypic data, can be used for complex disorders-psychiatric and non-psychiatric, as has been attempted by

others as well.[90],[91] What we are seeing across GWAS of complex disorders are not necessarily the same SNPs showing the strongest signal, but rather consistency at the level of genes and biological pathways. The distance from genotype to phenotype may be a bridge too far for genetic-only approaches, given genetic heterogeneity and the intervening complex layers of epigenetics and gene expression regulation.[92] Consistency is much higher at a gene expression level (Table 1-5),[93] and then at a biological pathway level. Using GWAS data in conjunction with gene expression data as part of CFG or integrative genomics[94] approaches, followed by pathway-level analysis of the prioritized candidate genes, can lead to the unraveling of the genetic code of complex disorders such as schizophrenia.

Third, our work provides additional integrated evidence focusing attention and prioritizing a number of genes as candidate blood biomarkers for schizophrenia, with an inherited genetic basis (Table 1-1 and Figure 1-2). While prior evidence existed as to alterations in gene expression levels of those genes in whole-blood samples or lymphoblastoid cell lines from schizophrenia patients, it was unclear prior to our analysis whether those alterations were truly related to the disorder or were instead related only to medication effects and environmental factors.

Fourth, we have put together a panel of SNPs, based on the top candidate genes we identified. We developed a GRPS based on our panel, and demonstrate how in four independent cohorts of two different ethnicities, the GRPS

differentiates between subjects with schizophrenia and normal controls. From a personalized medicine standpoint, genetic testing with highly prioritized panels of best SNP markers may have, upon further development (Figure 1-6) and calibration by ethnicity and gender, a role in informing decisions regarding early intervention and prevention efforts; for example, for classic age of onset schizophrenia before the illness fully manifests itself clinically, in young offspring from high-risk families. After the illness manifests itself, gene expression biomarkers and phenomic testing approaches, including clinical data, may have higher yield than genetic testing. A multi-modal integration of testing modalities would be the best approach to assess and track patients, as individual markers are likely to not be specific for a single disorder. The continuing re-evaluation in psychiatric nosology[84],[95] brought about by recent advances will have to be taken into account as well for final interpretation of any such testing. The complexity, heterogeneity, overlap and interdependence of major psychiatric disorders as currently defined by DSM suggests that the development of tests for dimensional disease manifestations (psychosis, mood and anxiety)[84] will ultimately be more useful and precise than developing tests for existing DSM diagnostic categories.

Finally, while we cannot exclude that rare genetic variants with major effects may exist in some individuals and families, we suggest a contextual cumulative combinatorics of common variants genetic model best explains our findings, and accounts for the thin genetic load margin between clinically ill subjects and normal controls, which leaves a major role to be played by gene

expression (including epigenetic changes) and the environment. This is similar to our conclusions when studying bipolar disorder,[29] and may hold true in general for complex medical disorders, psychiatric and non-psychiatric. Full-blown illness occurs when genetic and environmental factors converge, usually in young adulthood for schizophrenia. When they diverge, a stressful/hostile environment may lead to mild or transient illness even in normal genetic load individuals, whereas a favorable environment may lead to supra-normative functioning in certain life areas (such as creative endeavors) for individuals who carry a higher genetic load. The flexible interplay between genetic load, environment and phenotype may permit evolution to engender diversity, select and conserve alleles, and ultimately shape populations. Our emerging mechanistic understanding of psychosis as disconnectivity, mood as activity[29] and anxiety as reactivity[64] may guide such testing and understanding of population distribution as being on a multi-dimensional spectrum, from supra-normative to normal to clinical illness.

Chapter 2: Alcoholism

Alcohol is one of the most widely used recreational drugs in the U.S., with 82.1 percent of people over the age of 12 having drank alcohol at some point in their lifetime. There has been mixed data over potential health benefits of drinking alcohol. Light to moderate consumption has been associated in epidemiological investigations with reduced relative mortality risk[96], while heavy drinking is associated with increased risk, forming a distinctive and often replicated U or J-shaped curve when relative risk is plotted against average alcohol intake. [97] [98] While drinking itself can be socially normative and part of a healthy lifestyle, many people struggle with alcohol use disorders (AUD), and this can have a profoundly negative impact on both quality[99] and quantity of life[100].

This study sought to identify genes involved in alcohol use disorders in a similar fashion to prior work on schizophrenia. We began with an independent GWAS provided by collaborators in Germany. We again identified SNPs which were nominally significant with a $p < 0.05$, which were then converted into nearest gene. In this analysis we implemented an internal score based on purely on the data from the discovery cohort. This was calculated using the ratio of nominally significant SNPs associated with a gene to total SNPs tested for a gene. Genes in the top 0.1% of the distribution scored 4 points, genes in the top 5% received 3 points, and the remaining genes received 2 points. This provides additional weight to the primary analysis, and this internal score was added to the external score as described below.

Converted genes were then prioritized with a polyevidence CFG score. Keywords used to determine convergence for this project were alcohol and alcoholism, in addition to tissue or species relevant search terms. We used the Jackson Laboratory Mouse Phenome Database to find relevant transgenic animal models by searching the relevant phenotype categories as described below. This analysis used the variation as described previously by weighting the lines of evidence such that human evidence received twice as much as nonhuman evidence, and brain evidence received twice as many points as evidence from genetics or peripheral tissue. In this way, human brain evidence was given 4 points, human peripheral or genetic evidence would be given 2 points each, nonhuman brain evidence was given 2 points, and nonhuman peripheral, genetic, or transgenic evidence received 1 point each. In this weighting the maximum possible score from CFG was 12 (4 human brain, 2 human genetic, 2 human peripheral, 2 nonhuman brain, 1 nonhuman genetic, 1 nonhuman peripheral = 12). This was added to the internal score derived from the primary GWAS so that the maximum possible score overall could be 16 (4 from internal score, 12 from CFG).

Using the exact same methodology as the schizophrenia paper, we set the prioritization cutoff as being half the possible max score, in this case ≥8. 135 genes and 713 SNPs met this criterion, and were used to generate a polygenic GRPS. This showed a trend towards significance ($p=0.053$) in separating cases from controls in an additional independent German GWAS test cohort.

Prior work by the Niculescu lab had identified the D-Box binding protein (DBP) as a potential candidate gene for bipolar disorder[19]. D-box elements are important factors in circadian regulation. DBP is a transcription factor that binds to these elements, and plays a role in activating downstream circadian output genes. In previous studies of transgenic mice with a homozygous deletion of the DBP gene found that animals increased ethanol consumption in response to stress when compared to wild type animals. [58] Koob and many others have implicated stress systems in driving drug seeking behavior. [101] We sought to validate our top GWAS genes by finding the overlap with top candidate genes from the stress-reactive DBP knockout (KO) mouse model. Validation in the animal model produced 11 genes with 66 SNPs.

We used this smaller panel of genes and SNPs to create polygenic GRPS scores which were tested in the independent German GWAS as well as 2 additional independent GWAS from the United States. This smaller panel validated by the animal model showed nominally significant separation between cases and controls in the German GWAS (p=0.041) as well as in United States cohorts of alcohol dependence (p=1.3E-5) and alcohol abusers (p=1.2E-4).

We used a combination of three pathway enrichment tools (Ingenuity, KEGG, and GeneGO) to help further our understanding of the biology of our top genes. Pathway analysis explicitly identified addiction (cocaine) as well as implicating signaling pathways and neurogenesis.

Alcohol can be part of a healthy lifestyle. But for the many people who are susceptible to AUDs, alcohol can have a highly deleterious effect on health and mortality. It is important to recognize that while this study attempts to identify genetic predisposition, context and environmental stress plays a large role as evidenced by the enrichment by the stress reactive animal model. Reinforcing prior findings in schizophrenia, this study points to how we can begin to use genetic predisposition to inform the choices of an individual.

Genetic Risk Prediction and Neurobiological Understanding of Alcoholism

Introduction

Alcohol use and overuse (alcoholism) have deep historical and cultural roots, as well as important medical and societal consequences[102]. While there is evidence for roles for both genes and environment in alcoholism, a comprehensive biological understanding of the disorder has been elusive so far, despite extensive work in the field. Most notably, there has been until recently insufficient translational integration across functional and genetic studies, and across human and animal model studies, resulting in missed opportunities for a comprehensive understanding.

As part of a translational Convergent Functional Genomics (CFG) approach, developed by us over the last 15 years [30], and expanding upon our earlier work on identifying genes for alcoholism [22] [58] [65] , we set out to comprehensively identify candidate genes, pathways and mechanisms for alcoholism, integrating the available evidence in the field to date. We have used data from a published German genome-wide association study for alcoholism [103]. We integrated those data in a Bayesian-like fashion with other human genetic data (association or linkage) for alcoholism, as well as human gene expression data - postmortem brain gene expression data, and peripheral (blood, cell culture) gene expression data. We also used relevant animal model genetic data (transgenic, QTL), as well as animal model gene expression data (brain and blood) generated by our group and others

(Figures 2-1 and 2-2). Human data provides specificity for the illness, and animal model data provides sensitivity of detection. Together, they helped identify and prioritize candidate genes for the illness using a polyevidence CFG score, resulting in essence in a de facto field-wide integration putting together all the available evidence to date. Once that is done, biological pathway analyses can be conducted and mechanistic models can be constructed.

An obvious next step is developing a way of applying that knowledge to genetic testing of individuals to determine risk for the disorder. Based on our comprehensive identification of top candidate genes described in this paper, we have chosen all the nominally significant p-value SNPs corresponding to each of those 135 genes from the GWAS dataset used for discovery (top candidate genes prioritized by CFG with score of 8 and above (>=50% max. possible CFG score of 16), and assembled a Genetic Risk Prediction (GRP) panel out of those 713 SNPs. We then developed a Genetic Risk Prediction Score (GRPS) for alcoholism based on the presence or absence of the alleles of the SNPs associated with the illness from the discovery GWAS, and tested the GRPS in an independent German cohort [104], to see if it can differentiate alcohol dependent subjects from controls, observing a trend towards significance.

In order to validate and prioritize genes in this panel using a behavioral prism, we then looked at the overlap between our panel of 135 top candidate genes and genes changed in expression in a stress-reactive animal model for

alcoholism developed by our group, the DBP knock-out mouse[58,65]. We used this overlap to reduce our panel to 11 genes (66 SNPs).

This small panel of 11 genes was subsequently tested and shown to be able to differentiate between alcoholics and controls in the three independent test cohorts, one German [104] and two US based[105], suggesting that the animal model served in essence as a filter to identify from the larger list of CFG-prioritized genes the key behaviorally relevant genes. Our results indicate that panels of SNPs in top genes identified and prioritized by CFG analysis and by a behaviorally-relevant animal model can differentiate between alcoholics and controls at a population level (Figure 2-5), although at an individual level the margin may be small (Figure S2). The latter point suggests that, like for bipolar disorder [29] and schizophrenia [31], the contextual cumulative combinatorics of common gene variants and environment [106] plays a major role in risk for illness.

Lastly, we have looked at overlap with candidate genes for other major psychiatric disorders domains (bipolar disorders, anxiety disorders, schizophrenias) from our previous studies, and provide evidence for shared genes (Figures 2-3 and 2-4) as well as shared genetic risk (Figure 2-6).

Overall, this work sheds light on the genetic architecture and pathophysiology of alcoholism, provides mechanistic targets for therapeutic intervention, and has implications for genetic testing to assess risk for illness before the illness manifests itself clinically, opening the door for enhanced prevention strategies at a young age. As alcoholism is a disease that does not exist

if the exogenous agent (alcohol) is not consumed, the use of genetic information to inform lifestyle choices could be quite powerful.

Materials and Methods

Human subject cohorts

Discovery cohort (cohort 1): Data for the discovery CFG work (Cohort 1) were obtained from a GWAS of self-reported German descent subjects, consisting of 411 alcohol dependent male subjects and 1307 population-based controls (663 male and 644 female subjects). [103] Individuals were genotyped using HumanHap 550 BeadChips (Illumina Inc, San Diego, CA, USA). SNPs with a nominal allelic P-value <0.05 were selected for analysis. No Bonferroni correction was performed.

Test cohort 2 (alcohol dependence, Germany): An independent test cohort of German descent [104] consisting of 740 alcohol-dependent male subjects and 861 controls (276 male and 585 female subjects) was used for testing the results of the discovery analyses. Individuals were genotyped using Illumina Human 610 Quad or Illumina Human660w Quad BeadChips (Illumina Inc). The controls were genotyped using Illumina HumanHap550 Bead Chips.

Test cohort 3 (alcohol dependence, United States) and test cohort 4 (alcohol abuse, United States): The sample consisted of small nuclear families originally collected for linkage studies, and unrelated individuals, Caucasians and African-American, male and female subjects. The subjects were recruited at five US clinical sites: Yale University School of Medicine (APT Foundation; New Haven, CT, USA), the University of Connecticut Health Center (Farmington, CT, USA), the University

of Pennsylvania Perelman School of Medicine (Philadelphia, PA, USA), the Medical University of South Carolina (Charleston, SC, USA) and McLean Hospital (Belmont, MA, USA). All subjects were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism to derive diagnoses for lifetime alcohol dependence, alcohol abuse and other major psychiatric traits according to the DSM-IV criteria. There were 1687 male subjects with alcohol dependence, 366 male subjects with alcohol abuse and 475 male controls. There were 1081 female subjects with alcohol dependence, 234 female subjects with alcohol abuse and 786 female controls (Table 2-1). Individuals were genotyped on the Illumina HumanOmni1-Quad v1.0 microarray (988,306 autosomal SNPs). GWAS genotyping was conducted at the Yale Center for Genome Analysis and the Center for Inherited Disease Research. Genotypes were called using the GenomeStudio software V2011.1 and genotyping module version 1.8.4 (Illumina Inc). [105]

Table 2-1. Discovery and Test Cohorts

Discovery Cohort 1 GWAS Germany	**Alcohol Dependence**	**Control**
Male	411	663
Female	0	644
Ethnicity	All Caucasians	All Caucasians

Test Cohort 2 Germany	**Alcohol Dependence**	**Control**
Male	740	276
Female	0	585
Ethnicity	All Caucasians	All Caucasians

Test Cohorts 3 and 4	**Alcohol Dependence**	**Alcohol Abuse**	**Control**

United States			
Male	1687	366	475
Female	1081	234	786
Male Ethnicity (Caucasian/African-American)	802/885	201/165	168/307
Female Ethnicity (Caucasian/African-American)	471/610	123/111	220/566

Gene identification in discovery cohort 1

Quality control: Genotype data had been filtered using stringent quality-control criteria as described earlier[104] and accounted for call rate, population substructure, cryptic relatedness, minor allele frequency and batch effects.

Association test in discovery sample: Association testing was performed using PLINK 1.07 (http://pngu.mgh.harvard.edu/ ~ purcell) [107] software package. A logistic regression modelling approach was applied to correct for population stratification. Therefore, principal component analysis was conducted considering only independent autosomal SNPs with minor allele frequency >0.05 and pairwise R^2 <0.05 within a 200-SNP window. LD filtering resulted in a set of 28,505 SNPs used for principal component analysis, which was carried out using GCTA 1.04 (http://www.complex-traitgenomics.com/software/gcta/). [108] The first two principal components resulting from this analysis were included as covariates in the logistic regression model.

Assignment of SNPs to genes: Genes corresponding to SNPs were identified initially using the annotation file from the Illumina website (http://www.illumina.com, HumanHAP550v3_Gene_Annotation). Next, genes were cross-checked with GeneCards (http://www.genecards.org) to ensure that each gene symbol was current. Any gene symbol that matched to a different gene symbol in Gene Cards was checked to verify chromosome number and location match with the original gene, and was replaced with the current GeneCards gene symbol. SNPs from the original annotation files that had no gene matches in the

annotation file and UCSC Genome Browser (that is, not falling within an exon or intron of a known gene) were assumed to regulate and thus implicate the gene closest to the SNP location, using the refSNP database from NCBI (http://www.ncbi.nlm.nih.gov/snp/?SITE= NcbiHome&submit = Go).

Convergent functional genomic analyses

Databases and scoring were as previously described.

Prioritizing top alcoholism candidate genes that overlap with a stress-reactive animal model of alcoholism

Stress has been proposed as a driver of alcoholism, notably by Koob and colleagues[109] [110], as well as by Heilig and colleagues[111]. We have previously identified the circadian clock gene D-box Binding Protein *(DBP)* as a candidate gene for bipolar disorder [19], as well as for alcoholism [22], using a Convergent Functional Genomics (CFG) approach. In follow-up work, we established mice with a homozygous deletion of DBP (DBP KO) as a stress-reactive genetic animal model of bipolar disorder and alcoholism[58]. We reported that DBP KO mice have lower locomotor activity, blunted responses to stimulants, and gain less weight over time. In response to a stress paradigm that translationally mimics what can happen in humans (chronic stress-isolation housing for 4 weeks, with acute stress on top of that- experimental handling in week 3), the mice exhibit a diametric switch in these phenotypes. DBP KO mice are also activated by sleep deprivation, similar to bipolar patients, and that activation is prevented by treatment with the mood stabilizer drug valproate. Moreover, these mice show increased alcohol intake

following exposure to stress. Microarray studies of brain and blood revealed a pattern of gene expression changes that may explain the observed phenotypes. CFG analysis of the gene expression changes identified a series of candidate genes and blood biomarkers for bipolar disorder, alcoholism and stress reactivity. Subsequent studies by us showed that treatment with the omega-3 fatty acid docosahexaenoic acid (DHA) normalized the gene expression (brain, blood) and behavioral phenotypes of this mouse model, including reducing alcohol consumption[65].

We examined the overlap between the top candidate genes for alcoholism from the current analysis and the top candidate genes from the DBP KO stress mice, thus reducing the list from 135 to 11 (Figure 2-3).

Pathway Analyses

IPA 9.0 (Ingenuity Systems, www.ingenuity.com , Redwood City, CA) was used to analyze the biological roles, including top canonical pathways and diseases, of the candidate genes resulting from our work (Tables 2-3 and S2-2), as well as used to identify genes in our datasets that are the targets of existing drugs (Table S2-3). Pathways were identified from the IPA library of canonical pathways that were most significantly associated with genes in our data set. The significance of the association between the data set and the canonical pathway was measured in 2 ways: 1) a ratio of the number of molecules from the data set that map to the pathway divided by the total number of molecules that map to the canonical pathway is displayed; 2) Fisher's exact test was used to calculate a p-value

determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone. We also conducted a KEGG pathway analysis through the Partek Genomic Suites 6.6 software package, Partek Inc., Saint Louis, MO), and GeneGo MetaCore from Thomson Reuters, New York, NY) pathway analyses (https://portal.genego.com/).

Epistasis testing

The test cohort 2 data were used to test for epistatic interactions among the best p-value SNPs in the 11 top candidate genes from our work. SNP-SNP allelic epistasis was tested for each distinct pair of SNPs between genes, using the PLINK software package (Table S2-5).

Genetic Risk Prediction

The software package PLINK 1.07 (http://pngu.mgh.harvard.edu/~purcell) [107] was used to extract individual genotype information for each subject from the test cohorts 2, 3 and 4 data files. As we had previously done for bipolar disorder and schizophrenia, we developed a polygenic Genetic Risk Prediction Score (GRPS) for alcoholism based on the presence or absence of the alleles of the SNPs associated with illness in the discovery GWAS cohort 1, and tested the GRPS in three independent cohorts, from different geographic areas, ethnicities, and different types of alcoholism. We tested two panels: a larger panel containing all the nominally significant SNPs in top CFG scoring candidate genes (n=135) from the discovery GWAS1 in the top CFG prioritized genes (Tables S2-1 and S2-4) and

a smaller one (n=11) containing genes out of the larger panel that were cross-validated using an animal model of alcoholism.

Of note, our genes, SNP panels, and choice of affected alleles were based solely on analysis of the discovery GWAS1, which is our discovery cohort, completely independently from the test cohorts. Each SNP has two alleles (represented by base letters at that position). One of them is associated with the illness (affected), the other not (non-affected), based on the odds ratios from the discovery GWAS1. We assigned the affected allele a score of 1 and the non-affected allele a score of 0. A two-dimensional matrix of subjects by GRP panel alleles is generated, with the cells populated by 0 or 1. A SNP in a particular individual subject can have any permutation of 1 and 0 (1 and 1, 0 and 1, 0 and 0). By adding these numbers, the minimum score for a SNP in an individual subject is 0, and the maximum score is 2. By adding the scores for all the alleles in the panel, averaging that, and multiplying by 100, we generated for each subject an average score corresponding to a genetic loading for disease, which we call Genetic Risk Predictive Score (GRPS)[29,31]. To test for significance, a one-tailed t-test with unequal variance was performed between the alcoholic subjects and the control subjects, looking at differences in GRPS.

ROC Curves

ROC curves were plotted using IBM SPSS Statistics 21. Diagnosis was converted to a binary call of 0 (control) or 1(alcohol dependent or abuser) and

entered as the state variable, with calculated GRPS entered as the test variable (Figure S2-2).

Figures

Each figure in this chapter was completed by Daniel Levey and Helen Le-Niculescu. This work has been published. [112]

Results

Top candidate genes

To minimize false negatives, we initially cast a wide net, using as a filter a minimal requirement for a gene to have both some GWAS evidence and some additional independent evidence. Thus, out of the 6085 genes with at least a SNP at p<0.05 in the discovery GWAS cohort 1, we generated a list of 3142 genes that also had some additional evidence (human or animal model data), implicating them in alcoholism (CFG score >=2.5 (>=2 internal) + (>=0.5 external)). This suggests, using these minimal thresholds and requirements, that the repertoire of genes potentially involved directly or indirectly in alcohol consumption and alcoholism may be quite large, similar to what we have previously seen for bipolar disorder[113] and schizophrenia[31]. To minimize false positives, we used an internal score based on percent of SNPs in a gene that were nominally significant, with 4 points for those in the top 0.1% of the distribution (n= 77), 3 points for those in the top 5% of the distribution (n=561) and 2 points for the rest of the nominally significant SNPs (n=5447). We then used the CFG analysis and scoring integrating multiple lines of evidence to prioritize this list of genes (Figure 2-1),

and focused our subsequent analyses on only the top CFG scoring candidate genes. Overall, 135 genes had a CFG score of 8 and above (>=50% of maximum possible score of 16).

Table 2-2. Top candidate genes for alcoholism. Top genes with a CFG score of 8 and above that overlapped with top genes from the stress-reactive animal model are shown (n=11) (Figure 2-3). Best p-value SNP within the gene or flanking regions is depicted. A more complete list of genes with CFG score of 8 and above (n= 135) is available in the Supplementary Information section (Table S1). I - increased; D – decreased in expression; PFC - prefrontal cortex; AMY - amygdala; CP - caudate putamen; NAC - nucleus accumbens; VT - ventral tegmentum; TG- transgenic. P1- paradigm 1, P2- Paradigm 2, P3-paradigm 3 in the Rodd, Bertsch et al. 2007. Association- association evidence; Linkage- linkage evidence. Underlined gene symbol represents means gene is a blood biomarker candidate. **Bold p-values <0.001.**

Gene Symbol /Name	Discovery GWAS1 Best p-value SNP	Internal score Nominally Significant SNPs /Total SNPS tested (% significant)	Human Genetic Evidence	Human Postmortem Brain Expression Evidence	Human Peripheral Expression Evidence	Animal Model Genetic Evidence	Animal Model Brain Expression Evidence	Animal Model Peripheral (Blood) Expression Evidence	CFG Score	DBP Stresses[58]	DBP Stress DHA[65]

SNCA synuclein, alpha (non A4 component of amyloid precursor)	0.02363 rs17015982	2 4/68 (5.89%)	(Association) [114] [115]	(D) PFC [115, 116, 117] (I) FC [118]	(I) Blood [119] (I) Blood [120] (I) Serum [121] DNA hypermethylation[122]	QTL [123]	(D) FC, CP [124] (I) HIP P1 AMY P3 [22]	(I) Blood Male adult cynomolgus monkeys [125]	13	(D) AMY, Blood	(I) Blood
GFAP glial fibrillary acidic protein	0.01052 rs744281	3 4/12 (33.34%)	(Linkage) [126]	(D) FC [118] [116] [127]			(I) FC [128] HIP, NAC,		9.5	(I) AMY	

							PFC - P1[22]				
DRD2 dopamine receptor D2	0.03652 rs4938019	2 2/46 (4.35%)	(Association) [129, 130, 131, 132, 133]	(D) FC, CP [134] (D) [135]		(Transgenic) alcohol aversion decreased alcohol consumption [136]			9	(D) PFC	(I) AMY
GRM3 Glutamate Receptor, Metabotropic 3	0.001126 rs41440	2 15/133 (11.28%)	(Linkage) [137] [138]	(D) HIP [139]			(I) NAC [140] (D) FC [141] (D) CP P2 [22]		9	(I) AMY	

MBP myelin basic protein	0.006503 rs1124941	2 16/109 (14.68%)		(D) FC [118]			(I) PFC [142] (D) Cerebellum [143] Delayed expression [144]	(D) CG-4 glial cells Rat Brain [145]	8.5	(D) PFC AMY	(D) Blood (I) HIP
MOBP myelin-associated oligodendrocyte basic protein	0.01231 rs562545	2 2/43 (4.66%)		(D) FC [118]		QTL [146]	(I) PFC [142] (D) NAC [147] (I)Whole		8.5	(D) PFC (I) AMY	(D) Blood (I) HIP

							Brain [148] (I) PFC P1[22]				
GNAI1 Guanine Nucleotide Binding Protein (G Protein), Alpha Inhibiting Activity Polypeptide 1	**0.00059** rs12706724	2 11/66 (16.67%)	(Linkage) [149] [150]	(D) HIP [139]			(D) NAC P3 [22]		8	(D) PFC	
MOG myelin oligodendrocyte	0.01429 rs3117292	2 3/19 (15.79%)		(D) FC, HIP [118] [139]			(D) VTA [151] (I)		8	(D) PFC	(I) HIP

glycoprotein							PFC [142] (I) PFC P1 [22]				
RXRG retinoid X receptor, gamma	0.005815 rs10800098	2 1/37 (2.71%)	(Linkage) [152]	(D) FC [118]			(I) CP P3 [22]		8	(D) PFC	
SYT1 synaptotagmin I	0.04139 rs1245810	2 7/117 (5.99%)		(D) NAC [153]			(D) FC [141] (D) HIP [154];	(I) Cultured neurons [155] (I) Cortical neurons	8	(D) PFC	(D) AMY

								[156]			
TIMP2 TIMP metallopeptidase inhibitor 2	0.04309 rs7502935	2 1/25 (4%)	(Linkage) [152]	(D) FC, HIP, NAC [153] [157] [139]			(I) VTA [151]		8	(D) PFC	(D) Blood

Convergent Functional Genomics

Multiple Independent Lines of Evidence
For Cross-Validation of GWAS Data

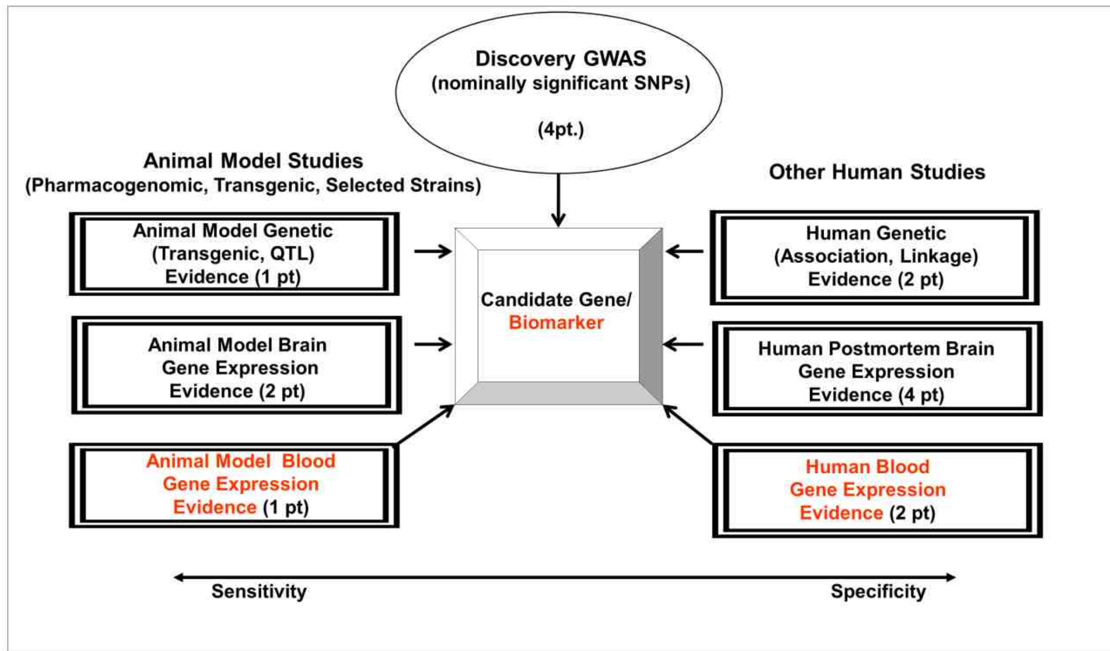


Figure 2-1. Convergent Functional Genomics. Schematic for Alcoholism study.

Of note, there was no correlation between CFG prioritization and gene size, thus excluding a gene-size effect for the observed enrichment (Supplementary Figure S2-1).

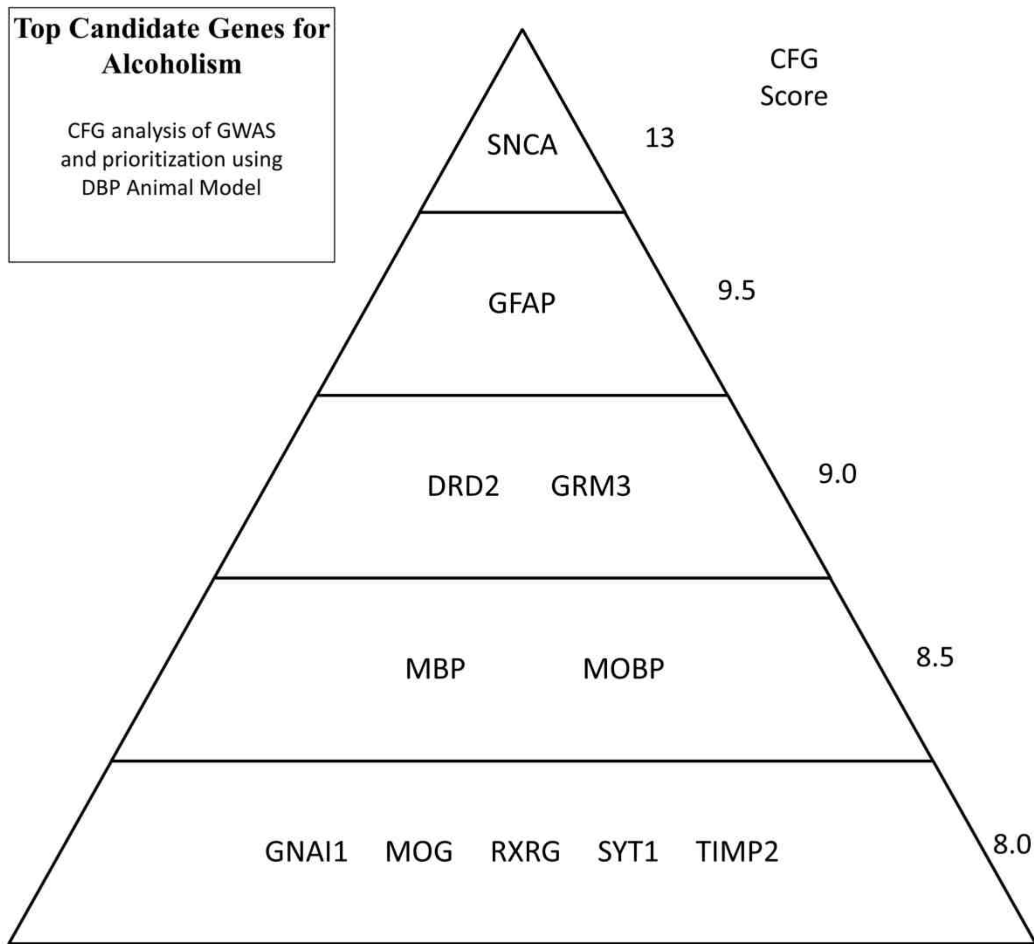


Figure 2-2. Top candidate genes for alcoholism.

Biological pathways and drug targets

Pathway analyses were carried out on the top candidate genes (Table 2-3). Notably, Gαi signaling, cocaine addiction, and transmission of nerve impulses, were the top biological pathways in alcoholism, which may be informative for treatments and drug discovery efforts by pharmaceutical companies. Of note, these top candidate genes were identified and prioritized only for evidence for alcoholism prior to pathway analyses, so the overlap with cocaine addiction is a completely independent result, suggesting a shared drive and neurobiology. Consistent with that, two of our 135 top candidate genes for alcoholism (CPE and VWF) had SNPs with p<10-5 in a recent GWAS of cocaine addiction[158].

Some of the top alcohol candidate genes have prior evidence of being modulated by the omega-3 fatty acid DHA in our DBP mouse animal model (Tables 2-2 and S2-1). That is of particular interest, as we have previously shown that treatment with the omega-3 fatty acid docosahexaenoic acid (DHA) decreased alcohol consumption in that animal model, as well as in another independent animal model, the alcohol preferring P rats[65]. Omega-3 fatty acids, particularly DHA, have been described to have alcoholism, mood, psychosis, and suicide modulating properties, in preclinical models as well as some human clinical trials and epidemiological studies. For example, deficits in omega-3 fatty acids have been linked to increased depression and aggression in animal models [159,160] and humans [161,162]. DHA prevents ethanol damage in vitro in rat hippocampal slices[163]. Omega-3 supplementation can prevent oxidative damage caused by prenatal

alcohol exposure in rats[164]. Of note, deficits in DHA have been reported in erythrocytes[165] and in the postmortem orbitofrontal cortex of patients with bipolar disorder, and were

Table 2-3. Pathway Analyses. Pathway Analyses of top candidate genes. A. Biological Pathways. B. Disease and Disorders.

A.	**Ingenuity Pathways**				**KEGG Pathways**			**GeneGO Pathways**		
GRPS-11 Top DBP KO genes out of CFG score >=8.0 genes (N=11 genes)	**#**	**Top Canonical Pathways**	**P-Value**	**Ratio**	**Pathway Name**	**Enrichment Score**	**Enrichment p-value**	**Networks**	**pValue**	**Ratio**
	1	**GαI Signaling**	4.68E-05	3/135 (0.022)	**Cocaine addiction**	11.9589	6.40226E-06	**Neurophysiological process_Transmission of nerve impulse**	6.010E-06	6/212
	2	cAMP-mediated signaling	2.64E-04	3/226 (0.013)	Gap junction	5.71209	0.00330577	Development_Neurogenesis in general	9.163E-04	4/192
	3	G-Protein	4.37E-04	3/276	Glutamatergic synapse	5.21128	0.00545466	Reproduction_GnRH signaling pathway	6.603E-03	3/166

		Coupled Receptor Signaling		(0.011)						
	4	14-3-3-mediated Signaling	2.14E-03	2/121 (0.017)	Dopaminergic synapse	5.0126	0.00665355	Transport_Synaptic vesicle exocytosis	7.762E-03	3/176
	5	Synaptic Long Term Depression	3.33E-03	2/161 (0.012)	Neuroactive ligand-receptor interaction	3.62169	0.0267375	Development_Neurogenesis_Synaptogenesis	8.258E-03	3/180

B.	**Ingenuity**	**GeneGO**

GRPS-11	**#**	**Diseases and Disorders**	**P-Value**	**Molecules**	**Diseases**	**P-Value**	**Ratio**
Top DBP KO genes out of CFG score >=8.0 genes (N=11 genes)	1	Hereditary Disorder	1.66E-08 – 1.29E-02	9	Schizophrenia	1.476E-10	11/1033
	2	Neurological Disease	1.66E-08 – 1.64E-02	10	Suicide	1.076E-09	6/151
	3	Psychological Disorders	1.66E-08 – 1.53-02	9	Depressive Disorder, Major	2.106E-09	10/966
	4	Cancer	1.90E-08 – 1.82E-02	10	Bipolar Disorder	2.983E-09	9/706
	5	Skeletal and Muscular Disorders	1.50E-07 – 1.34E-02	6	Alcoholism	3.468E-09	7/291

greater in those that had high vs. those that had low alcohol abuse [166]. Low DHA levels may be a risk factor for suicide [167,168]. Omega-3 fatty acids have been reported to be clinically useful in the treatment of both mood[169] [170] [171,172] and psychotic disorders [173] [174] [175].

Other existing pharmacological drugs that modulate alcohol candidate genes identified by us include, besides benzodiazepines, dopaminergic agents, glutamatergic agents, serotonergic agents, as well as statins (Table S2-3).

Genetic risk prediction score (GRPS)

Once the genes involved in a disorder are identified, and prioritized for likelihood of involvement, then an obvious next step is developing a way of applying that knowledge to genetic testing of individuals to determine risk for the disorder. Based on our identification of top candidate genes described above using CFG, we pursued a polygenic panel approach, with digitized binary scoring for presence or absence, similar to the one we have devised and used in the past for biomarkers testing [29],[176] and for genetic testing in bipolar disorder [29] and schizophrenia [31] . Somewhat similar approaches but without CFG prioritization, attempted by other groups, have been either unsuccessful[177] or have required very large panels of markers[178].

We chose all the nominally significant p-value SNPs ($p<0.05$) in each of our top CFG prioritized genes (n=135 with CFG score>=8) (Table S2-1) in the GWAS1 data set used for discovery, and assembled a GRPS-135 panel out of those SNPs (Table 2-4). We then tested the GRPS-135 in the independent German test cohort

2, based on the presence or absence of the alleles of the SNPs associated with the illness, comparing the alcoholic subjects to controls (Table 2-4), and showed that, while there was a trend, we were not able to distinguish alcoholics from controls in both independent test cohorts.

We then prioritized a smaller panel of 11 genes (Table 2-2) out of this larger panel, by using as a cross-validator the top genes from a stress-reactive mouse animal model for alcoholism, the DBP knock-out mouse[58] (Figure 2-3). The small panel (GRPS-11) showed more robust results than the larger panel (Table 2-4), suggesting that it captures key behaviorally-relevant genes.

Table 2-4. Genetic Risk Prediction Score (GRPS)- Panels from Discovery Cohort 1. Differentiation between alcoholics and controls in three independent test cohorts using : GRPS-135, a panel composed of all the nominally significant SNPs from GWAS1 in the top candidate genes prioritized by Convergent Functional Genomics (CFG); GRPS-11, a panel additionally prioritized by a stress-reactive animal model for alcoholism, the DBP KO stressed mouse; and GRPS-SNCA, the top candidate gene from our analyses. P-values depict one-tailed t-test results between alcoholics and controls.

	Test in Cohort 2 Alcohol Dependent vs. Control
GRPS-135 Genes with CFG score of >=8 all nominally significant SNPs in each gene (n=713)	**P=0.053** (135 genes, 713 SNPs)
GRPS-11 Top animal model (DBP mouse) prioritized genes out of genes with CFG score of >=8 all nominally significant SNPs in each gene (n=66)	**P= 0.041** (11 genes, 66 SNPs)
	Test in Cohort 3 Alcohol Dependent vs. Control
GRPS-11 Top animal model (DBP mouse) prioritized genes out of genes with CFG score of >=8	**P=0.00012** (10 genes, 34 SNPs present)

all nominally significant SNPs in each gene (n=66)	
GRPS-SNCA Top CFG gene all nominally significant SNPs in it (n=4)	**P= 0.000013** (1 gene, 1 SNP *rs17015888* present)
	Test **in Cohort 4** **Alcohol Abuse vs. Control**
GRPS-11 Top animal model (DBP mouse) prioritized genes out of genes with CFG score of >=8 all nominally significant SNPs in each gene (n=66)	**P=0.0094** (10 genes, 34 SNPs present)
GRPS-SNCA Top CFG gene all nominally significant SNPs in it (n=4)	**P= 0.023** (1 genes, 1 SNP *rs17015888* present)

Discussion

Our CFG approach helped prioritize a very rich in signal and biologically interesting set of genes (Tables 2-2 and S2-1). Some, such as SNCA, CPE, DRD2 and GRM3, have weaker evidence based on the GWAS data but strong independent evidence in terms of gene expression studies and other prior human or animal genetic work. Conversely, some of the top previous genetic findings in the field [179], such as ADH1C [180] (CFG score of 9), GABRA2 [181] (CFG score of 8), as well as AUTS2 (CFG score of 7), CHRM2 and KCNJ6 (CFG scores of 4) have fewer different independent lines of evidence, and thus received a lower CFG prioritization score in our analysis (Table S2-1), although they are clearly involved in alcoholism-related processes. While we cannot exclude that more recently discovered genes have had less hypothesis driven work done and thus might score lower on CFG, it is to be noted that the CFG approach integrates predominantly non-hypothesis driven, discovery-type datasets, such as GWAS data, linkage, quantitative traits loci, and particularly, gene expression. We also cap each line of evidence from an experimental approach (Figure 2-1), to minimize any 'popularity' bias, whereas multiple studies of the same kind are conducted on better-established genes. In the end, it is gene-level reproducibility across multiple approaches and platforms that is built into the approach and gets prioritized most by CFG scoring during the discovery process. Our top results subsequently show good reproducibility and predictive ability in independent cohort testing, the litmus test for any such work.

At the very top of our list of candidate genes for alcoholism, with a CFG score of 13, we have SNCA (synuclein alpha), a pre-synaptic chaperone that has been reported to be involved in modulating brain plasticity and neurogenesis, as well as neurotransmission, primarily as a brake [182], [183]. On the pathological side, low levels of SNCA might offer less protection against oxidative stress [184], while high levels of SNCA may play a role in neurodegenerative diseases, including in Parkinson Disease (PD). SNCA has been identified as a susceptibility gene for alcohol cravings [114] and response to alcohol cues [185]. The evidence provided by our data and other previous human genetic association studies suggest a genetic rather than purely environmental (alcohol consumption, stress) basis for its alteration in disease, and its potential utility as trait rather than purely state marker.

Alcoholics carry a genetic variant that leads to reduced baseline expression of *SNCA* (Janeczek et al. 2012). SNCA is also downregulated in expression in the frontal cortex and caudate-putamen of inbred alcohol preferring rats [124], as well as in the brain (amygdala) and blood of our stress-reactive DBP animal model of alcoholism, prior to exposure to any alcohol. SNCA is upregulated in expression in blood in human alcoholism [119,120], as well as in the blood of monkeys consuming alcohol, and in rats after alcohol administration (Rodd, Bertsch et al. 2007). Thus, it may serve as a blood biomarker. Overall, we may infer that, while low levels of SNCA may predispose to cravings for alcohol and consequent alcoholism, possibly mediated through increased neurobiological activity and drive (*the SNCA deficit*

hypothesis), excessive alcohol consumption then increases SNCA expression beyond that seen in non-alcohol consuming controls, potentially compounding risk for neurodegenerative diseases in individuals that have mutations that lead to its aggregation. This observation is also biologically consistent with the fact that dementia is often observed late in the course of alcohol dependence.

GFAP (glial fibrillary acidic protein), a top candidate gene with a CFG score of 9.5, is an astrocyte intermediate filament-type protein involved in neuron-astrocyte interactions, cell adhesion, process formation and cell-cell communication. It is decreased in expression in postmortem brain of alcoholics, but increased in expression in brains of animal models of predisposition to alcoholism, prior to exposure to alcohol (Table 2-2). This is consistent with a model for increased physiological robustness in individuals predisposed to alcoholism [22], as well as with the neurodegenerative consequences of protracted alcohol use.

DRD2 (dopamine receptor D2), another top candidate gene with a CFG score of 9, has prior human genetic association evidence. It is reduced in expression in the frontal cortex in human brain from alcoholics, as well as in the DBP animal model prior to any exposure to alcohol. One possible interpretation would be that lower levels of dopamine receptors are associated with reduced dopaminergic signaling and anhedonia, leading individuals to overcompensate by alcohol and drug abuse. Another interpretation, consistent with the low SNCA and consequently higher neurotransmitter (including dopamine) levels, would be that these individuals are in fact in a compulsive, hyperdopaminergic state, which drives

them to hedonic activities and leads to compensatory homeostatic downregulation of their DRD2 receptors. Consistent with this later scenario, mice that have a constitutive knock-out of their DRD2 receptors, not due to a hyperdopaminergic state, in fact consume less alcohol[136], unless they are exposed to stress[186].

Another top candidate gene, GRM3, is also involved in neurotransmitter signaling. Prior evidence in the field had implicated another metabotropic glutamate receptor, GRM2 [187].

Other top candidate genes in the panel (MOBP, MBP, MOG) are involved in myelination (Table 2-2). They are decreased in expression in the pre-frontal cortex of human alcoholics, as well as in our stress-reactive DBP animal model of alcoholism, prior to exposure to any alcohol. Decreased myelination may lead to decreased connectivity. Interestingly, MOBP and MBP are increased in expression in the amygdala in the DBP mice, opposite to the direction of change in the PFC, consistent with a frontal deactivation and a limbic hyperactivity, which could lead to impulsivity.

Epistasis testing of top candidate genes for alcoholism.

For the top 11 candidate genes, best p-value SNPs from GWAS1 were used to test for gene-gene interactions in GWAS2 (Table S2-5). Nominally significant interactions were found between SNPs in SNCA and RXRG, DRD2 and SYT1, MOBP and TIMP2. As a caveat, the p-value was not corrected for multiple comparisons. The corresponding genes merit future follow-up work to elucidate the biological and pathophysiological relevance of their interactions.

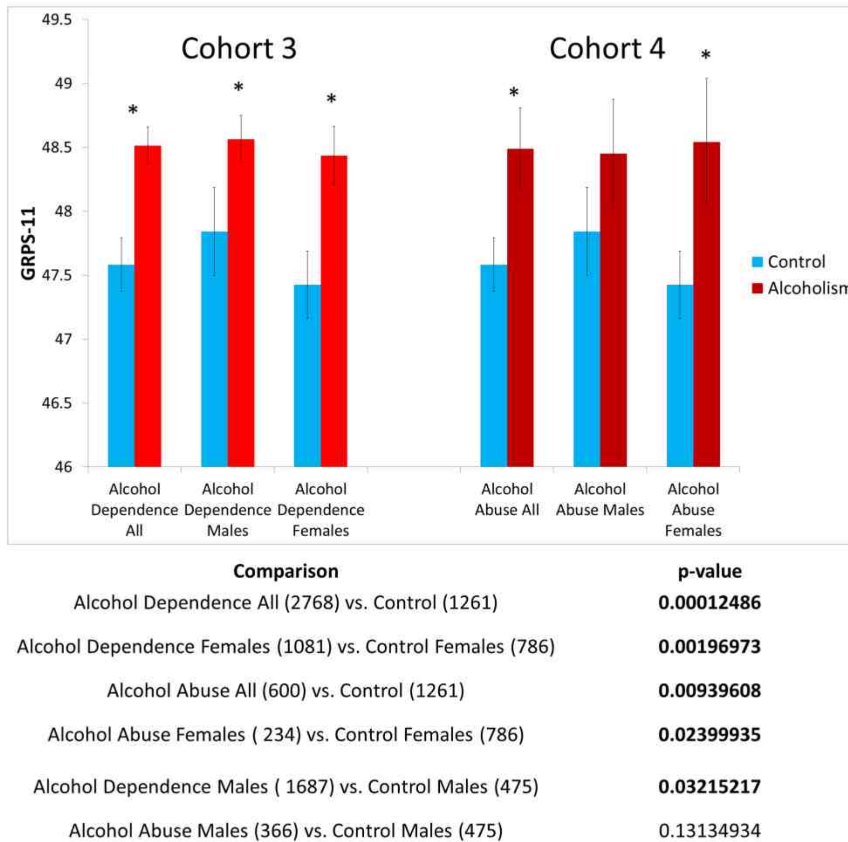
Pathways and mechanisms

Our pathway analyses (Tables 2-3 and S2-2) results are consistent with the accumulating evidence about the role of neuronal excitability and signaling in alcoholism[188,189] [178].

Overlap with other psychiatric disorders

Despite using lines of evidence for our CFG approach that have to do only with alcoholism, the list of genes identified has a notable overlap at a pathway analysis level (Tables 2B-2 and S2B-2.) and at a gene level (Figures 2-4 and 2-5) with other psychiatric disorders. This is a topic of major interest and debate in the field. We demonstrate an overlap between top candidate genes for alcoholism and top candidate genes for schizophrenia, anxiety and bipolar disorder, previously identified by us through CFG (Figure 2-4), thus providing a possible molecular basis for the frequently observed clinical co-morbidity and interdependence between alcoholism and those other major psychiatric disorders, as well as cross-utility of pharmacological agents. Moreover, we tested in alcoholics genetic risk predictive panels for bipolar disorder [29] and for schizophrenia [31] generated in previous studies by us, and show that they are significantly different in alcoholics vs. controls (Figure 2-6), beyond the overlap in genes with alcohol. There seems to be an increased genetic load for bipolar disorder, consistent with increased drive, and a decreased genetic load for schizophrenia, consistent with increased connectivity prior to alcohol use. These results led us to develop a heuristic, testable model of alcoholism (Figure 2-5). Some people may drink to be calm- mitigating the effects

of stress and anxiety, some people may drink to be happy- the common drive with bipolar disorder,



Comparison	p-value
Alcohol Dependence All (2768) vs. Control (1261)	**0.00012486**
Alcohol Dependence Females (1081) vs. Control Females (786)	**0.00196973**
Alcohol Abuse All (600) vs. Control (1261)	**0.00939608**
Alcohol Abuse Females (234) vs. Control Females (786)	**0.02399935**
Alcohol Dependence Males (1687) vs. Control Males (475)	**0.03215217**
Alcohol Abuse Males (366) vs. Control Males (475)	0.13134934

Figure 2-3. Genetic Risk Prediction using a panel of top candidate genes for alcoholism (GRPS-11). Testing in independent cohorts 3 and 4.

and some people may drink to be drunk- to disconnect from reality and/or get unstuck from internal obsessions and ruminations.

Genetic risk prediction

Of note, our SNP panels and choice of affected alleles were based solely on analysis of the discovery GWAS, completely independently from the test cohorts. Our results show that a relatively limited and well-defined panel of SNPs identified based on our CFG analysis could differentiate between alcoholism subjects and controls in three independent cohorts. The fact that our genetic testing worked for

both alcohol dependence and alcohol abuse suggests that these two diagnostic categories are actually overlapping, supporting the DSM-V reclassification of a single category of alcohol use disorders.

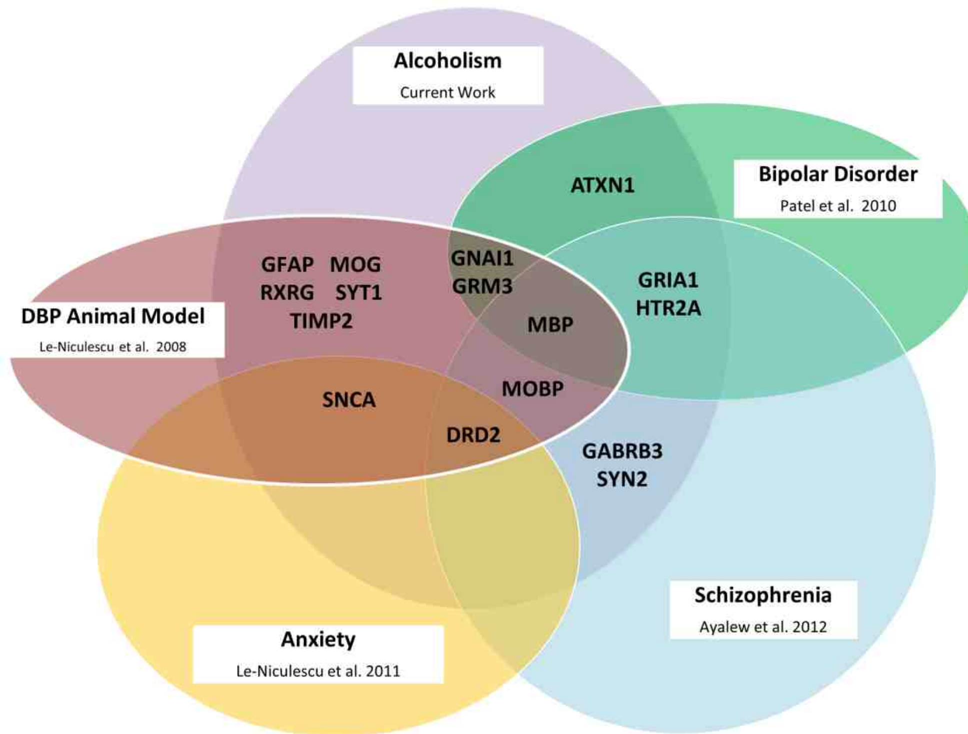
Reproducibility among studies

Our work provides striking evidence for the advantages, reproducibility and consistency of gene-level analyses of data, as opposed to SNP level analyses, pointing to the fundamental issue of genetic heterogeneity at a SNP level. In fact, it may be that the more biologically important a gene is for higher mental functions, the more heterogeneity it has at a SNP level and the more evolutionary divergence, for adaptive reasons. On top of that, CFG provides a way to prioritize genes based on disease relevance, not study-specific effects (that is, fit-to-disease as opposed to fit-to-cohort). Reproducibility of findings across different studies, experimental paradigms and technical platforms is deemed more important (and scored as such by CFG) than the strength of finding in an individual study (for example, P-value in a GWAS).

Potential limitations and confounds

The GWAS study (cohort 1) on which our discovery was based contained males as probands, but contained males and females as controls. This was the case for the German test cohort (cohort 2) as well. It is possible that some of the nominally significant SNPs identified in the discovery GWAS have to do with gender differences rather than to alcoholism per se, or at least may have to do with male alcoholism. Stratification across gender and ethnicities may also be a factor in our

test cohorts 3 and 4 (Table 2-1). The issue of possible ethnicity differences in alleles, genes, and the consequent neurobiology may need to be explored more in the future, with larger



sample sizes, and with environmental and cultural factors taken into account.

Figure 2-4. Overlap of alcoholism versus other major psychiatric disorders. Top candidate genes for alcoholism identified by CFG (n=135) in the current study versus top candidate genes for other psychiatric disorders and a stress-driven animal model of alcoholism (DBP knockout mouse) from our previous work.

ıf
e
ır
n
n

female controls (Figure 2-3). Moreover, a series of individual genes from the panel,

not just SNCA, separate alcoholics from controls in independent cohorts (Table 2-5).

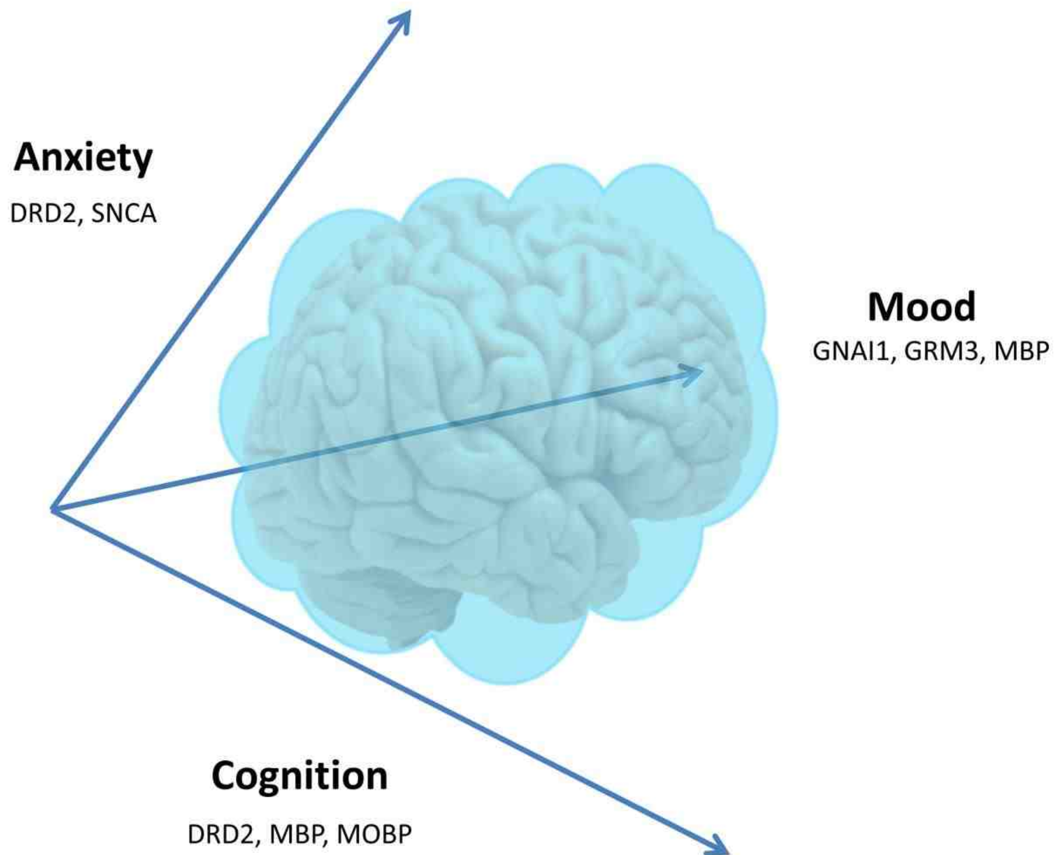


Figure 2-5. Mindscape (mental landscape)-dimensional view of genes that may be involved in alcoholism and other major psychiatric disorders.

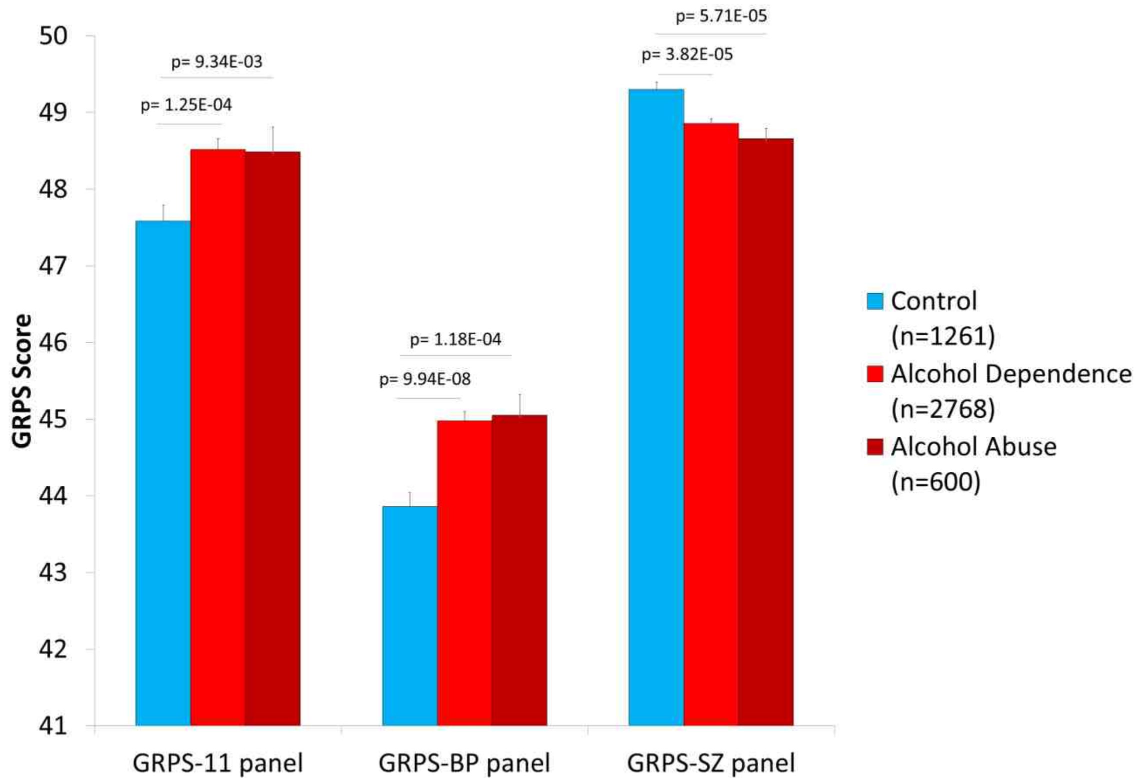


Figure 2-6. Genetic load for bipolar disorder and schizophrenia in alcoholism. A total of 34 out of 66 SNPs in our alcohol GRPS-11 panel (current work; in n =10 genes), 42 out of 224 SNPs in our bipolar GRPS53 (in n =34 genes) and 151 out of 542 SNPs in our schizophrenia GRPS54 (in n =35 genes) were present and tested in the alcohol cohorts 3 and 4. See also Supplementary Table S7.

Table 2-5. Individual top genes and genetic risk prediction in independent cohorts. A. Cohort 2. B. Cohorts 3 and 4. Italic- nominally significant, bold italic- survived Bonferroni correction.

A. **Test Cohort 2**		**Mean GRPS**		**t-test**
Gene/ **SNPs**	**CFG Score**	**Control (n=861)**	**Alcohol Dependence Cohort 2 (n=740)**	
Panel of 11 top genes 66 snps	>=8	53.98	54.61	*0.041*
SNCA rs7668883 rs17015888 rs17015982 rs6532183	13	93.93	92.84	0.086
GFAP rs3744473 rs3169733 rs736866 rs744281	9.5	63.99	64.69	0.303
DRD2	9	13.07	15.51	*0.024*

rs4648317 rs4938019				
GRM3 rs17160519 rs6944937 rs13236080 rs17315854 rs12668989 rs41440 rs2373124 rs13222675 rs2708553 rs12673599 rs4236502 rs1554888 rs10499898 rs1527769 rs17161018)	9	55.44	54.94	0.271
MBP rs470131 rs2282566 rs736421	8.5	44.92	47.07	***0.002***

rs1789094 rs9951586 rs1667952 rs1789105 rs1789103 rs1812680 rs1789139 rs4890912 rs9947485 rs1562771 rs1015820 rs1124941 rs11877526				
MOBP rs562545 rs2233204	8.5	49.88	49.93	0.487
GNAI1 rs4731111 rs6466884 rs7803811 rs17802148 rs7805663	8	72.97	72.71	0.393

rs10486920 rs2523189 rs2886611 rs2886609 rs12706724 rs4731302				
MOG rs3117292 rs2747442 rs3117294	8	34.53	34.56	0.493
RXRG rs10800098	8	6.04	5.27	0.174
SYT1 rs1569033 rs10735416 rs1245810 rs1245819 rs1268463 rs1245840 rs10861755	8	39.16	40.89	0.113
TIMP2 rs7502935	8	67.65	70.61	*0.038*

B. Test cohorts 3 and 4		Mean GRPS			t-test	
Gene/ SNPs	**CFG Score**	**Control (n=1261)**	**Alcohol Dependence Cohort 3 (n=2768)**	**Alcohol Abuse Cohort 4 (n=600)**	**Alcohol Dependence Cohort 3**	**Alcohol Abuse Cohort 4**
Panel of 10 top genes 34 snps	>=8	47.58	48.51	48.49	***0.00012***	*0.0094*
SNCA rs17015888	13	72.28	76.96	75.58	***0.000013***	*0.023*
GFAP rs3169733 rs736866	9.5	58.92	60.38	60.17	*0.042*	0.158
DRD2 rs4648317	9	15.38	14.92	15.61	0.293	0.429
GRM3	9	35.13	37.38	35.55	***0.000061***	0.309

rs17160519 rs6944937 rs17315854 rs4236502						
MBP rs470131 rs2282566 rs736421 rs1789094 rs9951586 rs1789103 rs4890912 rs9947485 rs1124941	8.5	47.23	48.01	48.31	*0.0443*	0.059
MOBP rs562545 rs2233204	8.5	50.28	50.80	50.75	0.233	0.337
GNAI1 rs4731111 rs6466884 rs17802148 rs10486920	8	61.58	63.03	62.87	*0.006435*	0.072

rs2523189 rs2886611 rs2886609						
MOG rs3117292 rs2747442 rs3117294	8	48.78	46.44	46.08	0.020216	0.056
RXRG rs10800098	8	2.62	3.42	3.33	*0.024279*	0.126
SYT1 rs1569033 rs1245819 rs1268463 rs10861755	8	41.56	42.48	44.25	0.11474	*0.0087*

The conversion from SNPs to genes as part of our discovery assumed the rule of proximity, i.e. an intragenic SNP implicates the gene inside which it falls, or if it falls into an intergenic region, it implicates the most proximal gene to it. That may not be true in reality in all cases, generating potentially false positives and false negatives. However, the convergent approach and focus on the top CFG scoring genes reduces the likelihood of false positives.

The only SNP for SNCA that was present/tested for in cohorts 3 and 4 (rs17015888) was relatively far away upstream (0.13 MB) from SNCA. However,

no other known genes are present in that region, SNCA is the closest gene, and the distance is well within range of known examples of regulatory regions (enhancers). Additionally, the risk allele for this SNP (G/G) seems to be the major variant in the population (Table S2-6), suggesting that this allele per se is evolutionarily advantageous, when not coupled with the exogenous ingestion of alcohol.

A relatively large list of genes (n= 6085) was implicated by nominally significant SNPs from the discovery GWAS. There is a risk that out of such a large list, CFG will find something to prioritize. We have tried to mitigate that by developing an internal score for each gene based on the proportion of SNPs tested in a gene that were nominally significant. Also, in the end, we tested the reproducibility and predictive ability of our top findings in multiple independent cohorts, which is the ultimate litmus test for any genetic or biomarker study.

Conclusion

Overall, while multiple mechanistic entry points may contribute to alcoholism pathogenesis, it is likely at its core a disease of an exogenous agent (alcohol) modulating different mind domains/dimensions (anxiety, mood, cognition)[84], precipitated by environmental stress on a background of genetic vulnerability (Figure 2-4). The degree to which various mind domains/dimensions are affected in different individuals is a fertile area for future research into subtypes of alcoholism, and lends itself to personalization of diagnosis and treatment, by integrating genetic data, blood gene expression biomarker data, and

clinical data. Lastly, it is important to note that individuals with a predisposition to alcoholism but no exposure to alcohol may in fact have a robust physiology and strong neurobiological drive that can be harnessed for other, more productive endeavors.

Chapter 3: Focus on Suicide

Historically, suicidal behavior, defined as a self-initiated sequence of behaviors by an individual who, at the time of initiation, expected that the set of actions would lead to his or her own death, has been considered a symptom of various psychiatric disorders such as Major Depressive Disorder (MDD) and Borderline Personality Disorder (BPD), rather than a distinct diagnostic entity. However, in advance of the publication of DSM V in May 2013, many clinicians and researchers advocated for the inclusion of suicidal behavior as an independent diagnosis. [190]

These individuals argued that while suicidal behavior occurs in the context of psychiatric conditions, this is not invariably the case. 10% of those who complete suicide in the United States have no identifiable psychiatric disorder and this number is estimated to be as high as 37% of suicide completers in China. Furthermore, even among disorders associated with a high risk for suicide the majority of those affected do not commit suicide. For example, only 29% of individuals with Bipolar Disorder (BD) report a lifetime history of suicidal behavior. Thus, the vast majority of those afflicted with disease do not engage in this behavior over the course of their lifespan. Consequently, it has been postulated that suicidal behavior should be considered a co-morbid condition of MDD, BPD as opposed to diagnostic criteria of these conditions. Finally, to include suicidal behavior as a symptom of MDD and BPD exclusively implies that this is not of central concern in the management of schizophrenia, post-traumatic stress

disorder (PTSD), and substance use disorders (SUD) which do not include suicidal behavior as diagnostic criteria, but nonetheless are associated with increased risk for suicide. [190] Ultimately, "Suicidal Behavior Disorder" was included in the "Conditions for Further Study" section of the DSM V as an independent diagnosis and for the purposes of this review will be considered as such while continuing to acknowledge the highly co-morbid nature of this behavioral phenotype.

Epidemiology

Per the most recent World Health Organization, suicide is the second leading cause of death amongst 18-29 year olds worldwide (tenth across all age groups) and results in 800,000 deaths yearly, the equivalent of one death every 40 seconds[191] . The age adjusted suicide rate in 2014 was 13.0 people per 100,000 in the general population, [192] while the suicide rate may be substantially higher in patients suffering from mental disorders. [193] The WHO estimates that global annual suicide fatalities could rise to 1.5 million by 2020. Suicide attempts are estimated to be approximately 20-25 times more frequent than completed suicides, which conservatively equates to 16 million attempts yearly.

In addition to age and gender disparities in rate of suicide, ethnic differences also exist. The highest U.S. suicide rate was among Whites and the second highest rate was among American Indians and Alaska Natives. Lower and similar rates were found among Hispanics, Asians and Pacific Islanders, and Blacks. [194]

Suicide via firearm, poisoning, and suffocation account for over 90% of suicide deaths in the United States. 49.9% of deaths are due to self-inflicted gunshot wounds, while 26.7% and 15.9% of suicides are a result of suffocation and poisoning respectively. [194]

Clinical Risk Assessment Scales

Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a current gold standard for assessment of suicidal ideation and behavior in clinical trials and was designed based on research into aspects of past suicidal ideation and behavior that predict the future risk of suicidal behavior. [195]

Convergent Functional Information for Suicide (CFI-S)

The CFI-S is new a 22 item scale and Android app containing known risk factors for suicidal behavior which, in a simple binary fashion (yes =1, no =1), integrates information about clinical history, mental and physical health, addictions and cultural factors without explicitly querying suicidal ideation. Positive responses are summed and then divided by the number of responses, a simple average, yielding a risk score between 0 and 1. This measure can be scored from a medical chart or an interview with a patient or family member. The scale has shown predictive validity for suicidal ideation and future hospitalizations due to suicidal behavior. [196] [197] The scale has been used posthumously in interviews of next of kin to evaluate suicide completers and is receiving ongoing validation in emergency room settings.

Hamilton Depression Rating Scale (HDRS) and Clinical Interview for Depression

The HDRS has been a gold standard measure for depression, first developed in the late 50s to assess the effectiveness of early anti-depressants. While depression itself can be considered a risk factor for suicide a plurality of those who experience it will not attempt suicide or experience suicidal ideation. Though lacking in specificity, scales like the HDRS may provide clinically meaningful context to the identification of suicide risk. In addition, where a scale like the HDRS may be particularly useful is in the pragmatic sense that it has widespread usage and contains an item that does specifically assess suicidal thoughts and behaviors.

Treatment

As discussed in the introduction, suicidal behavior has traditionally been considered a result of an underlying psychiatric condition that if treated appropriately will lead to reduced risk of suicide in the future. However, the unique genetic, epidemiologic, molecular, and neurobiological data presented in this review and elsewhere suggest that the pathophysiology of this behavior is distinct from the conditions it presents most commonly with. Thus while treatment of any comorbid psychiatric disturbances should be initiated promptly, it is important to highlight the pharmacologic, biological, and psychotherapeutic treatment modalities that have been demonstrated to be the most effective in addressing the unique pathophysiology of suicidal behavior, independent of their effects on any co-morbid conditions.

Lithium

Lithium was approved by FDA for treatment of manic episodes in bipolar disorder in 1974 and remains widely used for this indication. It is also used off label for treatment of bipolar depression and adjunctive treatment of unipolar depression. The anti-suicidal properties of lithium were first observed in the 1970's, and several studies in 1980's and 90's demonstrated decreased risk for suicidality in those with affective disorders, including independent studies both showing a 6-fold decrease in incidence of suicide attempts.

The first study to provide evidence that lithium's anti-suicidal properties are independent of its mood stabilization effect was published in 2001 by Conell et al. [198] This group investigated 33 patients with affective disorders who went through periods of discontinuation with lithium. They observed that those who had discontinued had poorer response relative to mood stabilization upon resuming lithium but the antisuicidal effects were unchanged by periods of discontinuation. Similarly, Ahrens and Müller-Oerlinghausen et. al [199], found a reduction in suicide attempts not only in the excellent lithium responders but also among patients with a moderate to poor response to lithium. An investigation of 12,662 Medicaid patients and demonstrated that lithium-treated bipolar patients had the lowest number of suicide attempts compared to patients treated with other mood stabilizers. [200] These results suggest that lithium can reduce the incidence of suicidality independent of its mood stabilization effects and that mood stabilization effects are not required for resolution of suicidality further suggesting

a unique neurobiology of suicidal behavior distinct from that of comorbid affective disorders. [201]

Clozapine

Clozapine is an atypical antipsychotic for the treatment of schizophrenia available since the 1960's. In December of 2002, based on data published from the International Suicide Prevention Trial, clozapine became the first medication FDA approved for the treatment of suicidality in schizophrenia. In this study clozapine reduced risk of suicide attempts by 25% when compared to the second generation antipsychotic, olanzapine . Similar to lithium, clozapine's antisuicide effects have been demonstrated to be independent of its effect on any other psychiatric symptoms. In the case of clozapine, risk for suicidal behavior decreased independently of its antipsychotic action . Despite clozapine's effect on suicidal behavior and potent antipsychotic action it remains under prescribed compared to other antipsychotic medications. The risk for agranulocytosis, which requires weekly white blood cell monitoring for the first six months of treatment, as well as myocarditis and seizures likely contributes to its relative under prescribing. Still, the potential benefits of reducing suicide risk may be worth this risk. [202]

Ketamine

Ketamine, known by its street name of "Special K", ketamine is an NMDA-receptor antagonist with potential as a treatment for mood, anxiety, and suicidal behavior has received significant attention in recent years as a result of its potential to rapidly treat mood symptoms and suicidality. Trials have shown it to have

promising anxiolytic and antidepressant properties. [203] In a small 2016 study of 14 subjects, repeated doses of open-label ketamine rapidly and robustly decreased suicidal ideation in pharmacologically treated outpatients with treatment-resistant depression with stable suicidal thoughts; this decrease was maintained for at least 3 months following the final ketamine infusion in 2 patients. Weaknesses to the existing literature include the small sample sizes of the studies and the exclusion of patients with significant suicidality at baseline from several of the studies. The evidence supporting the clinical use of ketamine for suicidality is certainly preliminary, but appears to be a promising therapeutic option worthy of further controlled trials to allow for meaningful clinical recommendations. [204]

Anti-depressants

Association of anti-depressants with suicidal behavior has been controversial. SSRI medications have been shown to increase experiences of suicidal ideation and behavior when compared to placebo. In May 2003, British pharmaceutical company GlaxoSmithKline submitted an analysis demonstrating a statistically significant increase in suicide-related adverse events in pediatric trials of the anti-depressant paroxetine compared to placebo. These findings prompted the FDA to review the risk for emergent suicidal behavior with administration of nine antidepressant medications used in pediatric populations. The studied drugs included fluoxetine, sertraline, paroxetine, fluvoxamine, citalopram, bupropion, venlafaxine, nefazodone, and mirtazapine. [205]

A meta-analyses of 372 randomized clinical trials of these antidepressants involving nearly 100,000 participants, demonstrated that the rate of suicidal thinking or suicidal behavior was 4% among patients assigned to receive an antidepressant, as compared with 2% among those assigned to receive placebo. Further analysis of this data demonstrated that this increased risk was only statistically significant among children and adolescents under the age of 18. Conversely, there was no evidence of increased suicidality in adults 24 years and older and a “clear protective effect” against the emergence of suicidal behavior in adults 65 years and older. [206] Thus while the appropriate treatment of any comorbid affective or anxiety disorders should include consideration of the initiation of SSRI’s, SNRI’s, or bupropion it is important to be aware the potential risk for exacerbation of suicidality in children and adolescents and that only in the geriatric population does antidepressant therapy appear to demonstrate a significant protective effect against future suicidality.

Omega 3 Fatty Acids

Low docosahexaenoic acid (DHA), has been shown to be associated with increased risk of suicide in the military. [207] DHA is the most abundant omega-3 fatty acid found in the brain, making up 10-20% of total fatty acids. The primary source of DHA is in the diet, with the largest quantity found in cold water fish. DHA supplementation promotes neurite outgrowth. [208] Prenatal DHA deficiencies increase learning and memory deficits in rodents, but this effect may be reversible with supplementation. [209] A small study has found that supplementation with

DHA along with additional essential omega-3 fatty acid eicosapentaenoic acid (EPA) was able to reduce surrogate markers of suicidal behavior. [210]

Psychotherapeutics

Cognitive Behavioral therapy developed by Aaron Beck in the 1970's is a therapeutic technique which focuses on the development of personal coping strategies to solve current problems and change unhelpful patterns in cognitions (e.g., thoughts, beliefs, and attitudes), behaviors, and emotional regulation . [211] It was originally designed to treat depression, and is now used for a number of mental health conditions. A Cochrane review examining over 17,000 patients and 55 trials demonstrated CBT's efficacy in reducing the risk for future self-harm in individuals who had previously engaged in self-harm.

This review also examined the effect of Dialectical Behavior Therapy (DBT) on reduction of future self-harm in individuals with Borderline Personality Disorder. DBT works towards helping people increase their emotional and cognitive regulation by learning about the triggers that lead to reactive states and helping to assess which coping skills to apply in the sequence of events, thoughts, feelings, and behaviors to help avoid undesired reactions. [212] The authors of the Cochrane review concluded that DBT may lead to a reduction in frequency of self-harm in "people with multiple episodes of SH/probable personality disorder". [213]

Biological Therapies

Both Electroconvulsive therapy (ECT) and more recently Transmagnetic Cranial Stimulation (TMS) have been proposed as alternatives to rapidly and safely

reduce risk for suicidal behavior. A randomized clinical trial comparing ECT and TMS (n = 73) found that ECT reduced depressive symptoms and suicidal ideation scores more rapidly and effectively than rTMS. Both objective (Hamilton Rating Scale for Depression) and subjective (Beck Depression Inventory) ratings of suicidal ideation were dramatically reduced after ECT, but only slightly with rTMS. [214] ECT may also be more effective than traditional antidepressant therapy in reducing suicidality. A 2003 study in which suicidal behavior was assessed in the 6 months after the treatment of depression in 519 patients demonstrated 0.8% rate of suicide attempts of the ECT-treated patients compared to 4.2% of those who had received "adequate" and 7% of those offered "inadequate" antidepressant medication treatment [215]

Etiology/genetics

Suicidal behavior encompasses a range of thoughts and acts centered on the act of intentionally ending one's own life. The causes can be very diverse, from long lasting psychological or physical pain to sudden traumatic events. It can be very difficult to detect. A patient may not want to admit to suicidal thoughts or feelings for fear of how others will judge them. They may not want to burden others with their problems. Still others, with a strong desire to commit suicide, may fear being thwarted in their attempt.

It is important to keep in mind that while we are talking about suicidal behaviors as singular outcome they may be the result of very distinctly different underlying and preceding pathology. A patient suffering from severe physical or

psychological pain may come to a very well thought out plan of action in what might be considered a 'cold' premeditated suicide attempt. Conversely, Impulsive rage or sudden grief could trigger could trigger a sudden and violent 'hot' suicide attempt. Command hallucinations could precipitate an attempt. An anecdotal feature found in a suicide note may imply a feeling of burden and guilt over causing hardship to others, which might be the more the result of delusion then a true assessment of the situation. These concepts have overlapping features with psychiatric disorders and may reveal subtle yet essential underlying subtypes of suicidal behavior. Many possible paths converge on a similar outcome, and addressing the path may forestall the outcome.

Paths to Pathology

A challenge in identifying suicide is the limitation to superficial observations that can be made of a patient. Internal and unobservable motivations often drive the observable suicidal behaviors we wish to prevent. Clinicians need to be perceptive of outward signs of risk and judicious with the use of objective tools such as clinical risk assessment scales and the emerging field of biomarkers.

Some of the known demographic factors (described above in epidemiology) that inform suicide risk, such a previous suicide attempts or presence of mental illness such as depression, are particularly informative. Other phenotypes have been shown to be associated more frequently with suicide attempts by those who meant to die. Assessment of feelings of emptiness or loneliness, hopelessness, and disconnection may indicate increased risk of suicide. [216] Visual analog scales

for mood and anxiety may also predict suicidal ideation or attempts. [217] [196] [197] The use of a synthesis of emotional and environmental intermediate phenotypes may give further evidence of suicidal risk.

The Biology of Suicide

A controversial subject, suicide may be a uniquely human characteristic. There is presently no animal model for suicidal behavior, and reports of this behavior in nonhuman animals are anecdotal and inconclusive, often calling on urban myth (lemmings leaping off of cliffs en masse) or behaviors which may not accurately fit the definition of suicide. For example, a honey bee must sacrifice itself to use its barbed stinger in defense of the hive, but the action of the bee is 'motivated' by defending the hive, not the intentional act of ending its life. Another example comes from various forms of parasite which makes the target organism engage in risky behavior. A rat infected with toxoplasma gondii will lose its instinctual aversion to cat urine and instead be attracted to it. This greatly increases the risk of being preyed upon by cats. While the neurobiology of this induced elevation in risky behavior could be relevant and interesting in the study of human suicide, the actual 'intent' comes from the toxoplasma gondii to complete its life cycle in the cat, not a 'desire' of the rat to end its own life.

On the other hand, a soldier throwing them self on a grenade, while an intentional act which may end the life of the soldier, may be more likely motivated by a desire to protect his unit rather than a desire to end his or herself. Thus motive or intent may be the crux of what defines and seperates suicide, the specific

desire with intent to end ones own life. An important challenge for the field of suicide then is finding endophenotypes which may explain neurobiological and neuropsychological underpinnings of suicidality.

There have been genetic studies indicating a heritable component to suicidal behavior, particularly linkage disequilbrium blocks found on chromosomes 2 and 6. [218] It remains to be clarified whether these associations represent direct links to suicidal behavior per se, or if these represent links to psychiatric conditions which indirectly influence or are positively correlated with suicidal behavior. And while linkage peaks are suggestive of targets for future study of association at a population level it doesn't provide clinical useful targets for individual patients.

Biomarkers have been defined by the National Institutes of Health Biomarkers Definitions Working Group as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention." Biomarkers as a means of identifying the pathogenic process underlying suicidal behavior are essential to the goal of understanding and preventing these tragedies. Several groups over the last decade have been working on identifying genetic risk factors and peripheral biomarkers for suicide. These findings begin to shed light on the underlying neuropathology involved in suicide. While these markers are still being evaluated for clinical relevance and validity, in the near future they may provide a quantitative lab testing method for identifying suicide risk.

While the target tissue for much of the putative pathology of suicide (and psychiatry generally) is the brain, this is not a candidate for biopsy for obvious reasons. This is why the studies of peripherally tissue that can be accessed peripherally such as blood or cerebral spinal fluid (CSF) have great potential utility due to relative ease of access. Imaging studies, on the other hand, hold the potential for identifying tissue specific (brain) biomarkers in vivo.

Hypothalamic-pituitary-adrenal (HPA) axis abnormalities

The hypothalamic-pituitary-adrenal (HPA) axis describes the interactions between the hypothalamus, the pituitary, and adrenal glands. The HPA axis is thought to play an important biological role in stress reactivity. Abnormalities in the HPA axis have been associated with pathology of numerous psychiatric disorders and phenomena, including suicidal behaviors.

Dexamethasone Suppression Test

Dexamethasone is an exogenous steroid that is often prescribed for relief of inflammation. The dexamethasone suppression test (DST) assesses adrenal gland function by measuring change of cortisol levels in response to acute injection of dexamethasone. Dexamethasone can inhibit adrenocorticotropic hormone (ACTH) secretion by the pituitary, leading to drops in serum cortisol in a patient with normative HPA axis function. In patients with HPA abnormalities this drop in cortisol may not be observed. These abnormalities are sometimes seen in patients with depression, though DST is not used in the diagnosis of depression. The DST has been proposed as a method to predict suicide risk. In a long-term follow-up

study of 78 inpatients with major depressive disorder or schizoaffective disorder, 32 of 78 patients were found to have abnormal DST results. Estimated risk of eventual suicide was 26.8% in the group of patients with abnormal DST results as compared to only 2.9% with normal results. [8] Another follow-up study of 106 depressed inpatients with a suicide attempt yielded an Area Under the Curve (AUC, a measure of the sensitivity and specificity of a test) of 0.636. [9] This is suggestive of a role for HPA axis abnormality in suicidal behavior, but may not rise to the level of clinical utility. It could be that HPA dysfunction is important in only certain specific subtypes of suicidal behavior and application of DST might be more informative in those individuals. Work is ongoing on identifying such subgroupings.

SKA2. Another gene and potential biomarker identified reproducibly associated with suicide is the spindle and kinetochore associated complex subunit 2 (SKA2). SKA2 has been implicated in HPA axis abnormalities associated with PTSD and with suicidal behavior. [219] Robust across gender groups and studies, SKA2 was first identified as a risk factor in genome-wide DNA methylation analysis of post-mortem brains. SKA2 was found to be differentially methylated, a finding that was extended to peripheral blood samples of additional cohorts. [220] In addition to increased DNA methylation, other studies have found decreased SKA2 expression in the blood to be associated with suicide. [196] [197] SKA2 may functionally explain some of the pathology underlying HPA axis abnormalities as well as serve as a biomarker for suicide.

SLC4A4. The solute carrier family member 4 (SLC4A4) gene codes for a sodium bicarbonate cotransporter. Mutations in this gene lead to proximal renal tubular acidosis, which is caused by the loss of the ability to reabsorb bicarbonate.[221] It has been found to be increased in expression in the blood of living psychiatric patients with suicidal ideation, and may be predictive of future hospitalizations in male psychiatric patients. [196] This gene is expressed in many tissues in the body, including blood, brain, and the kidneys. Functionally, abnormal expression in this gene may lead to pH dysregulation, which has been associated with the pathophysiology of acute panic attacks. Genome wide association studies have also provided evidence for the involvement common mutations in the gene for depressed suicide attempters. [222]

SAT1 and polyamines (SMOX). Polyamines are ubiquitous organic cations of low molecular mass and found in all living organisms and implicated in many biological processes. Specifically, putrescine, spermidine, and spermine are involved in apoptosis, cell growth and differentiation. Several of the enzymes involved in the catabolism of spermine to putrescine have been implicated in suicide. Differential gene expression of the spermine/spermidine acetyltransferase 1 (SAT1), the rate limiting step in the catabolism from spermine to spermidine and spermidine to putrescene, has been reproducibility associated with suicidal behavior. [223] Interestingly, there has been some discrepancy in the data regarding the direction of change. This could be due to tissue specific effects, as SAT1 was found to be decreased in many (though not all) brain regions in postmortem

tissue[224], while it was found to be increased in the peripheral blood of living bipolar male patients with suicidal ideation. [217] This difference could also be associated with the different suicidal subtypes, as decreased expression was associated with major depressive disorder while increased expression was found both in bipolar patients and non-depressed suicides. [217] [224]

Serotonin. The serotonergic system has been widely implicated in psychiatric disorders. This is particularly well known in the case of depression where one of the more classes of medication, the SSRI, targets the system. Disruption to the serotonergic system has also been frequently implicated by neuroanatomical studies of suicide, with some of the earliest findings showing diminished 5-hydroxyindoleacetic acid (5-HIAA) in the cerebral spinal fluid (CSF). [225] These findings have been replicated over the years with some consistency, with evidence centering around the raphae nucleus (RN). It was shown recently with positron emission tomography (PET) that greater serotonin1A binding potential (more available serotonin receptor binding sites) predict higher suicidal behavior and thinking. [226]

Inflammatory markers. Inflammatory markers have also been tied to suicide. IL-6 is increased in the CSF of suicide attempters. [227] This finding has also been reproduced in the peripheral blood of patients with suicidal ideation. [196] Pro-inflammatory cytokines like IL-6 may act by inducing enzymatic activity through the kynurenine pathway. Plasma kynurenine levels have been found to be elevated in suicide attempters with major depressive disorder, but not when

comparing patients with MDD to healthy volunteers. [228] A downstream outcome of this proinflammatory pathway could be the increased expression of quinolinic acid (QUIN), a potent neurotoxin with excitatory glutamatergic action through NMDA receptors. This effect would be reversible by the NMDA receptor antagonist and putative treatment for suicidal ideation ketamine. QUIN has indeed been found to be increased in the CSF of suicidal patients when compared to healthy controls. [229]

Identifying Biomarkers for Suicidal Behavior*.*

While prior suicide attempts are one of the greatest risk factors for a future suicide attempt, the majority of prior attempters will not eventually commit suicide. Data from a large web-based sample taken in Brazil provides evidence that the majority (72%) of suicide attempters did not have a real intention to die. [216] With regards to suicide completion, it may be that a substantial number of those who died might not have endorsed a real intention to do so had they survived. Early intervention and prevention of the first attempt may save a substantial number of lives. We embarked on a series of studies aimed at identifying biomarkers that might begin to provide for objective means to do so. The initial study was a pilot example of how a small male bipolar cohort of longitudinally followed individuals who showed a change in suicidal ideation over time.

Chapter 4: Discovery and Validation of Blood Biomarkers for Suicidality

Introduction

Whatever its evolutionary[230], teleological and cultural reasons for existing, suicidal behavior is in most cases pathological and leads to irreversible tragedies[231]. Paradoxically, given its importance, there are yet no reliable objective tools to assess and track changes in suicidal risk without asking the individuals directly. Such tools are desperately needed, as individuals at risk often choose not to share their ideation or intent with others, for fear of stigma, hospitalization, or that in fact their plans may be thwarted.

A convergence of methods assessing the persons' internal subjective feelings and thoughts, along with external more objective ratings of actions and behaviors, are used *de facto* in clinical psychiatry. Such an approach is insufficient, and lagging behind those used in other medical specialties. It lacks precision, objectivity and predictive ability.

Our group has previously provided first proof of principle for the use of blood gene expression biomarkers to predict mood state[232] and psychosis symptoms[233]. As the target organ in psychiatry- the brain- cannot be biopsied in live patients, it is essential to be able to identify and validate peripheral biomarkers for subsequent practical implementation in clinical settings. We now present a comprehensive and highly reductionist approach for discovering and validating blood biomarkers for suicidality.

We used a Convergent Functional Genomics approach to identify and prioritize biomarkers of relevance to suicidality. CFG is a powerful combined approach for extracting signal from noise in genetic and gene expression studies. The CFG methodology has already been applied to help identify and prioritize candidate genes, pathways and mechanisms for neuropsychiatric disorders such as bipolar disorder[234-237], alcoholism[238], anxiety[239], and schizophrenia[240], showing reproducibility and predictive ability in independent cohorts.

Methods

Human subjects

We present data from four cohorts: one live bipolar discovery cohort, one postmortem coroner's office test cohort, and two prospective follow-up live cohorts- one bipolar and one psychosis (schizophrenia/schizoaffective). Sample collection procedures are as described previously for subjects recruited through the Indianapolis VA Medical Center.

Our intra-subject discovery cohort, from which the biomarker data were derived, consisted of 9 male Caucasian subjects with bipolar disorder, with multiple visits, that had each had a diametric change in suicidal ideation scores from no suicidal ideation to high suicidal ideation from one testing visit to another testing visit. There were 6 subjects with 3 visits each, and 3 subjects with 2 visits each, resulting in a total of 24 blood samples for subsequent microarray studies (Table 4-1 and Figure 4-1).

Our postmortem cohort, in which the top biomarker findings were tested, consisted of an age-matched cohort of 9 male suicide completers obtained through the Marion County Coroner's Office (8 Caucasians, 1 African-American) (Table 4-1 and Table S4-2). We required a last observed alive post-mortem interval of 24 hours or less, and the cases selected had completed suicide by means other than overdose, which could affect gene expression. Next of kin signed informed consent at the coroner's office for donation of tissues and fluids for research. The samples were collected as part of our INBRAIN initiative (Indiana Center for Biomarker Research in Neuropsychiatry).

The bipolar follow-up cohort (n=42) (Table 4-1) consisted of male Caucasian subjects in which whole-genome blood gene expression data, including levels of SAT1, were obtained by us at testing visits over the years as part of our longitudinal study. If the subjects had multiple testing visits, the visit with the highest SAT1 level was selected for this analysis. The subjects' subsequent number of hospitalizations with or without suicidality was tabulated from electronic medical records. The psychosis (schizophrenia/schizoaffective) follow-up cohort (n=46) (Table S4-9) similarly consisted of Caucasian subjects in which whole-genome blood gene expression data, including levels of SAT1, were obtained by us at testing visits over the years as part of our longitudinal study. If the subjects had multiple testing visits, the visit with the highest SAT1 level was selected for this analysis. The subjects' subsequent number of hospitalizations with or without suicidality was tabulated from electronic medical records. A hospitalization was

deemed to be without suicidality if suicidality was not listed as a reason for admission, and no suicidal ideation was described in the admission and discharge medical notes. Conversely, a hospitalization was deemed to be due to suicidality if suicidal acts or intent was listed as a reason for admission, and suicidal ideation was described in the admission and discharge medical notes.

Table 4-1. Demographics A. Individual. B. Aggregate. Diagnosis established by comprehensive structured clinical interview (DIGS). NOS- not otherwise specified. Suicidal Ideation question is from the Hamilton Rating Scale for Depression obtained at the time of the blood draw for each subject. GSW-Gun Shot Wound.

A.

Cohort 1: Live Bipolar Subjects Discovery Cohort (n=9) (24 chips)

SubjectID-Visit	***Diagnosis***	***Age***	***Gender***	***Ethnicity***	***Suicidal Ideation***
phchp023v1	Bipolar Disorder NOS	52	M	Caucasian	0
phchp023v2	Bipolar Disorder NOS	52	M	Caucasian	3
phchp023v3	Bipolar Disorder NOS	52	M	Caucasian	0
phchp093v1	Bipolar I Disorder	51	M	Caucasian	0
phchp093v2	Bipolar I Disorder	51	M	Caucasian	0

phchp093v3	Bipolar I Disorder	52	M	Caucasian	3
phchp095v1	Bipolar I Disorder	28	M	Caucasian	3
phchp095v2	Bipolar I Disorder	29	M	Caucasian	0
phchp095v3	Bipolar I Disorder	29	M	Caucasian	2
phchp122v1	Bipolar Disorder NOS	51	M	Caucasian	0
phchp122v2	Bipolar Disorder NOS	51	M	Caucasian	2
phchp128v1	Bipolar I Disorder	45	M	Caucasian	2
phchp128v2	Bipolar I Disorder	45	M	Caucasian	0
phchp136v1	Bipolar I Disorder	41	M	Caucasian	0
phchp136v2	Bipolar I Disorder	41	M	Caucasian	0
phchp136v3	Bipolar I Disorder	41	M	Caucasian	3

phchp153v1	Bipolar II Disorder	55	M	Caucasian	0
phchp153v2	Bipolar II Disorder	55	M	Caucasian	2
phchp153v3	Bipolar II Disorder	56	M	Caucasian	0
phchp179v1	Bipolar Disorder NOS	36	M	Caucasian	0
phchp179v2	Bipolar Disorder NOS	37	M	Caucasian	0
phchp179v3	Bipolar Disorder NOS	37	M	Caucasian	3
phchp183v1	Bipolar I Disorder	48	M	Caucasian	3
phchp183v2	Bipolar I Disorder	48	M	Caucasian	0

Cohort 2: Coroner's Office Test Cohort- Suicide Completers (n=9) (9 chips)

SubjectID	Psychiatric Diagnosis	Age	Gender	Ethnicity	Suicide by

INBR009	Bipolar/ Schizophrenia	59	M	Caucasian	Hanging
INBR011	Depression/ADHD	26	M	Caucasian	GSW to chest
INBR012	Unknown	39	M	Caucasian	GSW to head
INBR013	Depression	68	M	African-American	GSW to mouth
INBR014	None	27	M	Caucasian	Hanging
INBR015	None	40	M	Caucasian	Hanging
INBR016	Anxiety/TBI	68	M	Caucasian	GSW to head
INBR017	Depression	56	M	Caucasian	GSW to chest
INBR018	None	65	M	Caucasian	Slit wrist

Cohort 3: Live Bipolar Subjects Prospective Follow-up Cohort (n=42)

SubjectID-Visit	**Diagnosis**	**Age**	**Gender**	**Ethnicity**	**SAT1 Levels at testing**	**Years since testing**	**Future Hosp. w/o suicidality**	**Future Hosp. due to suicidality**	**Frequency of Future Hosp. w/o suici**	**Frequency of Future Hosp. due to suici**

									dality	**dality**
phchp234v1	Bipolar II Disorder	44	M	Caucasian	1955.20	0.83	0	0	0.00	0.00
phchp053v2	Bipolar I Disorder	58	M	Caucasian	2178.30	5.67	4	0	0.71	0.00
phchp152v1	Bipolar I Disorder	45	M	Caucasian	2178.80	2.33	0	0	0.00	0.00
phchp122v1	Bipolar Disorder NOS	51	M	Caucasian	2245.60	0.58	0	0	0.00	0.00
phchp190v3	Bipolar Disor	50	M	Caucasian	2300.60	1.25	0	0	0.00	0.00

	der NOS									
phchp020v3	Bipolar Disorder NOS	63	M	Caucasian	2342.60	4.08	0	0	0.00	0.00
phchp113v1	Bipolar I Disorder	37	M	Caucasian	2437.40	3.00	0	0	0.00	0.00
phchp132v2	Bipolar I Disorder	51	M	Caucasian	2558.90	2.33	0	0	0.00	0.00
phchp184v3	Bipolar Disorder NOS	64	M	Caucasian	2575.40	1.33	0	0	0.00	0.00

phchp039v3	Bipolar I Disorder	52	M	Caucasian	2580.10	5.75	0	0	0.00	0.00
phchp147v1	Bipolar II Disorder	38	M	Caucasian	2582.80	2.25	0	0	0.00	0.00
phchp178v1	Bipolar I Disorder	49	M	Caucasian	2616.80	1.00	0	0	0.00	0.00
phchp136v3	Bipolar I Disorder	41	M	Caucasian	2635.90	2.00	0	0	0.00	0.00
phchp045v3	Bipolar I Disorder	36	M	Caucasian	2721.00	5.42	0	0	0.00	0.00

phchp 224v1	Bipolar I Disorder	59	M	Caucasian	2748.10	1.08	1	1	0.92	0.92
phchp 183v1	Bipolar I Disorder	48	M	Caucasian	2750.90	0.42	2	1	4.80	2.40
phchp 171v2	Bipolar Disorder NOS	36	M	Caucasian	2795.70	1.50	0	0	0.00	0.00
phchp 166v1	Bipolar Disorder NOS	56	M	Caucasian	2829.60	1.92	0	0	0.00	0.00
phchp 253v1	Bipolar Disor	25	M	Caucasian	2888.50	1.00	0	0	0.00	0.00

	der NOS									
phchp186v1	Bipolar II Disorder	43	M	Caucasian	2901.50	1.67	0	0	0.00	0.00
phchp079v2	Bipolar Disorder	44	M	Caucasian	3053.20	4.50	0	0	0.00	0.00
phchp128v1	Bipolar I Disorder	45	M	Caucasian	3118.60	2.67	0	0	0.00	0.00
phchp080v1	Bipolar I Disorder	44	M	Caucasian	3153.60	5.00	0	0	0.00	0.00
phchp088v1	Bipolar I Disorder	44	M	Caucasian	3194.10	4.58	0	10	0.00	2.18

phchp 109v1	Bipolar I Disorder	22	M	Caucasian	3200.80	3.00	1	2	0.33	0.67
phchp 134v3	Bipolar II Disorder	59	M	Caucasian	3202.30	1.92	0	0	0.00	0.00
phchp 153v1	Bipolar II Disorder	55	M	Caucasian	3304.90	2.00	0	0	0.00	0.00
phchp 274v2	Bipolar Disorder NOS	48	M	Caucasian	3349.00	0.50	0	0	0.00	0.00
phchp 140v3	Bipolar II Disorder	38	M	Caucasian	3393.80	1.92	0	0	0.00	0.00

phchp030v3	Bipolar I Disorder	49	M	Caucasian	3395.20	5.92	0	3	0.00	0.51
phchp124v1	Bipolar I Disorder	53	M	Caucasian	3660.90	2.50	0	6	0.00	2.40
phchp095v3	Bipolar I Disorder	29	M	Caucasian	3695.40	0.33	0	1	0.00	3.00
phchp100v1	Bipolar I Disorder	28	M	Caucasian	3767.80	1.58	0	0	0.00	0.00
phchp210v3	Bipolar I Disorder	44	M	Caucasian	3844.60	0.50	0	0	0.00	0.00

phchp219v1	Bipolar Disorder NOS	61	M	Caucasian	3845.10	1.17	0	0	0.00	0.00
phchp031v3	Bipolar I Disorder	52	M	Caucasian	4080.70	4.08	1	0	0.24	0.00
phchp093v3	Bipolar I Disorder	52	M	Caucasian	4137.40	2.67	0	1	0.00	0.38
phchp067v1	Bipolar II Disorder	39	M	Caucasian	4214.70	5.58	0	0	0.00	0.00
phchp142v3	Bipolar I Disorder	55	M	Caucasian	4310.70	1.92	0	0	0.00	0.00

phchp 112v2	Bipolar I Disorder	46	M	Caucasian	4410.40	1.33	0	0	0.00	0.00
phchp 149v2	Bipolar Disorder NOS	45	M	Caucasian	4586.90	2.00	1	0	0.50	0.00
phchp 117v1	Bipolar I Disorder	43	M	Caucasian	6531.10	3.00	0	0	0.00	0.00

B.

Live Bipolar Subjects Discovery Cohort (n=9)			
Suicidal Ideation (SI) (score)	***No SI (0)***	***High SI (2-4)***	***Overall***

Number of subjects (number of chips)	9(14)	9(10)	9(24)
Age *mean* *(SD)* *range*	46.1 (8.1) 29-56	43.8 (9.7) 28-55	45.1 (8.7) 28-56
Ethnicity (Caucasian/ African-American)	(9/0)	(9/0)	(9/0)

Coroner's Office Test Cohort- Suicide Completers (n=9)	
Number of subjects (number of chips)	9(9)
Age *mean* *(SD)* *range*	49.8 (17) 26-68
Ethnicity (Caucasian/ African-American)	(8/1)

Live Bipolar Subjects ***Prospective Follow-up Cohort (n=42)***			
SAT1 Levels	***Lower Tertile***	***Upper Tertile***	***Overall***
Number of subjects	14	14	42
Age ***mean*** ***(SD)*** ***range***	48.5 (9) 36-64	45.3 (9.5) 28-61	46.2 (9.9) 22-64
Ethnicity (Caucasian/ African-American)	(14/0)	(14/0)	(42/0)

Medications

The subjects in the discovery cohort were all diagnosed with bipolar disorder (Table 4-1). Their psychiatric medications are listed in Table S4-1. The subjects were on a variety of different psychiatric medications: mood stabilizer, antidepressants, antipsychotics, benzodiazepines, and others. Medications can have a strong influence on gene expression. However, our discovery of differentially expressed genes was based on intra-subject analyses, which factor out not only genetic background effects but also medication effects, as the subjects had no major medication changes between visits. Moreover, there was no consistent pattern in any particular type of medication, or between any change in medications and suicidal ideation, in the rare instances where there were changes in medications between visits.

Human blood gene expression experiments and analyses

RNA extraction, sample labeling, and procedures for microarray are previously described.

Analysis

We have used the subject's suicidal ideation (SI) scores at the time of blood collection (0—no SI compared to 2 and above- high SI). We looked at gene expression differences between the no SI and the high SI visits using both an intra-subject and an inter-subject design (Figure 4-1).

Differential Gene Expression Analyses in the Discovery Cohort

We imported all Affymetrix microarray data as .cel files into Partek Genomic Suites 6.6 software package (Partek Incorporated St.Louis, Missouri U.S.A). Using only the Perfect Match (PM) values, we ran a robust multi-array analysis (RMA), background corrected with quantile normalization and a median polish probeset summarization of all 24 chips, to obtain the normalized expression levels of all probesets for each chip. Then, to establish a list of differentially expressed probesets, we ran two analyses.

An intra-subject analysis using a fold change in expression of at least 1.2 between high and no SI visits within each subject was performed. There were in total 15 comparisons. Probesets that had a 1.2 fold change were then assigned either a 1 (increased in high SI) or a -1 (decreased in high SI) in each comparison. These values were then summed for each probeset across the 15 comparisons, yielding a range of scores between -11 and 12. The probesets in the top 5% (1,269 probesets, <5% of 54,675 total probesets) had an absolute (without sign) score value of 7 and greater, and received an internal CFG score of 1 point. The probesets in the top .1% (24 probesets, <.1% of 54,675 total probesets) had an absolute score of 11 and greater, and received an internal CFG score of 3 points.

Additionally, an inter-subject analysis using t-test (2-tailed, unequal variance) was performed to find probesets differential expressed between high SI and no SI chips (Figure 4-1), resulting in 648 probesets with P<.05. Probesets with a P<.05

received an internal CFG score of 1 point, while probesets with P<.001 received 3 points.

We further filtered results by only selecting probesets that overlapped between the intra-subject and the inter-subject analyses, resulting in 279 probesets corresponding to 246 unique genes. Gene names for the probesets were identified using Partek as well as NetAffyx (Affymetrix) for Affymetrix HG-U133 Plus 2.0 GeneChips, followed by GeneCards to confirm the primary gene symbol. In addition, for those probesets that were not assigned a gene name by Partek or NetAffyx, we used the UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) to directly map them to known genes. Genes were then scored using our manually curated CFG databases as described below (Figure 4-2).

Validation Analyses

We imported the 9 Affymetrix microarray data files from the suicide completers cohort as .cel files into Partek Genomic Suites 6.6 software package (Partek Incorporated St.Louis,Missouri U.S.A). We then ran a robust multi-array analysis (RMA), background corrected with quantile normalization and a median polish probeset summarization of all the chips from the discovery and validation cohort (24 +9=33 chips), to obtain the normalized expression levels of all probesets for each chip. Partek normalizes expression data into a log base of 2 for visualization purposes. We non-log transformed expression data by taking 2 to the power of the transformed expression value. We then used the non-log transformed expression data to compare expression levels of biomarkers in the different groups

(Figure 4-3). One-tail t-tests with unequal variance, one-way ANOVA, and Bonferonni corrections were used for statistical comparisons.

For live cohorts future hospitalization analyses in bipolar disorder and schizophrenia/schizoaffective, we similarly RMA normalized each cohort, prior to looking at biomarker levels in individual subjects. One-tail t-tests with equal variance were used for statistical comparisons. ROC curves were calculated using SPSS software for each of the 4 dimensional analyses, predicting the state variable of hospitalizations due to suicidality.

Figures

Each figure in this chapter was completed by Helen Le-Niculescu and Daniel Levey. This work has been published. [217]

Table 4-2. Top gene expression biomarkers for suicidality

Gene Symbol/ Gene name	**Probe sets**	**Change**	**Differential Expression Score**	**Prior Human Genetic Evidence**	**Prior Human Brain Expression Evidence**	**Total CFG Score**
SAT1 spermidine/spermine N1-acetyltransferase 1	203455_s_at	I	2	(Assoc) Suicide attempt[241]; Suicide [223]	Suicide in Depression (D) PFC [242] Suicide (D) AMY, PFC, HIP, THAL[243] Suicide (D) PFC[244] Suicide (D) PFC[245]	8

					Suicide (D) PFC[246] Suicide (D) PFC[247] Suicide (D) PFC[224] Suicide (D) PFC [223]	
CD24 CD24 molecule	209772_s_at	D	4		Suicide in mood disorders (D) NAC [248]	8
FOXN3 forkhead box N3	230790_x_at	I	2	(Assoc) Suicide [249]	Suicide (I) PFC [249]	8

GBP1 guanylate binding protein 1, interferon-inducible, 67kDa	231577_s_at 202269_x_at 202270_at	I	4 2 2		Suicide in mood disorders (D) NAC [248]	8 6 6
PIK3R5 Phosphoinositide-3-kinase, regulatory subunit 5	227553_at	I	4		Suicide in mood disorders (D) PFC [248]	8
APOL2 Apolipoprotein L2	221653_x_at	I	2		Suicide PFC (I)[250]	6
ATP13A2 ATPase type 13A2	218608_at	D	2		Suicide (D)[248]	6
ATP6V0E1 ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e1	214149_s_at 214244_s_at	I	2		Suicide (D) PFC [223]	6

EPHX1 epoxide hydrolase 1, microsomal (xenobiotic)	202017_at	D	2		Suicide in Schizophrenia (D) PFC[251]	6
GCOM1 GRINL1A complex locus 1	239099_at	I	2		Suicide in Depression (D) PFC[252]	6
HTRA1 HtrA serine peptidase 1	201185_at	D	2		Suicide (I)[248]	6
IL1B interleukin 1, beta	39402_at	I	2		Suicide (I) PFC[253]	6
LEPR leptin receptor	211354_s_at	D	2		Suicide (D) PFC[252] (D) PFC[254] (D) HIP[255] Suicide in Depression (I) PFC[256]	6

LHFP lipoma HMGIC fusion partner	218656_s_at	I	2		Suicide in mood disorders (I) NAC [248]	6
LIPA lipase A	236156_at	I	2		Violent Suicide (I) PFC[257]	6
MARCKS myristoylated alanine-rich protein kinase C substrate	213002_at	I	2		Suicide in Depression (I)[258]	6
PGLS 6-Phosphogluconolactonase	230699_at	I	2		Suicide PFC (D)[250]	6
PTEN phosphatase and tensin homolog	222176_at	I	2		Suicide PFC, HIP (I) [259]	6
RECK	216153_x_at	I	2		Suicide (I) PFC[248]	6

reversion-inducing-cysteine-rich protein with kazal motifs						
SPTBN1 spectrin, beta, non-erythrocytic 1	200671_s_at	D	2		Suicide in mood disorders (I) NAC [248]	6
TNFSF10 tumor necrosis factor (ligand) superfamily, member 10	202688_at 202687_s_at 214329_x_at	I	2		Suicide in Schizophrenia (I)PFC [251] Suicide in Depression (I) PFC[256]	6
ABCA1 ATP-binding cassette, sub-family A (ABC1), member 1	203504_s_at	I	4			4
ARHGEF40 (FLJ10357)	241631_at	I	4			4

Rho guanine nucleotide exchange factor (GEF) 40						
CASC1 cancer susceptibility candidate 1	220168_at	I	4			4
DHRS9 dehydrogenase/reductase (SDR family) member 9	219799_s_at	I	4			4
DISC1 disrupted in schizophrenia 1	244642_at	I	2	(Assoc) Suicide [249]		4
EIF2AK2 eukaryotic translation initiation factor 2-alpha kinase 2	204211_x_at	I	4			4
LOC727820 uncharacterized LOC727820	231247_s_at	I	4			4
MAP3K3	242117_at	I	4			4

mitogen-activated protein kinase kinase kinase 3						
MBNL2 muscleblind-like 2 (Drosophila)	205017_s_at	D	2	(Assoc) Suicide [249]		4
MT-ND6 (ND6) mitochondrially encoded NADH dehydrogenase 6	1553575_at	I	4			4
OR2J3 olfactory receptor, family 2, subfamily J, member 3	**217334_at**	**D**	**4**			**4**
RBM47 RNA binding motif protein 47	1565597_at	I	4			4
RHEB Ras homolog enriched in brain	227633_at	D	2	(Assoc) Suicide[260]		4
RICTOR	228248_at	I	4			4

RPTOR independent companion of MTOR, complex 2						
SAMD9L sterile alpha motif domain containing 9-like	243271_at; 230036_at	I	4			4
SCARF1 scavenger receptor class F, member 1	206995_x_at	I	4			4
SLC36A1 solute carrier family 36 (proton/amino acid symporter), member 1	213119_at	I	4			4
STAT1 signal transducer and activator of transcription 1, 91kDa	232375_at	I	4			4
UBA6 ubiquitin-like modifier activating enzyme 6	236879_at	I	4			4

ZC3HAV1 zinc finger CCCH-type, antiviral 1	1563075_s_at	I	4			4
COX5B cytochrome c oxidase subunit Vb	213736_at	I	2	(Linkage) 2q11.2[261]		3
SMARCA1 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	203874_s_at	I	2	(Linkage) Xq25[246]		3
DBP D-box binding protein	209782_s_at	D	2			2

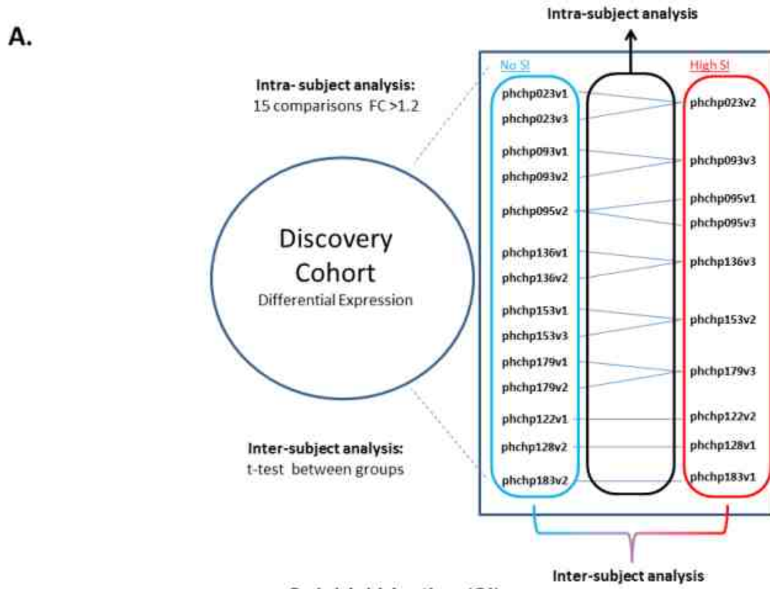
Results

Discovery

We conducted whole-genome gene expression profiling in blood samples from a longitudinally-followed homogeneous cohort of male subjects with a major mood disorder (bipolar disorder) that predisposes to suicidality. 1 in 3 individuals with bipolar disorder attempt suicide during their lifetime[262]. The samples were collected at repeated visits, 3 to 6 months apart. State information about suicidal ideation was collected from a questionnaire administered at the time of each blood draw (Table 4-1). Out of 75 bipolar subjects (with a total of 174 visits) followed longitudinally in our study, there were 9 subjects that switched from a no suicidal ideation (SI score of 0) to a high suicidal ideation state (SI score of 2 and above) at different visits, which was our intended study group. We used a powerful intra-subject design to analyze data from these 9 subjects and their 24 visits. An intra-subject design factors out genetic variability, as well as some medications, lifestyle and demographic effects on gene expression, permitting identification of relevant signal with Ns as small as 1[14]. An ancillary benefit of an intra-subject design may be accuracy/consistency of self-report of psychiatric symptoms ("phene expression"), similar in rationale to the signal-detection benefits it provides in gene expression. We also used an overall inter-subject case-case analysis, to identify genes differentially expressed in the blood in no suicidal ideation states vs. high suicidal ideation states (Figure 4-1). The number of subjects that met our criteria and were analyzed is small, but comparable to those in human postmortem brain

gene expression studies of suicide[248]. We are indeed treating the blood samples as surrogate tissue for brains, with the caveat that they are not the real target organ. However, with blood samples from live human subjects we have the advantages of in vivo accessibility, better knowledge of the mental state at the time of collection, less technical artifacts,

A.



B.

Suicidal Ideation (SI)
from Hamilton Rating Scale for Depression (HAMD17) .
No SI –score of 0; High SI- score of 2 or above.

SUICIDE
0= Absent
1= Feels life is not worth living
2= Wishes he were dead or any thoughts of possible death to self
3= Suicidal ideas or gesture
4= Attempts at suicide (any serious attempt rates 4)

C.

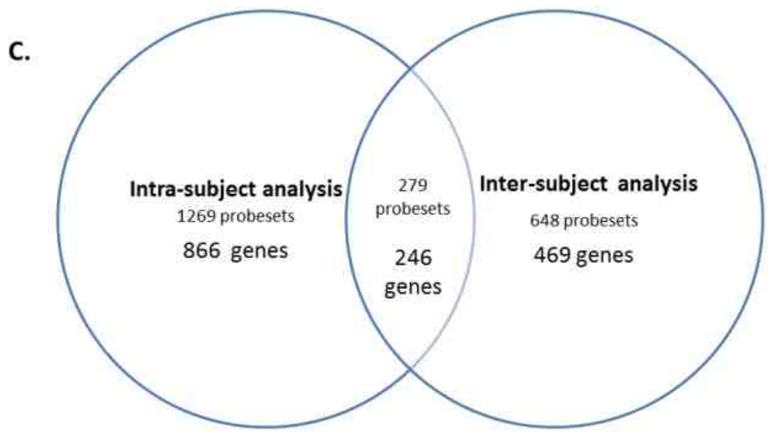


Figure 4-1. Discovery cohort: intrasubject and intersubject analyses. Phchp### is study ID for each subject. V# after it denotes visit number (1, 2 or 3). (a) Design and (b) suicidal ideation (SI) scoring.

and especially of being able to do powerful intra-subject analyses from visit to visit. We considered and differentially scored only the very top 0.1% and 5% of the gene expression probesets distributions, and also required overlap between the intra-subject and inter-subject analyses of gene expression changes. Such a restrictive approach was used as a way of minimizing false positives, even at the risk of having false negatives (Figure 4-1C). For example, there were genes on each of the two lists, from intra- and inter-subject analyses, that had clear prior evidence for involvement in suicidality, such as MT1E[248], respectively GSK3B[263], but were not included in our subsequent analyses because they were not in the overlap.

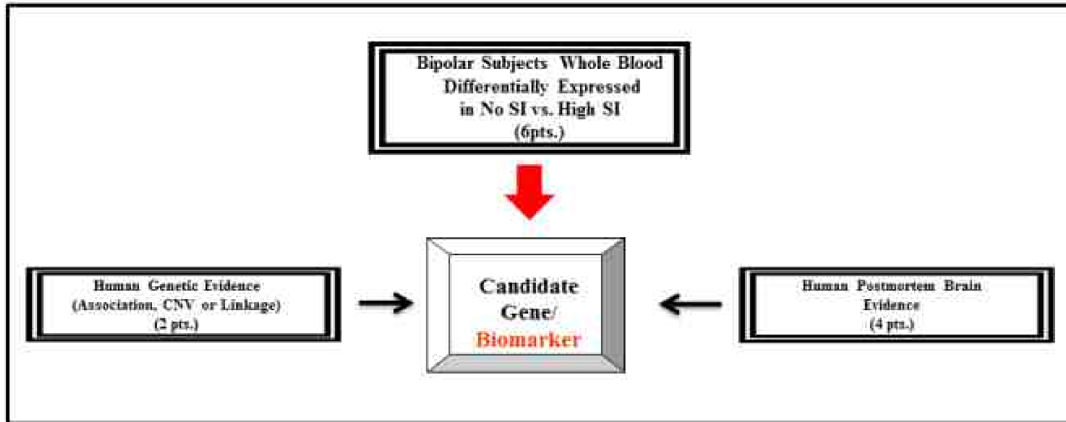
We then used a Convergent Functional Genomics approach (CFG) (Figure 4-2) to cross-match the list of 246 overlapping top differentially expressed genes from the blood samples with other key lines of evidence (human postmortem brain data, human genetic data) implicating them in suicidality, as a way of identifying and prioritizing disease-relevant genomic biomarkers, extracting generalizable signal out of potential cohort-specific residual noise and genetic heterogeneity. We have built in our lab manually curated databases of the psychiatric genomic and proteomic literature to date, for use in CFG analyses[21,264-266]. The CFG approach is thus a de facto field-wide collaboration. We use in essence, in a Bayesian fashion, the whole body of knowledge in the field to leverage findings from our discovery datasets. Unlike our use of CFG in previous studies, for the current one we did not use any human peripheral tissue evidence from the literature, as there was none

directly matching our genes, reflecting perhaps the dearth of peripheral gene expression work done so far on suicides, and the need for a study like ours. We also did not use animal model evidence, as there are to date no clear studies in animal models of self-harm or suicidality. SAT1 (spermidine/spermine N1-acetyltransferase 1) was the top scoring blood biomarker, with the most extensive convergent evidence, increased in suicidal states identified by our work (i.e. the top risk marker). CD24 (CD24 molecule/ small cell lung carcinoma cluster 4 antigen) was the top blood biomarker decreased in suicidal states (i.e. the top protective marker) (Figure 4-2 and Table 4-2).

Testing in suicide completers

In order to know whether our findings relate to actual completed suicide, we then tested SAT1 levels in blood samples from a heterogeneous cohort of 9 consecutive male suicide completers obtained from the coroner's office, with the following characteristics: we required that the cases included in our analysis had a postmortem interval from last observed alive under 24 hours, and that they had committed suicide by means other than overdoses, which could alter gene expression. Remarkably, we found SAT1 gene expression levels to be elevated in 9 out of 9 (100%) of subjects who committed suicide that we tested. In each of the suicide completers, the increase in SAT1 was at least three standard deviations above the average levels in high suicidal ideation subjects, which constitutes a very stringent threshold for use as a predictive biomarker (Figure 4-3).

A.



B.

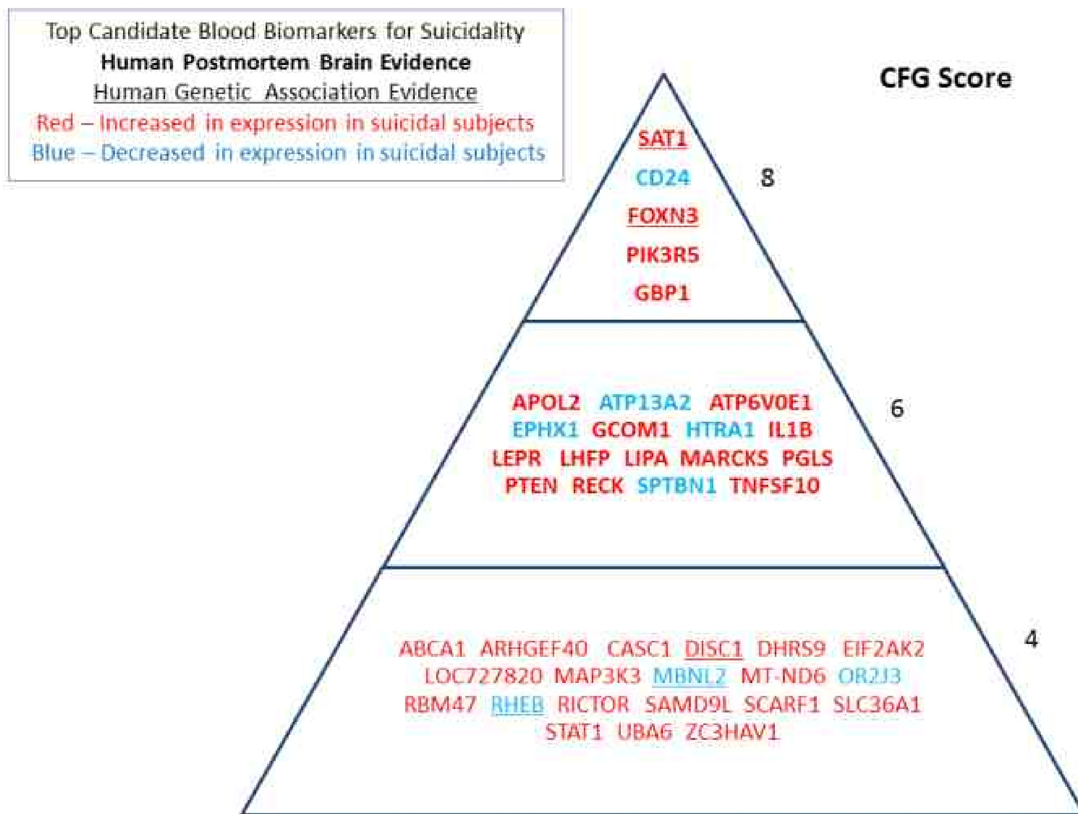
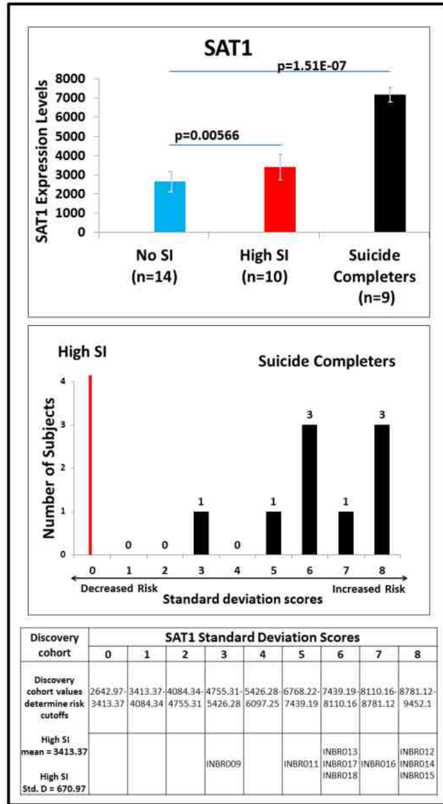


Figure 4-2. Convergent Functional Genomics approach for identification and prioritization of genomic biomarkers for suicidality.

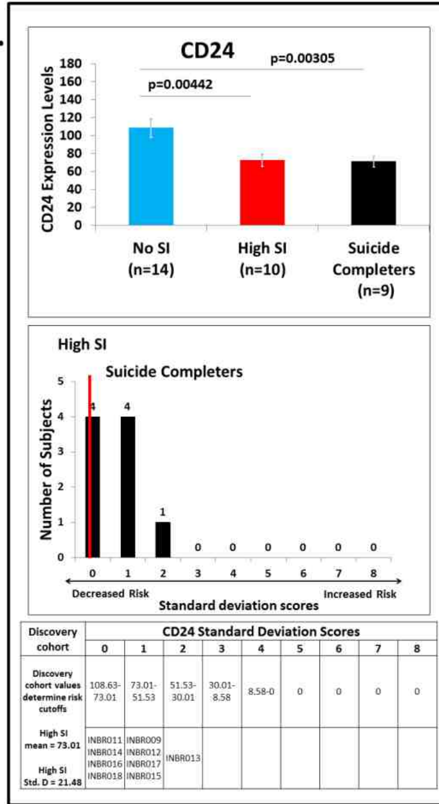
We also examined other top candidate biomarkers for suicidality (Figure 4-3 and Figure S4-3). Remarkably, 13 out of the 41 CFG top scoring biomarkers from Figure 4-2A (32%) showed step-wise significant change from No SI to High SI to the test Suicide Completers group. 6 out of them (15%) remained significant after strict Bonferroni correction for multiple comparisons (Figure 4-3). The top CFG scoring biomarker SAT1 remained the top biomarker after validation. Thirteen out of the 41 CFG top-scoring biomarkers from Figure 4-2b (32%) showed step-wise significant change from no SI to high SI, to the validation suicide completers group. Six out of them (15%) remained significant after strict Bonferroni correction for multiple comparisons. The top CFG scoring biomarker SAT1 remained the top biomarker after validation.

A.



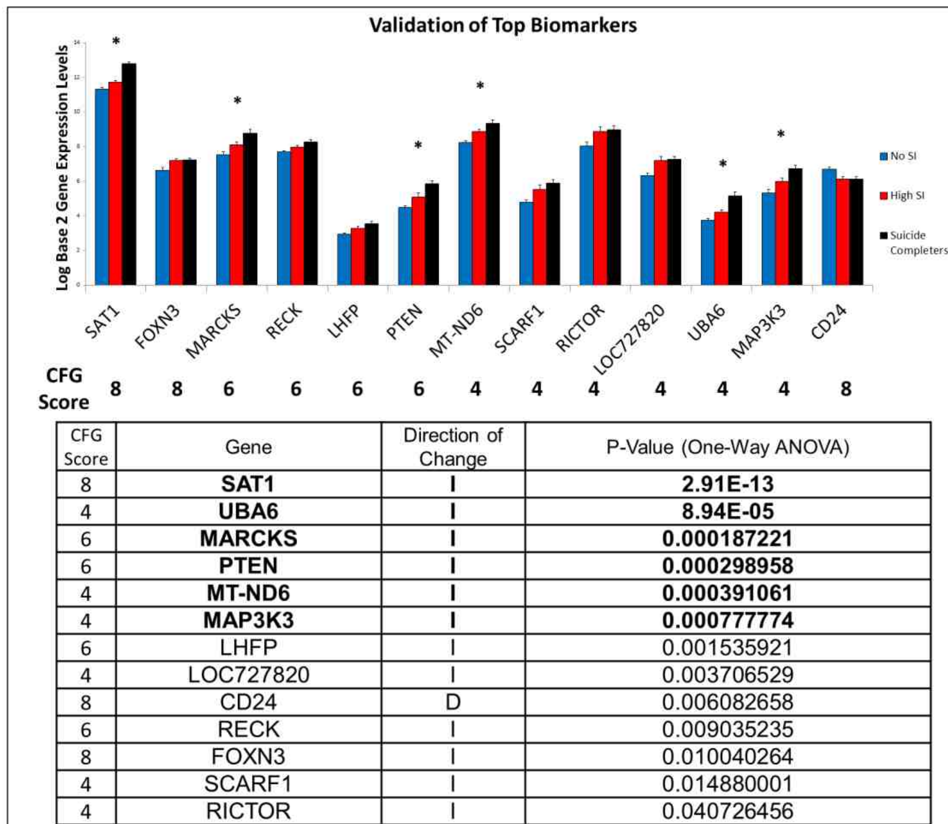
Discovery cohort	SAT1 Standard Deviation Scores								
	0	1	2	3	4	5	6	7	8
Discovery cohort values determine risk cutoffs	2642.97-3413.37	3413.37-4084.34	4084.34-4755.31	4755.31-5426.28	5426.28-6097.25	6768.22-7439.19	7439.19-8110.16	8110.16-8781.12	8781.12-9452.1
High SI mean = 3413.37 High SI Std. D = 670.97				INBR009		INBR011	INBR013 INBR017 INBR018	INBR016	INBR012 INBR014 INBR015

B.



Discovery cohort	CD24 Standard Deviation Scores								
	0	1	2	3	4	5	6	7	8
Discovery cohort values determine risk cutoffs	108.63-73.01	73.01-51.53	51.53-30.01	30.01-8.58	8.58-0	0	0	0	0
High SI mean = 73.01 High SI Std. D = 21.48	INBR011 INBR014 INBR016 INBR018	INBR009 INBR012 INBR017 INBR015	INBR013						

C.



CFG Score	Gene	Direction of Change	P-Value (One-Way ANOVA)
8	**SAT1**	**I**	**2.91E-13**
4	**UBA6**	**I**	**8.94E-05**
6	**MARCKS**	**I**	**0.000187221**
6	**PTEN**	**I**	**0.000298958**
4	**MT-ND6**	**I**	**0.000391061**
4	**MAP3K3**	**I**	**0.000777774**
6	LHFP	I	0.001535921
4	LOC727820	I	0.003706529
8	CD24	D	0.006082658
6	RECK	I	0.009035235
8	FOXN3	I	0.010040264
4	SCARF1	I	0.014880001
4	RICTOR	I	0.040726456

Figure 4-3. Testing of biomarkers in suicide completers. (a) Upper: SAT1 (spermidine/spermine N1–acetyltransferase 1) expression is significantly increased (P ¼ 0.0057) in our discovery work between subjects with high suicidal ideation (SI) (mean ¼ 3413.37) and those reporting no SI (mean ¼ 2642.97). Our test cohort of suicide completers (mean ¼ 7171.51) showed significantly greater expression of SAT1 than both high SI (P ¼ 7.27e-07) and no SI (P ¼ 1.51e-07) groups from the discovery cohort. Lower: a suicide risk score was calculated by scoring the s.d. band a subject fell within as derived from the high SI discovery cohort, starting from the mean of the high-SI discovery cohort. A score of 0 indicates the subject falling between the means of the high SI and no SI subjects in the discovery cohort. A score of 1 means between the mean of the high SI and the first s.d. above it, score of 2 between the first and second s.d., score of 3 between the second and third s.d., and so on. Red line marks where the average SAT1 gene expression in high SI subjects would fall. (b) Upper: CD24 (CD24 molecule/small cell lung carcinoma cluster 4 antigen) expression was significantly decreased (P ¼ 0.0044) within the discovery cohort between subjects reporting high SI (mean ¼ 73.01) and no SI (mean ¼ 108.634). The test cohort of suicide completers (mean ¼ 71.61) was also significantly decreased (P ¼ 0.0031) when compared with subjects reporting no SI. Lower: suicide risk score defined as the s.d. band in which the subject expression fell below the mean of the high-SI discovery cohort. Red line marks where the average CD24 gene expression in high SI subjects would fall. (c) Testing of top candidate biomarkers for suicidality.

Mechanistic understanding

Pathway analyses of our suicidality biomarker data identified among the top pathways the omega-3 docosahexaenoic acid (DHA) signaling pathway. Low omega-3 levels have been correlated with increased suicidality in human epidemiological studies [167] [168]. Several of the biomarkers from our current study (SAT1, S100A8, IL1B and 16 others) were changed in expression by omega-3 treatment in the blood of the circadian clock gene DBP (D-box binding protein) knock-out mouse model in opposite direction to our human suicidality data (Table S4-6). DBP is also one of the biomarkers identified to be decreased in high suicidal states in the current analysis. Serendipitously, previous work by our group has implicated DBP in mood disorders[267], psychosis[28], alcoholism[238], and anxiety disorders[239]. Mice engineered to lack DBP were stress-reactive and displayed a behavioral phenotype similar to bipolar disorder and co-morbid alcoholism[268]. In addition to bipolar disorder, alcoholism increases risk for suicide [269]. PTEN, a biomarker increased in suicidality in the current study in the blood, as well as in the brain of suicide completers[259], was also increased in the amygdala and decreased in the pre-frontal cortex of DBP knock-out mice subjected to stress[269]. S100A8, another biomarker increased in suicidality in the current study, was also increased in the blood of DBP stressed mice. Treatment with omega-3 fatty acids normalized the phenotype of those mice[270].

Other circadian clock-modulated genes identified by our analysis as biomarkers for suicidality were PIK3R5, MARCKS, IL1B, CASC1, CCRN4L, H3F3B,

RBCK1, TNK2, and UBE2B. Circadian genes are involved in sleep-wake cycles, as well as mood regulation[235,236,267,271,272]. Abnormal sleep (insomnia) has been identified as a risk factor for suicide[273] . IL1B is also an inflammatory marker, and has previously been implicated by us in anxiety disorders[239].

Table 4-3. Underlying Biology. A. Pathways. B. Disease and Disorders.

A.	**INGENUITY Pathways**				**KEGG Pathways**		
	#	**Top Canonical Pathways**	**P-Value**	**Ratio**	**Pathway Name**	**Enrichment Score**	**Enrichment p-value**
CFG score>=6.0 N=21 genes	1	Role of Tissue Factor in Cancer	2.63 E-04	3/115 (0.026)	**Apoptosis**	6.69102	0.001242
	2	Dendritic Cell Maturation	9.83 E-04	3/207 (0.014)	Measles	6.06369	0.002326
	3	Melanoma Signaling	1.13 E-03	2/46 (0.043)	Endometrial cancer	4.96787	0.006958
	4	**Docosahexaenoic Acid (DHA) Signaling**	1.18	2/49 (0.041)	Influenza A	4.90223	0.00743

	#	Top Canonical Pathways	P-Value	Ratio	Pathway Name	Enrichment Score	Enrichment p-value
			E-03				
	5	Endometrial Cancer Signaling	1.69 E-03	2/57 (0.035)	Phosphatidylinositol signaling system	4.85448	0.007793
	#	**Top Canonical Pathways**	**P-Value**	**Ratio**	**Pathway Name**	**Enrichment Score**	**Enrichment p-value**
CFG score>=4.0 N=41 genes	1	NF-kB Signaling	4.42 E-04	4/175 (0.023)	Measles	8.7667	0.000156
	2	Dendritic Cell Maturation	5.38 E-04	4/207 (0.019)	Influenza A	6.87308	0.001035
	3	PDGF Signaling	7.5E-04	3/85 (0.035)	mTOR signaling pathway	6.34986	0.001747

	#						
	4	Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	1.14E-03	3/106 (0.028)	**Apoptosis**	4.75687	0.008592
	5	Role of Tissue Factor in Cancer	1.78E-03	3/115 (0.026)	Toll-like receptor signaling pathway	4.37269	0.012617

B.	**INGENUITY**			
CFG score>= 6.0 N=21 genes	**#**	**Diseases and Disorders**	**P-Value**	**Number of Molecules**
	1	**Cancer**	1.22E-06 - 4.54E-03	14
	2	Connective Tissue Disorders	2.19E-04 - 3.41E-03	8
	3	**Inflammatory Disease**	2.19E-04 - 4.54E-03	8
	4	Skeletal and Muscular Disorders	2.19E-04 - 4.42E-03	9
	5	Gastrointestinal Disease	2.22E-04 - 4.54E-03	12

	#	Diseases and Disorders	P-Value	Number of Molecules
CFG score>= 4.0 N=41 genes	1	**Cancer**	4.51E-06 - 6.45E-03	20
	2	**Inflammatory Response**	2.70E-05 - 6.45E-03	12
	3	Antimicrobial Response	9.95E-05 - 6.45E-03	4
	4	Infectious Disease	1.25E-04 - 5.52E-03	6
	5	Connective Tissue Disorders	1.53E-04 - 6.45E-03	11

Additionally, S100A8, MBNL2 and 3 other biomarkers had evidence for modulation by clozapine in blood in opposite direction to our human suicidality data in previous independent animal model pharmacogenomics studies conducted by us [28,233] (Table S4-6). Clozapine is the only FDA approved treatment for suicidality[274].

Thus, the convergent evidence for our biomarkers is strong in translational ways beyond those used for their discovery and selection. S100A8 may be a key biomarker to monitor in terms of response to treatment with classic (clozapine)

and complementary (omega-3) agents. Other potential drugs to be studied for modulating suicidality were revealed by our analyses (Tables S4-5 and S4-6).

SAT1, FOXN3, DISC1, MBNL2 and RHEB had genetic association evidence for suicidality, suggesting that they are not only state biomarkers but also trait factors influencing suicidal risk. DISC1 is also one of the top candidate genes for schizophrenia based on a large scale CFG analysis of schizophrenia GWAS we recently conducted[240], while DISC1 and MBNL2 are also among of the top candidate genes for bipolar disorder based on a large scale CFG analysis of bipolar disorder GWAS we previously conducted[236]. Additionally, DISC1 has clear animal model data for the role of its interaction with environmental stress in the pathophysiology of psychotic depression[275]. DISC1 and MBNL2 may thus be key state and trait factors for suicidality risk in psychotic mood disorder subjects, and an indication for clozapine treatment in such subjects.

We also looked at the overlap of our suicide biomarkers with our previous mood biomarker[232] and psychosis biomarker[233] work (Table S4-7), as well as with the human postmortem brain literature for other psychiatric disorders (Table S4-8). DOCK5 and 4 other biomarkers were changed in high suicidal states in opposite direction to their change in high mood states, and DOCK5 and 6 other biomarkers were changed in the same direction as their change in high psychosis states, suggesting that suicidality could be viewed as a psychotic dysphoric state, and that DOCK5 may be an additional key biomarker reflecting that state. This molecularly-informed view is consistent with the emerging clinical evidence in the field[276]. The

convergence of evidence then suggests that, at least in the population we studied, suicidality may be associated with dysphoric mood, as well as increased psychosis, anxiety and stress. In our own data, SAT1 blood gene expression levels showed a trend towards increase in low mood, high psychosis, high anxiety, and high stress in our bipolar subjects (Figure 4-4).

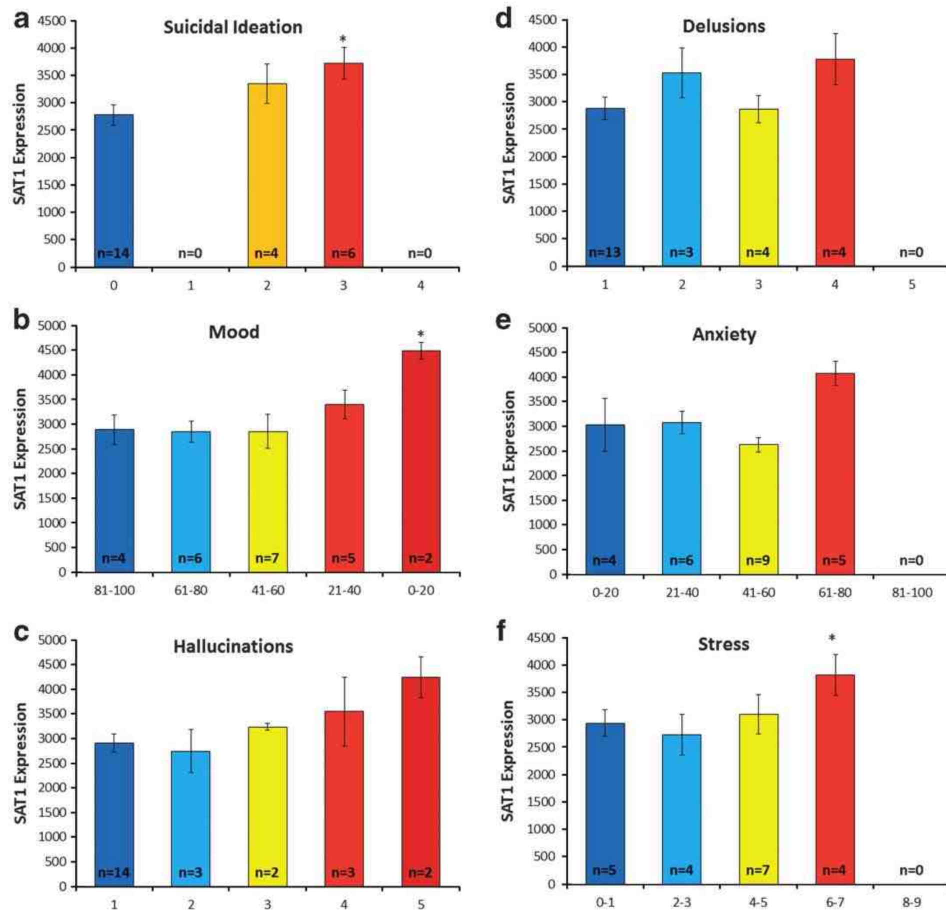


Figure 4-4. SAT1 (spermidine/spermine N1–acetyltransferase 1) expression in the bipolar discovery cohort: relationship with suicidal ideation (SI), mood, psychosis, anxiety and stress. (a) SAT1 expression and SI item from Hamilton Rating Scale for Depression (HAMD) (scores of 0–4). (b) SAT1 expression and visual-analog scale for mood (0–100). High mood is to the left on the x-axis, low mood is to the right. (c) SAT1 expression and Hallucinations item from Positive and Negative Symptoms Scale (PANSS; scores of 1–7). Higher score indicates higher symptoms.

Prospective Validation

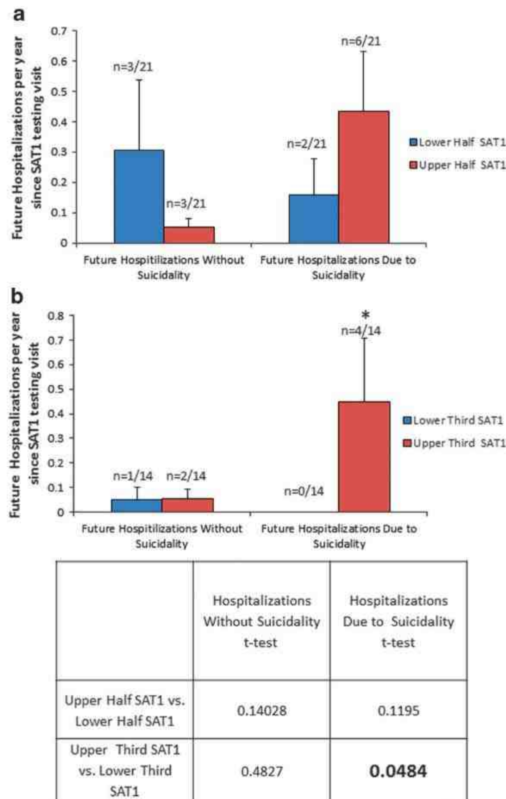
To further validate SAT1, our top marker, we also looked at subsequent hospitalizations with and without suicidality (Table 4-1 and Table S4-9), and previous hospitalizations with and without suicidality (Table S10-3), in two live cohorts, one bipolar (n=42) and one psychosis (schizophrenia/schizoaffective) (n=46). Higher SAT1 levels compared to lower SAT1 levels at time of testing differentiated future as well as past hospitalizations due to suicidality in the bipolar disorder subjects (Figure 4-5). A similar but weaker pattern was exhibited in the psychosis (schizophrenia/schizoaffective) subjects (Figure S4-2). Remarkably, besides SAT1, three other (PTEN, MARCKS and MAP3K3) of the six biomarkers that survived Bonferroni correction in the suicide completers cohort validation step also showed similar but weaker results (Table S4-11 and Figure S4-3). Taken together, the prospective and retrospective hospitalization data suggests SAT1, PTEN, MARCKS and MAP3K3 might be not only state markers but perhaps trait markers as well.

We also examined whether using a multi-dimensional approach enhanced our ability to predict future hospitalizations, by adding data about mood, anxiety, and psychosis to the data about SAT1 expression levels (Figure 4-6). We found that the ROC curve improved in a step-wise fashion, from an AUC of .640 with SAT1 alone, to and AUC of .798 with SAT1 and anxiety, AUC of .813 with SAT1, anxiety and mood, and AUC of .835 with SAT1, anxiety, mood and psychosis. From our preliminary work, we identified levels of SAT1 that provide different levels of

sensitivity and specificity (Table S4-12). The anxiety and mood information was obtained from simple visual analog scales, previously described by us[277]. The psychosis information is based on the combining of the scores on the hallucinations and delusions in the PANSS (Figure S4-5). Of note, this simple clinical-genomic approach does not directly ask about suicidal ideation, which some individuals may deny or choose not to share with clinicians. Similar data were obtained for the panel of six top markers (Figure S4-6).

Discussion

Using discovery in live subjects and validation in suicide completers, we found possible biomarkers for suicidality. Our top biomarker finding, SAT1, as well as PTEN, MARCKS and MAP3K3, were additionally validated by prospective and retrospective analyses in live subjects, looking at ability to predict and differentiate future and past hospitalizations due to suicidality in bipolar disorder and psychosis (schizophrenia/schizoaffective) (Table S4-11).



	Hospitalizations Without Suicidality t-test	Hospitalizations Due to Suicidality t-test
Upper Half SAT1 vs. Lower Half SAT1	0.14028	0.1195
Upper Third SAT1 vs. Lower Third SAT1	0.4827	**0.0484**

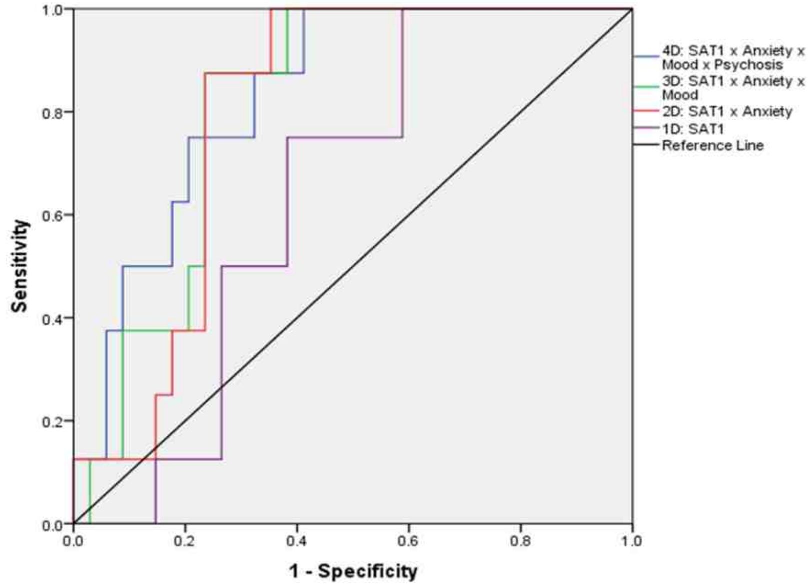
Figure 4-5. Prospective validation of SAT1 (spermidine/spermine N1–acetyltransferase 1): follow-up of future psychiatric hospitalizations due to suicidality. We analyzed in 42 bipolar subjects whether their SAT1 levels at the time of initial testing differentiated those who had subsequent hospitalizations due to suicidality in the years since the testing occurred. Range was 0.33–5.92 years of follow-up, average 2.48 years. (a) Upper half of SAT1 scores versus lower half of SAT1 scores. Twenty-one subjects in each group. There were six psychiatric hospitalizations not due to suicidality, and eight psychiatric hospitalizations due to suicidality. (b) Upper tertile of SAT1 scores versus lower tertile of SAT1 scores.

Fourteen subjects in each group. There were three psychiatric hospitalizations not due to suicidality, and four psychiatric hospitalizations due to suicidality.

Apoptosis

Beyond predictions, as a window into the biology of suicidality, the current work shows overlap at a gene and pathway level with apoptosis (Table 4-3, Table S4-3, Table S4-4). SAT1, for example, is a key catabolic enzyme for polyamines. Polyamine levels within cells control cell viability, and significant decreases in polyamine levels can result in apoptosis[278]. They seem to reflect an endowment for cellular and organismal activity and growth, key characteristics of mood[84,232,236]. SAT1, which is increased in live suicidal ideation subjects and in suicide completers in our studies, is highly inducible by a variety of stimuli, including toxins, cytokines, heat shock, ischemia, and other stresses. SAT1 overexpressing mice had alterations in their polyamine pool, hair loss, infertility and weight loss[279,280]. Turecki and colleagues have provided compelling evidence for changes in the polyamine system in the brain of suicide completers[281]. CD24, our top biomarker decreased in suicidal subjects, also has roles in apoptosis. Mice lacking CD24 show an increased rate of apoptosis[282]. It could be that simpler mechanisms related to cellular survival and programed cell-death decision have been recruited by evolution for higher mental functions such as feelings, thoughts, actions and behaviors leading to suicidality.

A



B

D	Test Result Variable(s)	Area Under the Curve	Std. Error	Significance	95% Confidence Interval	
					Lower Bound	Upper Bound
1D	SAT1	.640	.086	.224	.471	.808
2D	SAT1 x Anxiety	.798	.068	.009	.665	.931
3D	SAT1 x Anxiety x Mood	.813	.066	.006	.683	.942
4D	SAT1 x Anxiety x Mood x Psychosis	.835	.066	.004	.706	.964

C

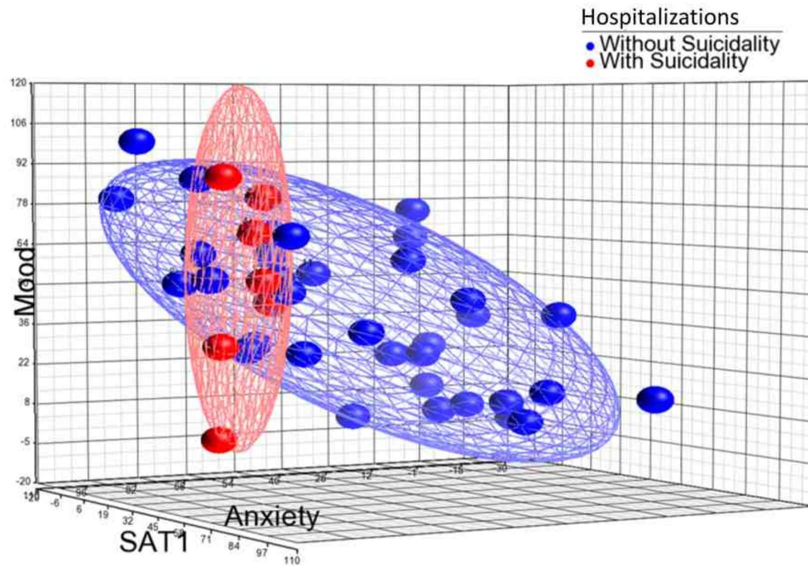


Figure 4-6. Multi-dimensional prediction of future psychiatric hospitalizations due to suicidality. We analyzed in 42 bipolar subjects whether their SAT1 (spermidine/spermine N1–acetyltransferase 1), anxiety, mood and psychosis levels at the time of initial testing differentiated from those who had subsequent hospitalizations due to suicidality in the years since the testing occurred. Data in each dimension was normalized to a 0–100 scale (with the mood visual-analog scale (VAS) inverted, as the assumption was made that depressed mood states would more closely correlate with suicidality). The angle between dimensions was assumed to be 901, and a simple Pythagorean distance from origin score was calculated. The distribution of this score in the test cohort was used to generate a receiver-operating characteristic curve for hospitalizations due to suicidality. (a) ROC curve. (b) Detailed results. (c) Three-dimensional visualization.

In that sense, suicidality could be viewed as whole-organism self-poptosis. Apoptosis mechanisms have previously been implicated in mood disorders and their inhibition in affective resilience[283]. Interestingly, lithium, a medication with clinical evidence for preventing suicidality in bipolar disorder[284], has anti-apoptotic effects at a cellular level[285]. Imaging studies have shown reduced gray matter volume in the brain of individuals with bipolar disorder and history of suicide attempts. Long-term lithium treatment was associated with increased gray matter volumes in the same areas where suicide was associated with decreased gray matter[286].

Conclusions and future directions

Taken together, our results have implications for the understanding of suicide, as well as for the development of objective laboratory tests and tools to track suicidal risk and response to treatment. First, our results open empirical avenues for future field trials, clinical testing and validation in various at-risk populations, including studies in individuals with major depressive disorder. The current work was based on subjects with bipolar disorder, psychosis (schizophrenia/ schizoaffective disorder), and coroner's office cases, which may represent a more externalizing or impulsive population and type of suicidality. Other types are likely to exist. Second, more work also needs to be done to examine potential gender and ethnicity differences. Our current work is based on male Caucasian subjects. Third, predicting suicidal feelings and thoughts (ideation) may be different than predicting suicidal actions and behaviors. Our current work

has focused on suicide completers and hospitalizations, admittedly a more emergent concern. Fourth, state vs. trait issues, and sensitivity vs. specificity for suicidality, for the individual markers identified by us, as well as for panels of markers and multi-modal approaches, need to be studied more extensively in different populations. Fifth, past individual and family history, as well as environmental context, may help improve predictive approaches. Our approach was very focused and reductionist, albeit with good results.

Given the fact that approximately one million people die of suicide worldwide each year, and this is a potentially preventable cause of death, the need for, urgency and importance of efforts such as ours cannot be overstated.

Chapter 5: An INTRA n-of-1 approach to biomarker discovery

The next study represents an evolution in approach to the challenge of biomarker discovery. While the previous study used a discovery cohort design consisting of individuals who had a change in suicidal ideation over time, this was done so at a group level, comparing group means based on the SI phenotype. This is certainly not a wrong approach, but it may not be the best way to get at relative changes that are happening within an individual. [287]

We have designed an algorithm that scores assigns simple differential expression scores that tracks gene expression over time as it relates to each individual persons phenotype. This methodology is described in greater detail below. Briefly, each successive clinic visit by a participant is scored for change or stability in phenotype. At the same time gene expression is scored in the same way. The score is tabulated for each succession of visits for each individual and then summed across all individuals to provide a simple score that indicates how well gene expression tracks changes in phenotype. Methods of this type may be more powerful for discovering biomarkers as it helps to minimize noise artifact differences *between* individuals while magnifying true signal differences *within* individuals.

Because of this innovation we felt confident in our ability to identify reproducible signal beyond the singular diagnosis of bipolar disorder used previously to also include major depressive disorder, schizophrenia, schizoaffective disorder, post traumatic stress disorder, and mood disorder NOS. This more than

quadrupled our n from 9 participants to 37 for discovery and more than doubled the size of the independent testing cohorts. This also enabled us to look at diagnostic differences in biomarker performance.

While there is emerging evidence for biological risk factors for suicidal behavior, it is likely that clinical risk assessments and a clinicians judgement will remain vital in identifying and preventing suicides, now and in the future. Gene expression or metabolites may lack the specificity to illness without the context of clinical assessments and patient history. Implicated neurotransmitter systems are an especially good example. Dysregulation in the serotonin system has been reproducibly indicated in suicide, but this has been found for a multitude of psychiatric conditions. Serotonin dysregulation without the context of patient history and clinical data, like other purely biological markers, may never be sufficient to assess a patient's suicidal risk. The last major evolution in methodology was the incorporation of contextual phenotypic data in the form of visual analog scales for mood and anxiety, along with the suicide risk assessment checklist, the CFI-S. Integration of clinical information into a universal predictive measure (UP-Suicide) showed greatly enhanced predictive ability to identify present state suicidal ideation and future hospitalizations due to suicidal behavior.

Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach

Introduction

Predicting suicidal behavior in individuals is one of the hard problems in psychiatry, and in society at large. Improved, objective, and quantitative ways to do it are needed. One cannot always ask individuals if they are suicidal, as desire not to be stopped or future impulsive changes of mind may make their self-report of feelings, thoughts and plans to be unreliable. We had previously provided proof of principle of how first generation blood biomarkers for suicide discovered in male bipolar participants, alone or in combination with clinical symptoms data for anxiety and mood, could have predictive ability for future hospitalizations for suicidality. We now present comprehensive new data for discovery, prioritization, validation, and testing of next generation broader-spectrum blood biomarkers for suicidal ideation and behavior, across psychiatric diagnoses. We also describe two clinical information questionnaires in the form of apps, one for affective state (Simplified Affective State Scale, SASS) and one for suicide risk factors (Convergent Functional Information for Suicide, CFI-S), and show their utility in predicting suicidality. Both these instruments do not directly ask about suicidal ideation. Lastly, we demonstrate how our apriori primary endpoint, a comprehensive universal predictor for suicide (UP-Suicide), composed of the combination of top biomarkers (from discovery, prioritization and validation), along

with CFI-S, and SASS, predicts in independent test cohorts suicidal ideation and future psychiatric hospitalizations for suicidality.

Methods

Human participants

We present data from four cohorts: one live psychiatric participants discovery cohort; one postmortem coroner's office validation cohort; and two live psychiatric participants test cohorts–one for predicting suicidal ideation and one for predicting future hospitalizations for suicidality (Figure 5-1).

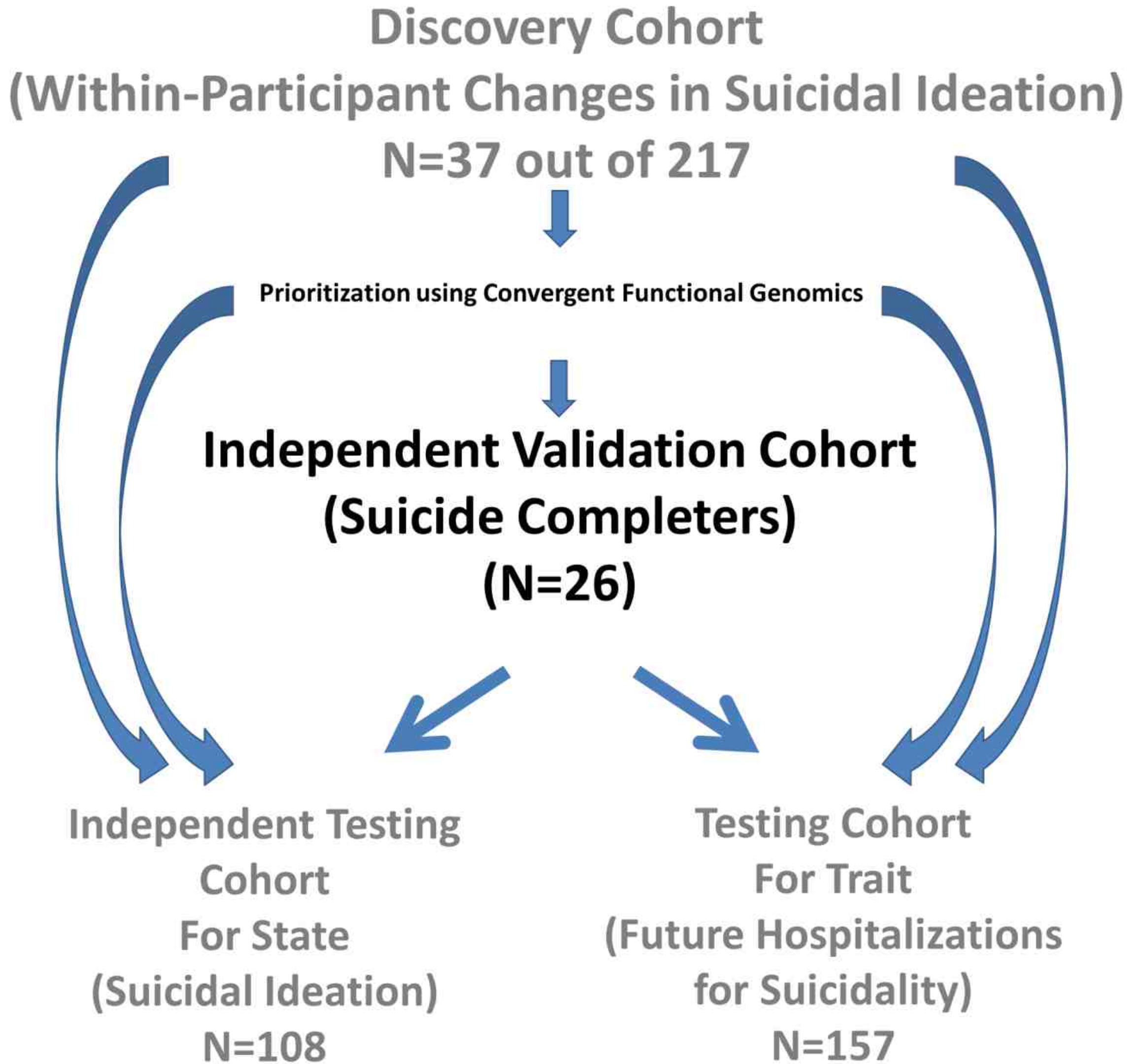


Figure 5-1. Cohorts used in study depicting flow of discovery, prioritization, validation and testing of biomarkers from each step.

The live psychiatric participants are part of a larger longitudinal cohort being collected and studied by us. Participants are recruited from the patient population at the Indianapolis VA Medical Center. The participants are recruited largely through referrals from care providers, the use of brochures left in plain sight in public places and mental health clinics, and through word of mouth. All participants understood and signed informed consent forms detailing the research goals, procedure, caveats and safeguards. Participants completed diagnostic assessments by an extensive structured clinical interview—Diagnostic Interview for Genetic Studies—at a baseline visit, followed by up to six testing visits, 3–6 months apart or whenever a hospitalization occurred. At each testing visit, they received a series of psychiatric rating scales, including the Hamilton Rating Scale for Depression-17, which includes a suicidal ideation (SI) rating item (Figure 2), and the blood was drawn. Whole blood (10 ml) was collected in two RNA-stabilizing PAXgene tubes, labeled with an anonymized ID number, and stored at -80 degrees C in a locked freezer until the time of future processing. Whole-blood (predominantly lymphocyte) RNA was extracted for microarray gene expression studies from the PAXgene tubes, as detailed below. We focused this study on a male population because of the demographics of our catchment area (primarily male in a VA Medical Center), and to minimize any potential gender-related effects on gene expression, which would have decreased the discriminative power of our analysis given our relatively small sample size.

Our within-participant discovery cohort, from which the biomarker data were derived, consisted of 37 male participants with psychiatric disorders, with multiple visits in our lab, who each had at least one diametric change in SI scores from no SI to high SI from one testing visit to another testing visit. There was 1 participant with 6 visits, 1 participant with 5 visits, 1 participant with 4 visits, 23 participants with 3 visits each, and 11 participants with 2 visits each, resulting in a total of 106 blood samples for subsequent microarray studies (Figure 5-2 and Table 5-1).

Our postmortem cohort, in which the top biomarker findings were validated, consisted of a demographically matched cohort of 26 male violent suicide completers obtained through the Marion County coroner's office (Table 5-1 and Supplementary Table S5-2). We required a last observed alive postmortem interval of 24 h or less, and the cases selected had completed suicide by means other than overdose, which could affect gene expression. 15 participants completed suicide by gunshot to head or chest, 9 by hanging, 1 by electrocution and 1 by slit wrist. Next of kin signed informed consent at the coroner's office for donation of blood for research. The samples were collected as part of our INBRAIN initiative (Indiana Center for Biomarker Research in Neuropsychiatry).

A.

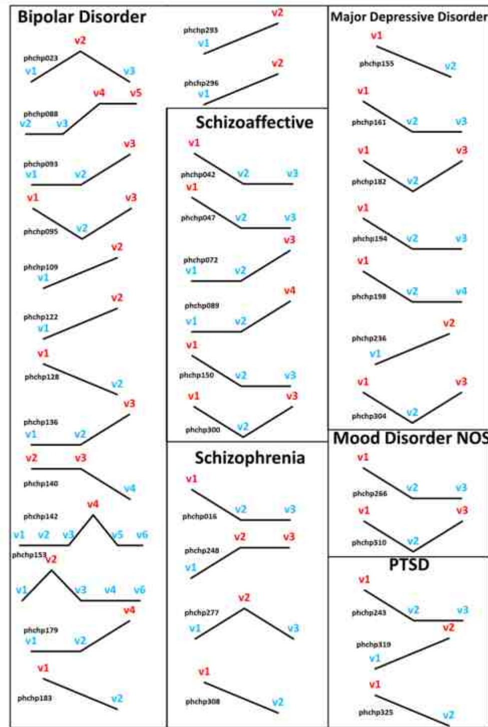
Suicidal Ideation (SI)
from Hamilton Rating Scale for Depression (HAMD).
No SI - score of 0; High SI - score of 2 or above

Suicide
0 = Absent
1 = Feels life is not worth living
2 = Wishes he were dead or any thoughts of possible death to self
3 = Suicidal ideas or gesture
4 = Attempts at suicide (any serious attempt rates 4)

B.

Discovery Cohort:

37 psychiatric participants who have at least one switch between a No SI state visit and a High SI state visit.



C.

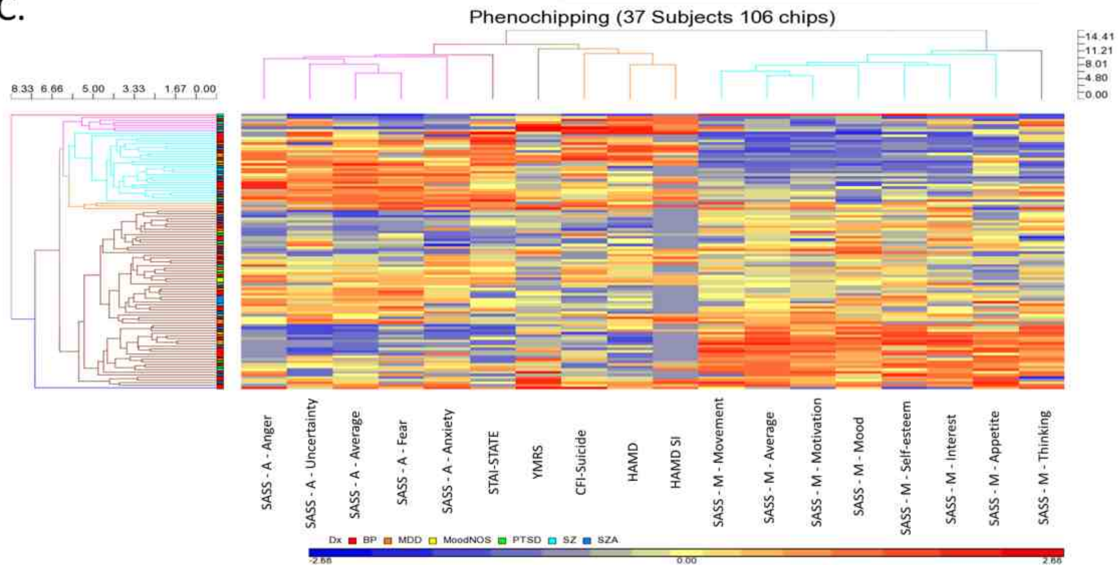


Figure 5-2. Discovery cohort: longitudinal within-participant analysis. Phchp### is study ID for each participant. V# denotes visit number (1, 2, 3, 4, 5 or 6). (a) Suicidal ideation (SI) scoring. (b) Participants and visits. (c) PhenoChipping: two-way unsupervised hierarchical clustering of all participant visits in the discovery cohort vs 18 quantitative phenotypes measuring affective state and suicidality. A—anxiety items (anxiety, uncertainty, fear, anger, average). M—mood items (mood, motivation, movement, thinking, self-esteem, interest, appetite, average). SASS, simplified affective state scale; STAI-STATE, state trait anxiety inventory, state subscale; YMRS, Young Mania Rating Scale.

Our independent test cohort for predicting suicidal ideation (Table 5-1) consisted of 108 male participants with psychiatric disorders, demographically matched with the discovery cohort, with one or multiple testing visits in our lab, with either no SI, intermediate SI, or high SI, resulting in a total of 223 blood samples in whom whole-genome blood gene expression data were obtained (Table 5-1 and Table S5-1).

Our test cohort for predicting future hospitalizations (Table 5-1 and Table S5-1) consisted of male participants in whom whole-genome blood gene expression data were obtained by us at testing visits over the years as part of our longitudinal study. If the participants had multiple testing visits, the visit with the highest marker (or combination of markers) levels was selected for the analyses (so called "high watermark" or index visit). The participants' subsequent number

of psychiatric hospitalizations, with or without suicidality, was tabulated from electronic medical records.

Table 1. Cohorts used in study

	Subjects	*Diagnosis*	*Ethnicity*	*Age mean s.d.*	*T-test for age*	
Discovery cohort (within-participant changes in suicidal ideation)	37	BP = 15 MDD = 7 SZA = 6 SZ = 4 PTSD = 3 Mood NOS = 2	EA = 29 AA = 8 Other = 0	47.25 8.59		
Independent validation cohort-gene expression (suicide completers)	26	NP = 13 MDD = 8 BP = 2 SZ = 1 AX = 1 Alcoholism = 1	EA = 21 AA = 4 Other = 1	40.81 17.47		*T-test for age with discovery cohort 0.114*
Independent validation cohort-CFI-S (suicide completers)	35	NP = 14 MDD = 16 BP = 2 SZ = 1 AX = 1 Alcoholism = 1	EA = 29 AA = 4 Other = 2	42.46 17.82		*T-test for age with discovery cohort 0.156*
Independent testing cohort for state (suicidal ideation)	108	No SI BP = 17 MDD = 17 SZA = 19 SZ = 20 Intermediate SI BP = 5 MDD = 0 SZA = 3 SZ = 4 High SI BP = 7 MDD = 8 SZA = 6 SZ = 2	EA = 71 AA = 36 Other = 1	47.1 9.6 No SI = 47.8 High SI = 45.7	*T-test for age between no and high SI 0.554*	*T-test for age with discovery cohort* P *= 0.919*
Testing cohort for trait (first year hospitalizations for suicidality)	157 (No Hosp for Suicidality = 139 Hosp for Suicidality = 18)	No Hosp for Suicidality BP = 43 MDD = 20 SZA = 41 SZ = 35 Hosp for Suicidality BP = 7 MDD = 3 SZA = 3 SZ = 5	No hosp for SI EA = 90 AA = 47 Other = 2 Hosp for SI EA = 13 AA = 5	49.6 9.5 No hosp for SI = 49.56 Hosp for SI = 49.92	*T-test for age between no Hosp for suicidality and Hosp for suicidality 0.886*	*T-test for age with discovery cohort 0.149*
Testing cohort for trait (all future hospitalizations for suicidality)	157 (No Hosp for Suicidality = 122 Hosp for Suicidality = 35)	No Hosp for Suicidality BP = 41 MDD = 20 SZA = 29 SZ = 32 Hosp for Suicidality BP = 9 MDD = 3 SZA = 15 SZ = 8	No hosp for Suicidality EA = 78 AA = 43 Other = 1 Hosp for Suicidality EA = 25 AA = 9 Other = 1	49.6 9.5 No Hosp for suicidality = 49.9 Hosp for suicidality = 48.4	*T-test for age between no Hosp for suicidality and Hosp for suicidality 0.436*	*T-test for age with discovery cohort 0.149*

Abbreviations: AX, anxiety disorder nos; BP, bipolar; CFI-S, Convergent Function Information for Suicide; MDD, major depressive disorder; NP, non-psychiatric; PTSD, post-traumatic stress disorder; SZA, schizoaffective; SZ, schizophrenia; SI, suicidal ideation.

All participants had at least one year of follow-up or more at our VA Medical Center since the time of the testing visits in the lab. Participants were evaluated for the presence of future hospitalizations for suicidality, and for the frequency of such hospitalizations. A hospitalization was deemed to be without suicidality if suicidality was not listed as a reason for admission, and no SI was described in the admission and discharge medical notes. Conversely, a hospitalization was deemed to be because of suicidality if suicidal acts or intent was listed as a reason for admission, and/or SI was described in the admission and discharge medical notes.

Medications

The participants in the discovery cohort were all diagnosed with various psychiatric disorders (Table 5-1). Their psychiatric medications were listed in their electronic medical records, and documented by us at the time of each testing visit. The participants were on a variety of different psychiatric medications: mood stabilizers, antidepressants, antipsychotics, benzodiazepines and others (data not shown). Medications can have a strong influence on gene expression. However, our discovery of differentially expressed genes was based on within- participant analyses, which factor out not only genetic background effects but also medication effects, as the participants had no major medication changes between visits. Moreover, there was no consistent pattern in any particular type of medication, or between any change in medications and SI, in the rare instances where there were changes in medications between visits.

Human blood gene expression experiments and analyses

RNA extraction. Whole blood (2.5–5 ml) was collected into each PaxGene tube by routine venipuncture. PaxGene tubes contain proprietary reagents for the stabilization of RNA. RNA was extracted and processed as previously described[217].

Microarrays. Biotin-labeled aRNAs were hybridized to Affymetrix HG-U133 Plus 2.0 GeneChips (Affymetrix; with over 40,000 genes and expressed sequence tags), according to the manufacturer's protocols. Arrays were stained using standard Affymetrix protocols for antibody signal amplification and scanned on an Affymetrix GeneArray 2500 scanner with a target intensity set at 250. Quality-control measures, including 30/50 ratios for glyceraldehyde 3-phosphate dehydrogenase and b-actin, scale factors, and background, were within acceptable limits.

Analysis. We have used the participant's SI scores at the time of blood collection (0—no SI compared with 2 and above—high SI). We looked at gene expression differences between the no SI and the high SI visits, using a within-participant design, then an across-participants summation (Figure 5-2).

Gene expression analyses in the discovery cohort

We analyzed the data in two ways: an Absent-Present (AP) approach, as in previous work by us on mood biomarkers[15] and on psychosis biomarkers[24], and a differential expression (DE) approach, as in previous work by us on suicide biomarkers[217]. The AP approach may capture turning on and off of genes, and the DE approach may capture gradual changes in expression. For the AP approach,

we used Affymetrix Microarray Suite Version 5.0 (MAS5) to generate Absent (A), Marginal (M), or Present (P) calls for each probeset on the chip (Affymetrix U133 Plus 2.0 GeneChips) for all participants in the discovery cohort. For the DE approach we imported all Affymetrix microarray data as .cel files into Partek Genomic Suites 6.6 software package (Partek Incorporated, St Louis, MI, USA). Using only the perfect match values, we ran a robust multi-array analysis (RMA), background corrected with quantile normalization and a median polish probeset summarization, to obtain the normalized expression levels of all probesets for each chip. RMA was performed independently for each of the 6 diagnoses used in the study, to avoid potential artefacts due to different ranges of gene expression in different diagnoses[288]. Then the participant's normalized data was extracted from these RMAs and assembled for the different cohorts used in the study.

A/P analysis. For the longitudinal within-participant AP analysis, comparisons were made within-participant between sequential visits to identify changes in gene expression from Absent to Present that track changes in phene expression (suicidal ideation) from No SI to High SI. For a comparison, if there was a change from A to P tracking a change from No SI to High SI, or a change from P to A tracking a change from High SI to No SI, that was given a score of +1 (increased biomarker in High SI). If the change was in opposite direction in the gene vs the phene (SI), that was given a score of -1 (decreased biomarker in High SI). If there was no change in gene expression between visits despite a change of phene expression (suicidal ideation), or a change in gene expression between visits

despite no change in phene expression (suicidal ideation), that was given a score of 0 (not tracking as a biomarker). If there was no change in gene expression and no change in suicidal ideation between visits, that was given a score of +1 if there was concordance (P-P with High SI-High SI, or A-A with No SI-No SI), or a score of -1 if there was the opposite (A-A with High SI-High SI, or P-P with No SI-No SI). If the changes were to M (moderate) instead of P, the values used were 0.5 or –0.5. These values were then summed up across the comparisons in each participant, resulting in an overall score for each gene/probeset in each participant. We also used a perfection bonus. If the gene expression perfectly tracked the suicidal ideation in a participant that had at least two comparisons (3 visits), that probeset was rewarded by a doubling of its overall score. Additionally, we used a non-tracking correction. If there was no change in gene expression in any of the comparisons for a particular participant, that overall score for that probeset in that participant was zero.

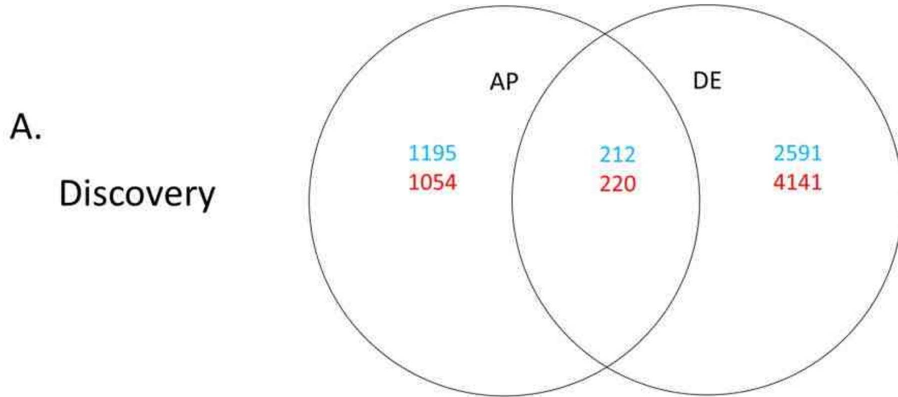
DE analysis. For the longitudinal within-participant DE analysis, fold changes (FC) in gene expression were calculated between sequential visits within each participant. Scoring methodology was similar to that used above for AP. Probesets that had a FC ≥ 1.2 were scored + 1 (increased in High SI) or -1 (decreased in High SI). FC ≥ 1.1 were scored +0.5 or -0.5. FC lower than 1.1 was considered no change. The only difference between the DE and the AP analyses was when scoring comparisons where there was no phene expression (SI) change between visits and no change in gene expression between visits (FC lower than 1.1). In that

case, the comparison received the same score as the nearest preceding comparison where there was a change in SI from visit to visit. If no preceding comparison with a change in SI was available, then it was given the same score as the nearest subsequent comparison where there was a change in SI. For DE also we used a perfection bonus and a non-tracking correction. If the gene expression perfectly tracked the suicidal ideation in a participant that had at least two comparisons (3 visits), that probeset was rewarded by a doubling of its score. If there was no change in gene expression in any of the comparisons for a particular participant, that overall score for that probeset in that participant was zero.

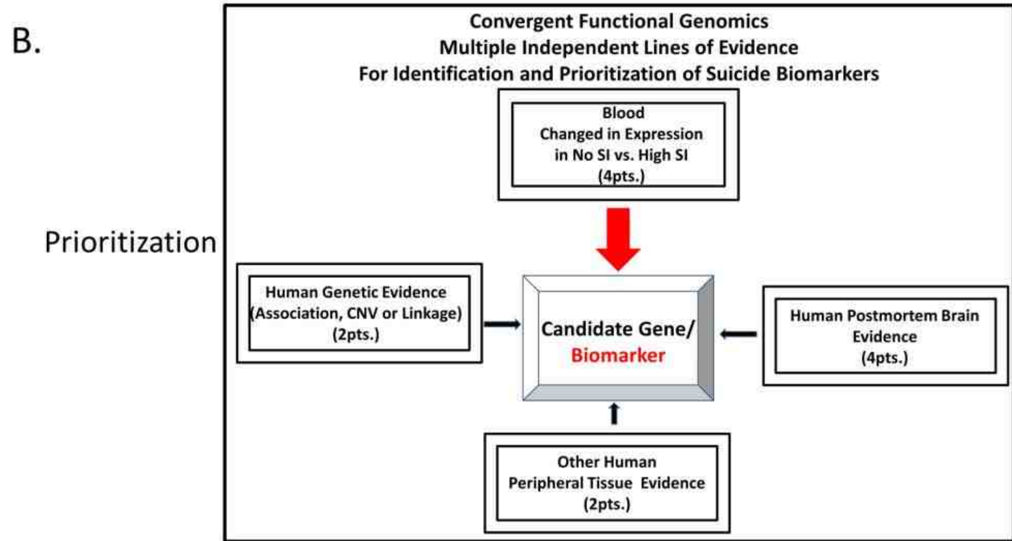
Internal score. Once scores within each participant were calculated, an algebraic sum across all participants was obtained, for each probeset. Probesets were then given internal points based upon these algebraic sum scores. Probesets with scores above the 33.3% of the distribution (for increased probesets and decreased probesets) received 1 point, those above 50% of the distribution received 2 points, and those above 80% of the distribution received 4 points. For AP analyses, we have 23 probesets which received 4 points, 581 probesets with 2 points, and 2077 probesets with 1 point, for a total of 2681 probesets. For DE analyses, we have 31 probesets which received 4 points, 1294 probesets with 2 points, and 5839 probesets with 1 point, for a total of 7164 probesets. The overlap between the two discovery methods is shown in Figure 5-3. Different probesets may be found by the two methods due to differences in scope (DE capturing genes

that are present in both visits of a comparison, i.e. PP, but are changed in expression), thresholds (what makes the 33.3% change cutoff across participants varies between methods), and technical detection levels (what is considered in the noise range varies between the methods).

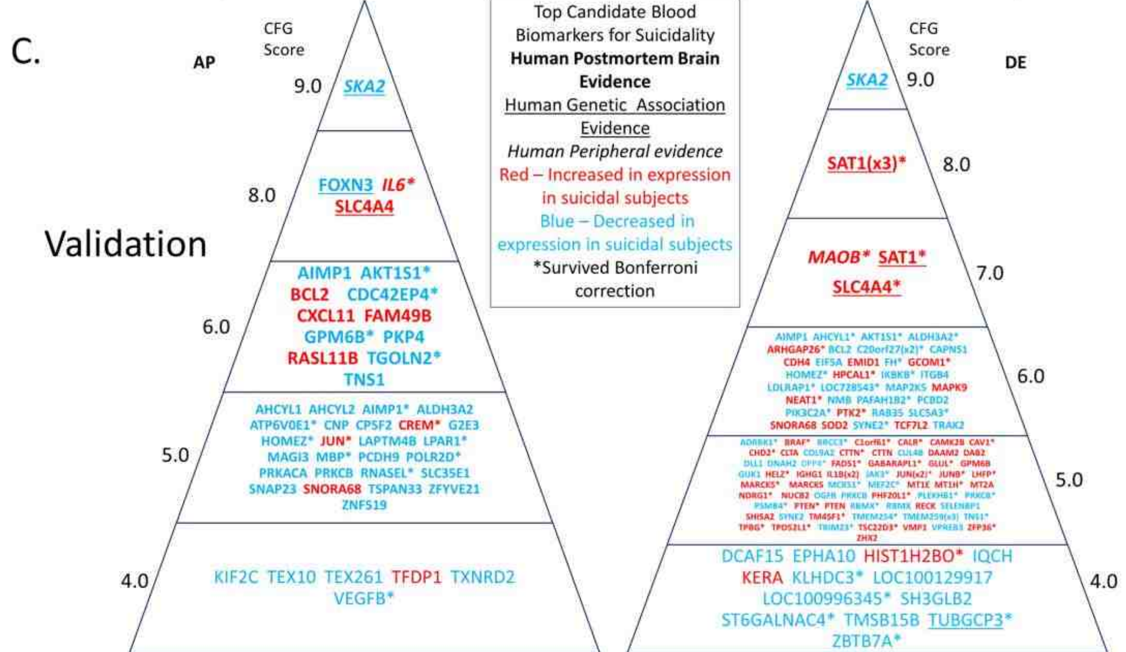
In total, we identified 9413 probesets with internal CFG score of 1. Gene names for the probesets were identified using NetAffyx (Affymetrix) and Partek for Affymetrix HG-U133 Plus 2.0 GeneChips, followed by GeneCards to confirm the primary gene symbol. In addition, for those probesets that were not assigned a gene name by NetAffyx or Partek, we used the UCSC Genome Browser to directly map them to known genes, with the following limitations: 1. in case the probeset fell in an intron, that particular gene was assumed to be implicated; 2. only one gene was assigned to each probeset. Genes were then scored using our manually curated CFG databases as described below (Figure 5-3).



A.
Discovery
AP
DE
1195
1054
212
220
2591
4141



B.
Prioritization
Convergent Functional Genomics
Multiple Independent Lines of Evidence
For Identification and Prioritization of Suicide Biomarkers
Blood
Changed in Expression
in No SI vs. High SI
(4pts.)
Human Genetic Evidence
(Association, CNV or Linkage)
(2pts.)
Candidate Gene/
Biomarker
Human Postmortem Brain
Evidence
(4pts.)
Other Human
Peripheral Tissue Evidence
(2pts.)



C.
Validation
AP
CFG
Score
9.0
SKA2
8.0
FOXN3 IL6*
SLC4A4
6.0
AIMP1 AKT1S1*
BCL2 CDC42EP4*
CXCL11 FAM49B
GPM6B* PKP4
RASL11B TGOLN2*
TNS1
5.0
AHCYL1 AHCYL2 AIMP1* ALDH3A2
ATP6V0E1* CNP CPSF2 CREM* G2E3
HOMEZ* JUN* LAPTM4B LPAR1*
MAGI3 MBP* PCDH9 POLR2D*
PRKACA PRKCB RNASEL* SLC35E1
SNAP23 SNORA68 TSPAN33 ZFYVE21
ZNF519
4.0
KIF2C TEX10 TEX261 TFDP1 TXNRD2
VEGFB*
Top Candidate Blood
Biomarkers for Suicidality
Human Postmortem Brain
Evidence
Human Genetic Association
Evidence
Human Peripheral evidence
Red – Increased in expression
in suicidal subjects
Blue – Decreased in
expression in suicidal subjects
*Survived Bonferroni
correction
DE
CFG
Score
SKA2
9.0
SAT1(x3)*
8.0
MAOB* SAT1*
SLC4A4*
7.0
6.0
5.0
DCAF15 EPHA10 HIST1H2BO* IQCH
KERA KLHDC3* LOC100129917
LOC100996345* SH3GLB2
ST6GALNAC4* TMSB15B TUBGCP3*
ZBTB7A*
4.0

Figure 5-3. Biomarker discovery, prioritization and validation. (a) Discovery—number of probesets carried forward from the absent–present and differential expression analyses, with an internal score of 1 and above. Red-increased in expression in high suicidal ideation, blue-decreased in expression in high suicidal ideation. (b) Prioritization—convergent functional genomics integration of multiple lines of evidence to prioritize suicide-relevant genes from the discovery step. (c) Validation—top convergent functional genomics genes, with a total score of 4 and above, validated in the cohort of suicide completers. All the genes shown were significantly changed in analysis of variance from no suicidal ideation to high suicidal ideation to suicide completers. *Survived Bonferroni correction. SAT1 (x3) had three different probesets with the same total score of 8.

Convergent Functional Genomics

As described in Universal Methods.

Pathway Analyses

IPA 9.0 (Ingenuity Systems, www.ingenuity.com, Redwood City, CA, USA), GeneGO MetaCore (Encinitas, CA), and Kyoto Encyclopedia of Genes and Genomes (KEGG) (through the Partek Genomics Suite 6.6 software package) were used to analyze the biological roles, including top canonical pathways, and diseases, of the candidate genes resulting from our work, as well as to identify genes in our dataset that are the target of existing drugs (Table 5-3 and S5-3). We ran the analyses together for all the AP and DE probesets with a total CFG score ≥ 4, then for those of them that showed stepwise change in the suicide completer's validation cohort, then for those of them that were nominally significant, and finally for those of them that survived Bonferroni correction.

Validation analyses

For validation of our candidate biomarker genes, we examined which of the top candidate genes (CFG score of 4 or above) were stepwise changed in expression from the No SI group to the High SI group to the suicide completers group. We used an empirical cutoff of 33.3% of the maximum possible CFG score of 12, which also permits the inclusion of potentially novel genes with maximal internal CFG score but no external CFG score. Statistical analyses were performed in SPSS using one-way ANOVA and Bonferroni corrections.

For the AP analyses, we imported the Affymetrix microarray data files from the participants in the validation cohort of suicide completers into MAS5, alongside the data files from the participants in the discovery cohort

For the DE analyses, we imported .cel files into Partek Genomic Suites. We then ran a RMA, background corrected with quantile normalization, and a median polish probeset summarization of all the chips from the validation cohort to obtain the normalized expression levels of all probesets for each chip. Partek normalizes expression data into a log base of 2 for visualization purposes. We non-log-transformed expression data by taking 2 to the power of the transformed expression value. We then used the non-log-transformed expression data to compare expression levels of biomarkers in the different groups (Figure S5-1).

Clinical measures

The Simplified Affective State Scale (SASS) is an 11 item scale for measuring mood and anxiety, previously developed and described by us as TASS (Total Affective State Scale) [277]. The SASS has a set of 11 visual analog scales (7 for mood, 4 for anxiety) that ends up providing a number ranging from 0 to 100 for mood state, and the same for anxiety state. We have now developed an Android app version (Figure S5-2).

Convergent Functional Information for Suicidality (CFI-S) (Table 5-4) is a new 22 item scale and Android app (Figure S5-2) for suicide risk, which integrates, in a simple binary fashion (Yes-1, No-0), similar to a polygenic risk score, information about known life events, mental health, physical health, stress, addictions, and

cultural factors that can influence suicide risk[289] [290]. For live psychiatric participants, the scale was administered at participant testing visits (n= 57), or scored based on retrospective electronic medical record information and Diagnostic Interview for Genetic Testing (DIGS) information (n=269). For suicide completers (n=35), the scale was based on answers provided by next of kin, and corroborated by coroner's office reports and medical record information. When information was not available for an item, it was not scored (NA).

Combining gene expression and clinical measures

The Universal Predictor for Suicide (UP-Suicide) construct was decided upon as part of our apriori study design to be broad- spectrum, and combine our top biomarkers from each step (discovery, prioritization, validation) with the phenomic (clinical) markers (SASS and CFI-S). That was our primary endpoint. Had we done it post-hoc with only the markers that showed the best predictive ability in our testing analyses, the results would be even better, but not independent.

Testing analyses

The test cohort for suicidal ideation and the test cohort for future hospitalizations analyses were assembled out of data that was RMA normalized by diagnosis. Phenomic (clinical) and gene expression markers used for predictions were z scored by diagnosis, to be able to combine different markers into panels and to avoid potential artefacts due to different ranges of phene expression and gene expression in different diagnoses. Markers were combined by computing the

average of the increased risk markers minus the average of the decreased risk markers. Predictions were performed using R-studio.

Predicting Suicidal Ideation. Receiver-operating characteristic (ROC) analyses between marker levels and suicidal ideation (SI) were performed by assigning participants with a HAMD-SI score of 0-1 into the no SI category, and participants with a HAMD-SI score of 2 and greater into the SI category. Additionally, ANOVA was performed between no (HAMD-SI 0), intermediate (HAMD-SI 1), and high SI participants (HAMD-SI 2 and above) and Pearson R (one-tail) was calculated between HAMD-SI scores and marker levels (Table 5B-5 and Figure 5-5).

Predicting Future Hospitalizations for Suicidality. We conducted analyses for hospitalizations in the first year following testing, on the participants for which we had at least a year of follow-up data. For each participant in the test cohort for future hospitalizations, the study visit with highest levels for the marker or combination of markers was selected as index visit (or with the lowest levels, in the case of decreased markers). ROC analyses between marker levels and future hospitalizations were performed based on assigning if participants had been hospitalized for suicidality (ideation, attempts) or not following the index testing visit. Additionally, a one tailed t-test with unequal variance was performed between groups of participants with and without hospitalizations for suicidality. Pearson R (one-tail) correlation was performed between hospitalization frequency (number of hospitalizations for suicidality divided by duration of follow-up) and biomarker score. We also conducted only the correlation analyses for

hospitalizations frequency for all future hospitalizations due to suicidality, beyond one year, as this calculation, unlike the ROC and t-test, accounts for the actual length of follow-up, which varied beyond one year from participant to participant.

Figures

Each figure in this chapter was completed by Daniel Levey and Helen Le-Niculescu. This work has been published. [196]

Table 5-2 Top biomarkers for suicidality from Discovery, Prioritization and Validation.

Gene symbol/ Gene Name	Probesets	**Discovery** (Change) Method/ Score	Prior human genetic evidence	Prior human brain expression evidence	Prior human peripheral expression evidence	**Prioritization** Total CFG Score For Suicide	**Validation** ANOVA p-value	**Comment**
SKA2 spindle and kinetochore associated	225686_at	**(D)** DE/1 AP/1	Suicide [220]	(D) PFC [220]	(I) Blood	**9**	0.006 0.027	Top Decreased BioM

complex subunit 2					Suicide [220]			In Prioritization from AP and DE
IL6 interleukin 6 (interferon, beta 2)	205207_at	**(I)** AP/2		(I) Prefrontal cortex (BA-10) [291] Hippocampus [292]	(I) CSF [293] [294] (D) Blood [295]	**8**	1.44E-08	Top Increased BioM in Validation from AP
SAT1 spermidine/spermine N1-acetyltransferase 1	213988_s_at 210592_s_at 230333_at 203455_s_at	**(I)** DE/2 DE/1	Suicide [241] [296]	(I) PFC BA46	(I) **Blood** [217]	**8**	1.08E-44 1.24E-40 6.93E-12 3.09E-38	Top Increased BioM in Prioritization from DE

								Top biomarker in our previous work
SLC4A4 solute carrier family 4 (sodium bicarbonate cotransporter), member 4	211494_s_at 210739_x_at	**(I)** AP/2 DE/1	Suicide [222]	(D) PFC BA46/10 [297]		**8**	5.84E-05 0.002	Top Increased BioM in Prioritization from AP
MAOB monoamine oxidase B	204041_at	**(I)** DE/1		(I) PFC [298]	(D) Blood [299]	**7**	8.11E-08	Top Pharmacological Target
JUN jun proto-oncogene	201464_x_at 213281_at	**(I)** DE/1 AP1		(D) HIP[300]		**5**	2.63E-51 1.02E-41	Top Increased BioM

	201466_s_at						2.21E-08	in Validation from DE
MARCKS myristoylated alanine-rich protein kinase C substrate	213002_at 201670_s_at	**(I)** DE/1		(I) HIP, PFC [301]	(I) **Blood**[217]	**5**	1.51E-06; 0.0004	A top biomarker in our previous study
MBP myelin basic protein	225408_at	**(D)** AP/1		(I) NAC [248]		**5**	6.74E-10	Top Decreased BioM in Validation from AP
PTEN phosphatase and tensin homolog	204053_x_at 222176_at	**(I)** DE/1		(I) PFC, HIP[259]; [302]	(I) **Blood**[217]	**5**	7.66E-17 0.0003	A top biomarker in our

								previou s study
CADM1 cell adhesion molecule 1	237259_ at	**(I)** DE/ 4				**4**	NC	Top Increas ed BioM in Discove ry from DE
CLIP4 CAP-GLY domain containing linker protein family, member 4	219944_ at	**(D)** DE/ 4				**4**	NC	Top Decreas ed BioM in Discove ry from DE
DTNA dystrobrevin, alpha	211493_ x_at	**(I)** AP/ 4				**4**	NC	Top Increas ed BioM

								in Discovery from AP
KIF2C kinesin family member 2C	211519_s_at	**(D)** AP/4				**4**	0.00056	Top Decreased BioM in Discovery from AP
KLHDC3 kelch domain containing 3	214383_x_at	**(D)** DE/4			(D) **Blood**[217]	**4**	1.57E-17	Top Decreased BioM in Validation from DE A top biomark

								er in our previou s study

Results

Discovery of biomarkers for suicidal ideation

We conducted whole-genome gene expression profiling in the blood samples from a longitudinally followed cohort of male participants with psychiatric disorders that predispose to suicidality. The samples were collected at repeated visits, 3–6 months apart. State information about suicidal ideation (SI) was collected from a questionnaire (HAMD) administered at the time of each blood draw (Table S5-1). Out of 217 psychiatric participants (with a total of 531 visits) followed longitudinally in our study, there were 37 participants that switched from a no SI (SI score of 0) to a high SI state (SI score of 2 and above) at different visits, which was our intended discovery group (Figure 5-2). We used a powerful within-participant design to analyze data from these 37 participants and their 106 visits. A within-participant design factors out genetic variability, as well as some medications, lifestyle, and demographic effects on gene expression, permitting identification of relevant signal with Ns as small as 1[14]. Another benefit of a within-participant design may be accuracy/consistency of self-report of psychiatric symptoms ('phene expression'), similar in rationale to the signal detection benefits it provides in gene expression. The number of participants that met our criteria and were analyzed is small, but comparable to those in human postmortem brain gene expression studies of suicide. We are indeed treating the blood samples as surrogate tissue for brains, with the caveat that they are not the real target organ. However, with the blood samples from live human participants we have the

advantages of in vivo accessibility, better knowledge of the mental state at the time of collection, less technical artifacts and especially of being able to do powerful within-participant analyses from visit to visit.

For discovery, we used two differential expression methodologies: Absent/Present (reflecting on/off of transcription), and Differential Expression (reflecting more subtle gradual changes in expression levels). The genes that tracked suicidal ideation in each participant were identified in our analyses. We used three thresholds for increased in expression genes and for decreased in expression genes: ≥ 33.3% (low), ≥50% (medium), and ≥ 80% (high) of the maximum scoring increased and decreased gene across participants. Such a restrictive approach was used as a way of minimizing false positives, even at the risk of having false negatives. For example, there were genes on each of the two lists, from AP and DE analyses, that had clear prior evidence for involvement in suicidality, such as OLR1[303,248] (32%) and LEPR[252, 217] (32%) for AP, and OPRM1[304, 305] (32%) and CD24 [248, 217] (33%) from DE, but were not included in our subsequent analyses because they did not meet our apriori set 33.3% threshold.

Prioritization of biomarkers based on prior evidence in the field

These differentially expressed genes were then prioritized using a Bayesian-like Convergent Functional Genomics (CFG) approach (Figure 5-3) integrating all the previously published human genetic evidence, postmortem brain gene expression evidence, and peripheral fluids evidence for suicide in the field available

at the time of our final analyses (September 2014).This is a way of identifying and prioritizing disease relevant genomic biomarkers, extracting generalizable signal out of potential cohort-specific noise and genetic heterogeneity. We have built in our lab manually curated databases of the psychiatric genomic and proteomic literature to date, for use in CFG analyses. The CFG approach is thus a de facto field-wide collaboration. We use in essence, in a Bayesian fashion, the whole body of knowledge in the field to leverage findings from our discovery data sets. Unlike our use of CFG in many previous studies, for the current one we did not use any animal model evidence, as there are to date no clear animal models of self-harm or suicidality published to date.

Validation of biomarkers for behavior in suicide completers

For validation in suicide completers, we used 412 genes that had a CFG score of 4 and above, from AP and DE, reflecting either maximum internal score from discovery or additional external literature cross-validating evidence. Out of these, 208 did not show any stepwise change in suicide completers (NC- non-concordant). As such, they may be involved primarily in ideation and not in behavior (Table S5-6). The remaining 204 genes (49.5%) had levels of expression that were changed stepwise from no suicidal ideation to high suicidal ideation to suicide completion. 143 of these genes (34.7%) were nominally significant, and 76 genes (18.4%) survived Bonferroni correction for multiple comparisons (Figures 5-3 and S5-1). These genes are likely involved in suicidal ideation *and*

suicidal behavior. (You can have suicidal ideation without suicidal behavior, but you cannot have suicidal behavior without suicidal ideation).

Selection of biomarkers for testing of predictive ability

For testing, we decided apriori to select the top scoring increased and decreased biomarkers from each step (discovery, prioritization, validation), so as to avoid potential false negatives in the prioritization step due to lack of prior evidence in the literature, or false negatives in validation step due to possible postmortem artefacts. The top scoring genes after the discovery step were DTNA and KIF2C from AP, CADM1 and CLIP4 from DE. The top genes after the prioritization with CFG step were SLC4A4 and SKA2 from AP, SAT1 and SKA2 from DE. The top genes after the validation in suicide completers step were IL6 and MBP from AP, JUN and KLHDC3 from DE (Figure 5-3). Notably, our SAT1 finding is a replication and expansion of our previously reported results identifying SAT1 as a blood biomarker for suicidality in bipolars (Le-Niculescu et al. 2013), and our SKA2 finding is an independent replication of a previous report identifying SKA2 as a blood biomarker for suicidality by Kaminsky and colleagues[220]. We also replicated in this larger cohort other top biomarkers from our previous work in bipolar disorder, notably MARCKS and PTEN (Table 5-2, Figure S5-4). A number of other genes we identified (CADM1, KIF2C, DTNA, CLIP4) are completely novel in terms of their involvement in suicidality.

Biological understanding

We also sought to understand the biology represented by the biomarkers identified by us, and derive some mechanistic and practical insights. We conducted: 1. unbiased biological pathway analyses and hypothesis driven mechanistic queries, 2. overall disease involvement and specific neuropsychiatric disorders queries, and 3. overall drug modulation along with targeted queries for omega-3, lithium and clozapine[306] (Tables 5-3, S5-3, S5-4). Administration of omega-3s in particular may be a mass- deployable therapeutic and preventive strategy[167].

The sets of biomarkers identified have biological roles in immune and inflammatory response, growth factor regulation, mTOR signaling, stress, and perhaps overall the switch between cell survival and proliferation vs. apoptosis (Tables 5-3 and S5-3). 14% of the candidate biomarkers in Table S5-3 have evidence for involvement in psychological stress response, and 19% for involvement in programmed cell death/ cellular suicide (apoptosis). An extrapolation can be made and model proposed whereas suicide is a whole body apoptosis (or "self-poptosis") in response to perceived stressful life events.

We also examined evidence for the involvement of these biomarkers for suicidality in other psychiatric disorders, permitting us to address issues of context and specificity (Table S5-3). SKA2, HADHA, SNORA68, RASL11B, CXCL11, HOMEZ, LOC728543, AHCYL1, LDLRAP1, NEAT1 and PAFAH1B2 seem to be relatively specific for

Table 5-3. Biological pathways and diseases

A.	**Ingenuity Pathways**				**KEGG Pathways**			**GeneGO Pathways**		
Prioritization on CFG score ≥4 (n = 412	**#**	**Top Canonical Pathways**	**P-Value**	**Ratio**	**Pathway Name**	**Enrichment Score**	**Enrichment p-value**	**Process Networks**	**Ratio**	**p-value**
	1	G-Protein Coupled Receptor Signaling	6.27E-14	10.6% 28/264	GABAergic synapse	10.8524	1.94E-05	Cell adhesion_ Amyloid proteins	27/195	4.78E-09
	2	cAMP-mediated signaling	2.04	10.3% 23/223	Amoebiasis	10.7231	2.20E-05	Reproduction_ Gonadotropin regulation	27/1	7.49E-09

genes)			E-11						99	
	3	Glucocorticoid Receptor Signaling	6.84E-11	8.9% 25/281	Melanogenesis	10.2992	3.37E-05	Reproduction_ GnRH signaling pathway	22/166	3.00E-07
	4	CREB Signaling in Neurons	1.01E-10	11.2% 20/179	Pathogenic Escherichia coli infection	9.03249	0.000119	Development_ Hedgehog signaling	28/254	3.65E-07
	5	Cardiac Hypertrophy Signaling	2.88E-10	9.5% 22/232	Chemokine signaling pathway	8.82088	0.000148	Cytoskeleton_ Regulation of cytoskeleton rearrangement	23/183	4.22E-07

		Top Canonical Pathways	**P-Value**	**Ratio**	**Pathway Name**	**Enrichment Score**	**Enrichment p-value**	**Process Networks**	**Ratio**	**p-value**
Non-Validated in Suicide Completers	1	G-Protein Coupled Receptor Signaling	2.28E-08	5.7% 15/264	Pathogenic Escherichia coli infection	7.19808	0.000748	Cytoskeleton_**Regulation of cytoskeleton rearrangement**	16/183	5.75E-07
	2	cAMP-mediated signaling	1.51E-07	5.8% 13/223	Amoebiasis	5.51218	0.004037	Development_**Neurogenesis_Axonal guidance**	17/230	2.65E-06

Stepwise (n=208 genes)	3	CREB Signaling in Neurons	6.20E-06	5.6% 10/179	Dorso-ventral axis formation	4.7856	0.008349	Development_ **Hedgehog signaling**	17/254	1.01E-05
	4	Cardiac Hypertrophy Signaling	1.02E-05	4.7% 11/232	Melanogenesis	4.31121	0.013417	Reproduction_ Progesterone signaling	14/214	8.25E-05
	5	Synaptic Long Term Potentiation	2.26E-05	6.3% 8/127	Influenza A	4.23564	0.014471	Cardiac development_ Wnt_beta-catenin, Notch, VEGF, IP3 and	11/150	0.0001819

								integrin signaling		
		Top Canonical Pathways	**P-Value**	**Ratio**	**Pathway Name**	**Enrichment Score**	**Enrichment p-value**	**Process Networks**	**Ratio**	**p-value**
Validation Stepwise in Suicid	1	B Cell Receptor Signaling	1.01E-08	7.2% 13/181	Focal adhesion	10.5307	2.67E-05	Signal transduction_ **WNT signaling**	19/177	8.10E-10
	2	Ovarian Cancer Signaling	3.31E-08	8.3% 11/133	Colorectal cancer	10.3054	3.35E-05	Cell cycle_ **G1-S Growth factor regulation**	18/195	2.62E-08

e Co mp let ers (n =2 04 ge ne s)	3	Glucocorticoid Receptor Signaling	3.97E-08	5.3% 15/281	GABAergic synapse	8.60276	0.000184	Reproduction_ **Gonadotropin regulation**	18/199	3.60E-08
	4	Colorectal Cancer Metastasis Signaling	4.00E-08	5.8% 14/241	mTOR signaling pathway	8.47678	0.000208	Reproduction_ GnRH signaling pathway	16/166	9.05E-08
	5	G12/13 Signaling	1.12E-07	8.5% 10/118	Chagas disease (American trypanosomiasis)	7.66796	0.000468	Neurophysiological process_ Transmission of nerve impulse	18/212	9.58E-08
		Top Canonical	**P-**	**Ratio**	**Pathway Name**	**Enrichment**	**Enrichment**	**Process Networks**	**Ra**	**p-val**

		Pathways	Value			Score	p-value		tio	ue
Validation Nominally significant In Suicide Comp	1	B Cell Receptor Signaling	1.95E-07	5.5% 10/181	Focal adhesion	8.91242	0.000135	Cell cycle_ G1-S Growth factor regulation	16/195	3.62E-09
	2	Cholecystokinin/Gastrin-mediated Signaling	3.18E-07	7.5% 8/106	mTOR signaling pathway	8.34274	0.000238	Inflammation_ Histamine signaling	14/213	6.04E-07
	3	G12/13 Signaling	7.25E-07	6.8% 8/118	Wnt signaling pathway	7.1443	0.000789	Signal transduction_ WNT signaling	12/177	2.99E-06

letters (n = 143 genes)	4	Glucocorticoid Receptor Signaling	1.46E-06	3.9% 11/281	Amphetamine addiction	6.67296	0.001265	Cell cycle_ G1-S Interleukin regulation	10/128	6.05E-06
	5	Ovarian Cancer Signaling	1.80E-06	6.0% 8/133	Neurotrophin signaling pathway	6.54296	0.00144	Cell adhesion_ Amyloid proteins	12/195	8.17E-06
		Top Canonical Pathways	**P-Value**	**Ratio**	**Pathway Name**	**Enrichment Score**	**Enrichment p-value**	**Process Networks**	**Ratio**	**p-value**
Validation	1	**B Cell Receptor Signaling**	2.38	3.9% 7/181	**mTOR signaling**	11.496	1.02E-05	Cell cycle_ **G1-S Growth**	9/1	2.34E-05

Bonferroni significant in Suicide Completers (n=76ge			E-06		**pathway**			**factor regulation**	95	
	2	G12/13 Signaling	2.70E-06	5.1% 6/118	Arginine and proline metabolism	8.02409	0.000327	Reproduction_ Gonadotropin regulation	8/199	1.82E-04
	3	IL-17A Signaling in Airway Cells	2.71E-06	7.6% 5/66	Focal adhesion	7.79535	0.000412	Inflammation_ IL-4 signaling	6/115	3.12E-04
	4	IL-8 Signaling	3.06E-06	3.7% 7/188	Pathways in cancer	7.05537	0.000863	Cell cycle_ G1-S Interleukin regulation	6/128	5.55E-04

ne s)	5	Integrin Signaling	4.76E-06	3.5% 7/201	Renal cell carcinoma	6.07809	0.002293	Inflammation_ IL-12,15,18 signaling	4/59	1.28E-03

suicide, based on the evidence to date in the field. SAT1, IL6 , FOXN3 and FKBP5 are less specific for suicide, having equally high evidence for involvement in suicide and in other psychiatric disorders, possibly mediating stress response as a common denominator[248,307]. These boundaries and understanding will likely change as additional evidence in the field accumulates. For example, CADM1, discovered in this work as a top biomarker for suicide, had previous evidence for involvement in other psychiatric disorders, such as autism and bipolar disorder. Interestingly, it was identified in a previous study by us as a blood biomarker increased in expression in low mood states in bipolar participants, and it is increased in expression in the current study in high suicidal ideation states. Increased expression of CADM1 is associated with decreased cellular proliferation and with apoptosis, and this gene is decreased in expression or silenced in certain types of cancers.

A number of other genes besides CADM1 are changed in opposite direction in suicide in this study vs. high mood in our previous mood biomarker study-CHD2, MBP, LPAR1, IGHG1, TEX261 (Table S5-3), suggesting that suicidal participants are in a low mood state. Also, some of the top suicide biomarkers are changed in expression in the same direction as in high psychosis participants in a previous psychosis biomarker study of ours -PIK3C2A, GPM6B, PCBD2, DAB2, IQCH, LAMB1, TEX261 (Table S5-3), suggesting that suicidal participants may be in a psychosis-like state. TEX261 in particular appears in all three studies, decreased in expression in suicide and high hallucinations, and increased in expression in high mood. This

protective marker may be an interesting target for future biological studies and drug development. Taken together, the data indicates that suicidality could be viewed as a psychotic dysphoric state, and that TEX261 may be a key biomarker reflecting that state. This molecularly informed view is consistent with the emerging clinical evidence in the field[308].

Lastly, we conducted biological pathway analyses on the genes that, after discovery and prioritization, were stepwise changed in suicide completers (n=204) and may be involved in ideation *and* behavior, vs. those that were not stepwise changed (n=208), and that may only be involved in ideation (Table S5-6). The genes involved in ideation map to pathways related to neuronal connectivity (cytoskeleton rearrangement, axonal guidance) and schizophrenia. The genes involved in behavior map to pathways related to neuronal activity (WNT, growth factors) and mood disorders. This is consistent with ideation being related to psychosis, and behavior being related to mood. Of note, clinically, the risk for suicide behavior/completion is higher in mood disorders than in psychotic disorders.

Clinical information

We also developed a simple new 22 item scale and app for suicide risk, Convergent Functional Information for Suicidality (CFI-S), which scores in a simple binary fashion and integrates, information about known life events, mental health, physical health, stress, addictions, and cultural factors that can influence suicide risk[289] [290]. Clinical risk predictors and scales are of high interest in the military[309]

and in the general population at large[310]. Our scale builds on those excellent prior achievements, while aiming for comprehensiveness, simplicity and quantification similar to a polygenic risk score. CFI-S is able to distinguish between individuals who committed suicide (coroner's cases n=35, information obtained from the next of kin) and those high risk participants who did not but had experienced changes in suicidal ideation (our discovery cohort of psychiatric participants) (Figure 5-4). We analyzed which items of the CFI-S scale were the most significantly different between high suicidal ideation live participants and suicide completers. We identified 7 items that were significantly different, 5 of which survived Bonferroni correction: lack of coping skills when faced with stress (p= 3.35E-11), dissatisfaction with current life (p=2.77E-06), lack of hope for the future (4.58E-05), current substance abuse (p=1.25E-04), and acute loss/grief (p= 9.45E-5). It is highly interesting that the top item was inability to cope with stress, which is independently consistent with our biological marker results.

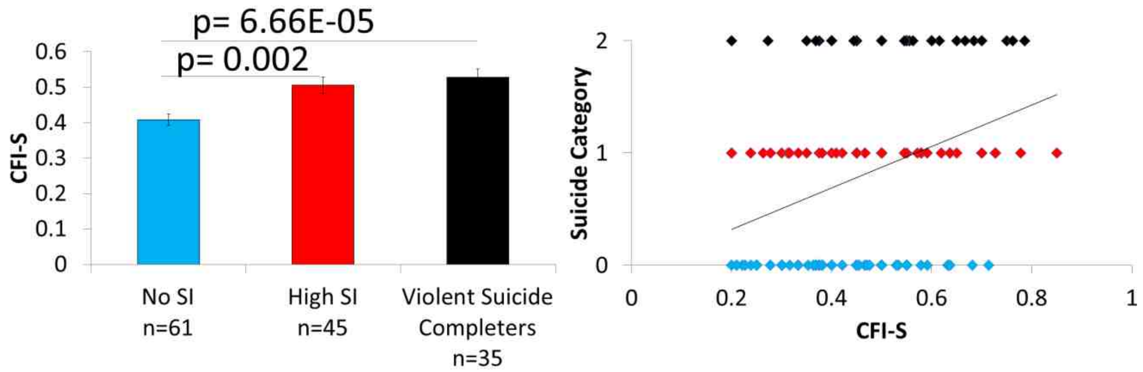
We also simplified the wording (and developed a new app for) an 11 item scale for measuring mood and anxiety, the Simplified Affective State Scale (SASS), previously developed and described by us as TASS (Total Affective State Scale) [277]. The SASS is a set of 11 visual analog scales (7 for mood, 4 for anxiety) that ends up providing a number ranging from 0 to 100 for mood state, and the same for anxiety state.

Table 4. Convergent Functional Information for Suicide (CFI-S) Scale

Items	*Yes*	*No*	*NA*	*Domain*	*Type* *Increased Reasons (IR)* *Decreased Barriers (DB)*
1. Psychiatric illness diagnosed and treated				Mental health	IR
2. With poor treatment compliance				Mental health	DB
3. Family history of suicide in blood relatives				Mental health	IR
4. Personally knowing somebody who committed suicide				Cultural factors	DB
5. History of abuse: physical, sexual, emotional, neglect				Life satisfaction	IR
6. Acute/severe medical illness, including acute pain ("I just can't stand this pain anymore.") (within last 3 months)				Physical health	IR
7. Acute stress: Losses, grief (within last 3 months)				Environmental stress	IR
8. Chronic stress: perceived uselessness, not feeling needed, burden to extended kin				Environmental stress	IR
9. History of excessive introversion, conscientiousness (including planned suicide attempts)				Mental health	IR
10. Dissatisfaction with life at this moment in time				Life satisfaction	IR
11. Lack of hope for the future				Life satisfaction	IR
12. Current substance abuse				Addictions	DB
13. Past history of suicidal acts/gestures				Mental health	DB
14. Lack of religious beliefs				Cultural factors	DB
15. Acute stress: Rejection (within last 3 months)				Environmental stress	IR
16. Chronic stress: lack of positive relationships, social isolation				Environmental stress	DB
17. History of excessive extroversion and impulsive behaviors (including rage, anger, physical fights, seeking revenge)				Mental health	DB
18. Lack of coping skills when faced with stress (cracks under pressure)				Mental health	DB
19. Lack of children. If has children, not in touch/not helping take care of them				Life satisfaction	DB
20. History of command hallucinations of self-directed violence				Mental health	IR
21. Age: older >60 or younger <25				Age	IR
22. Gender: male				Gender	DB

Abbreviations: CFI-S, Convergent Functional Information for Suicide; DB, decreased barrier; IR, increased reasons; NA, not available. Items are scored 1 for Yes, 0 for No. Total Score has a maximum possible of 22. Final Score (normalized) is Total Score divided by number of items that were scored, as for some items information might be NA, so they are not scored.

A. Validation of scale



Predictor	ANOVA	t-test (Completers vs. High SI)	Correlation R	Correlation p-value
CFI-S	6.66E-05	0.223	0.344	1.49E-05

B. Validation of items

Item	Description	P-Value (One-Way ANOVA) No SI vs High SI vs Completers	Stepwise	T-Test (two tailed) High SI vs Completers
18	**Lack of coping skills (cracks under pressure)**	**3.35E-11**	**Y**	**2.42E-05**
10	**Dissatisfaction with present life**	**2.77E-06**	**Y**	0.06804
11	**Lack of hope for the future**	**3.28E-05**	**Y**	**7.28E-05**
12	**Current substance abuse**	**0.000125**	**Y**	0.01273
7	**Acute stress: losses, grief**	**0.000945**	**Y**	0.07253
16	Chronic stress: lack of positive relationships, social isolation	0.0149	Y	0.2897
15	Acute stress: rejection	0.03	Y	0.02242
17	History of excessive extroversion and impulsive behaviors (including rage, anger, physical fights, seeking revenge)	0.0607	Y	0.2097
6	Acute/severe medical illness, pain	0.0892	Y	0.1113
19	Lack of children	0.365	Y	0.2479
22	Gender: Male	No females	No females	No females
4	Personally knowing somebody who committed suicide	No data for completers	No data for completers	No data for completers
1	Psychiatric illness diagnosed and treated	5.06E-15	N	6.02E-06
13	Past history of suicidal acts/gestures	2.40E-05	N	4.46E-05
21	Age: Older >60 or Younger <25	8.89E-05	N	0.000267
5	History of abuse: physical, sexual, emotional, neglect	0.0165	N	0.004361
20	History of command hallucinations of self-directed violence	0.0397	N	0.009659
3	Family history of suicide in blood relatives	0.0797	N	0.0242
2	With poor treatment compliance	0.147	N	0.07321
14	Lack of religious beliefs	0.117	N	0.06151
9	History of excessive introversion, conscientiousness	0.303	N	0.2439
8	Chronic stress: perceived uselessness, not feeling needed, burden to extended kin	0.42	N	0.2097

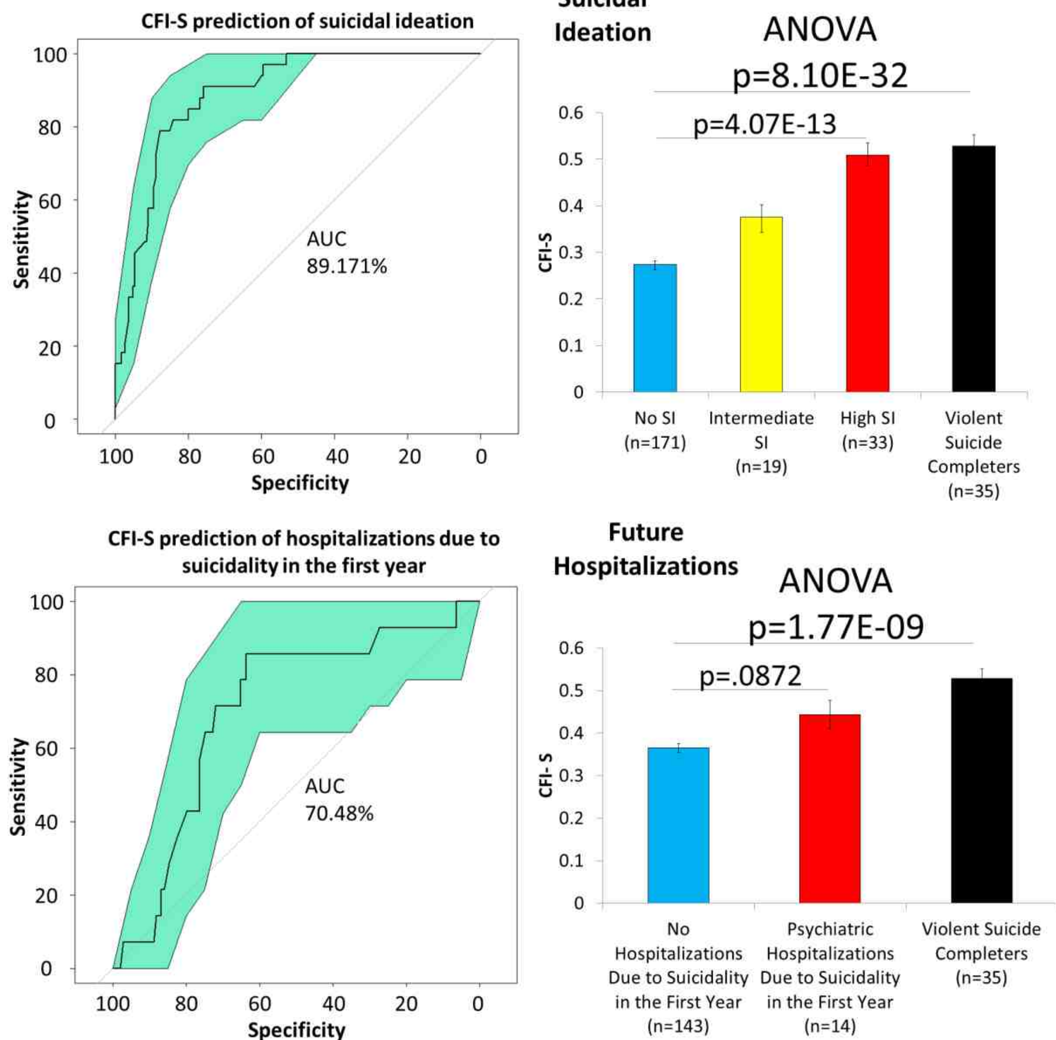


Figure 5-4. Convergent Functional Information for Suicide (CFI-S) Scale. (a) Validation of scale. Convergent Functional Information for Suicide levels in the discovery cohort and suicide completers. (b) Validation of items. Convergent Functional Information for Suicide was developed independently of any data from this study, by compiling known sociodemographic and clinical risk factors for suicide. It is composed of 22 items that assess the influence of mental health factors, as well as of life satisfaction, physical health, environmental stress, addictions, cultural factors known to influence suicidal behavior, and two

demographic factors, age and gender. These 22 items are shown here validated in the discovery cohort and suicide completers in a manner similar to that for biomarkers. Additionally, a student's t-test was used to evaluate items that were increased in suicide completers when compared to living participants with high suicidal ideation. (c) Predictions. Convergent Functional Information for Suicide predicting SI in the independent test cohort, and predicting future hospitalizations due to suicidality.

Testing for predictive ability

The best single biomarker predictor for suicidal ideation state across all diagnostic groups is SLC4A4 (ROC AUC 0.72, p-value 2.41E-05), the top increased biomarker from our prioritization with CFG of discovery data from AP (Table 5-5). Within diagnostic groups, the accuracy is even higher. SLC4A4 has very good accuracy at predicting future high suicidal ideation in bipolar participants (AUC 0.93, p-value 9.45E-06) and good accuracy in schizophrenia participants (AUC 0.76, p-value 0.030). SLC4A4 is a sodium-bicarbonate co-transporter that regulates intracellular pH, and possibly apoptosis. Very little is known to date about its roles in the brain, thus representing a completely novel finding. Brain pH has been reported by Wemmie and colleagues to play a role in pain, fear and panic attacks[311], which clinically share features with acute suicidal ideation states.

SKA2, the top decreased biomarker from prioritization with CFG of discovery data from AP and DE, has good accuracy at predicting suicidal ideation across all diagnostic groups (AUC 0.69, p-value 0.00018), and even better accuracy in bipolar participants (AUC 0.76, p-value 0.0045) and schizophrenia participants (AUC 0.82, p-value 0.011).

The best single biomarker predictor for future hospitalizations for suicidal behavior in the first year across all diagnostic groups was SAT1, the top increased biomarker from the prioritization with CFG of discovery data from DE (AUC 0.55, p-value 0.28). The results across all diagnoses are modest, likely due to the significant variation of markers by diagnostic group (Table 5-5 and Figure S5-4).

This seems to be even more of an issue for trait than for state predictions. Within diagnostic groups, in bipolar disorder, the SAT1 prediction accuracy for future hospitalizations is higher (AUC 0.63, p-value 0.18), consistent with our previous work[217]. CADM1 (AUC 0.72, p-value 0.076), SKA2 (AUC 0.71, p-value 0.056), and SLC4A4 (AUC 0.70, p-value 0.08) are even better predictors than SAT1 in bipolar disorder.

CFI-S has very good accuracy (AUC 0.89, p-value 3.53E-13) at predicting suicidal ideation in psychiatric participants across diagnostic groups (Figure 4C-5). Within diagnostic groups, in affective disorders, the accuracy is even higher. CFI-S has excellent accuracy at predicting high suicidal ideation in bipolar participants (AUC 0.97, p-value 1.75E-06) and in depression participants (AUC 0.95, p-value 7.98E-06). CFI-S has good accuracy (AUC 0.71, p-value 0.006) at predicting future hospitalizations for suicidality in the first year, across diagnostic groups.

SASS has very good accuracy (AUC 0.85, 9.96E-11) at predicting suicidal ideation in psychiatric participants across diagnostic groups. Within diagnostic groups, in bipolar disorder, the accuracy is even higher (AUC 0.87, p-value 0.00011). SASS also has good accuracy (AUC 0.71, p-value 0.008) at predicting future hospitalizations for suicidality in the first year following testing.

Our apriori primary endpoint was a combined universal predictor for suicide (UP-Suicide), composed of the top increased and decreased biomarkers (n=11) from the discovery for ideation (CADM1, CLIP4, DTNA, KIF2C), prioritization with CFG for prior evidence (SAT1, SKA2, SLC4A4), and validation for behavior in

suicide completers (IL6, MBP, JUN, KLHDC3) steps, along with CFI-S, and SASS. UP-Suicide is an excellent predictor of suicidal ideation across all disorders in the independent cohort of psychiatric

Table 5. Predictions

A. Best predictors

Predictors ROC AUC/P-value	*All participants*	*BP participants*	*MDD participants*	*SZA participants*	*SZ participants*
Suicidal ideation cohort $N=108$ participants	UP-Suicide 0.92/7.94e − 15	UP-Suicide 0.98/ 1.19E-6	UP-Suicide 0.95/2.96E-7	UP-Suicide 0.81/0.0018	Mood 0.94/0.00075 UP-Suicide 0.91/0.0015
First year hospitalizations for suicidality cohort $N=157$ participants	SASS 0.71/0.0080 UP-Suicide 0.71/0.0094	SASS 0.95/0.0016 UP-Suicide 0.94/0.0021	CFI-S 0.78/0.066 UP-Suicide 0.70/0.16	Anxiety 0.65/0.21 UP-Suicide 0.52/0.47	UP-Suicide 0.68/0.17

B. All predictions

Predictors ROC AUC/P-value		*All participants* *No SI = 73* *Intermediate SI = 12* *High SI = 23*	*BP participants* *No SI = 17* *Intermediate SI = 5* *High SI = 7*	*MDD participants* *No SI = 17* *Intermediate SI = 0* *High SI = 8*	*SZA participants* *No SI = 19* *Intermediate SI = 3* *High SI = 6*	*SZ participants* *No SI = 20* *Intermediate SI = 4* *High SI = 2*
Suicidal ideation cohort $N=108$ participants	Biomarkers	SKA2 0.69/0.00018	SKA2 0.76/0.0045	SKA2 0.54/0.34	SKA2 0.68/0.06	SKA2 0.82/0.011
		SLC4A4 0.72/2.41E-5	SLC4A4 0.93/9.45E-6	SLC4A4 0.55/0.33	SLC4A4 0.64/0.11	SLC4A4 0.76/0.03
		KIF2C 0.42/0.92	KIF2C 0.33/0.96	KIF2C 0.52/0.45	KIF2C 0.41/0.78	KIF2C 0.43/0.71
		DTNA 0.54/0.22	DTNA 0.61/0.15	DTNA 0.53/0.41	DTNA 0.53/0.41	DTNA 0.45/0.66
		MBP 0.53/0.30	MBP 0.54/0.35	MBP 0.61/0.15	MBP 0.43/0.74	MBP 0.58/0.28
		IL6 0.66/0.0017	IL6 0.66/0.06	IL6 0.76/0.0057	IL6 0.58/0.24	IL6 0.62/0.19
		SAT1 0.35/1	SAT1 0.19/1	SAT1 0.39/0.86	SAT1 0.48/0.59	SAT1 0.37/0.84
		CLIP4 0.52/0.37	CLIP4 0.76/0.0050	CLIP4 0.21/1	CLIP4 0.54/0.38	CLIP4 0.61/0.21
		CADM1 0.59/0.045	CADM1 0.73/0.013	CADM1 0.63/0.11	CADM1 0.48/0.56	CADM1 0.49/0.54
		KLHDC3 0.47/0.72	KLHDC3 0.52/0.41	KLHDC3 0.47/0.60	KLHDC3 0.38/0.86	KLHDC3 0.49/0.53
		JUN 0.46/0.76	JUN 0.39/0.86	JUN 0.54/0.37	JUN 0.54/0.38	JUN 0.37/0.84
		BIOM6 0.64/0.0042	BIOM6 0.69/0.028	BIOM6 0.72/0.017	BIOM6 0.61/0.18	BIOM6 0.49/0.55
		BIOM5 0.54/0.23	BIOM5 0.69/0.029	BIOM5 0.44/0.73	BIOM5 0.44/0.69	BIOM5 0.61/0.21
		BIOM11 0.63/0.0088	BIOM11 0.75/0.0070	BIOM11 0.57/0.26	BIOM11 0.51/0.46	BIOM11 0.64/0.16
	Clinical	Anxiety 0.78/2.3E-7	Anxiety 0.86/0.00018	Anxiety 0.81/0.0015	Anxiety 0.75/0.12	Anxiety 0.62/0.19
		Mood 0.82/1.62E-9	Mood 0.81/0.00091	Mood 0.81/0.0015	Mood 0.77/0.0080	Mood 0.94/0.00075
		SASS 0.85/9.96E-11	SASS 0.87/0.00011	SASS 0.87/6.01E-5	SASS 0.81/0.0019	SASS 0.85/0.0058
		CFI-S 0.89/3.53E-13	CFI-S 0.97/ 1.75E-6	CFI-S 0.95/7.98E-6	CFI-S 0.74/0.016	CFI-S 0.85/0.0049
	Combined	UP-Suicide 0.92/ 7.94E-15	UP-Suicide 0.98/ 1.19E-6	UP-Suicide 0.95/ 2.96E-7	UP-Suicide 0.81/ 0.0018	UP-Suicide 0.91/ 0.0015

Predictors ROC AUC/P-value		*All participants* *No Hosp = 139* *Hosp = 18*	*BP participants* *No Hosp = 43* *Hosp = 7*	*MDD participants* *No Hosp = 20* *Hosp = 3*	*SZA participants* *No Hosp = 41* *Hosp = 3*	*SZ participants* *No Hosp = 35* *Hosp = 5*
First year hospitalizations for suicidality $N=157$ participants	Biomarkers	SKA2 0.44/0.78	SKA2 0.71/0.056	SKA2 0.048/0.99	SKA2 0.41/0.70	SKA2 0.13/0.99
		SLC4A4 0.47/0.66	SLC4A4 0.70/0.08	SLC4A4 0.048/0.99	SLC4A4 0.37/0.78	SLC4A4 0.39/0.74
		KIF2C 0.54/0.30	KIF2C 0.59/0.26	KIF2C 0.45/0.61	KIF2C 0.42/0.67	KIF2C 0.67/0.19
		DTNA 0.44/0.77	DTNA 0.61/0.21	DTNA 0.29/0.83	DTNA 0.13/0.96	DTNA 0.46/0.61
		MBP 0.38/0.92	MBP 0.30/0.90	MBP 0.42/0.68	MBP 0.29/0.88	MBP 0.53/0.46
		IL6 0.48/0.60	IL6 0.45/0.65	IL6 0.76/0.090	IL6 0.28/0.90	IL6 0.55/0.41
		SAT1 0.55/0.28	SAT1 0.63/0.18	SAT1 0.62/0.29	SAT1 0.37/0.76	SAT1 0.44/0.63
		CLIP4 0.31/0.99	CLIP4 0.26/0.91	CLIP4 0.25/0.92	CLIP4 0.31/0.87	CLIP4 0.41/0.69
		CADM1 0.53/0.36	CADM1 0.72/0.076	CADM1 0.74/0.17	CADM1 0.048/0.99	CADM1 0.56/0.39
		KLHDC3 0.31/0.98	KLHDC3 0.41/0.72	KLHDC3 0.31/0.81	KLHDC3 0.29/0.88	KLHDC3 0.16/0.95
		JUN 0.40/0.89	JUN 0.36/0.85	JUN 0.37/0.77	JUN 0.58/0.35	JUN 0.30/0.90
		BIOM6 0.51/0.46	BIOM6 0.62/0.23	BIOM6 0.68/0.18	BIOM6 0.14/0.96	BIOM6 0.43/0.63
		BIOM5 0.35/0.96v	BIOM5 0.50/0.52	BIOM5 0.24/0.88	BIOM5 0.28/0.90	BIOM5 0.40/0.72
		BIOM11 0.42/0.82	BIOM11 0.63/0.24	BIOM11 0.48/0.55	BIOM11 0.23/0.94	BIOM11 0.40/0.72
	Clinical	Anxiety 0.64/0.066	Anxiety 0.69/0.14	Anxiety 0.52/0.48	Anxiety 0.65/0.21	Anxiety 0.58/0.34
		Mood 0.58/0.16	Mood 0.70/0.059	Mood 0.60/0.32	Mood 0.45/0.63	Mood 0.5/0.51
		SASS 0.71/0.0080	SASS 0.95/0.0016	SASS 0.77/0.083	SASS 0.59/0.31	SASS 0.63/0.25
		CFI-S 0.71/0.0058	CFI-S 0.86/0.01	CFI-S 0.78/0.066	CFI-S 0.75/0.12	CFI-S 0.54/0.40
	Combined	UP-Suicide 0.71/ 0.0094	UP-Suicide 0.94/ 0.0021	UP-Suicide 0.7/0.16	UP-Suicide 0.52/0.47	UP-Suicide 0.68/0.17

Abbreviations: AUC, area under curve; BP, bipolar; CFI-S, convergent functional information for suicide; MDD, major depressive disorder; ROC, receiver operating characteristic; SASS, simplified affective state scale; SI, suicidal ideation; SZA, schizoaffective; SZ, schizophrenia; UP, universal predictive measure. ROC AUC/*P*-values. UP-Suicide is composed of increased markers (CFI-S, anxiety, BioM-6 panel of increased biomarkers) and decreased markers (mood, BioM-5 panel of decreased biomarkers); SASS is composed of increased marker (anxiety), and decreased marker (mood).

participants (AUC 0.92, p-value 7.94E-15) (Figure 5-6). UP-Suicide also has good predictive ability for future psychiatric hospitalizations for suicidality in the first year of follow-up (AUC 0.71, p-value 0.0094). The predictive ability of UP-Suicide is notably higher in affective disorder participants (bipolar, depression) (Table 5-5 and Figure 5-5).

Discussion

We carried out systematic studies to identify clinically useful predictors for suicide. Our work focuses on identifying markers involved in suicidal ideation *and* suicidal behavior, including suicide completion. Markers involved in behavior may be on a continuum with some of the markers involved in ideation, varying in the degree of expression changes from less severe (ideation) to more severe (behavior). One cannot have suicidal behavior without suicidal ideation, but it may be possible to have suicidal ideation without suicidal behavior.

As a first step, we sought to use a powerful but difficult to conduct within-participant design for discovery of blood biomarkers. Such a design is more informative than case-control, case-case, or even identical twins designs. The power of a within-participants longitudinal design for multi-omic discovery was first illustrated by Snyder and colleagues[14] in a landmark paper with an n=1. We studied a cohort of male participants with major psychiatric disorders (n=217 participants) followed longitudinally (2 to 6 testing visits, at 3 to 6 months interval). In a smaller (n=37) but very valuable subset of these participants, we captured

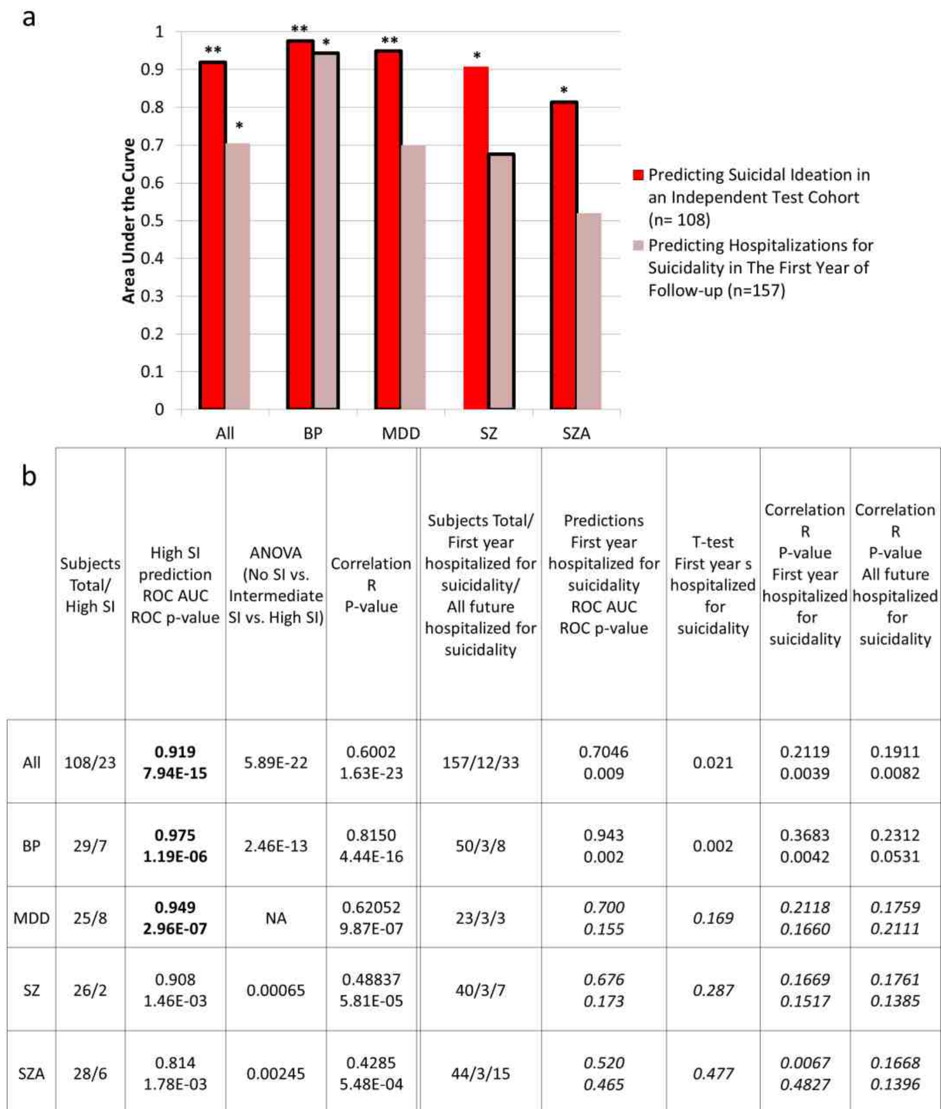
one or more major switches from a no suicidal ideation state to a high suicidal ideation state at the time of the different testing visits (Figure 5-1 and 5-2).

Second, we conducted whole-genome gene expression discovery studies in the participants that exhibited the switches, using a longitudinal within-participant design, which factors out genetic variability and reduces environmental variability as well. We have demonstrated the power of such a design in our previous work on suicide biomarkers with an n=9[217]. Our current n=37 was four-fold higher, and consequently our power to detect signal was commensurately increased (Figure 5-2). Genes whose levels of expression tracked suicidal ideation within each participant were identified.

Third, the lists of top candidate biomarkers for suicidal ideation from the discovery and prioritization step (genes with a CFG score of 4 and above, reflecting genes that have maximal experimental internal evidence from this study and/or additional external literature cross-validating evidence), were additionally validated for involvement in suicidal behavior in a cohort of demographically matched suicide completers from the coroner's office (n=26) (Figure 5-3).

Given that we used two methods (AP, DE), three steps (discovery for ideation, prioritization based on literature evidence, validation for behavior in completers), and two types of markers (increased, decreased), we anticipated having 2 x 3 x 2 =12 top markers. We ended up with 11 due to overlap (Table 5-2). Of note, 8 of these 11 markers (SAT1, SKA2, SLC4A4, KIF2C, MBP, IL6, JUN, KLHDC3), were significant in validation for behavior in terms of being changed

even more in suicide completers, and 5 of them survived Bonferroni correction (SAT1, SLC4A4, MBP, IL6, KLHDC3). The 3 out of 11 markers that were not validated for behavior (DTNA, CLIP4 and CADM1) seemed indeed better in the independent test cohorts at predicting suicidal ideation than at predicting suicidal behavior (hospitalizations) (Table 5-5B).



	Subjects Total/ High SI	High SI prediction ROC AUC ROC p-value	ANOVA (No SI vs. Intermediate SI vs. High SI)	Correlation R P-value	Subjects Total/ First year hospitalized for suicidality/ All future hospitalized for suicidality	Predictions First year hospitalized for suicidality ROC AUC ROC p-value	T-test First year s hospitalized for suicidality	Correlation R P-value First year hospitalized for suicidality	Correlation R P-value All future hospitalized for suicidality
All	108/23	**0.919 7.94E-15**	5.89E-22	0.6002 1.63E-23	157/12/33	0.7046 0.009	0.021	0.2119 0.0039	0.1911 0.0082
BP	29/7	**0.975 1.19E-06**	2.46E-13	0.8150 4.44E-16	50/3/8	0.943 0.002	0.002	0.3683 0.0042	0.2312 0.0531
MDD	25/8	**0.949 2.96E-07**	NA	0.62052 9.87E-07	23/3/3	*0.700 0.155*	*0.169*	*0.2118 0.1660*	*0.1759 0.2111*
SZ	26/2	0.908 1.46E-03	0.00065	0.48837 5.81E-05	40/3/7	*0.676 0.173*	*0.287*	*0.1669 0.1517*	*0.1761 0.1385*
SZA	28/6	0.814 1.78E-03	0.00245	0.4285 5.48E-04	44/3/15	*0.520 0.465*	*0.477*	*0.0067 0.4827*	*0.1668 0.1396*

Figure 5-5. Testing of universal predictor for suicide (UP-Suicide). UP-Suicide is a combination of our best gene expression biomarkers (top increased and decreased biomarkers from discovery, prioritization by CFG, and validation in suicide completers steps), and phenomic data (CFI-S and SASS). (a) Area Under the Curve (AUC) for the UP-Suicide predicting suicidal ideation and hospitalizations within the first year in all participants, as well as separately in bipolar (BP), major depressive disorder (MDD), schizophrenia (SZ), and schizoaffective (SZA)

participants. **Indicates the comparison survived Bonferroni correction for multiple comparisons. *Indicates nominal significance of Po0.05. Bold outline indicates that the UP-Suicide was synergistic to its components, i.e., performed better than the gene expression biomarkers or phenomic data individually. (b) Table containing descriptive statistics for all participants together, as well as separately in BP, MDD, SZ, and SZA. Bold indicates the measure survived Bonferroni correction for 200 comparisons (20 genomic and phenomic markers/combinations × 2 testing cohorts for SI and future hospitalizations in the first year × 5 diagnostic categories–all, BP, MDD, SZA, SZ). We also show Pearson correlation data in the suicidal ideation test cohort for HAMD-SI vs. UP-Suicide, as well as Pearson correlation data in the hospitalization test cohort for frequency of hospitalizations for suicidality in the first year, and for frequency of hospitalizations for suicidality in all future available follow-up interval (which varies among participants, from 1 year to 8.5 years).

Fourth, we describe a novel, simple and comprehensive phenomic (clinical) risk assessment scale, the Convergent Functional Phenomics for Suicidality (CFI-S) scale, as well as a companion app to it for use by clinicians and individuals (Figure S5-2). CFI-S was developed independently of any data from this study, by integrating known risk factors for suicide from the clinical literature. It has a total of 20 items (scored in a binary fashion- 1 for present, 0 for absent, NA for information not available) that assess the influence of mental health factors, as well as of life satisfaction, physical health, environmental stress, addictions, and cultural factors known to influence suicidal behavior. It also has 2 demographics risk factors items: age and gender. The result is a simple polyphenic risk score with an absolute range of 0 to 22, normalized by the number of items on which we had available information, resulting in a score in the range from 0 to 1 (Table 5-4). We present data validating the CFI-S in our discovery cohort of live psychiatric participants and in suicide completers from the coroner's office (Figure 5-4). We acknowledge the possibility of a potential upward bias in next-of-kin reporting post-suicide completion, although each item of the scale was scored factually by a trained rater on its own merits. We believe it is still illustrative and informative to compare the CFI-S in live participants with ideation vs. suicide completers, and identify which items are most different (such as inability to cope with stress, which is consistent with biological data from the biomarker side of our study).

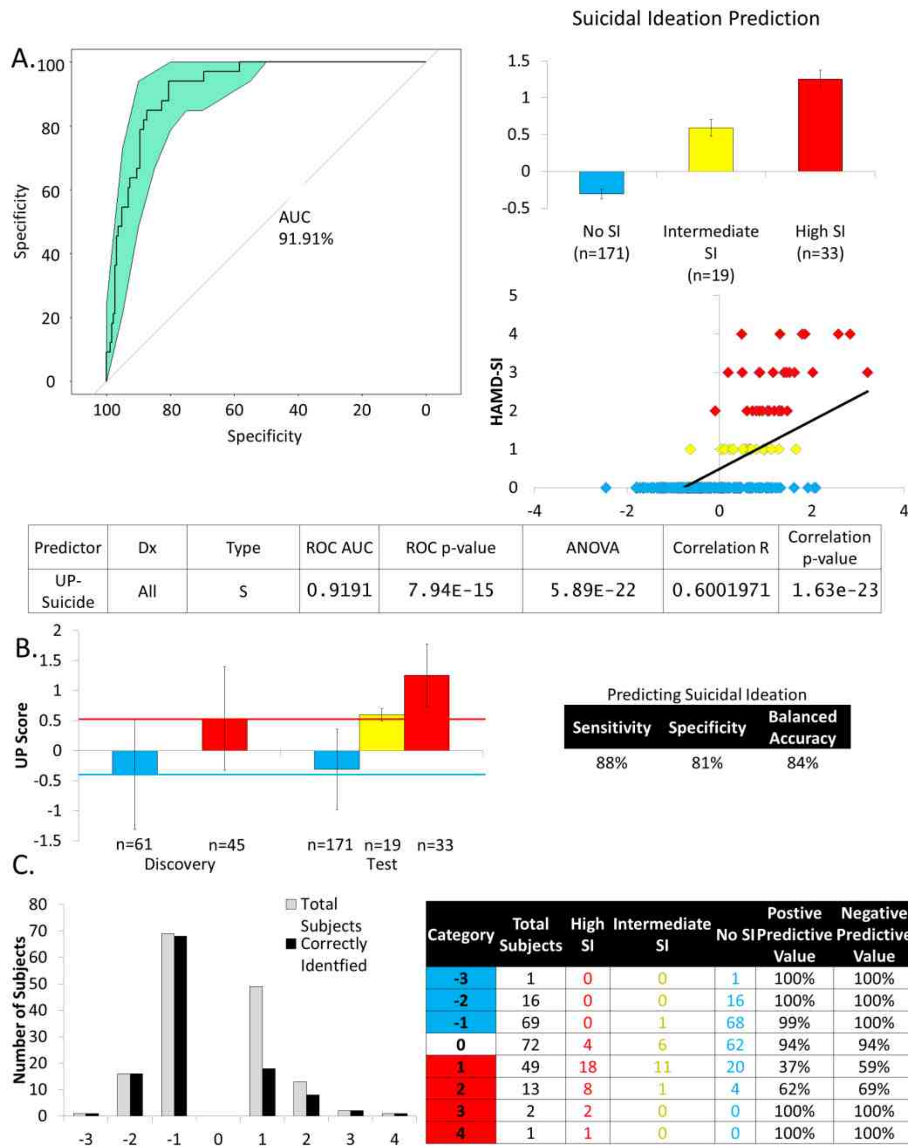
Fifth, we have also assessed anxiety and mood, using a visual analog Simplified Affective State Scale (SASS), previously described by us (Niculescu et al. 2006), for which we now have developed an app version (Figure S5-2). Using a PhenoChipping approach[277] in our discovery cohort of psychiatric participants, we show that anxiety measures cluster with suicidal ideation and CFI-S, and mood measures are in the opposite cluster, suggesting that our participants have high suicidal ideation when they have high anxiety and low mood (Figure 5-2). We would also like to include in the future measures of psychosis, and of stress, to be more comprehensive.

Sixth, we examined how the biomarkers identified by us are able to predict *state* (suicidal ideation) in a larger independent cohort of psychiatric participants (n= 108 participants).

Seventh, we examined whether the biomarkers are able to predict *trait* (future hospitalizations for suicidal behavior) in psychiatric participants (n=157) in the short term (first year of follow-up) as well as overall (all data for future hospitalizations available for each patient).

Last but not least, we demonstrate how our apriori primary endpoint, a comprehensive universal predictor for suicide (UP-Suicide), composed of the combination of the top increased and decreased biomarkers (n=11) from the discovery, prioritization and validation steps, along with CFI-S, and SASS, predicts state (suicidal ideation) and trait (future psychiatric hospitalizations for suicidality).

Figure 6 Prediction of Suicidal Ideation by UP-Suicide



Predictor	Dx	Type	ROC AUC	ROC p-value	ANOVA	Correlation R	Correlation p-value
UP-Suicide	All	S	0.9191	7.94E-15	5.89E-22	0.6001971	1.63e-23

Sensitivity	Specificity	Balanced Accuracy
88%	81%	84%

Category	Total Subjects	High SI	Intermediate SI	No SI	Postive Predictive Value	Negative Predictive Value
-3	1	0	0	1	100%	100%
-2	16	0	0	16	100%	100%
-1	69	0	1	68	99%	100%
0	72	4	6	62	94%	94%
1	49	18	11	20	37%	59%
2	13	8	1	4	62%	69%
3	2	2	0	0	100%	100%
4	1	1	0	0	100%	100%

Figure 5-6. Prediction of suicidal ideation by universal predictive measure-suicide. (a) (top left) Receiver-operating curve identifying participants with suicidal ideation against participants with no suicidal ideation or intermediate SI. (top right) Y axis contains the average UP-Suicide scores with standard error of mean for no suicidal ideation, intermediate suicidal ideation and high suicidal ideation. (bottom right) Scatter plot depicting HAMDSI score on the Y axis and universal predictive

measure-suicide score on the X axis with linear trend line. (bottom) Table summarizing descriptive statistics. Analysis of variance was performed between groups with no suicidal ideation, intermediate suicidal ideation and high suicidal ideation. (b) Predictions in test cohort based on thresholds in the discovery cohort - average UP-Suicide scores with standard deviation. (c) Number of participants correctly identified in the test cohort by categories based on thresholds in the discovery cohort. Category 1 means within 1 s.d. above the average of high suicidal ideation participants in the discovery cohort, category 2 means between 1 and 2 s.d. above, and so on. Category 1 means within 1 s.d. below the average of the no suicidal ideation participants in the discovery cohort, category 2 means between 1 and 2 s.d. below and so on.

The rationale for identifying blood biomarkers as opposed to brain biomarkers is a pragmatic one- the brain cannot be readily accessed in live individuals. Other peripheral fluids, such as CSF, require more invasive and painful procedures. Nevertheless, it is likely that many of the peripheral blood transcriptomic changes are not necessarily mirroring what is happening in the brain, and vice-versa. The keys to finding peripheral biomarkers[288] are, first, to have a powerful discovery approach, such as our within-participant design, that closely tracks the phenotype you are trying to measure and reduces noise. Second, cross-validating and prioritizing the results with other lines of evidence, such as brain gene expression and genetic data, are important in order to establish relevance and generalizability of findings. Third, it is important to validate for behavior in an independent cohort with a robust and relevant phenotype, in this case suicide completers. Fourth, testing for predictive ability in independent/prospective cohorts is a must.

Biomarkers that survive such a rigorous step-wise discovery, prioritization, validation and testing process are likely directly relevant to the disorder studied. As such, we endeavored to study their biology, whether they are involved in other psychiatric disorders or are relatively specific for suicide, and whether they are the modulated by existing drugs in general, and drugs known to treat suicidality in particular. We have identified a series of biomarkers that seem to be changed in opposite direction in suicide vs. in treatments with omega-3 fatty acids, lithium,

clozapine, or MAOIs. These biomarkers could potentially be used to stratify patients to different treatment approaches, and monitor their response (Supplementary Table S5-4).

We also conducted predictive studies, across all participants and by diagnosis, as a way of assessing how generalizable and how particular to a diagnosis biomarkers are. Different diagnostic groups have different disease biology and are on different medications, which may modify the levels of the biomarkers. We observe a significant variation in the predictive ability of biomarkers by diagnosis, which has important practical applications for future work on diagnostic-specific predictors (Table 5-5). Of note, a number of biomarkers from the current larger study reproduce our previous work in a smaller, bipolar cohort (SAT1, MARCKS, PTEN, as well as FOXN3, GCOM1, RECK, IL1B, LHFP, ATP6V0E1, and KLHDC3) (Table S5-2). In the current datasets, we have also post-hoc carried out biomarker discovery within each diagnosis, which revealed a diversity of top markers, but should be interpreted with caution given the smaller N within each diagnostic group (Table S5-5).

Prior to any testing, we planned to use a comprehensive combination of genomic data (specifically, the top increased and decreased biomarkers from discovery, prioritization, and validation) and phenomic data (specifically, the CFI-S and the SASS) as the primary endpoint measure, a broad-spectrum universal predictor (UP-Suicide) for state suicidal ideation and trait future hospitalizations. It has not escaped our attention that certain single biomarkers, particular

phenotypic items, or combinations thereof seem to perform better than the UP-Suicide in one or another type of prediction or diagnostic group (see Table 5-5). However, since such markers and combinations were not chosen by us apriori and such insights derive from testing, we cannot exclude a fit to cohort effect for them and reserve judgement as to their robustness as predictors until further testing in additional independent cohorts, by us and others. What we can put forward for now based on the current work is the UP-Suicide, which seems to be a robust predictor across different scenarios and diagnostic groups.

Overall, our predictive ability for trait future hospitalizations is somewhat less than for state suicidal ideation (Figure 5-5, Tables 5-5). However, clinically, events may indeed be driven by state, and the immediate concern is preventing immediate or short term adverse outcomes.

Our study has a number of limitations. All this work was carried out in psychiatric patients, a high risk group, and it remains to be seen how such predictors apply to non-psychiatric participants. Additionally, the current studies were carried out exclusively in males. Similar work is needed in females, with and without psychiatric disorders. Such work is ongoing in our group. Lastly, for the UP-Suicide testing, the prevalence rate for high suicidal ideation in our independent test cohort was a relatively low 21% (23 out of 108), and the incidence of future hospitalizations for suicidality was even lower: 7.6% in the first year (12 out of 157), and 21.0% overall (33 out of 157) (Figure 5-5). While this is fortunate for the participants enrolled and may reflect the excellence of

clinical care they were receiving in our hospital independent of this study, it may bias the predictions. Studies with larger numbers and longer follow-up, currently ongoing, as well as studies in different clinical settings, may provide more generalizability. It is to be noted, however, that the incidence of suicidality in the general population is lower, for example at 1.5% in adolescents in an European cohort[312] and estimates of 0.2 to 2% in the US[231], which underlines the rationale of using a very high risk group like we did for magnifying and enabling signal detection with a relatively small N.

In conclusion, we have advanced the biological understanding of suicidality, highlighting behavioral and biological mechanisms related to inflammation, mTOR signaling, growth factors, stress response and apoptosis. mTOR signaling has been identified as necessary for the rapid antidepressant response of ketamine[313]. The fact that this biological pathway was identified in an unbiased fashion by our work as the top pathway changed in suicide in the validated biomarkers from our analyses (Table 5-3 and Figure S5-3) is scientifically interesting, and provides a biological rationale for studying ketamine as a potential treatment in acutely suicidal individuals[314]. Of equal importance, we developed instruments (biomarkers and apps) for predicting suicidality, that do not require asking the person assessed if they have suicidal thoughts, as individuals who are truly suicidal often do not share that information with people close to them or with clinicians. We propose that the widespread use of such risk prediction tests as part of routine or targeted healthcare assessments will lead to early disease interception followed

by preventive lifestyle modifications or treatment. Given the magnitude and urgency of the problem, the importance of efforts to implement such tools cannot be overstated.

Chapter 6: The Gender Paradox in Suicidal Behavior

The gender paradox in suicide alludes to the fact that while women have higher rates of suicidal ideation and behavior, men have significantly higher rates of mortality from suicide. In the United States, males are 3.5 times more likely to die by suicide. Despite the gender disparities in suicide completion rates, women actually attempt suicide at three times the rate of men[194].

This finding has been fairly stable across most countries where data is readily available, and has been consistently observed for more than a decade. [315] The most recent data from the WHO Global Heath Observatory data repository from 2012 shows that men commit suicide with greater frequency than women in 166 out of 171 countries with data available when standardized for age. A common thread across these different nations may be found in reduced social role opportunities, culturally accepted codes of expressiveness, and reluctance for men to seek help. [316] In addition to social factors, men tend to use more violent methods, such as firearms. Given the disparity in suicide outcomes by gender, it may prove worthwhile to consider gender differences in pathophysiology as well. [196] [197]

Previous work had focused entirely on male study participants. This was partially a pragmatic decision, as the catchment at Indianapolis VA Medical Center skews more male, and we didn't have a sufficient sample size for the analysis. Another approach could have been to pool male and female samples. Given the striking epidemiological differences in suicidal behavior, however, the decision was

made to instead use a segregated approach within each gender and continue to collect more female participants. We have carried out a pilot study in women, finding some similarities but also novel differences between the genders.

Towards understanding and predicting suicidality in women: biomarkers and clinical risk assessment

Predicting suicidality (suicidal ideation, suicide attempts, and suicide completion) in individuals is a difficult task, which is even more challenging in an understudied population like women. Although women have a lower rate of suicide completion than men, due in part to the less-violent methods used, they have a higher rate of suicide attempts[317]. It is reasonable to assume that genetic and biological differences may exist in suicidality between men and women. Studies by gender are a first step towards individualized medicine. We have previously shown in men with psychiatric disorders how blood biomarkers for suicide, alone or in combination with quantitative phenomic data for anxiety and mood, the Simplified Affective State Scale (SASS), and with a risk profile scale we have developed, Convergent Functional Information for Suicide (CFI-S), collected in the form of apps, could have predictive ability for suicidal ideation, and for future hospitalizations for suicidality[196]. We now present data for discovery, prioritization, validation, and testing of blood biomarkers for suicidality in women, across psychiatric diagnoses. We also show the utility of SASS and CFI–S in predicting suicidality in women. Both these type of tools, biomarkers and phenomic data apps, do not directly ask about suicidal ideation. We demonstrate how our apriori primary endpoint, a comprehensive universal predictor for suicide (UP-Suicide), composed of the combination of 50 top Bonferroni validated biomarkers, along with SASS, and CFI-S, predicts in independent test cohorts suicidal ideation and

future psychiatric hospitalizations for suicidality. Lastly, we uncover biological pathways involved in suicide in women, and potential therapeutics.

Methods

Human participants

We derived our data from four cohorts: one live psychiatric participants discovery cohort; one postmortem coroner's office validation cohort; and two live psychiatric participants test cohorts–one for predicting suicidal ideation, and one for predicting future hospitalizations for suicidality (Figure 6-1).

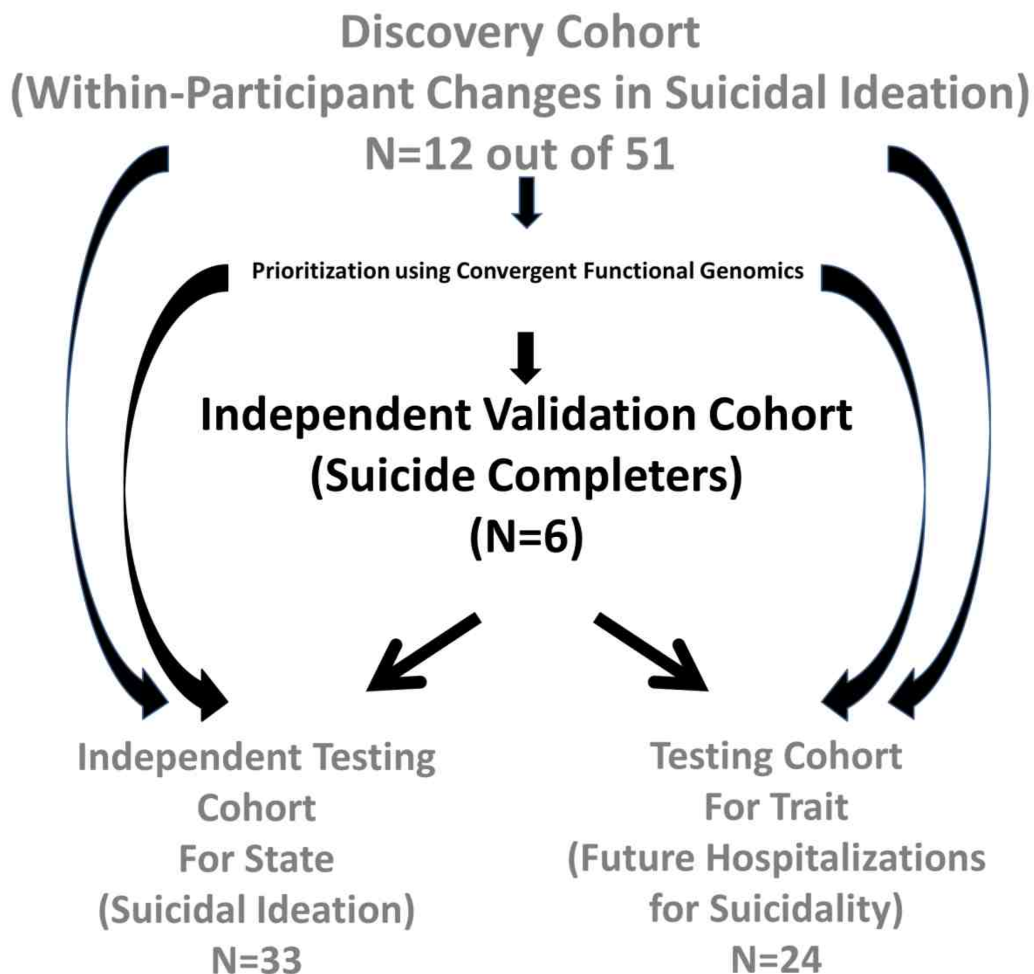


Figure 6-1. Cohorts used in study depicting, flow of discovery, prioritization, validation and testing of biomarkers from each step.

Our within participant discovery cohort, from which the biomarker data were derived, consisted of 12 female participants with psychiatric disorders and multiple visits in our lab, who each had at least one diametric change in SI scores from no SI to high SI from one testing visit to another. There were 7 participants with 3 visits each, and 5 participants with 2 visits each, resulting in a total of 31 blood samples for subsequent microarray studies (Figure 6-2 and Table S6-1).

Our postmortem cohort, in which the top biomarker findings were validated for behavior, consisted of a demographically matched cohort of 6 female violent suicide completers obtained through the Marion County coroner's office (Table 6-1 and Supplementary Table S6-1). We required a last observed alive postmortem interval of 24 h or less, and the cases selected had completed suicide by means other than overdose, which could affect gene expression. 5 participants completed suicide by gunshot to head or chest, and 1 by asphyxiation. Next of kin signed informed consent at the coroner's office for donation of blood for research. The samples were collected as part of our INBRAIN initiative (Indiana Center for Biomarker Research in Neuropsychiatry).

Our independent test cohort for predicting suicidal ideation (Table 6-1) consisted of 33 female participants with psychiatric disorders, demographically matched with the discovery cohort, with one or multiple testing visits in our lab, with either no SI, intermediate SI, or high SI, resulting in a total of 74 blood samples in whom whole-genome blood gene expression data were obtained (Table 6-1 and Table S6-1).

Our test cohort for predicting future hospitalizations (Table 6-1 and Table S6-1) consisted of 24 female participants in whom whole-genome blood gene expression data were obtained by us at testing visits over the years as part of our longitudinal study. If

Table 1. Cohorts used in study

	Participants	*Diagnosis*	*Ethnicity*	*Age mean (s.d.)*	*T-test for age*	
Discovery cohort (within-participant changes in suicidal ideation)	12	BP = 4 MDD = 4 SZA = 3 SZ = 1	EA = 9 AA = 2 Asian = 1	All = 44.39 (11.65) No SI = 44.56 High SI = 44.15	*T-test for age between no SI and high SI 0.926*	
Independent validation cohort for gene expression (suicide completers)	6	BP = 1 MDD = 3 PTSD = 1 Non-psychiatric = 1	EA = 5 AA = 1	43.5 (14.24)		*T-test for age with discovery cohort P = 0.890*
Independent testing cohort for state predictions (suicidal ideation)	33	All BP = 17 MDD = 7 SZA = 7 SZ = 2 No SI BP = 13 MDD = 4 SZA = 6 SZ = 2 Intermediate SI BP = 3 SZA = 1 High SI BP = 3 MDD = 3 SZA = 1	EA = 26 AA = 5 Asian = 1 Mixed = 1	All = 44.05 (8.81) No SI = 43.98 High SI = 41.28	*T-test for age between no SI and high SI 0.553*	*T-test for age with discovery cohort P = 0.887*
Combined discovery and testing cohort for state (suicidal ideation) used for CFI-S analysis (Figure 3)	45	All BP = 21 MDD = 11 SZA = 10 SZ = 3 No SI BP = 17 MDD = 8 SZA = 9 SZ = 3 Intermediate SI BP = 3 SZA = 1 High SI BP = 7 MDD = 7 SZA = 4 SZ = 1	EA = 35 AA = 7 Asian = 2 Mixed = 1	All = 44.15 (9.68) No SI = 44.12 High SI = 43.15	*T-test for age between no SI and high SI 0.727*	
Testing cohort for trait predictions (future hospitalizations for suicidality)	24	All BP = 10 MDD = 9 SZA = 3 SZ = 2 No Hosp for SI BP = 8 MDD = 8 SZA = 1 SZ = 2 Hosp for SI BP = 2 MDD = 1 SZA = 2 SZ = 0	EA = 19 AA = 4 Mixed = 1	All = 46.51 (6.66) No Hosp for SI = 47.2 Hosp for SI = 43.4	*T-test for age between no Hosp for SI and Hosp for SI 0.0430*	*T-test for age with discovery cohort P = 0.354*

Abbreviation: BP, bipolar; MDD, major depressive disorder; PTSD, post-traumatic stress disorder; SZ, schizophrenia; SZA, schizoaffective disorder; SI, suicidal ideation.

the participants had multiple testing visits, the visit with the highest marker (or combination of markers) levels was selected for the analyses (so called "high watermark" or index visit). The participants' subsequent number of psychiatric hospitalizations, with or without suicidality (ideation or attempt), was tabulated from electronic medical records. Participants were evaluated for the presence of future hospitalizations for suicidality, and for the frequency of such hospitalizations. A hospitalization was deemed to be without suicidality if suicidality was not listed as a reason for admission, and no SI was described in the admission and discharge medical notes. Conversely, a hospitalization was deemed to be because of suicidality if suicidal acts or intent was listed as a reason for admission, and/or SI was described in the admission and discharge medical notes.

Medications

The participants in the discovery cohort were all diagnosed with various psychiatric disorders (Table 6-1). Their psychiatric medications were listed in their electronic medical records, and documented by us at the time of each testing visit. The participants were on a variety of different psychiatric medications: mood stabilizers, antidepressants, antipsychotics, benzodiazepines and others (data not shown). Medications can have a strong influence on gene expression. However, our discovery of differentially expressed genes was based on within- participant analyses, which factor out not only genetic background effects but also medication effects, as the participants had no major medication changes between visits. Moreover, there was no consistent pattern in any particular type of medication, or

between any change in medications and SI, in the rare instances where there were changes in medications between visits.

Figure 2

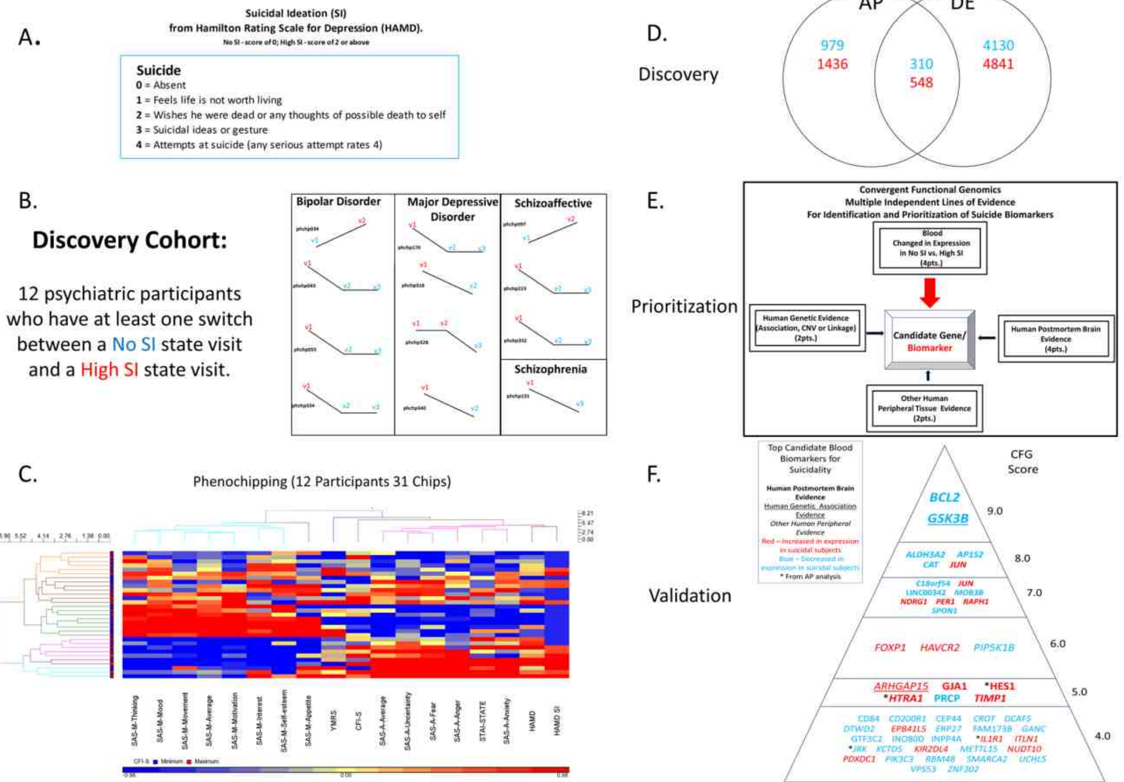


Figure 6-2. Biomarker discovery, prioritization and validation. Discovery cohort: longitudinal within-participant analysis. Phchp### is study ID for each participant. V# denotes visit number (1, 2 or 3). (a) Suicidal ideation (SI) scoring. (b) Participants and visits. (c) PhenoChipping: two-way unsupervised hierarchical clustering of all participant visits in the discovery cohort vs 18 quantitative phenotypes measuring affective state and suicidality. SASS, Simplified Affective State Scale. A—Anxiety items (Anxiety, Uncertainty, Fear, Anger, Average). M—Mood items (Mood, Motivation, Movement, Thinking, Self-esteem, Interest, Appetite, Average). STAI-STATE is State Trait Anxiety Inventory, State Subscale. YMRS is Young Mania Rating Scale. (d) Discovery—number of probesets carried forward from the Absent-Present (AP) and differential expression (DE) analyses,

with an internal score of 1 and above. Red—increased in expression in high SI and blue—decreased in expression in high SI; (e) Prioritization—CFG integration of multiple lines of evidence to prioritize suicide—relevant genes from the discovery step. (f) Validation—Top CFG genes, with a total score of 4 and above, validated in the cohort of suicide completers. All the genes shown were significantly changed and survived Bonferroni correction in ANOVA from no SI to high SI to suicide completers.

Clock Gene Database

We compiled a database of genes associated with circadian function, by using a combination of review papers (Zhang et al. 2009, McCarthy and Welsh 2012[318,319]) and searches of existing databases CircaDB (http://circadb.hogeneschlab.org), GeneCards (http://www.genecards.org), and GenAtlas (http://genatlas.medecine.univ-paris5.fr). Using the data we compiled from these sources we identified a total of 1468 genes that show circadian functioning. We further classified genes into "core" clock genes, i.e. those genes that are the main engine driving circadian function (n=18), "immediate" clock genes, i.e. the genes that directly input or output to the core clock (n=331), and "distant" clock genes, i.e. genes that directly input or output to the immediate clock genes (n=1,119).

Clinical measures

The Simplified Affective State Scale (SASS) is an 11 item scale for measuring mood and anxiety, previously developed and described by us [277,320]. The SASS has a set of 11 visual analog scales (7 for mood, 4 for anxiety) that ends up providing a number ranging from 0 to 100 for mood state, and the same for anxiety state. We have developed an Android app version.

Convergent Functional Information for Suicidality (CFI-S) (Figures 6-3 and S6-2) is a 22 item scale and Android app for suicide risk[320], which integrates, in a simple binary fashion (Yes-1, No-0), similar to a polygenic risk score, information about known life events, mental health, physical health, stress, addictions, and

cultural factors that can influence suicide risk[289 290]. The scale was administered at participant testing visits (n= 39), or scored based on retrospective electronic medical record information and Diagnostic Interview for Genetic Testing (DIGS) information (n=48). When information was not available for an item, it was not scored (NA).

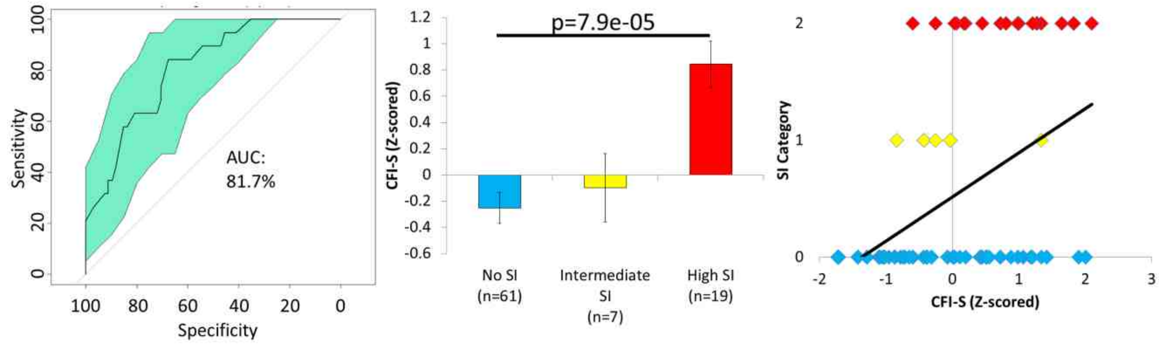
Combining gene expression biomarkers and clinical measures

The Universal Predictor for Suicide (UP-Suicide) construct, our primary endpoint, was decided upon as part of our apriori study design to be broad-spectrum, and combine our top Bonferroni validated biomarkers with the phenomic (clinical) markers (SASS and CFI-S). It is calculated as the average of three increased markers (BioM-18 averaged increased Bonferroni biomarkers, Anxiety, CFI-S), minus the average of two decreased markers (BioM-32 averaged decreased Bonferroni biomarkers, Mood). All individual markers are Z-scored by diagnosis, to account for different ranges and be able to combine them into a composite predictor.

Testing analyses

The test cohort for suicidal ideation and the test cohort for future hospitalizations analyses were assembled out of data that was RMA normalized by diagnosis. Phenomic (clinical) and gene expression markers used for predictions were z scored by diagnosis, to be able to combine different markers into panels and to avoid potential artefacts due to different ranges of phene expression and gene expression in different diagnoses. Markers were combined by computing the

average of the increased risk markers minus the average of the decreased risk markers. Predictions were performed using R-studio.



Predictor	ROC AUC	AUC p-value	ANOVA	Correlation R	Correlation p-value
CFI-S	0.817	1.29e-05	7.9e-5	0.441	9.35e-06

CFI-S Item	Description	Correct direction	T-test (one tailed) High SI vs. No SI p-value
16	**Chronic stress: lack of positive relationships, social isolation**	**Y**	**0.0040**
12	**Current substance abuse**	**Y**	**0.0071**
17	**History of excessive extroversion and impulsive behaviors (including rage, anger, physical fights, seeking revenge)**	**Y**	**0.0147**
14	**Lack of religious beliefs**	**Y**	**0.0175**
13	**Past history of suicidal acts/gestures**	**Y**	**0.0253**
15	**Acute stress: Rejection (within last 3 months)**	**Y**	**0.0294**
20	**History of command hallucinations of self-directed violence**	**Y**	**0.0453**
10	Dissatisfaction with life at this moment in time	Y	*0.0583*
8	Chronic stress: perceived uselessness, not feeling needed, burden to extended kin.	Y	*0.0635*
4	Personally knowing somebody who committed suicide	Y	*0.0733*
7	Acute stress: Losses, grief (within last 3 months)	Y	*0.0748*
3	Family history of suicide in blood relatives	Y	0.1422
21	Age: Older >60 or Younger <25	Y	0.2374
2	With poor treatment compliance	Y	0.2477
6	Acute/severe medical illness, including acute pain ("I just can't stand this pain anymore.") (within last 3 months)	Y	0.2714
5	History of abuse: physical, sexual, emotional, neglect	Y	0.3348
9	History of excessive introversion, conscientiousness (including planned suicide attempts)	Y	0.3388
18	Lack of coping skills when faced with stress (cracks under pressure)	Y	0.3723
19	Lack of children. If has children, not in touch /not helping take care of them.	N	0.0714
1	Psychiatric illness diagnosed and treated	All have dx	All have dx
11	Lack of hope for the future	No difference	1
22	Gender: Male	All females	All females

Figure 6-3. Convergent Functional Information for Suicide (CFI-S) scale testing in women. Prediction of high suicidal ideation in women in a larger cohort that combines the discovery and test cohorts used for biomarker work. CFI-S was developed independently of any data from this study, by compiling known socio-demographic and clinical risk factors for suicide. It is composed of 22 items that

assess the influence of mental health factors, as well as of life satisfaction, physical health, environmental stress, addictions, cultural factors known to influence suicidal behavior, and two demographic factors, age and gender. Table depicts individual items and their ability to differentiate between no SI and high SI.

Predicting Suicidal Ideation. Receiver-operating characteristic (ROC) analyses between genomic and phenomic marker levels and suicidal ideation (SI) were performed by assigning participants with a HAMD-SI score of 2 and greater into the high SI category. We used the pROC function of the R studio. We used the z-scored biomarker and app scores, running them in this ROC generating program against the "diagnostic" groups in the independent test cohort (high SI vs. the rest of subjects).

Additionally, ANOVA was performed between no SI (HAMD-SI 0), intermediate (HAMD-SI 1), and high SI participants (HAMD-SI 2 and above) and Pearson R (one-tail) was calculated between HAMD-SI scores and marker levels (Table 6-4 and Figure 6-4).

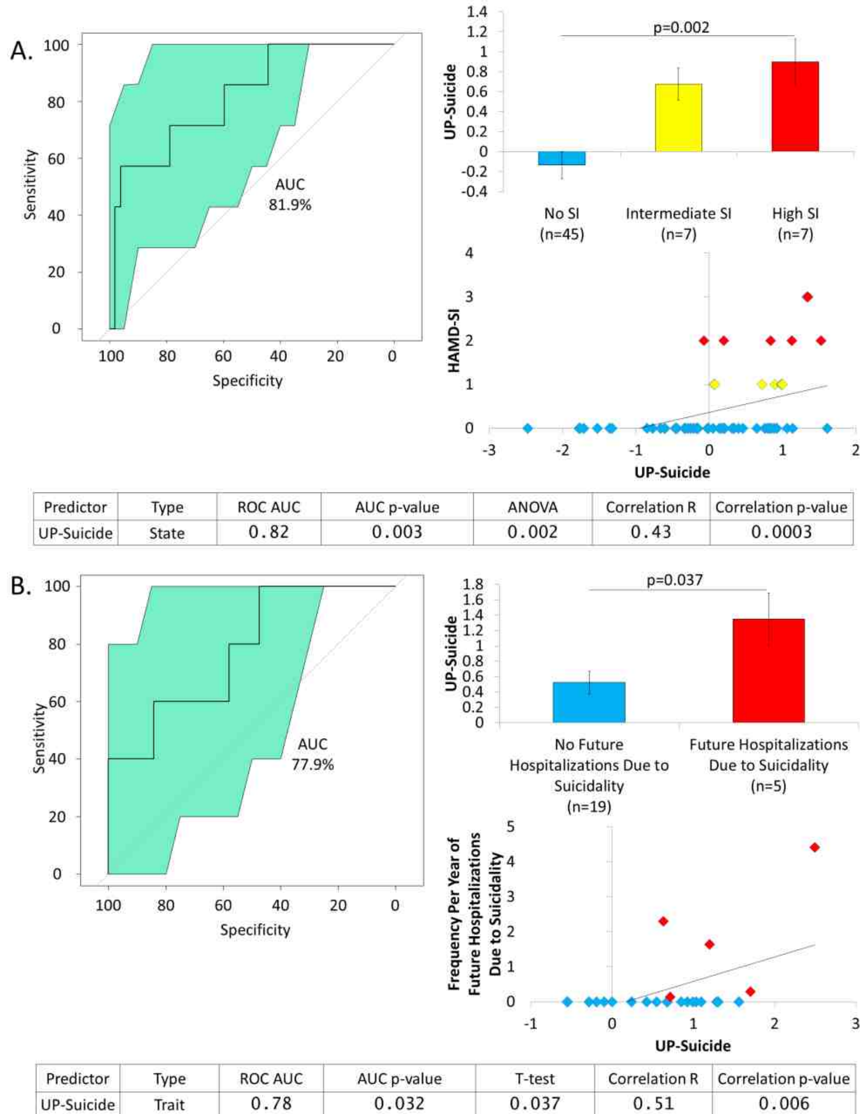


Figure 6-4. UP-Suicide predicting suicidal ideation in the independent test cohort, and predicting future hospitalizations due to suicidality. UP-Suicide is composed of the 50 Bonferroni validated biomarkers along with CFI-S scores and SASS (Mood and Anxiety scores). n= number of testing visits. (a) Top left: Receiver operating curve identifying participants with suicidal ideation against participants with no SI or intermediate SI. Top right: Y axis contains the average UP-Suicide scores with standard error of mean for no SI, intermediate SI and high SI. Bottom right:

Scatter plot depicting HAMD-SI score on the Y axis and UP-Suicide score on the X axis with linear trend line. Bottom: Table summarizing descriptive statistics. (b) Top left: Receiver operating curve identifying participants with future hospitalizations due to suicidality against participants without future hospitalizations due to suicidality. Top right: Y axis contains the average UP-Suicide scores with standard error of mean for no future hospitalizations due to suicidality and participants with future hospitalizations due to suicidality. Bottom right: Scatter plot depicting frequency of future hospitalizations due to suicidality on the Y axis and UP-Suicide score on the X axis with linear trend line. Bottom: Table summarizing descriptive statistics.

***Predicting Future Hospitalizations for Suicidality*.**

We conducted analyses for hospitalizations in the years following testing (on average 2.75 years, range 0.3 to 7.5 years; see Table S6-1). For each participant in the test cohort for future hospitalizations, the study visit with highest levels for the marker or combination of markers was selected as index visit (or with the lowest levels, in the case of decreased markers). ROC analyses between genomic and phenomic marker levels and future hospitalizations were performed as described above, based on assigning if participants had been hospitalized for suicidality (ideation, attempts) or not following the index testing visit. Additionally, a one tailed t-test with unequal variance was performed between groups of participants with and without hospitalizations for suicidality. Pearson R (one-tail) correlation was performed between hospitalization frequency (number of hospitalizations for suicidality divided by duration of follow-up) and marker scores. We conducted correlation analyses for hospitalizations frequency for all future hospitalizations due to suicidality as this calculation, unlike the ROC and t-test, accounts for the actual length of follow-up at our VA, which varied from participant to participant. The ROC and t-test might in fact, if anything, under-represent the power of the markers to predict, as the more severe psychiatric patients are more likely to move geographically and/or be lost to follow-up.

Figures

Each figure in this chapter was completed by Daniel Levey and Helen Le-Niculescu. This work has been published. [197]

Table 6-2.

Gene symbol/ Gene Name	Probesets	**Discovery** (Change) Method/ Score	Prior human genetic evidence	Prior human brain expression evidence	Prior human peripheral expression evidence	**Prioritization** Total CFG Score For Suicide	**Validation** ANOVA p-value	**Predictions** ROC/ p-value	**Clock function**
Best Predictive Biomarkers Out of Validated Biomarkers (Bonferroni) (49 genes, 50 probesets)									

BCL2 B-cell CLL	203684_s_at	(D) DE/2	Linkage [261]	(D) PFC [321]	(D) Blood [320]	9.00	**3.95E-06**	SI: 0.48/0.56 Hosp: 0.89/0.007	
<u>ALDH3A2</u> aldehyde dehydrogenase 3 family, member A2	202053_s_at	(D) DE/2		(I) BA 4, BA 44, Thalamus [243]	(D) Blood [320]	8.00	**1.62E-06**	SI: 0.63/0.14 Hosp: 0.6/0.29	
MOB3B MOB kinase activator 3B	229568_at	(D) DE/1		(I) ACC [248]	(I) Blood [320]	7.00	**4.69E-06**	SI: 0.55/0.35	

								Hosp: 0.85/0.015	
PER1 period circadian clock 1	202861_at	(I) DE/1		(D) DLFPC 248	(D) Blood 320	7.00	**5.32E-12**	SI: 0.45/0.66 Hosp: 0.84/0.018	Clock Core
<u>HAVCR2</u> hepatitis A virus cellular receptor 2	1555629_at	(I) DE/4			(D) Blood 320	6.00	**1.69E-12**	SI: 0.62/0.15 Hosp: 0.8/0.022	

<u>ARHGAP15</u> Rho GTPase activating protein 15	1561489_at	(I) DE/1	Suicide [322]		(I) Blood [320]	5.00	**3.05E-06**	SI: 0.55/0.34 Hosp: 0.79/0.041	
HTRA1 HtrA serine peptidase 1	201185_at	(I) AP/1		(I) NAC [248]		5.00	**3.17E-07**	SI: 0.36/0.89 Hosp: 0.84/0.01	Clock Distant Output
EPB41L5 erythrocyte membrane protein band 4.1 like 5	229292_at	(I) DE/1	Linkage [323]		(I) Blood [320]	4.00	**4.58E-14**	**SI: 0.68/0.062** Hosp:	

								0.63/ 0.24	
GTF3C2 general transcription factor IIIC, polypeptide 2, beta 110kDa	210620_s_at	(D) DE/2			(D) Blood [320]	4.00	**1.68E-07**	SI: 0.64/ 0.12 Hosp: 0.81/ 0.05	
PDXDC1 pyridoxal-dependent decarboxylase domain containing 1	1560013_at	(I) DE/2			(I) Blood [320]	4.00	**1.03E-05**	SI: 0.51/ 0.46 Hosp:0.81/0.018	
PIK3C3 phosphatidylinositol 3-kinase,	232086_at	(D) DE/1	Suicide, Antidepressants [260]		(I) Blood [320]	4.00	**3.14E-08**	**SI: 0.65/0.098 Hosp:**	

catalytic subunit type 3								**0.9/ 0.011**	
Best Predictive Biomarkers Out of Top Discovery and Prioritization Biomarkers(Non Bonferroni Validated, 65 genes)									
LRRC8B leucine rich repeat containing 8 family, member B	212976_at	(D) DE/4		(I) BA 11 [223]		8.00	0.231881	SI: 0.60/ 0.19 Hosp: 0.69/ 0.14	
<u>ACTR3</u> ARP3 actin-related protein 3 homolog (yeast)	213102_at	(D) DE/4	Linkage [323]		(I) Blood [320]	7.00	*0.00045239*	SI: 0.62/ 0.15 Hosp: 0.73/ 0.12	
<u>ASPH</u> aspartate	242037_a	(I)			(I) Blood [320]	6.00	*0.001087*	SI: 0.65/ .098	

beta-hydroxylase	t	DE/4						Hosp: 0.8/.022	
CSNK1A1 casein kinase 1, alpha 1	235464_at	(D) DE/4			(D) Blood [320]	6.00	*NC*	SI: 0.56/0.32 **Hosp: 0.96/0.0007**	Clock Immediate Input
DPCD deleted in primary ciliary dyskinesia homolog (mouse)	226009_at	(I) DE/4			(D) Blood [320]	6.00	*NC*	SI: 0.67/0.067 Hosp: 0.76/0.044	

GTF3C3 general transcription factor IIIC, polypeptide 3, 102kDa	1555439_at	(I) AP/4			(I) Blood 320	6.00		SI: 0.67/ 0.075 Hosp: 0.75/ .067	
KLHL28 kelch-like family member 28	220374_at	(I) AP/4			(I) Blood 320	6.00	NC	SI: 0.64/ 0.11 Hosp: 0.91/ 0.002	
LARP4 La ribonucleoprotein domain family, member 4	214155_s_at	(D) DE/4			(D) Blood 320	6.00	*0.014911*	SI: 0.49/ 0.55 Hosp:	

								0.9/0.005	
NUDT6 nudix (nucleoside diphosphate linked moiety X)-type motif 6	220183_s_at	(D) AP/4			(D) Blood 320	6.00		SI: 0.62/0.16 Hosp: 0.71/0.14	
SNX27 sorting nexin family member 27	244349_at	(I) AP/4			(I) Blood 320	6.00		SI: 0.63/0.13 Hosp: 0.85/0.015	
UIMC1 ubiquitin interaction motif containing 1	233596_at	(I) DE/4			(I) Blood 320	6.00		SI: 0.41/0.76 Hosp:	

								0.86/ 0.00 6	
ZNF548 zinc finger protein 548	1553 718_ at	(D) DE/4			(D) Blood [320]	6.00	*0.0 00 46 1*	SI: 0.40/ 0.82 Hosp : 0.83/ 0.01 2	

Results

Discovery of biomarkers for suicidal ideation

We conducted whole-genome gene expression profiling in the blood samples from a longitudinally followed cohort of female participants with psychiatric disorders that predispose to suicidality. The samples were collected at repeated visits, 3–6 months apart. State information about suicidal ideation (SI) was collected from a questionnaire (HAMD) administered at the time of each blood draw (Table S6-1). Out of 51 female psychiatric participants (with a total of 123 visits) followed longitudinally in our study, with a diagnosis of BP, MDD, SZ and SZA, there were 12 participants that switched from a no SI (SI score of 0) to a high SI state (SI score of 2 and above) at different visits, which was our intended discovery group (Figure 6-2). We used a powerful within-participant design to analyze data from these 12 participants and their 31 visits. A within-participant design factors out genetic variability, as well as some medications, lifestyle, and demographic effects on gene expression, permitting identification of relevant signal with Ns as small as 1[14]. Another benefit of a within-participant design may be accuracy/consistency of self-report of psychiatric symptoms ('phene expression'), similar in rationale to the signal detection benefits it provides in gene expression.

For discovery, we used two methodologies: Absent/Present (reflecting on/off of transcription), and Differential Expression (reflecting more subtle gradual changes in expression levels). The genes that tracked suicidal ideation in each

participant were identified in our analyses. We used three thresholds for increased in expression genes and for decreased in expression genes: ≥ 33.3% (low), ≥50% (medium), and ≥ 80% (high) of the maximum scoring increased and decreased gene across participants. Such a restrictive approach was used as a way of minimizing false positives, even at the risk of having false negatives. For example, there were genes on each of the two lists, from AP and DE analyses, that had clear prior evidence for involvement in suicidality, such as AKAP10 ([248]) (31.7%) and MED28 [248] (31.8%) from AP, and S100B[248,324] (31.7%) and SKA2[220] (31.4%) for DE, but were not included in our subsequent analyses because they did not meet our apriori set 33.3% threshold. Notably, SKA2 reproduces our results in males (Niculescu et al. 2015), as well as the work from Kaminsky and colleagues[219,220].

Prioritization of biomarkers based on prior evidence in the field

These differentially expressed genes were then prioritized using a Bayesian-like Convergent Functional Genomics (CFG) approach (Figure 6-2) integrating all the previously published human genetic evidence, postmortem brain gene expression evidence, and peripheral fluids evidence for suicide in the field available at the time of our analyses (September 2015).This is a way of identifying and prioritizing disease relevant genomic biomarkers, extracting generalizable signal out of potential cohort-specific noise and genetic heterogeneity. We have built in our lab manually curated databases of the psychiatric genomic and proteomic literature to date, for use in CFG analyses. The CFG approach is thus a de facto

field-wide collaboration. We use in essence, in a Bayesian fashion, the whole body of knowledge in the field to leverage findings from our discovery data sets.

Validation of biomarkers for behavior in suicide completers

For validation in suicide completers, we used 1471 genes that had a CFG score of 4 and above, from AP and DE, reflecting either maximum internal score from discovery or additional external literature cross-validating evidence. Out of these, 882 did not show any stepwise change in suicide completers (NC- non-concordant). As such, they may be involved primarily in ideation and not in behavior (Table S6-5). The remaining 589 genes (40.0%) had levels of expression that were changed stepwise from no suicidal ideation to high suicidal ideation to suicide completion. 396 of these genes (26.9%) were nominally significant, and 49 genes (50 probesets- two for JUN) (3.33%) survived Bonferroni correction for multiple comparisons (Figure 6-2f). These genes are likely involved in suicidal ideation *and* suicidal behavior. (A person can have suicidal ideation without suicidal behavior, but cannot have suicidal behavior without suicidal ideation).

Selection of biomarkers for testing of predictive ability

For testing, we decided apriori to focus on the Bonferroni validated biomarkers (49 genes, 50 probesets). We also examined in a secondary analysis the top scoring biomarkers from both discovery and prioritization (65 genes), so as to avoid potential false negatives in the validation step due to possible postmortem artefacts or extreme stringency of statistical cutoff (Figure S6-1). The top CFG scoring genes after the Bonferroni validation step were BCL2 and GSK3B.

The top CFG scoring genes from the discovery and prioritization steps were FAM214A, CLTA, HSPD1, and ZMYND8. Notably, all have co-directional gene expression changes evidence in brains of suicide completers in studies from other groups (Figure 6-2, Tables 6-2 and S6-2).

Biological understanding

We also sought to understand the biology represented by the biomarkers identified by us, and derive some mechanistic and practical insights. We conducted: 1. unbiased biological pathway analyses and hypothesis driven mechanistic queries, 2. overall disease involvement and specific neuropsychiatric disorders queries, and 3. overall drug modulation along with targeted queries for omega-3, lithium and clozapine[306] (Tables 6-3, S6-3, S6-4). Administration of omega-3s in particular may be a mass- deployable therapeutic and preventive strategy[167,207].

Table 6-3 Biological Pathways and Diseases

A.	**Ingenuity Pathways**				**KEGG Pathways**			**GeneGO Pathways**		
Prioritization CFG score ≥4 (n=1904 probe sets/ 1471 genes)	**#**	**Top Canonical Pathways**	**P-Value**	**Ratio**	**Pathway Name**	**Ratio**	**Enrichment p-value**	**Process Networks**	**Ratio**	**p-value**
	1	B Cell Receptor Signaling	2.88 E-13	22.9% 41/179	Morphine addiction	19/ 239	9.27 E-06	Immune response_ BCR pathway	42/ 137	4.332E-11
	2	Protein Kinase A Signaling	3.61 E-13	16.6%	Phosphatidyl inosito	18/ 245	4.20 E-05	Apoptosis_ Anti-Apoptosis	45/ 179	1.0

				66/398	l signaling system			mediated by external signals via MAPK and JAK/STAT		70E-08
	3	PI3K Signaling in B Lymphocytes	5.80E-12	24.8% 33/133	Neurotrophin signaling pathway	29/545	7.23E-05	Reproduction_Gonadotropin regulation	48/199	1.452E-08
	4	IGF-1 Signaling	7.76E-12	28.3% 28/99	Amoebiasis	22/363	9.46E-05	Cell cycle_G1-S Growth factor regulation	47/195	2.115E-

										08
	5	Glucocorticoid Receptor Signaling	1.96E-11	17.8% 50/281	Insulin signaling pathway	27/520	0.0001855	Development_Hemopoiesis, Erythropoietin pathway	37/136	2.393E-08
	#	**Top Canonical Pathways**	**P-Value**	**Ratio**	**Pathway Name**	**Ratio**	**Enrichment p-value**	**Process Networks**	**Ratio**	**p-value**
Validation	1	Glucocorticoid Receptor Signaling	2.86E-06	7.8% 22	Morphine addiction	9/249	0.0006493	Reproduction_Gonadotropin regulation	24/199	9.84

Stepwise in Suicide Completers (n=589 genes)				/281						3E-07
	2	IGF-1 Signaling	7.18E-06	12.1% 12/99	Colorectal cancer	9/287	0.0016932	Reproduction_GnRH signaling pathway	20/166	8.256E-06
	3	Renin-Angiotensin Signaling	8.72E-06	11.0% 13/118	Cocaine addiction	6/155	0.0037291	Reproduction_Progesterone signaling	23/214	1.194E-

										05
	4	Protein Kinase A Signaling	1.02 E-05	6.5% 26/398	Insulin signaling pathway	12/535	0.004728 4	<u>Signal transduction_NOTCH signaling</u>	24/236	1.962E-05
	5	Melanocyte Development and Pigmentation Signaling	1.02 E-05	12.8% 11/86	Inositol phosphate metabolism	6/193	0.010198 6	<u>Signal transduction_Androgen receptor signaling cross-talk</u>	12/72	2.241E-05

	#	Top Canonical Pathways	P-Value	Ratio	Pathway Name	Ratio	Enrichment p-value	Process Networks	Ratio	p-value
Validation Nominally significant In Suicide Completers (n= 396 genes)	1	Neurotrophin/TRK Signaling	3.48 E-06	12.5% 9/72	Cocaine addiction	6/155	0.000454	Reproduction_Gonadotropin regulation	16/199	7.748E-05
	2	Glucocorticoid Receptor Signaling	2.68 E-05	5.7% 16/281	Colorectal cancer	7/289	0.002226	Reproduction_GnRH signaling pathway	14/166	1.323E

										-04
	3	Melanocyte Development and Pigmentation Signaling	1.06E-04	9.3% 8/86	Wnt signaling pathway	9/495	0.003675	Reproduction_Progesterone signaling	16/214	1.822E-04
	4	G-Protein Coupled Receptor Signaling	1.79E-04	5.3% 14/264	Notch signaling pathway	5/170	0.004315	Signal transduction_NOTCH signaling	16/236	5.495E-04

	5	Corticotropin Releasing Hormone Signaling	2.23 E-04	7.4% 9/121	Adherens junction	7/340	0.005315	Signal transduction_WNT signaling	13/177	8.738E-04
	#	**Top Canonical Pathways**	**P-Value**	**Ratio**	**Pathway Name**	**Ratio**	**Enrichment p-value**	**Process Networks**	**Ratio**	
Validation Bonferroni significant	1	IL-17 Signaling	1.34 E-05	5.6% 4/72	Inositol phosphate metabolism	3/196	0.000383	Cell cycle_G1-S Interleukin regulation	6/128	6.4000E

in Suicide Completers (n= 49 genes)										-06
	2	p53 Signaling	4.52E-05	4.1% 4/98	Phosphatidylinositol signaling system	3/260	0.000863	Immune response_ BCR pathway	5/137	1.387E-04
	3	Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	8.71E-05	2.2% 5/225	Colorectal cancer	3/293	0.001214	Immune response_ Th17-derived cytokines	4/98	4.589E-04

	4	Docosahexaenoic Acid (DHA) Signaling	1.02 E-04	6.7 % 3/45	Tryptophan metabolism	2/132	0.004229	Inflammation_IL-2 signaling	4/104	5.752E-04
	5	Ovarian Cancer Signaling	1.48 E-04	3.0 % 4/133	Neurotrophin signaling pathway	3/571	0.007844	Cell cycle_G1-S Growth factor regulation	5/195	7.127E-04

B.	**Ingenuity**	**GeneGO**

	#	Diseases and Disorders	P-Value	# Molecules	Diseases	Ratio	P-Value
Prioritization on CFG score ≥4 (n= 1904 probesets / 1471 genes)	1	Cancer	2.25E-06 - 2.21E-45	1242	Mental Disorders	256/1610	1.890E-35
	2	Organismal Injury and Abnormalities	2.25E-06 - 2.21E-45	1242	Psychiatry and Psychology	284/1904	2.194E-34
	3	Gastrointestinal Disease	1.02E-06 - 5.07E-31	905	Depressive Disorder, Major	120/543	2.660E-29
	4	Reproductive System Disease	1.43E-06 - 2.31E-24	617	Central Nervous System Diseases	379/3060	8.770E-29

	5	Infectious Diseases	8.30E-07 - 1.15E-17	246	Depressive Disorder	120/557	3.125E-28
	#	**Diseases and Disorders**	**P-Value**	**# Molecules**	**Diseases**	**Ratio**	**P-Value**
Validation Step wise in Suicide Completers	1	Cancer	6.57E-04 - 6.34E-17	487	Breast Neoplasms	356/8894	3.727E-15
	2	Organismal Injury and Abnormalities	6.57E-04 - 6.34E-17	492	Breast Diseases	356/8895	3.798E-15
	3	Gastrointestinal Disease	6.23E-04 - 2.76E-10	355	Psychiatry and Psychology	115/1904	2.268E-14

(n=589 genes)	4	Reproductive System Disease	6.50E-04 - 8.34E-09	240	Pathological Conditions, Signs and Symptoms	207/4433	1.078E-13
	5	Infectious Diseases	6.57E-04 - 6.95E-08	104	Mental Disorders	101/1610	1.146E-13
	#	**Diseases and Disorders**	**P-Value**	**# Molecules**	**Diseases**	**Ratio**	**P-Value**
Validation Nominally significant	1	Cancer	2.36E-03 - 1.87E-10	325	Depressive Disorder, Major	40/543	1.045E-12
	2	Organismal Injury and Abnormalities	2.47E-03 - 1.87E-10	330	Pathological Conditions, Signs and Symptoms	150/4433	2.002E-12

In Suicide Completers (n=396 genes)	3	Tumor Morphology	2.29E-03 - 1.17E-07	36	Depressive Disorder	40/557	2.333E-12
	4	Developmental Disorder	2.47E-03 - 1.40E-06	69	Breast Neoplasms	245/8894	2.770E-11
	5	Gastrointestinal Disease	2.44E-03 - 2.43E-06	230	Breast Diseases	245/8895	2.806E-11
	#	**Diseases and Disorders**	**P-Value**	**# Molecules**	**Diseases**	**Ratio**	**P-Value**
Validation Bonferron	1	Immunological Disease	4.03E-03 - 1.27E-06	14	Lymphoma, Mantle-Cell	8/196	3.430E-08

i significant in Suicide Completers (n=49 genes)	2	Cancer	4.15E-03 - 3.97E-06	42	Psychiatry and Psychology	19/1904	1.209E-07
	3	Dermatological Diseases and Conditions	4.03E-03 - 3.97E-06	10	Lymphoma, Non-Hodgkin	12/726	2.323E-07
	4	Hematological Disease	4.03E-03 - 3.97E-06	5	Mental Disorders	17/1610	3.079E-07
	5	Organismal Injury and Abnormalities	4.15E-03 - 3.97E-06	42	Leukemia, Myeloid	16/1436	3.667E-07

The sets of biomarkers identified have biological roles in inflammation, neurotrophins, inositol signaling, stress response, and perhaps overall the switch between cell survival and proliferation vs. apoptosis (Tables 6-3 and S6-5).

We also examined evidence for the involvement of these biomarkers for suicidality in other psychiatric disorders, permitting us to address issues of context and specificity (Table S6-3). FAM214A, MOB3B, ZNF548, and ARHGAP35 seem to be relatively specific for suicide, based on the evidence to date in the field. BCL2, GSK3B, HSPD1, and PER1 are less specific for suicide, having equally high evidence for involvement in suicide and in other psychiatric disorders.

These boundaries and understanding will likely change as additional evidence in the field accumulates. For example, HSPD1, discovered in this work as a top biomarker increased in expression in suicidality, is also increased in expression in the blood following anti-depressant treatment[325,326] , and thus might be a useful biomarker for treatment-emergent suicidal ideation (TESI).

A number of the genes are changed in expression in opposite direction in suicide in this study vs. high mood in our previous mood biomarker study[15]- SSBP2, ZNF596 (Table S6-3), suggesting that suicidal participants are in a low mood state. Also, some of the top suicide biomarkers are changed in expression in the same direction as in high psychosis participants in a previous psychosis biomarker study of ours[24] – HERC4, PIP5K1B, SLC35B3, SNX27, KIR2DL4, NUDT10 (Table S6-3), suggesting that suicidal participants may be in a psychosis-like state. Taken together, the data indicates that suicidality could be viewed as a psychotic dysphoric state. This molecularly informed view is consistent with the emerging clinical evidence in the field[308].

A number of top biomarkers identified by us have biological roles that are related to the core circadian clock (such as PER1), or modulate the circadian clock (such as CSNK1A1), or show at least some circadian pattern (such as HTRA1). To be able to ascertain all the genes in our dataset that were circadian and do estimates for enrichment, we compiled from the literature a database of all the known genes that fall into these three categories, numbering a total of 1468 genes. Using an estimate of about 21,000 genes in the human genome, that gives about 7% of genes having some circadian pattern. Out of our 49 Bonferroni validated biomarker genes, 7 had circadian evidence (14.3%) (Table S6-3), suggesting a two-fold enrichment for circadian genes. Circadian clock abnormalities are related to mood disorders[268,319], and sleep abnormalities have been implicated in suicide[327].

Table 6-4. Predictions. UP-Suicide is composed of 50 validated biomarkers (18 increased in expression, 32 decreased in expression), along with clinical measures app scores (CFI-S, SASS). SASS is composed of Mood scale and Anxiety scale.

		Marker	Participants with Suicidality / Participants Total	ROC AUC/p-value	Pearson's Correlation R/p-value	Student's t-test p-value
Suicidal Ideation Cohort	Best Blood Biomarkers	**Best Validated Biomarkers (Bonferroni) (49 genes, 50 probesets)**				
		EPB41L5	7 | 33	*0.68/0.062*	**0.22/0.028**	*0.091*
		HAVCR2	7 | 33	0.62/0.15	*0.17/0.069*	0.18

n=33 Participants		**ARHGAP15**	7 | 33	0.55/0.34	0.12/0.15	0.22
		PIK3C3	7 | 33	0.65/0.098	**-0.21/0.037**	*0.084*
		GTF3C2	7 | 33	0.64/0.11	-0.11/0.18	*0.072*
		ALDH3A2	7 | 33	0.62/0.14	**-0.21/0.036**	0.14
		Best Discovery and Prioritization Biomarkers(Non Bonferroni Validated)				
		DPCD	7 | 33	*0.67/0.067*	**0.21/0.040**	0.12
		GTF3C3	7 | 33	*0.67/0.075*	**0.23/0.024**	0.11
		ASPH	7 | 33	0.65/0.098	0.073/0.27	0.13
		ACTR3	7 | 33	0.62/0.15	**-0.19/0.054**	0.13

		NUDT6	7 | 33	0.62/0.15	-0.072/0.27	0.19
		LRRC8B	7 | 33	0.60/0.19	**-0.15/0.11**	0.13
		Panels of Validated Biomarkers (Increased, Decreased, Combined)				
		BioM-18	7 | 33	0.37/0.87	0.032/0.39	0.59
		BioM-32	7 | 33	0.43/0.72	-0.0031/0.49	0.68
		BioM-50	7 | 33	0.50/0.51	0.021/0.43	0.47
	Clinical measures	**Anxiety**	7 | 33	**0.72/0.029**	**0.26/0.011**	**0.0083**
		Mood	7 | 33	**0.78/0.0078**	**-0.37/0.00062**	**0.020**

		SASS	7 | 33	**0.81/0.0035**	**0.38/0.00049**	**5.04E-05**		
		CFI-S	7 | 29	**0.84/0.0020**	**0.39/0.0013**	**0.0031**		
		CFI-S + SASS	7 | 29	**0.87/0.00088**	**0.48/5.22E-05**	**0.0027**		
	Combined	**UP-Suicide**	7 | 29	**0.82/0.0034**	**0.43/0.00029**	**0.0015**		
Future Hospitalizations for Suicidality n=24 Participants	Best Blood Biomarkers	Marker	Participants Hospitalized for Suicidality / Participant	ROC AUC/p-value	Pearson's Correlation R/p-value	Student's t-test p-value	Cox Regression Hazard Ratio	Cox Regression p-value

			s Hospitalized Total					
		Best Validated Biomarkers (Bonferroni) (49 genes, 50 probesets)						
		HTRA1	5 | 24	**0.84/0.0096**	**0.62/0.00058**	**0.024**	**4.6**	**0.012**
		PER1	4 | 24	**0.84/0.018**	**0.39/0.029**	**0.13**	1.5	0.16
		PDXDC1	5 | 24	**0.81/0.018**	**0.64/0.00043**	**0.042**	**2.4**	**0.015**
		PIK3C3	3 | 24	**0.90/0.011**	-0.25/0.12	**0.026**	6.0	0.12
		BCL2	4 | 24	**0.89/0.0067**	**-0.35/0.047**	**0.054**	**3.1**	**0.012**
		MOB3B	4 | 24	**0.85/0.015**	**-0.34/0.053**	**0.0046**	*9.6*	*0.089*

		Best Discovery and Prioritization Biomarkers(Non Bonferroni Validated)						
		KLHL28	5 | 24	**0.91/0.0020**	**0.51/0.0051**	**0.0026**	**7.3**	**0.036**
		UIMC1	5 | 24	**0.86/0.0060**	0.40/0.025	**0.037**	**2.3**	**0.028**
		SNX27	4 | 24	**0.85/0.015**	0.74/1.83E-05	**0.052**	**4.5**	**0.015**
		CSNK1A1	4 | 24	**0.96/0.00066**	-0.27/0.10	**0.0066**	**620.5**	**0.022**
		LARP4	4 | 24	**0.90/0.0050**	-0.30/0.077	**6.30E-05**	37.0	0.11
		ZNF548	5 | 24	**0.83/0.012**	-0.31/0.071	**0.0077**	**15.9**	**0.019**
		Panels of Validated Biomarkers (Increased, Decreased, Combined)						
		BioM-18	4 | 24	**0.88/0.0088**	**0.46/0.011**	**0.033**	**27.6**	**0.021**

		BioM-32	4 | 24	**0.71/0.11**	**-0.34/0.053**	**0.16**	**10.6**	**0.23**
		BioM-50	5 | 24	**0.94/0.0017**	**0.54/0.0033**	**0.0058**	**89.5**	**0.023**
	Clinical measures	**Anxiety**	4 | 24	**0.86/0.014**	**0.43/0.017**	**0.0039**	*16.0*	*0.054*
		Mood	3 | 24	0.68/0.18	-0.21/0.16	0.22	33.4	0.10
		SASS	4 | 24	**0.83/0.023**	**0.40/0.027**	**0.034**	*4.0*	*0.062*
		CFI-S	3 | 24	0.50/0.52	0.25/0.12	0.38	1.2	0.79
		CFI-S + SASS	4 | 24	*0.74/0.079*	**0.40/0.026**	*0.083*	*4.9*	*0.054*
	Combined	**UP-Suicide**	5 | 24	**0.78/0.032**	**0.51/0.0057**	**0.037**	**9.6**	**0.010**

Lastly, we conducted biological pathway analyses on the genes that, after discovery and prioritization, were stepwise changed in suicide completers (n=882) and may be involved in ideation *and* behavior, vs. those that were not stepwise changed (n=589), and that may only be involved in ideation (Table S5-6). The genes involved in ideation map to pathways related to PI3K signaling. The genes involved in behavior map to pathways related to glucocorticoid receptor signaling. This is consistent with ideation *without* behavior being related to neurotrophic factors, and ideation *with* behavior being related to stress.

Clinical information

We used a simple new 22 item scale and app for suicide risk, Convergent Functional Information for Suicidality (CFI-S), which scores in a simple binary fashion and integrates information about known life events, mental health, physical health, stress, addictions, and cultural factors that can influence suicide risk[289 290,320]. Clinical risk predictors and scales are of high interest in the military[309] and in the general population at large[310]. Our scale aims for comprehensiveness, simplicity and quantification similar to a polygenic risk score, and may provide context to the blood biomarker signals. We analyzed which items of the CFI-S scale were the most significantly different between no and high suicidal ideation live participants (Figure 6-3). We identified 7 items that were significantly different: lack of positive relationships/social isolation (p=0.004), substance abuse (p=0.0071), history of impulsive behaviors (p=0.015), lack of religious beliefs (p=0.018), past history of suicidal acts/gestures (p=0.025), rejection (p=0.029),

and history of command auditory hallucinations (p=0.045). Social isolation increases vulnerability to stress, which is independently consistent with our biological marker results.

We also used an 11 item scale for measuring mood and anxiety, the Simplified Affective State Scale (SASS)[320]. The SASS is a set of 11 visual analog scales (7 for mood, 4 for anxiety) that ends up providing a number ranging from 0 to 100 for mood state, and the same for anxiety state.

Testing for predictive ability

The best single increased (risk) biomarker predictor for suicidal ideation state is EPB41L5 (ROC AUC 0.68, p-value 0.06; Pearson Correlation 0.22, p-value 0.03), an increased in expression, Bonferroni validated biomarker (Tables 6-2 and 6-4). This biomarker was also identified co-directionally in our previous male work[320], and has no evidence for involvement in other psychiatric disorders. The best single decreased (protective) biomarker predictor for suicidal ideation is PIK3C3 (ROC AUC 0.65, p-value 0.1; Pearson Correlation -0.21, p-value 0.037), a decreased in expression, Bonferroni validated biomarker (Tables 6-2 and 6-4). PIK3C3 is also decreased in expression in postmortem brains in depression [328].
The best single increased (risk) biomarker predictor for future hospitalizations for suicidality is HTRA1 (ROC AUC 0.84, p-value 0.01; Cox Regression Hazard Ratio 4.55, p- value 0.01), an increased in expression, Bonferroni validated biomarker (Tables 6-2 and 6-4). HTRA1 is also increased in expression in the blood of schizophrenics[329]. The best single decreased (protective) biomarker predictor for

future hospitalizations for suicidality is CSNK1A1 (ROC AUC 0.96, p-value 0.0007; Cox Regression Hazard Ratio 620.5, p-value 0.02), a top discovery and prioritization, non-Bonferroni validated biomarker (Tables 6-2 and 6-4). This biomarker was also identified co-directionally in our previous male work[320]. CSNK1A1 (casein kinase 1, alpha 1) is a circadian clock gene, part of the input into the core clock. It is decreased in expression in suicidality in our work, and decreased in postmortem brains of alcoholics[330]. Interestingly, it is increased in expression by mood stabilizers [331] and by omega-3 fatty acids [65]. PIK3C3 is also a good predictor for future hospitalizations for suicidality (ROC AUC 0.9, p-value 0.011).

Gene symbol / Gene Name	Probesets	**<u>Males</u> Discovery** (Direction of Change) Method/ Internal Score (%)	**<u>Males</u> Participants Tested** With Suicidality / Total	**<u>Males</u> Predictions** ROC/ p-value		**<u>Females</u> Discovery** (Direction of Change) Method/ Internal Score	**<u>Females</u> Participants Tested** With Suicidality / Total	**<u>Females</u> Predictions** ROC/ p-value

Top Biomarkers from Males that were co-directional in Females								
SLC4A4 solute carrier family 4 (sodium bicarbonate cotransporter), member 4	210739_x_at	(I) AP/2 (71%)	SI: 33 | 108 Hosp: 32 | 157	**SI: 0.72/ 2.41E-05** Hosp: 0.44/ 0.87		(I) DE/0 (20%)	SI: 7 | 33 Hosp: 3 | 24	SI: 0.62/0.15 **Hosp: 0.86/ 0.03**
SKA2 spindle and kinetochore associated	225686_at	(D) DE/1 (34%) AP/1 (42%)	SI: 33 | 108 Hosp: 32 | 157	**SI: 0.69/ 0.0002** Hosp: 0.46/ 0.75		(D) DE/0 (6%)	SI: 7 | 33 Hosp: 3 | 24	SI: 0.50/0.51 *Hosp: 0.78/0.07*

complex subunit 2								
Top Biomarkers from Females that were co-directional in Males								
PIK3C3 phosphatidylinositol 3-kinase, catalytic subunit type 3	232086_at	(D) DE/0 (14%)	SI: 33 | 108 Hosp: 32 | 157	**SI: 0.62/0.01** Hosp: 0.5/0.49		(D) DE/1 (49%)	SI: 7 | 33 Hosp: 3 | 24	*SI: 0.65/0.098* **Hosp: 0.9/0.011**
CSNK1A1 casein kinase 1, alpha 1	235464_at	(D) AP/0 (21%)	SI: 33 | 108 Hosp: 31 | 157	**SI: 0.63/0.01** Hosp: 0.5/0.53		(D) DE/4 (86%) AP/1 (36%)	SI: 7 | 33 Hosp: 3 | 24	SI: 0.56/0.316 **Hosp: 0.96/0.001**

Table 6-5. Cross-prediction in the other gender. Examples of top predictive biomarkers from men (Niculescu et al. 2015)[196] and from women (current study) that were changed in expression in the same direction in both genders. The

markers were discovered in just one gender, as they were below the apriori set threshold for discovery (33.3%) in the other gender. Yet they show some ability to predict in the other gender as well. SI- predicting suicidal ideation. Hosp- predicting future hospitalizations for suicidality. **Bold -p-value is significant.**

BCL2, the top CFG scoring biomarker from validation, has good accuracy at predicting future hospitalizations for suicidality (ROC AUC 0.89, p-value 0.007; Cox Regression Hazard Ratio 3.08, p-value 0.01). The panel of 50 validated biomarkers, BioM-50, had even better accuracy at predicting future hospitalizations for suicidality (ROC AUC 0.94, p-value 0.002; Cox Regression Hazard Ratio 89.46, p-value 0.02). Overall, in women, blood biomarkers seemed to perform better for predicting future hospitalizations for suicidality (trait) than for predicting suicidal ideation (state). This is different than the trend we saw in men[320], where blood biomarkers were somewhat better predictors of state than of trait. This gender differences are interesting, and merit exploration in additional future comparative studies.

CFI-S has very good accuracy (ROC AUC 0.84, p-value 0.002; Pearson Correlation 0.39, p-value 0.001) at predicting suicidal ideation in psychiatric participants across diagnostic groups. The other app, SASS, also has very good accuracy (ROC AUC 0.81, p-value 0.003; Pearson Correlation 0.38, p-value 0.0005) at predicting suicidal ideation in women psychiatric participants. The combination of the apps is synergistic (ROC AUC 0.87, p-value 0.0009; Pearson Correlation 0.48, p-value 0.0001). Thus, even without the benefit of potentially more costly, invasive and labor intensive blood biomarker testing, clinically useful predictions could be made with the apps.

Our apriori primary endpoint was a combined universal predictor for suicide (UP-Suicide), composed of the scores in CFI-S and in SASS (Mood, Anxiety), along

with the Bonferroni validated biomarkers (n=50) resulting from the sequential discovery for ideation, prioritization with CFG , and validation for behavior in suicide completers steps. UP-Suicide is a good predictor of suicidal ideation (ROC AUC 0.82, p-value 0.003; Pearson Correlation 0.43, p-value 0.0003) (Table 6-4 and Figure 6-4). UP-Suicide also has good predictive ability for future psychiatric hospitalizations for suicidality (ROC AUC 0.78, p-value 0.032; Cox Regression Hazard Ratio 9.61, p-value 0.01). Overall, while there may post-hoc appear to be better individual predictors for suicidal ideation and for future hospitalizations (Table 6-4), our apriori primary broad spectrum endpoint (UP-Suicide) has been successful, may be more robust to effects of fit to cohort, and might be more generalizable to other populations.

Discussion

We carried out systematic studies to identify clinically useful predictors for suicide in women, an understudied population to date. Our work focuses on identifying markers involved in suicidal ideation *and* suicidal behavior, including suicide completion. Markers involved in behavior may be on a continuum with some of the markers involved in ideation, varying in the degree of expression changes from less severe (ideation) to more severe (behavior). One cannot have suicidal behavior without suicidal ideation, but it may be possible to have suicidal ideation without suicidal behavior.

As a first step, we sought to use a powerful but difficult to conduct within-participant design for discovery of blood biomarkers. Such a design is more

informative than case-control, case-case, or even identical twins designs. The power of a within-participants longitudinal design for multi-omic discovery was first illustrated by Snyder and colleagues[14] in a landmark paper with an n=1. We also have previously demonstrated its power in an initial pilot study in male bipolar participants (n=9 out of 75 showed a switch from a no suicidal ideation to a high suicidal ideation state)[217], and then a larger studies in males with major psychiatric disorders (n=37 out of 217) [320]. In this small (n= 12 out of 51) but very valuable pilot study in women, we followed a similar path.

Second, we conducted whole-genome gene expression discovery studies in the participants that exhibited the switches, using a longitudinal within-participant design, which factors out genetic variability and reduces environmental variability as well. We have demonstrated the power of such a design in our earlier successful pilot work on suicide biomarkers in men with an n=9[217]. Our current n=12 is comparable (Figure 6-2). Genes whose levels of expression tracked suicidal ideation within each participant were identified.

Third, the lists of top candidate biomarkers for suicidal ideation from the discovery and prioritization step (genes with a CFG score of 4 and above, reflecting genes that have maximal experimental internal evidence from this study and/or additional external literature cross-validating evidence), were additionally validated for involvement in suicidal behavior in a cohort of demographically matched suicide completers from the coroner's office (n=6) (Figure 6-2).

We ended up with 50 biomarkers that survived Bonferroni correction (49 genes; one gene, JUN, had two different probesets that validated). Additionally, we tested 65 other biomarkers that were non Bonferroni validated but had maximum internal score of 4 in discovery *and* a CFG score of 6 and above, which means that in addition to strong evidence in this study they also had prior independent evidence of involvement in suicide from other studies. These additional biomarkers are likely involved in suicide but did not make our Bonferroni validation cutoff due to its stringency or potential technical/postmortem artefact reasons (Table 6-2 and S6-2).

Fourth, we describe the use in a female population of the simple and comprehensive phenomic (clinical) risk assessment scale, Convergent Functional Phenomics for Suicidality (CFI-S) scale[320], as well as of the companion app to it for use by clinicians and individuals (Figure S6-2). CFI-S was developed independently of any data from this study, by integrating known risk factors for suicide from the clinical literature. It has a total of 20 items (scored in a binary fashion- 1 for present, 0 for absent, NA for information not available) that assess the influence of mental health factors, as well as of life satisfaction, physical health, environmental stress, addictions, and cultural factors known to influence suicidal behavior. It also has 2 demographics risk factors items: age and gender. The result is a simple polyphenic risk score with an absolute range of 0 to 22, normalized by the number of items on which we had available information, resulting in a score in the range from 0 to 1 (Figures 6-3 and S6-2). We present data validating the CFI-

S in women, in the combined discovery and test cohort of live psychiatric participants (Figure 6-3). We identified the chronic stress of lack of positive relationships/social isolation as the top differential item between no and high SI in women, which is consistent with biological data from the biomarker side of our study.

Fifth, we also assessed anxiety and mood, using a visual analog Simplified Affective State Scale (SASS), previously described by us[320] (Niculescu et al. 2006), for which we now have developed an app version (Figure S6-2). Using a PhenoChipping approach[277] in our discovery cohort of psychiatric participants, we show that anxiety measures cluster with suicidal ideation and CFI-S, and mood measures are in the opposite cluster, suggesting that our participants have high suicidal ideation when they have high anxiety and low mood (Figure 6-2). We would also like to include in the future measures of psychosis, and of stress, to be more comprehensive.

Sixth, we examined how the biomarkers identified by us are able to predict *state* (suicidal ideation) in a larger independent cohort of women psychiatric participants (n= 33 participants).

Seventh, we examined whether the biomarkers are able to predict *trait* (future hospitalizations for suicidal behavior) in women psychiatric participants (n=24).

Last but not least, we demonstrate how our apriori primary endpoint, a comprehensive universal predictor for suicide (UP-Suicide), composed of the

combination of the Bonferroni validated biomarkers (n=50), along with the scores from CFI-S and SASS, predicts state (suicidal ideation) and trait (future psychiatric hospitalizations for suicidality).

The rationale for identifying blood biomarkers as opposed to brain biomarkers is a pragmatic one- the brain cannot be readily accessed in live individuals. Other peripheral fluids, such as cerebrospinal fluid, require more invasive and painful procedures. Nevertheless, it is likely that many of the peripheral blood transcriptomic changes are not necessarily mirroring what is happening in the brain, and vice-versa. The keys to finding peripheral biomarkers[288] are, first, to have a powerful discovery approach, such as our within-participant design, that closely tracks the phenotype you are trying to measure and reduces noise. Second, cross-validating and prioritizing the results with other lines of evidence, such as brain gene expression and genetic data, are important in order to establish relevance to disease and generalizability of findings. Third, it is important to validate for behavior in an independent cohort with a robust and relevant phenotype, in this case suicide completers. Fourth, testing for predictive ability in independent/prospective cohorts is a must (Figure S6-1).

Biomarkers that survive such a rigorous step-wise discovery, prioritization, validation and testing process are likely directly relevant to the disorder studied. As such, we endeavored to study their biology, whether they are involved in other psychiatric disorders or are relatively specific for suicide, and whether they are

modulated by existing drugs in general, and drugs known to treat suicidality in particular.

We have identified a series of biomarkers that seem to be changed in opposite direction in suicide vs. in treatments with omega-3 fatty acids, lithium, clozapine (Table S6-4). These biomarkers could potentially be used to stratify patients to different treatment approaches, and monitor their response. BCL2, JUN, GHA1, ENTPD1, ITIH5, MBNL1, and SSBP2 are changed in expression by two of these three treatments, suggesting they may be core to the anti-suicidal mechanism of these drugs. Interestingly, MBNL1 which is decreased in expression in suicidality, was identified as increased in expression in longevity/ healthy aging[332] . BCL2, CAT, and JUN may be useful blood pharmacogenomic markers of response to lithium. CD84, MBNL1, and RAB22A may be useful blood pharmacogenomic markers of response to clozapine. NDRG1, FOXP1, AFF3, ATXN1, CSNK1A1, ENTPD1, ITIH5, PRDX3, and SSBP2 may be useful blood pharmacogenomic markers of response to omega-3 fatty acids. Three existing drugs used for other indications have been identified as targeting the top suicide biomarkers identified by us (Table S6-4), and could potentially be re-purposed for testing in treatment of acute suicidality: anakinra (inhibiting ILR1), enzastaurin (inhibiting AKT3), and tesevatinib (inhibiting EPHB4). Additionally, Connectivity Map[333] analyses (Table S6-6) identified novel compounds that induce gene expression signatures that are the opposite of those present in suicide, and might generate leads and/or be tested for use to treat/prevent suicidality, including

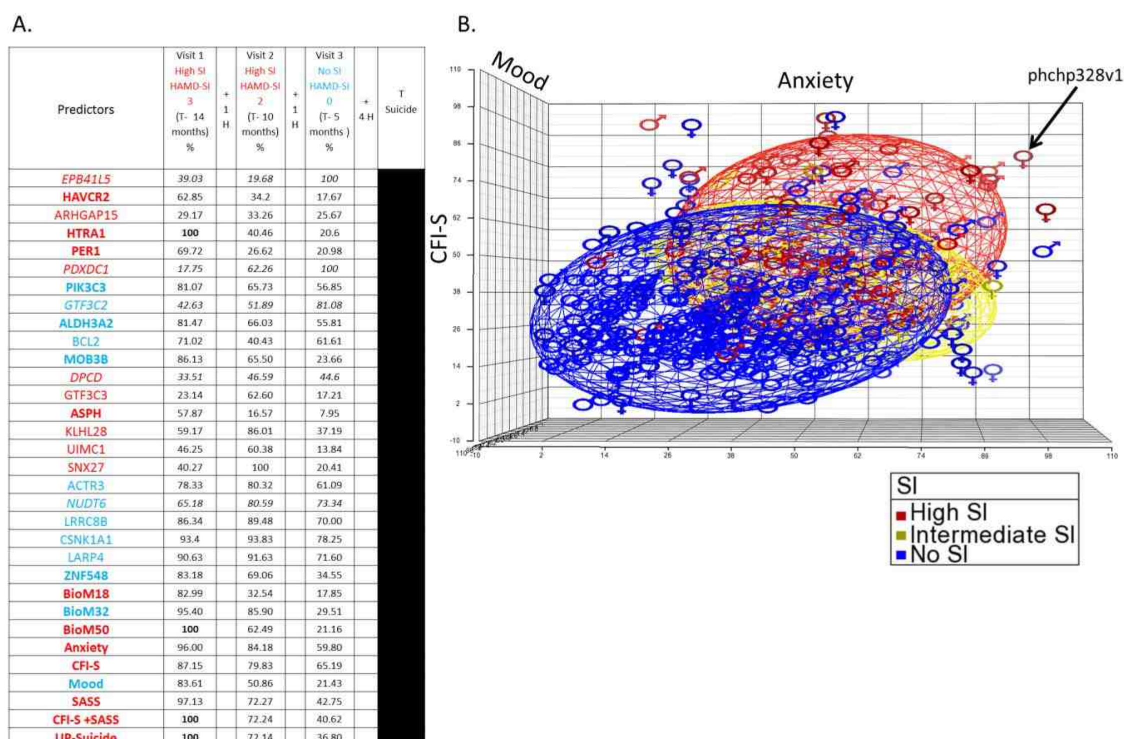
mifepristone, LY294002, acetylsalicylic acid, estradiol, buspirone, corticosterone, metformin, diphenhydramine, haloperidol, and fluoxetine (Table S6-6).

Of note, a number of biomarkers from the current study in women reproduce and are co-directional with our previous findings in men (Table 6-5, Table 6-2 and Table S6-2), whereas others had changes in opposite directions (Table 6-2 and Table S6-2), underlying the issue of biological context and differences in suicidality between the two genders. This avenue merits attention in the field, and detailed future comparative studies, as do studies by diagnostic groups.

Prior to any testing, we planned to use a comprehensive combination of genomic data (specifically, the top validated biomarkers) and phenomic data (specifically, the CFI-S and the SASS) as the primary endpoint measure, a broad-spectrum universal predictor (UP-Suicide) for state suicidal ideation and trait future hospitalizations. It has not escaped our attention that certain single biomarkers, particular phenotypic items, or combinations thereof seem to perform better than the UP-Suicide in one or another type of prediction (see Table 4-6). However, since such markers and combinations were not chosen by us apriori and such insights derive from testing, we cannot exclude a fit to cohort effect for them and reserve judgement as to their robustness as predictors until further testing in additional independent cohorts, by us and others. What we can put forward for now based on the current work is the UP-Suicide, which seems to be a robust predictor across different scenarios and diagnostic groups.

Our study has a number of limitations. All this work was carried out in psychiatric patients, a high risk group, and it remains to be seen how such predictors apply to non-psychiatric participants. For the UP-Suicide testing, the prevalence rate for suicidality in our test cohorts was 21% (7 out of 33 for suicidal ideation, 5 out of 24 for future hospitalizations) (Table 4-6). Of note, this rate was remarkably similar to our previous work in men[320]. It is to be noted that the incidence of suicidality in the general population is lower, for example at 1.5% in adolescents in an European cohort[312] and estimates of 0.2 to 2% in the US[231], which underlines the rationale of using a very high risk group like we did for magnifying and enabling signal detection with a relatively small N. Over 40% of the live participants from the discovery cohort (5 out of 12) and independent test cohort (14 out of 33) are non-VA, and all the suicide completers used for validation are from the general population, not VA, so we believe our results have broader relevance. Studies with larger numbers and longer follow-up, currently ongoing, as well as studies in different clinical settings, may provide more generalizability.

The current studies were carried out exclusively in females. Similar work is needed in larger meta-analyses across gender, in participants with and without psychiatric disorders, to find generalizable predictors. Conversely, a narrow focus by gender, diagnosis (or lack of), and perhaps age, may be needed to find more individualized predictors. Such work is ongoing in our group.


A.

Predictors	Visit 1 High SI HAMD-SI 3 (T- 14 months) %	+ 1 H	Visit 2 High SI HAMD-SI 2 (T- 10 months) %	+ 1 H	Visit 3 No SI HAMD-SI 0 (T- 5 months) %	+ 4 H	T Suicide
EPB41L5	*39.03*		*19.68*		*100*		
HAVCR2	62.85		34.2		17.67		
ARHGAP15	29.17		33.26		25.67		
HTRA1	**100**		40.46		20.6		
PER1	69.72		26.62		20.98		
PDXDC1	*17.75*		*62.26*		*100*		
PIK3C3	81.07		65.73		56.85		
GTF3C2	*42.63*		*51.89*		*81.08*		
ALDH3A2	81.47		66.03		55.81		
BCL2	71.02		40.43		61.61		
MOB3B	86.13		65.50		23.66		
DPCD	*33.51*		*46.59*		*44.6*		
GTF3C3	23.14		62.60		17.21		
ASPH	57.87		16.57		7.95		
KLHL28	59.17		86.01		37.19		
UIMC1	46.25		60.38		13.84		
SNX27	40.27		100		20.41		
ACTR3	78.33		80.32		61.09		
NUDT6	*65.18*		*80.59*		*73.34*		
LRRC8B	86.34		89.48		70.00		
CSNK1A1	93.4		93.83		78.25		
LARP4	90.63		91.63		71.60		
ZNF548	83.18		69.06		34.55		
BioM18	82.99		32.54		17.85		
BioM32	95.40		85.90		29.51		
BioM50	**100**		62.49		21.16		
Anxiety	96.00		84.18		59.80		
CFI-S	87.15		79.83		65.19		
Mood	83.61		50.86		21.43		
SASS	97.13		72.27		42.75		
CFI-S +SASS	**100**		72.24		40.62		
UP-Suicide	**100**		72.14		36.80		

Figure 6-5. Study participant who committed suicide. Subject phchp328 was a 38-year-old divorced Caucasian female with a long history of MDD, PTSD, BP and polysubstance abuse/dependence. She had multiple psychiatric hospitalizations due to suicidal ideation (n =21) and due to suicidal attempts (n=3), in the 5 years before her suicide. She committed suicide by overdose with pills, leaving behind a suicide note addressed to her mother. (a) Percentile for scores on top predictors in all the female subjects in this study (n=105 for biomarkers and n =88 for apps and UP-Suicide). Her panel of Bonferroni validated biomarkers (BioM50) score, apps score (CFI-S+SASS), and UP-Suicide predictor score at a study visit (Visit 1) were at the 100% of the scores of all the psychiatric participant visits tested in this current study. Of note, that testing was conducted during an inpatient hospitalization due to suicidal ideation. While her scores did improve at subsequent

outpatient testing visits (Visits 2 and 3), this high watermark score indicated her high risk. After the last testing visit in our study, she had four subsequent psychiatric hospitalizations: three due to suicidal ideation, one for opioid withdrawal/detox (the last one), ending 2 weeks before date of committing suicide (T). For decreased biomarkers, a higher percentile corresponds to lower expression values. Only 5 of the 32 predictors (biomarkers, clinical, combined) were discordant between the highest and lowest SI visit (italicized). In all, 17 of the 32 predictors (bold) were stepwise decreased corresponding to her SI scores. One of the biomarkers (HTRA1) was in the 100% of the subjects tested, as was the panel of 50 validated markers (BioM-50), the combination of the clinical measures/apps (CFI-S+SASS), and the combined biomarker panels and clinical/ apps predictor (UP-Suicide). (b) Tri-dimensional representation of the percentilized scores of the combination of the two apps, CFI-S and SASS (Anxiety and Mood) of all the female participant visits tested in the current study (n= 87) and all the male participant visits in our previous work (n=317). A tri-dimensional scatter plot was created using Partek. Tri-dimensional 95% confidence intervals were inserted as ellipsoids, color coded blue, yellow and red for No SI, Intermediate SI and High SI, respectively. Subject phchp328visit1 had the highest Euclidian D (distance from origin), as indicated by the arrow. This is the only subject that completed suicide as far as we know, as of the end of this study in November 2015. BP, bipolar disorder; MDD, major depressive disorder; PTSD, post-traumatic stress disorder.

In conclusion, we have advanced the biological understanding of suicidality in women, highlighting behavioral and biological mechanisms related to inflammation, neurotrophic factors, circadian clock, stress response, and apoptosis. Biomarkers that may track treatment response to lithium and intriguingly, omega-3 fatty acids, have been identified. Of equal importance, we developed instruments (biomarkers and apps) for predicting suicidality, that do not require asking the person assessed if they have suicidal thoughts, as individuals who are truly suicidal often do not share that information with people close to them or with clinicians. We propose that the widespread use of such risk prediction tests as part of routine or targeted healthcare assessments will lead to early disease interception followed by preventive lifestyle modifications or treatment. Given the magnitude and urgency of the problem, the importance of efforts to implement such tools cannot be overstated. We note that we have sadly lost one study participant to suicide (Figure 6-5), which in retrospect was highlighted by UP-Suicide as being the highest participant risk in our cohort.

Chapter 7: Conclusions

These works indicate how the field may move forward by integrating data from genomics, transcriptomics, and phenomics, which could yield a more precision approach to psychiatric medicine. While evidence from any one approach will provide information on an individual patients risk for disease, a single line of evidence is often insufficient. In the cases of the genetic studies of schizophrenia and alcoholism discussed above in chapter 1 and chapter 2, respectively, we find convincing evidence for a panel of SNPs, when combined into a polygenic GRP score, are involved in the genetic risk for each disease. Further, we are able to use the discovered panel of SNPs and replicate in completely independent cohorts and demonstrate a statistical separation between groups. Despite the strength of these findings, this panel alone still had insufficient sensitivity and specificity to discriminate individual subjects. Genetic polymorphisms clearly play an important role but more information from an individual is necessary to improve predictive accuracy and provide more precise and targeted treatment to patients.

CFG

Using a Bayesian integration of evidence from studies in human patients and multiple animal models in the form of Convergent Functional Genomics, as discussed in previous chapters, is one way to prioritize disease-relevant genes. Following our initial experiments, CFG cross-validates experimental findings across approaches and tissue types. Animal studies provide the sensitivity to detect mechanistic changes in developed models of disease. Human studies provide the

specificity of the target disease. We used this approach to prioritize our own experimental findings.

In the schizophrenia study, thirteen SNPs in proximity to DISC1 were identified as significantly changed in schizophrenia, but they were otherwise unremarkable compared to the other 45,972 SNPs identified as 'significant'. CFG allowed us to integrate numerous hypothesis driven findings such as prior knowledge that DISC1 expression has been shown to be elevated during acute psychotic episodes and to be reduced by treatment[44]. In the suicide work SLC4A4 tracked suicidal ideation in our longitudinal discovery cohort and it was also shown to be increased in suicide completers, but it was not the 'strongest' finding in these discovery analyses. When prioritized with what is already known in the field by using CFG we find that a SNP associated with SLC4A4 has previously been associated with suicide[222], and that expression is altered in the prefrontal cortex of schizophrenia patients who committed suicide when compated to schizophrenia patients who died of other causes[297]. SLC4A4 had strong evidence from our own experiments for tracking suicidal ideation, was found to be concordantly increased in our cohort of suicide completers, but it was the addition of CFG that made SLC4A4 our top prioritized marker. In follow up testing it was our strongest predictor of suicidal ideation, with a 93% AUC for bipolar patients. Prior evidence in the published literature is underleveraged. This work provides additional pragmatic proof of principle as to how this rich and abundant resource could be better used by the field.

Within subject longitudinal design

Another powerful approach that improves power and precision is the use of within subject design of a longitudinal cohort of participants who provided a wealth of dimensional quantitative phenotypes in addition to gene expression. This allows us to focus on how genes track and change along with an individuals phenotype, filtering out artifacts produced by inter individual differences. It's true that a large focus of this work used CFG to prioritize findings, but where we utilized a within subject analysis in the suicide project we also sought to test the novel markers that were the best at tracking suicidal ideation in the discovery cohort. In the male suicide work, reported in Chapter 5, CLIP4 was the most significantly decreased biomarker in tracking suicidal ideation within individuals. In an independent cohort of male bipolar patients this marker was able to distinguish no SI from high SI with an AUC of 0.76. Even without the filtering power derived from CFG, this approach is able to identify biomarkers with potential clinical utility.

It could be that previous findings in the literature and under developed drugs could see new utility within certain subsets of individuals, and this data may be lost without a more precise approach. Several papers in recent years have highlighted the power of carefully designed trials within even a single individual[14 334]. This is an approach and design which should see wider implementation in the field.

Dimensionality

Perhaps the most important message that should be taken from this work is the dimensionality of an individual patient and how this influences suicide risk. The genes identified to be associated with suicidal ideation may not be specific to suicide per se, but may be related to impulsive, stressed, or agitated states that influence suicidal ideation. Providing additional contextual precision is essential to improving diagnostic and predictive accuracy.

We have in our lab quantitative scales, developed by Dr. Niculescu, for assessing mood and anxiety state. Dr. Niculescu has also developed a 22-item scale and Android app for assessing suicide risk called the CFI-S. These scales alone have shown the ability to discriminate between no and high SI states. When combined so that we are using an individual's mood, anxiety, and suicidal risk factors together in the same model with the biomarkers we get a much more powerful ability to distinguish those with and without suicidal ideation.

This approach should be applied broadly. It is the rare exception that a single gene or biomarker is sufficiently sensitive or specific to define an illness or behavior. The fields of psychiatry and psychology are ripening with increasing information and knowledge from areas such as genetics, transcriptomics, and imaging. Harvesting this information through dynamic integrations at an individual level will enable personalized precision medicine, improving our understanding of psychiatric disease and improving patient treatment and outcomes.

Limitations

While these studies represent a comprehensive approach to integrating genetics, functional genomics and phenomics, there are several limitations which should be addressed by future studies. Firstly, the criteria for assigning of SNPs or probes to genes is a decision that could have profound impact on the analysis bioinformatics analyses that occur downstream. For GWAS study we followed the assumption that the most proximal gene should be assigned to a given SNP. This may not be true in all cases, particularly for intergenic SNPs. This could open up the possibility of false positive or negative findings in the initial discovery analysis. Using the convergent approach employed in the preceding chapters reduces the risk of false positives through the use of prior peer reviewed evidence. This is also addressed by using outside cohorts to test the initial findings, showing reproducibility.

The issue of false negatives is a more difficult one. A more expansive selection criteria could be employed. One popular method has been to select genes which fall within a set distance from the identified SNP (often measured in base pairs). This method doesn't necessarily eliminate the problem of false negatives, as the 'true positive' gene impacted by a SNP could still fall outside of the selected range. The methodology we selected is a trade-off which might increase the rate of false negatives.

Secondly, while the weighting of the CFG follows a logical and a priori locked algorithm, it is true that altering the weighting of the different lines of evidence

could alter the prioritization of top genes. While this would have absolutely no impact on the list of genes discovered in initial experiments it is still not a trivial matter. The method employed is simple and straight forward and consist of assuming that evidence in human studies is more specific to human illness (and therefore is weighed twice) and evidence found in the target tissue, the brain, to be weighed twice as much as evidence found in the periphery or by genetics. This decision to weight the CFG was made following the schizophrenia work, and the algorithm was locked for all of the work which followed.

A danger in not using a locked algorithm is an increase in the possibility of overfitting findings to the cohorts being studied and losing generalizability. All experiments, of course, require fitting of data derived from the cohort being studied. The challenge is optimizing how much of the fit is due to the variable being studied (suicidal ideation) and how much is due to uncontrolled confounding variables in the cohort. The former provides the signal of gene expression that tracks suicidal ideation, the latter creates noise that reduces generalizability and confounds findings. The goal of this work has been to identify genes which may serve as biomarkers for psychiatry. Altering the algorithm with goal of developing more empirically derived weights introduces an additional variable, and it might become difficult to discern whether an improvement in predictive accuracy is due to better fitting the signal or the noise in the data. Available additional outside independent cohorts were used for the purpose of testing those findings and providing evidence for the generalizability of biomarkers. Future meta analyses

seeking to refine the algorithm across multiple independent data sets could be an important step to improving biomarker discovery. Future work should examine the utility of more empirical methods of assigning weight to the different lines of evidence.

Future Directions

Moving forward the exponential increase in available data provides great opportunity for enabling precision medicine in psychiatry. This massive increase in data also presents a great challenge in how to best integrate that information and how to screen it for quality. Convergent Functional Genomics is a tool with great potential to separate high quality information from the vast quantity that is being produced on a daily basis. This work highlights how it can be used to prioritize reproducible biomarkers for disease states and traits. Room for improvement exists, however, as discussed in the limitations section. More empirical algorithmic means for weighting the individual lines of evidence should improve the predictive ability of identified biomarkers, but great care will need to be taken to ensure that the new weighting reflects high quality signal in the data and not just a fit to noise in the individual cohorts used to create the weighting.

Another challenge that must be addressed is the relative non-specificity of any one individual biomarker. Implications of glutamate, serotonin, polyamines, and SKA2 among others are a common theme identified not only by us but by many other labs to associate with suicidal behavior. The reproducibility of these findings is not necessarily evidence of a role in suicide specifically but of other

associated states and conditions which are tied to risk for suicidal behavior. Future work should continue to try to place these findings into a greater dimensional context with recognition that these risk factors alone are neither necessary nor are they sufficient to predict suicidal outcomes.

Bibliography

1. Roger, V.L. Epidemiology of heart failure. *Circ Res* **113**, 646-59 (2013).
2. Meltzer, H.Y. Clozapine: balancing safety with superior antipsychotic efficacy. *Clin Schizophr Relat Psychoses* **6**, 134-44 (2012).
3. McEvoy, J.P. *et al.* Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior atypical antipsychotic treatment. *Am J Psychiatry* **163**, 600-10 (2006).
4. Fahlke, C. *et al.* Rearing experiences and stress-induced plasma cortisol as early risk factors for excessive alcohol consumption in nonhuman primates. *Alcohol Clin Exp Res* **24**, 644-50 (2000).
5. Walker, E.F. *et al.* Longitudinal changes in cortisol secretion and conversion to psychosis in at-risk youth. *J Abnorm Psychol* **119**, 401-8 (2010).
6. Nugent, C.A., Nichols, T. & Tyler, F.H. Diagnosis of Cushing's Syndrome; Single Dose Dexamethasone Suppression Test. *Arch Intern Med* **116**, 172-6 (1965).
7. Raff, H. Cushing syndrome: update on testing. *Endocrinol Metab Clin North Am* **44**, 43-50 (2015).
8. Coryell, W. & Schlesser, M. The dexamethasone suppression test and suicide prediction. *Am J Psychiatry* **158**, 748-53 (2001).

9. Jokinen, J., Nordstrom, A.L. & Nordstrom, P. ROC analysis of dexamethasone suppression test threshold in suicide prediction after attempted suicide. *J Affect Disord* **106**, 145-52 (2008).

10. Lahiri, D.K. & Maloney, B. The "LEARn" (Latent Early-life Associated Regulation) model integrates environmental risk factors and the developmental basis of Alzheimer's disease, and proposes remedial steps. *Exp Gerontol* **45**, 291-6 (2010).

11. Riedel, B.C., Thompson, P.M. & Brinton, R.D. Age, APOE and sex: Triad of risk of Alzheimer's disease. *J Steroid Biochem Mol Biol* **160**, 134-47 (2016).

12. Genin, E. *et al.* APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* **16**, 903-7 (2011).

13. Thambisetty, M. *et al.* Effect of complement CR1 on brain amyloid burden during aging and its modification by APOE genotype. *Biol Psychiatry* **73**, 422-8 (2013).

14. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293-307 (2012).

15. Le-Niculescu, H. *et al.* Identifying blood biomarkers for mood disorders using convergent functional genomics. *Mol Psychiatry* **14**, 156-74 (2009).

16. Keri, S., Szabo, C. & Kelemen, O. Blood biomarkers of depression track clinical changes during cognitive-behavioral therapy. *J Affect Disord* **164**, 118-22 (2014).

17. Matise, T.C. *et al.* A second-generation combined linkage physical map of the human genome. *Genome Res* **17**, 1783-6 (2007).

18. Ellaithy, A., Younkin, J., Gonzalez-Maeso, J. & Logothetis, D.E. Positive allosteric modulators of metabotropic glutamate 2 receptors in schizophrenia treatment. *Trends Neurosci* **38**, 506-16 (2015).

19. Niculescu, A.B., 3rd *et al.* Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. *Physiol Genomics* **4**, 83-91 (2000).

20. Ogden, C.A. *et al.* Candidate genes, pathways and mechanisms for bipolar (manic-depressive) and related disorders: an expanded convergent functional genomics approach. *Mol Psychiatry* **9**, 1007-29 (2004).

21. Bertsch, B. *et al.* Convergent functional genomics: a Bayesian candidate gene identification approach for complex disorders. *Methods* **37**, 274-9 (2005).

22. Rodd, Z.A. *et al.* Candidate genes, pathways and mechanisms for alcoholism: an expanded convergent functional genomics approach. *Pharmacogenomics J* **7**, 222-56 (2007).

23. Le-Niculescu, H. *et al.* Towards understanding the schizophrenia code: an expanded convergent functional genomics approach. *Am J Med Genet B Neuropsychiatr Genet* **144B**, 129-58 (2007).

24. Kurian, S.M. *et al.* Identification of blood biomarkers for psychosis using convergent functional genomics. *Mol Psychiatry* **16**, 37-58 (2011).

25. International Schizophrenia, C. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).

26. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753-7 (2009).

27. Brennand, K.J. *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221-5 (2011).

28. Le-Niculescu, H. *et al.* Towards understanding the schizophrenia code: an expanded convergent functional genomics approach. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **144B**, 129-58 (2007).

29. Patel, S.D. *et al.* Coming to grips with complex disorders: genetic risk prediction in bipolar disorder using panels of genes identified through convergent functional genomics. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 850-77 (2010).

30. Niculescu, A.B. & Le-Niculescu, H. Convergent Functional Genomics: what we have learned and can learn about genes, pathways, and mechanisms. *Neuropsychopharmacology* **35**, 355-6 (2010).

31. Ayalew, M. *et al.* Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol Psychiatry* **17**, 887-905 (2012).

32. Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* **104**, 19428-33 (2007).

33. Tsutsumi, A. *et al.* The genetic validation of heterogeneity in schizophrenia. *Behav Brain Funct* **7**, 43 (2011).

34. Ikeda, M. *et al.* Genome-wide association study of schizophrenia in a Japanese population. *Biol Psychiatry* **69**, 472-8 (2011).

35. Rasetti, R. *et al.* Altered cortical network dynamics: a potential intermediate phenotype for schizophrenia and association with ZNF804A. *Arch Gen Psychiatry* **68**, 1207-17 (2011).

36. Ishizuka, K. *et al.* DISC1-dependent switch from progenitor proliferation to migration in the developing cortex. *Nature* **473**, 92-6 (2011).

37. Brandon, N.J. & Sawa, A. Linking neurodevelopmental and synaptic theories of mental illness through DISC1. *Nat Rev Neurosci* **12**, 707-22 (2011).

38. Porteous, D.J., Millar, J.K., Brandon, N.J. & Sawa, A. DISC1 at 10: connecting psychiatric genetics and neuroscience. *Trends Mol Med* **17**, 699-706 (2011).

39. Hattori, T. *et al.* DISC1 regulates cell-cell adhesion, cell-matrix adhesion and neurite outgrowth. *Mol Psychiatry* **15**, 778, 798-809 (2010).

40. Callicott, J.H. *et al.* Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proc Natl Acad Sci U S A* **102**, 8627-32 (2005).

41. Carless, M.A. *et al.* Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. *Mol Psychiatry* **16**, 1096-104 (2011).

42. Singh, K.K. *et al.* Common DISC1 polymorphisms disrupt Wnt/GSK3beta signaling and brain development. *Neuron* **72**, 545-58 (2011).

43. Nakata, K. *et al.* DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proc Natl Acad Sci U S A* **106**, 15873-8 (2009).

44. Olincy, A. *et al.* Elevated DISC1 transcript levels in PBMCs during acute psychosis in patients with schizophrenia. *Transl Biomed* **2**(2011).

45. Ayhan, Y. *et al.* Differential effects of prenatal and postnatal expressions of mutant human DISC1 on neurobehavioral phenotypes in transgenic mice: evidence for neurodevelopmental origin of major psychiatric disorders. *Mol Psychiatry* **16**, 293-306 (2011).

46. Millar, J.K. *et al.* Disrupted in schizophrenia 1 and phosphodiesterase 4B: towards an understanding of psychiatric illness. *J Physiol* **584**, 401-5 (2007).

47. Wang, Q. *et al.* The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. *Mol Psychiatry* **16**, 1006-23 (2011).

48. Kang, E. *et al.* Interaction between FEZ1 and DISC1 in Regulation of Neuronal Development and Risk for Schizophrenia. *Neuron* **72**, 559-71 (2011).

49. Singh, K.K. *et al.* Dixdc1 is a critical regulator of DISC1 and embryonic cortical development. *Neuron* **67**, 33-48 (2010).

50. Lennertz, L. *et al.* Novel schizophrenia risk gene TCF4 influences verbal learning and memory functioning in schizophrenia patients. *Neuropsychobiology* **63**, 131-6 (2011).

51. Li, T. *et al.* Common variants in major histocompatibility complex region and TCF4 gene are significantly associated with schizophrenia in Han Chinese. *Biol Psychiatry* **68**, 671-3 (2010).

52. Quednow, B.B. *et al.* The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia spectrum and healthy volunteers. *J Neurosci* **31**, 6684-91 (2011).

53. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744-7 (2009).

54. Mudge, J. *et al.* Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PLoS One* **3**, e3625 (2008).

55. Baruch, K. *et al.* Association between golli-MBP and schizophrenia in the Jewish Ashkenazi population: are regulatory regions involved? *Int J Neuropsychopharmacol* **12**, 885-94 (2009).

56. Tkachev, D. *et al.* Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* **362**, 798-805 (2003).

57. Glatt, S.J. *et al.* Similarities and differences in peripheral blood gene-expression signatures of individuals with schizophrenia and their first-degree biological relatives. *Am J Med Genet B Neuropsychiatr Genet* **156B**, 869-87 (2011).

58. Le-Niculescu, H. *et al.* Phenomic, convergent functional genomic, and biomarker studies in a stress-reactive genetic animal model of bipolar disorder and co-morbid alcoholism. *Am J Med Genet B Neuropsychiatr Genet* **147B**, 134-66 (2008).

59. Haroutunian, V. & Davis, K.L. Introduction to the special section: Myelin and oligodendrocyte abnormalities in schizophrenia. *Int J Neuropsychopharmacol* **10**, 499-502 (2007).

60. Takahashi, N., Sakurai, T., Davis, K.L. & Buxbaum, J.D. Linking oligodendrocyte and myelin dysfunction to neurocircuitry abnormalities in schizophrenia. *Prog Neurobiol* **93**, 13-24 (2011).

61. Konradi, C., Sillivan, S.E. & Clay, H.B. Mitochondria, oligodendrocytes and inflammation in bipolar disorder: Evidence from transcriptome studies points to intriguing parallels with multiple sclerosis. *Neurobiol Dis* **45**, 37-47 (2012).

62. Pae, C.U. *et al.* The impact of heat shock protein 70 gene variations on clinical presentation and outcome in schizophrenic inpatients. *Neuropsychobiology* **59**, 135-41 (2009).

63. Arion, D., Unger, T., Lewis, D.A., Levitt, P. & Mirnics, K. Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. *Biol Psychiatry* **62**, 711-21 (2007).

64. Le-Niculescu, H. *et al.* Convergent functional genomics of anxiety disorders: translational identification of genes, biomarkers, pathways and mechanisms. *Transl Psychiatry* **1**, e9 (2011).

65. Le-Niculescu, H. *et al.* Convergent functional genomic studies of omega-3 fatty acids in stress reactivity, bipolar disorder and alcoholism. *Transl Psychiatry* **1**, e4 (2011).

66. Roussos, P. *et al.* Molecular and genetic evidence for abnormalities in the nodes of Ranvier in schizophrenia. *Arch Gen Psychiatry* **69**, 7-15 (2012).

67. Schwarz, E. *et al.* Identification of a biological signature for schizophrenia in serum. *Mol Psychiatry* **17**, 494-502 (2012).

68. Ujike, H. *et al.* CNR1, central cannabinoid receptor gene, associated with susceptibility to hebephrenic schizophrenia. *Mol Psychiatry* **7**, 515-8 (2002).

69. Xu, B. *et al.* Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. *Proc Natl Acad Sci U S A* **106**, 16746-51 (2009).

70. Eggan, S.M., Hashimoto, T. & Lewis, D.A. Reduced cortical cannabinoid 1 receptor messenger RNA and protein expression in schizophrenia. *Arch Gen Psychiatry* **65**, 772-84 (2008).

71. Newsom, R.J. *et al.* Cannabinoid receptor type 1 antagonism significantly modulates basal and loud noise induced neural and hypothalamic-pituitary-adrenal axis responses in male Sprague-Dawley rats. *Neuroscience* **204**, 64-73 (2012).

72. Murray, R.M., Morrison, P.D., Henquet, C. & Di Forti, M. Cannabis, the mind and society: the hash realities. *Nat Rev Neurosci* **8**, 885-95 (2007).

73. Moghaddam, B. & Javitt, D. From Revolution to Evolution: The Glutamate Hypothesis of Schizophrenia and its Implication for Treatment. *Neuropsychopharmacology* **37**, 4-15 (2012).

74. Auerbach, B.D., Osterweil, E.K. & Bear, M.F. Mutations causing syndromic autism define an axis of synaptic pathophysiology. *Nature* **480**, 63-8 (2011).

75. Cisse, M. *et al.* Reversing EphB2 depletion rescues cognitive functions in Alzheimer model. *Nature* **469**, 47-52 (2011).

76. Javitt, D.C. *et al.* Translating glutamate: from pathophysiology to treatment. *Sci Transl Med* **3**, 102mr2 (2011).

77. Levinson, D.F. *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* **168**, 302-16 (2011).

78. Vacic, V. *et al.* Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* **471**, 499-503 (2011).

79. Tregellas, J.R. *et al.* Effects of an alpha 7-nicotinic agonist on default network activity in schizophrenia. *Biol Psychiatry* **69**, 7-11 (2011).

80. Charych, E.I., Jiang, L.X., Lo, F., Sullivan, K. & Brandon, N.J. Interplay of palmitoylation and phosphorylation in the trafficking and localization of phosphodiesterase 10A: implications for the treatment of schizophrenia. *J Neurosci* **30**, 9027-37 (2010).

81. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* **17**, 142-53 (2012).

82. Niculescu, A.B. *et al.* PhenoChipping of psychotic disorders: a novel approach for deconstructing and quantitating psychiatric phenotypes. *Am J Med Genet B Neuropsychiatr Genet* **141B**, 653-62 (2006).

83. Niculescu, A.B., 3rd. Polypharmacy in oligopopulations: what psychiatric genetics can teach biological psychiatry. *Psychiatr Genet* **16**, 241-4 (2006).

84. Niculescu, A.B., 3rd, Schork, N.J. & Salomon, D.R. Mindscape: a convergent perspective on life, mind, consciousness and happiness. *J Affect Disord* **123**, 1-8 (2010).

85. Hennah, W. *et al.* DISC1 association, heterogeneity and interplay in schizophrenia and bipolar disorder. *Mol Psychiatry* **14**, 865-73 (2009).

86. Bord, L. *et al.* Primate disrupted-in-schizophrenia-1 (DISC1): high divergence of a gene for major mental illnesses in recent evolutionary history. *Neurosci Res* **56**, 286-93 (2006).

87. Craddock, N. & Owen, M.J. The Kraepelinian dichotomy - going, going... but still not gone. *Br J Psychiatry* **196**, 92-5 (2010).

88. Williams, H.J. *et al.* Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. *Hum Mol Genet* **20**, 387-91 (2011).

89. Buckley, P.F., Miller, B.J., Lehrer, D.S. & Castle, D.J. Psychiatric comorbidities and schizophrenia. *Schizophr Bull* **35**, 383-402 (2009).

90. Allen, N.C. *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* **40**, 827-34 (2008).

91. Jia, P., Sun, J., Guo, A.Y. & Zhao, Z. SZGR: a comprehensive schizophrenia gene resource. *Mol Psychiatry* **15**, 453-62 (2010).

92. Tan, H.Y., Callicott, J.H. & Weinberger, D.R. Intermediate phenotypes in schizophrenia genetics redux: is it a no brainer? *Mol Psychiatry* **13**, 233-8 (2008).

93. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519-23 (2011).

94. Degnan, J.H. *et al.* Genomics and genome-wide association studies: an integrative approach to expression QTL mapping. *Genomics* **92**, 129-33 (2008).

95. O'Donovan, M.C., Craddock, N.J. & Owen, M.J. Genetics of psychosis; insights from views across the genome. *Hum Genet* **126**, 3-12 (2009).

96. Renaud, S. & de Lorgeril, M. Wine, alcohol, platelets, and the French paradox for coronary heart disease. *Lancet* **339**, 1523-6 (1992).

97. Di Castelnuovo, A. *et al.* Alcohol dosing and total mortality in men and women: an updated meta-analysis of 34 prospective studies. *Arch Intern Med* **166**, 2437-45 (2006).

98. Gmel, G., Gutjahr, E. & Rehm, J. How stable is the risk curve between alcohol and all-cause mortality and what factors influence the shape? A precision-weighted hierarchical meta-analysis. *Eur J Epidemiol* **18**, 631-42 (2003).

99. Foster, J.H., Powell, J.E., Marshall, E.J. & Peters, T.J. Quality of life in alcohol-dependent subjects--a review. *Qual Life Res* **8**, 255-61 (1999).

100. Kendler, K.S., Ohlsson, H., Sundquist, J. & Sundquist, K. Alcohol Use Disorder and Mortality Across the Lifespan: A Longitudinal Cohort and Co-relative Analysis. *JAMA Psychiatry* **73**, 575-81 (2016).

101. Koob, G.F. A role for brain stress systems in addiction. *Neuron* **59**, 11-34 (2008).

102. Schuckit, M.A. Alcohol-use disorders. *Lancet* **373**, 492-501 (2009).

103. Treutlein, J. *et al.* Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry* **66**, 773-84 (2009).

104. Frank, J. *et al.* Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict Biol* **17**, 171-80 (2012).

105. Gelernter, J. *et al.* Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry* **19**, 41-9 (2014).

106. Niculescu, A.B. & Le-Niculescu, H. The P-value illusion: how to improve (psychiatric) genetic studies. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 847-9 (2010).

107. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575 (2007).

108. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).

109. Edwards, S. *et al.* Traumatic stress reactivity promotes excessive alcohol drinking and alters the balance of prefrontal cortex-amygdala activity. *Transl Psychiatry* **3**, e296 (2013).

110. Koob, G.F. *et al.* Addiction as a stress surfeit disorder. *Neuropharmacology* **76 Pt B**, 370-82 (2014).

111. Schank, J.R., Ryabinin, A.E., Giardino, W.J., Ciccocioppo, R. & Heilig, M. Stress-related neuropeptides and addictive behaviors: beyond the usual suspects. *Neuron* **76**, 192-208 (2012).

112. Levey, D.F. *et al.* Genetic risk prediction and neurobiological understanding of alcoholism. *Transl Psychiatry* **4**, e391 (2014).

113. Le-Niculescu, H. *et al.* Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 155-81 (2009).

114. Foroud, T. *et al.* Association of alcohol craving with alpha-synuclein (SNCA). *Alcohol Clin Exp Res* **31**, 537-45 (2007).

115. Janeczek, P., Mackay, R.K., Lea, R.A., Dodd, P.R. & Lewohl, J.M. Reduced expression of alpha-synuclein in alcoholic brain: influence of SNCA-Rep1 genotype. *Addict Biol* (2012).

116. Mayfield, R.D. *et al.* Patterns of gene expression are altered in the frontal and motor cortices of human alcoholics. *J Neurochem* **81**, 802-13 (2002).

117. Lewohl, J.M. *et al.* The application of proteomics to the human alcoholic brain. *Ann N Y Acad Sci* **1025**, 14-26 (2004).

118. Lewohl, J.M. *et al.* Gene expression in human alcoholism: microarray analysis of frontal cortex. *Alcohol Clin Exp Res* **24**, 1873-82 (2000).

119. Taraskina, A.E. *et al.* High level of alpha-synuclein mRNA in peripheral lymphocytes of patients with alcohol dependence syndrome. *Bull Exp Biol Med* **146**, 609-11 (2008).

120. Bonsch, D. *et al.* Elevated alpha synuclein mRNA levels are associated with craving in patients with alcoholism. *Biol Psychiatry* **56**, 984-6 (2004).

121. Bonsch, D. *et al.* Alpha-synuclein protein levels are increased in alcoholic patients and are linked to craving. *Alcohol Clin Exp Res* **29**, 763-5 (2005).

122. Bonsch, D., Lenz, B., Kornhuber, J. & Bleich, S. DNA hypermethylation of the alpha synuclein promoter in patients with alcoholism. *Neuroreport* **16**, 167-70 (2005).

123. Liang, T. *et al.* alpha-Synuclein maps to a quantitative trait locus for alcohol preference and is differentially expressed in alcohol-preferring and -nonpreferring rats. *Proc Natl Acad Sci U S A* **100**, 4690-5 (2003).

124. Liang, T. *et al.* Candidate genes for alcohol preference identified by expression profiling in alcohol-preferring and -nonpreferring reciprocal congenic rats. *Genome Biol* **11**, R11 (2010).

125. Walker, S.J. & Grant, K.A. Peripheral blood alpha-synuclein mRNA levels are elevated in cynomolgus monkeys that chronically self-administer ethanol. *Alcohol* **38**, 1-4 (2006).

126. Dick, D.M. *et al.* Linkage analyses of IQ in the collaborative study on the genetics of alcoholism (COGA) sample. *Behav Genet* **36**, 77-86 (2006).

127. Liu, J. *et al.* Gene expression profiling of individual cases reveals consistent transcriptional changes in alcoholic human brain. *J Neurochem* **90**, 1050-8 (2004).

128. Kimpel, M.W. *et al.* Functional gene expression differences between inbred alcohol-preferring and -non-preferring rats in five brain regions. *Alcohol* **41**, 95-132 (2007).

129. Ponce, G. *et al.* The ANKK1 kinase gene and psychiatric disorders. *Neurotox Res* **16**, 50-9 (2009).

130. Dick, D.M. *et al.* Family-based association analyses of alcohol dependence phenotypes across DRD2 and neighboring gene ANKK1. *Alcohol Clin Exp Res* **31**, 1645-53 (2007).

131. Kraschewski, A. *et al.* Association of the dopamine D2 receptor gene with alcohol dependence: haplotypes and subgroups of alcoholics as key factors for understanding receptor function. *Pharmacogenet Genomics* **19**, 513-27 (2009).

132. Munafo, M.R., Matheson, I.J. & Flint, J. Association of the DRD2 gene Taq1A polymorphism and alcoholism: a meta-analysis of case-control studies and evidence of publication bias. *Mol Psychiatry* **12**, 454-61 (2007).

133. Joe, K.H. *et al.* Genetic association of DRD2 polymorphisms with anxiety scores among alcohol-dependent patients. *Biochem Biophys Res Commun* **371**, 591-5 (2008).

134. Noble, E.P., Blum, K., Ritchie, T., Montgomery, A. & Sheridan, P.J. Allelic association of the D2 dopamine receptor gene with receptor-binding characteristics in alcoholism. *Arch Gen Psychiatry* **48**, 648-54 (1991).

135. Volkow, N.D. *et al.* Decreases in dopamine receptors but not in dopamine transporters in alcoholics. *Alcohol Clin Exp Res* **20**, 1594-8 (1996).

136. Phillips, T.J. *et al.* Alcohol preference and sensitivity are markedly reduced in mice lacking dopamine D2 receptors. *Nat Neurosci* **1**, 610-5 (1998).

137. Zhu, X., Cooper, R., Kan, D., Cao, G. & Wu, X. A genome-wide linkage and association study using COGA data. *BMC Genet* **6 Suppl 1**, S128 (2005).

138. Wang, S. *et al.* Whole-genome linkage analysis in mapping alcoholism genes using single-nucleotide polymorphisms and microsatellites. *BMC Genet* **6 Suppl 1**, S28 (2005).

139. McClintick, J.N. *et al.* Stress-response pathways are altered in the hippocampus of chronic alcoholics. *Alcohol* (2013).

140. Sommer, W., Arlinde, C. & Heilig, M. The search for candidate genes of alcoholism: evidence from expression profiling studies. *Addict Biol* **10**, 71-9 (2005).

141. Worst, T.J. *et al.* Transcriptome analysis of frontal cortex in alcohol-preferring and nonpreferring rats. *J Neurosci Res* **80**, 529-38 (2005).

142. Tapocik, J.D. *et al.* Coordinated dysregulation of mRNAs and microRNAs in the rat medial prefrontal cortex following a history of alcohol dependence. *Pharmacogenomics J* **13**, 286-96 (2013).

143. Zoeller, R.T., Butnariu, O.V., Fletcher, D.L. & Riley, E.P. Limited postnatal ethanol exposure permanently alters the expression of mRNAS encoding myelin basic protein and myelin-associated glycoprotein in cerebellum. *Alcohol Clin Exp Res* **18**, 909-16 (1994).

144. Chiappelli, F., Taylor, A.N., Espinosa de los Monteros, A. & de Vellis, J. Fetal alcohol delays the developmental expression of myelin basic protein and transferrin in rat primary oligodendrocyte cultures. *Int J Dev Neurosci* **9**, 67-75 (1991).

145. Bichenkov, E. & Ellingson, J.S. Ethanol alters the expressions of c-Fos and myelin basic protein in differentiating oligodendrocytes. *Alcohol* **43**, 627-34 (2009).

146. Weng, J., Symons, M.N. & Singh, S.M. Ethanol-responsive genes (Crtam, Zbtb16, and Mobp) located in the alcohol-QTL region of chromosome 9 are associated with alcohol preference in mice. *Alcohol Clin Exp Res* **33**, 1409-16 (2009).

147. Bell, R.L. *et al.* Gene expression changes in the nucleus accumbens of alcohol-preferring rats following chronic ethanol consumption. *Pharmacol Biochem Behav* **94**, 131-47 (2009).

148. Treadwell, J.A. & Singh, S.M. Microarray analysis of mouse brain gene expression following acute ethanol treatment. *Neurochem Res* **29**, 357-69 (2004).

149. Bergen, A.W. *et al.* Genomic regions linked to alcohol consumption in the Framingham Heart Study. *BMC Genet* **4 Suppl 1**, S101 (2003).

150. Foroud, T. *et al.* Alcoholism susceptibility loci: confirmation studies in a replicate sample and further mapping. *Alcohol Clin Exp Res* **24**, 933-45 (2000).

151. McBride, W.J. *et al.* Changes in gene expression within the ventral tegmental area following repeated excessive binge-like alcohol drinking by alcohol-preferring (P) rats. *Alcohol* **47**, 367-80 (2013).

152. Hill, S.Y. *et al.* A genome wide search for alcoholism susceptibility genes. *Am J Med Genet B Neuropsychiatr Genet* **128**, 102-13 (2004).

153. Flatscher-Bader, T. *et al.* Alcohol-responsive genes in the frontal cortex and nucleus accumbens of human alcoholics. *J Neurochem* **93**, 359-70 (2005).

154. Kabbaj, M., Evans, S., Watson, S.J. & Akil, H. The search for the neurobiological basis of vulnerability to drug abuse: using microarrays to investigate the role of stress and individual differences. *Neuropharmacology* **47 Suppl 1**, 111-22 (2004).

155. Pignataro, L. *et al.* Alcohol regulates gene expression in neurons via activation of heat shock factor 1. *J Neurosci* **27**, 12957-66 (2007).

156. Varodayan, F.P., Pignataro, L. & Harrison, N.L. Alcohol induces synaptotagmin 1 expression in neurons via activation of heat shock factor 1. *Neuroscience* **193**, 63-71 (2011).

157. Liu, J. *et al.* Patterns of gene expression in the frontal cortex discriminate alcoholic from nonalcoholic individuals. *Neuropsychopharmacology* **31**, 1574-82 (2006).

158. Gelernter, J. *et al.* Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol Psychiatry* (2013).

159. DeMar, J.C., Jr. *et al.* One generation of n-3 polyunsaturated fatty acid deprivation increases depression and aggression test scores in rats. *J Lipid Res* **47**, 172-80 (2006).

160. Rao, J.S. *et al.* n-3 polyunsaturated fatty acid deprivation in rats decreases frontal cortex BDNF via a p38 MAPK-dependent mechanism. *Mol Psychiatry* **12**, 36-46 (2007).

161. Zanarini, M.C. & Frankenburg, F.R. omega-3 Fatty acid treatment of women with borderline personality disorder: a double-blind, placebo-controlled pilot study. *Am J Psychiatry* **160**, 167-9 (2003).

162. Lin, P.Y., Huang, S.Y. & Su, K.P. A Meta-Analytic Review of Polyunsaturated Fatty Acid Compositions in Patients with Depression. *Biol Psychiatry*.

163. Collins, M.A., Moon, K.H., Tajuddin, N., Neafsey, E.J. & Kim, H.Y. Docosahexaenoic Acid (DHA) Prevents Binge Ethanol-Dependent Aquaporin-4 Elevations While Inhibiting Neurodegeneration: Experiments in Rat Adult-Age Entorhino-Hippocampal Slice Cultures. *Neurotox Res* **23**, 105-10 (2013).

164. Patten, A.R., Brocardo, P.S. & Christie, B.R. Omega-3 supplementation can restore glutathione levels and prevent oxidative damage caused by prenatal ethanol exposure. *J Nutr Biochem* (2012).

165. McNamara, R.K. DHA deficiency and prefrontal cortex neuropathology in recurrent affective disorders. *J Nutr* **140**, 864-8.

166. McNamara, R.K. *et al.* Deficits in docosahexaenoic acid and associated elevations in the metabolism of arachidonic acid and saturated fatty acids in the postmortem orbitofrontal cortex of patients with bipolar disorder. *Psychiatry Res* **160**, 285-99 (2008).

167. Sublette, M.E., Hibbeln, J.R., Galfalvy, H., Oquendo, M.A. & Mann, J.J. Omega-3 polyunsaturated essential fatty acid status as a predictor of future suicide risk. *Am J Psychiatry* **163**, 1100-2 (2006).

168. Lewis, M.D. *et al.* Suicide deaths of active-duty US military and omega-3 fatty-acid status: a case-control comparison. *The Journal of clinical psychiatry* **72**, 1585-90 (2011).

169. Stoll, A.L. *et al.* Omega 3 fatty acids in bipolar disorder: a preliminary double-blind, placebo-controlled trial. *Arch Gen Psychiatry* **56**, 407-12 (1999).

170. Parker, G. *et al.* Omega-3 Fatty acids and mood disorders. *Am J Psychiatry* **163**, 969-78 (2006).

171. Osher, Y. & Belmaker, R.H. Omega-3 fatty acids in depression: a review of three studies. *CNS Neurosci Ther* **15**, 128-33 (2009).

172. Clayton, E.H. *et al.* Reduced mania and depression in juvenile bipolar disorder associated with long-chain omega-3 polyunsaturated fatty acid supplementation. *Eur J Clin Nutr* **63**, 1037-40 (2009).

173. Peet, M. & Stokes, C. Omega-3 fatty acids in the treatment of psychiatric disorders. *Drugs* **65**, 1051-9 (2005).

174. Berger, G.E. *et al.* Ethyl-eicosapentaenoic acid in first-episode psychosis. A 1H-MRS study. *Neuropsychopharmacology* **33**, 2467-73 (2008).

175. Amminger, G.P. *et al.* Long-chain omega-3 fatty acids for indicated prevention of psychotic disorders: a randomized, placebo-controlled trial. *Arch Gen Psychiatry* **67**, 146-54.

176. Kurian, S.M. *et al.* Identification of blood biomarkers for psychosis using convergent functional genomics. *Mol Psychiatry* (2009).

177. Yan, J. *et al.* Using genetic information from candidate gene and genome-wide association studies in risk prediction for alcohol dependence. *Addict Biol* (2013).

178. Kos, M.Z. *et al.* Common biological networks underlie genetic risk for alcoholism in African- and European-American populations. *Genes Brain Behav* **12**, 532-42 (2013).

179. Edenberg, H.J. & Foroud, T. Genetics and alcoholism. *Nat Rev Gastroenterol Hepatol* **10**, 487-94 (2013).

180. Biernacka, J.M. *et al.* Replication of genome wide association studies of alcohol dependence: support for association with variation in ADH1C. *PLoS One* **8**, e58798 (2013).

181. Villafuerte, S., Strumba, V., Stoltenberg, S.F., Zucker, R.A. & Burmeister, M. Impulsiveness mediates the association between GABRA2 SNPs and lifetime alcohol problems. *Genes Brain Behav* **12**, 525-31 (2013).

182. Winner, B. *et al.* Role of alpha-synuclein in adult neurogenesis and neuronal maturation in the dentate gyrus. *J Neurosci* **32**, 16906-16 (2012).

183. Lundblad, M., Decressac, M., Mattsson, B. & Bjorklund, A. Impaired neurotransmission caused by overexpression of alpha-synuclein in nigral dopamine neurons. *Proc Natl Acad Sci U S A* **109**, 3213-9 (2012).

184. Musgrove, R.E., King, A.E. & Dickson, T.C. alpha-Synuclein protects neurons from apoptosis downstream of free-radical production through modulation of the MAPK signalling pathway. *Neurotox Res* **23**, 358-69 (2013).

185. Wilcox, C.E., Claus, E.D., Blaine, S.K., Morgan, M. & Hutchison, K.E. Genetic variation in the alpha synuclein gene (SNCA) is associated with BOLD response to alcohol cues. *J Stud Alcohol Drugs* **74**, 233-44 (2013).

186. Delis, F. *et al.* Chronic mild stress increases alcohol intake in mice with low dopamine D2 receptor levels. *Behav Neurosci* **127**, 95-105 (2013).

187. Zhou, Z. *et al.* Loss of metabotropic glutamate receptor 2 escalates alcohol consumption. *Proc Natl Acad Sci U S A* **110**, 16963-8 (2013).

188. Heilig, M., Goldman, D., Berrettini, W. & O'Brien, C.P. Pharmacogenetic approaches to the treatment of alcohol addiction. *Nat Rev Neurosci* **12**, 670-84 (2011).

189. Zhao, Z. *et al.* Multi-species data integration and gene ranking enrich significant results in an alcoholism genome-wide association study. *BMC Genomics* **13 Suppl 8**, S16 (2012).

190. Oquendo, M.A. & Baca-Garcia, E. Suicidal behavior disorder as a diagnostic entity in the DSM-5 classification system: advantages outweigh limitations. *World Psychiatry* **13**, 128-30 (2014).

191. Saxena, S., Krug, E.G., Chestnov, O. & World Health Organization. Department of Mental Health and Substance Abuse. *Preventing suicide : a global imperative*, 89 pages (World Health Organization, Geneva, 2014).

192. Curtin, S.C., Warner, M. & Hedegaard, H. Increase in Suicide in the United States, 1999-2014. *NCHS Data Brief*, 1-8 (2016).

193. Harris, E.C. & Barraclough, B. Suicide as an outcome for mental disorders. A meta-analysis. *Br J Psychiatry* **170**, 205-28 (1997).

194. WISQARS. Web-based Injury Statistics Query and Reporting System (WISQARS) [online]. *Available from URL: www.cdc.gov/injury/wisqars* (2014).

195. Posner, K. *et al.* The Columbia-Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults. *American Journal of Psychiatry* **168**, 1266-1277 (2011).

196. Niculescu, A.B. *et al.* Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach. *Mol Psychiatry* **20**, 1266-85 (2015).

197. Levey, D.F. *et al.* Towards understanding and predicting suicidality in women: biomarkers and clinical risk assessment. *Mol Psychiatry* **21**, 768-85 (2016).

198. Coryell, W. *et al.* Lithium and suicidal behavior in major affective disorder: a case-control study. *Acta Psychiatrica Scandinavica* **104**, 193-197 (2001).

199. Ahrens, B. & Muller-Oerlinghausen, B. Does lithium exert an independent antisuicidal effect? *Pharmacopsychiatry* **34**, 132-136 (2001).

200. Collins, J.C. & McFarland, B.H. Divalproex, lithium and suicide among Medicaid patients with bipolar disorder. *Journal of Affective Disorders* **107**, 23-28 (2008).

201. Lewitzka, U. *et al.* The suicide prevention effect of lithium: more than 20 years of evidence-a narrative review. *Int J Bipolar Disord* **3**, 32 (2015).

202. Sinyor, M. & Remington, G. Is psychiatry ignoring suicide? The case for clozapine. *J Clin Psychopharmacol* **32**, 307-8 (2012).

203. Irwin, S.A. *et al.* Daily oral ketamine for the treatment of depression and anxiety in patients receiving hospice care: a 28-day open-label proof-of-concept trial. *J Palliat Med* **16**, 958-65 (2013).

204. Wilkinson, S.T. & Sanacora, G. Ketamine: A Potential Rapid-Acting Antisuicidal Agent? *Depress Anxiety* **33**, 711-7 (2016).

205. Hammad, T.A., Laughren, T. & Racoosin, J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry* **63**, 332-9 (2006).

206. Rahn, K.A., Cao, Y.J., Hendrix, C.W. & Kaplin, A.I. The role of 5-HT1A receptors in mediating acute negative effects of antidepressants: implications in pediatric depression. *Transl Psychiatry* **5**, e563 (2015).

207. Lewis, M.D. *et al.* Suicide deaths of active-duty US military and omega-3 fatty-acid status: a case-control comparison. *J Clin Psychiatry* **72**, 1585-90 (2011).

208. Calderon, F. & Kim, H.Y. Docosahexaenoic acid promotes neurite growth in hippocampal neurons. *J Neurochem* **90**, 979-88 (2004).

209. Moriguchi, T. & Salem, N., Jr. Recovery of brain docosahexaenoate leads to recovery of spatial task performance. *J Neurochem* **87**, 297-309 (2003).

210. Hallahan, B., Hibbeln, J.R., Davis, J.M. & Garland, M.R. Omega-3 fatty acid supplementation in patients with recurrent self-harm. Single-centre double-blind randomised controlled trial. *Br J Psychiatry* **190**, 118-22 (2007).

211. Beck, J.S. & Beck, J.S. *Cognitive behavior therapy : basics and beyond*, xix, 391 p. (Guilford Press, New York, 2011).

212. Long, B. & Witterholt, S. An Overview of Dialectical Behavior Therapy. *Psychiatric Annals* **43**, 152-157 (2013).

213. Hawton, K. *et al.* Psychosocial interventions following self-harm in adults: a systematic review and meta-analysis. *Lancet Psychiatry* **3**, 740-50 (2016).

214. Keshtkar, M., Ghanizadeh, A. & Firoozabadi, A. Repetitive Transcranial Magnetic Stimulation Versus Electroconvulsive Therapy for the Treatment of Major Depressive Disorder, A Randomized Controlled Clinical Trial. *Journal of Ect* **27**, 310-314 (2011).

215. Khan, A., Khan, S., Kolts, R. & Brown, W.A. Suicide rates in clinical trials of SSRIs, other antidepressants, and placebo: Analysis of FDA reports. *American Journal of Psychiatry* **160**, 790-792 (2003).

216. de Araujo, R.M., Mazzochi, L., Lara, D.R. & Ottoni, G.L. Thinking about dying and trying and intending to die: results on suicidal behavior from a large Web-based sample. *J Clin Psychiatry* **76**, e359-65 (2015).

217. Le-Niculescu, H. *et al.* Discovery and validation of blood biomarkers for suicidality. *Mol Psychiatry* **18**, 1249-64 (2013).

218. Willour, V.L. *et al.* A genome-wide association study of attempted suicide. *Mol Psychiatry* **17**, 433-44 (2012).

219. Kaminsky, Z. *et al.* Epigenetic and genetic variation at SKA2 predict suicidal behavior and post-traumatic stress disorder. *Transl Psychiatry* **5**, e627 (2015).

220. Guintivano, J. *et al.* Identification and replication of a combined epigenetic and genetic biomarker predicting suicide and suicidal behaviors. *Am J Psychiatry* **171**, 1287-96 (2014).

221. Igarashi, T. *et al.* Mutations in SLC4A4 cause permanent isolated proximal renal tubular acidosis with ocular abnormalities. *Nat Genet* **23**, 264-6 (1999).

222. Perlis, R.H. *et al.* Genome-wide association study of suicide attempts in mood disorder patients. *Am J Psychiatry* **167**, 1499-507 (2010).

223. Sequeira, A. *et al.* Implication of SSAT by gene expression and genetic variation in suicide and major depression. *Arch Gen Psychiatry* **63**, 35-48 (2006).

224. Klempan, T.A. *et al.* Profiling brain expression of the spermidine/spermine N1-acetyltransferase 1 (SAT1) gene in suicide. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 934-43 (2009).

225. Asberg, M., Traskman, L. & Thoren, P. 5-HIAA in the cerebrospinal fluid. A biochemical suicide predictor? *Arch Gen Psychiatry* **33**, 1193-7 (1976).

226. Oquendo, M.A. *et al.* Positron Emission Tomographic Imaging of the Serotonergic System and Prediction of Risk and Lethality of Future Suicidal Behavior. *JAMA Psychiatry* (2016).

227. Lindqvist, D. *et al.* Interleukin-6 is elevated in the cerebrospinal fluid of suicide attempters and related to symptom severity. *Biol Psychiatry* **66**, 287-92 (2009).

228. Sublette, M.E. *et al.* Plasma kynurenine levels are elevated in suicide attempters with major depressive disorder. *Brain Behav Immun* **25**, 1272-8 (2011).

229. Erhardt, S. *et al.* Connecting inflammation with glutamate agonism in suicidality. *Neuropsychopharmacology* **38**, 743-52 (2013).

230. Berngruber, T.W., Lion, S. & Gandon, S. Evolution of suicide as a defence strategy against pathogens in a spatially structured environment. *Ecol Lett* (2013).

231. Nock, M.K. *et al.* Suicide and suicidal behavior. *Epidemiol Rev* **30**, 133-54 (2008).

232. Le-Niculescu, H. *et al.* Identifying blood biomarkers for mood disorders using convergent functional genomics. *Molecular psychiatry* **14**, 156-74 (2009).

233. Kurian, S.M. *et al.* Identification of blood biomarkers for psychosis using convergent functional genomics. *Molecular psychiatry* **16**, 37-58 (2011).

234. Niculescu, A. *et al.* Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. *Physiological Genomics* **4**, 83-91 (2000).

235. Ogden, C.A. *et al.* Candidate genes, pathways and mechanisms for bipolar (manic-depressive) and related disorders: an expanded convergent functional genomics approach. *Molecular psychiatry* **9**, 1007-29 (2004).

236. Patel, S.D. *et al.* Coming to grips with complex disorders: genetic risk prediction in bipolar disorder using panels of genes identified through convergent functional genomics. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **153B**, 850-77 (2010).

237. Le-Niculescu, H. *et al.* Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **150B**, 155-81 (2009).

238. Rodd, Z.A. *et al.* Candidate genes, pathways and mechanisms for alcoholism: an expanded convergent functional genomics approach. *The pharmacogenomics journal* **7**, 222-56 (2007).

239. Le-Niculescu, H. *et al.* Convergent functional genomics of anxiety disorders: translational identification of genes, biomarkers, pathways and mechanisms. *Translational Psychiatry* **1**, e9 (2011).

240. Ayalew, M. *et al.* Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Molecular psychiatry* **17**, 887-905 (2012).

241. Fiori, L.M. *et al.* Association of polyaminergic loci with anxiety, mood disorders, and attempted suicide. *PLoS One* **5**, e15146 (2010).

242. Chen, G.G. *et al.* Evidence of altered polyamine concentrations in cerebral cortex of suicide completers. *Neuropsychopharmacology* **35**, 1477-84 (2010).

243. Fiori, L.M. *et al.* Global gene expression profiling of the polyamine system in suicide completers. *Int J Neuropsychopharmacol* **14**, 595-605 (2011).

244. Fiori, L.M. & Turecki, G. Epigenetic regulation of spermidine/spermine N1-acetyltransferase (SAT1) in suicide. *J Psychiatr Res* **45**, 1229-35 (2011).

245. Fiori, L.M., Mechawar, N. & Turecki, G. Identification and characterization of spermidine/spermine N1-acetyltransferase promoter variants in suicide completers. *Biol Psychiatry* **66**, 460-7 (2009).

246. Fiori, L.M., Zouk, H., Himmelman, C. & Turecki, G. X chromosome and suicide. *Mol Psychiatry* **16**, 216-26 (2011).

247. Guipponi, M. *et al.* Genetic and epigenetic analysis of SSAT gene dysregulation in suicidal behavior. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 799-807 (2009).

248. Sequeira, A. *et al.* Gene expression changes in the prefrontal cortex, anterior cingulate cortex and nucleus accumbens of mood disorders subjects that committed suicide. *PLoS One* **7**, e35367 (2012).

249. Galfalvy, H. *et al.* A pilot genome wide association and gene expression array study of suicide with and without major depression. *World J Biol Psychiatry* (2011).

250. Kekesi, K.A. *et al.* Altered functional protein networks in the prefrontal cortex and amygdala of victims of suicide. *PloS one* **7**, e50532 (2012).

251. Kim, S., Choi, K.H., Baykiz, A.F. & Gershenfeld, H.K. Suicide candidate genes associated with bipolar disorder and schizophrenia: an exploratory gene expression profiling analysis of post-mortem prefrontal cortex. *BMC Genomics* **8**, 413 (2007).

252. Klempan, T.A. *et al.* Altered expression of genes involved in ATP biosynthesis and GABAergic neurotransmission in the ventral prefrontal cortex of suicides with and without major depression. *Mol Psychiatry* **14**, 175-89 (2009).

253. Pandey, G.N. *et al.* Proinflammatory cytokines in the prefrontal cortex of teenage suicide victims. *J Psychiatr Res* **46**, 57-63 (2012).

254. Lalovic, A., Klempan, T., Sequeira, A., Luheshi, G. & Turecki, G. Altered expression of lipid metabolism and immune response genes in the frontal cortex of suicide completers. *J Affect Disord* **120**, 24-31 (2010).

255. Sequeira, A. *et al.* Patterns of gene expression in the limbic system of suicides with and without major depression. *Mol Psychiatry* **12**, 640-55 (2007).

256. Zhurov, V. *et al.* Molecular pathway reconstruction and analysis of disturbed gene expression in depressed individuals who died by suicide. *PloS one* **7**, e47581 (2012).

257. Freemantle, E., Mechawar, N. & Turecki, G. Cholesterol and phospholipids in frontal cortex and synaptosomes of suicide completers: relationship with endosomal lipid trafficking genes. *Journal of psychiatric research* **47**, 272-9 (2013).

258. Pandey, G.N. *et al.* Altered expression and phosphorylation of myristoylated alanine-rich C kinase substrate (MARCKS) in postmortem brain of suicide victims with or without depression. *J Psychiatr Res* **37**, 421-32 (2003).

259. Dwivedi, Y. *et al.* Modulation in activation and expression of phosphatase and tensin homolog on chromosome ten, Akt1, and 3-phosphoinositide-dependent kinase 1: further evidence demonstrating altered

phosphoinositide 3-kinase signaling in postmortem brain of suicide subjects. *Biol Psychiatry* **67**, 1017-25 (2010).

260. Menke, A. *et al.* Genome-wide association study of antidepressant treatment-emergent suicidal ideation. *Neuropsychopharmacology* **37**, 797-807 (2012).

261. Butler, A.W. *et al.* A genomewide linkage study on suicidality in major depressive disorder confirms evidence for linkage to 2p12. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 1465-73 (2010).

262. Novick, D.M., Swartz, H.A. & Frank, E. Suicide attempts in bipolar I and bipolar II disorder: a review and meta-analysis of the evidence. *Bipolar disorders* **12**, 1-9 (2010).

263. Karege, F. *et al.* Alteration in kinase activity but not in protein levels of protein kinase B and glycogen synthase kinase-3beta in ventral prefrontal cortex of depressed suicide victims. *Biol Psychiatry* **61**, 240-5 (2007).

264. Le-Niculescu, H. *et al.* Convergent Functional Genomics of bipolar disorder: from animal model pharmacogenomics to human genetics and biomarkers. *Neurosci Biobehav Rev* **31**, 897-903 (2007).

265. Le-Niculescu, H., Patel, S.D. & Niculescu, A.B. Convergent integration of animal model and human studies of bipolar disorder (manic-depressive illness). *Current opinion in pharmacology* **10**, 594-600 (2010).

266. Niculescu, A.B. & Le-Niculescu, H. Convergent Functional Genomics: what we have learned and can learn about genes, pathways, and mechanisms.

Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology **35**, 355-6 (2010).

267. Niculescu, A.B., 3rd *et al.* Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. *Physiological genomics* **4**, 83-91 (2000).

268. Le-Niculescu, H. *et al.* Phenomic, convergent functional genomic, and biomarker studies in a stress-reactive genetic animal model of bipolar disorder and co-morbid alcoholism. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **147B**, 134-66 (2008).

269. Oquendo, M.A. *et al.* Increased risk for suicidal behavior in comorbid bipolar disorder and alcohol use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *The Journal of clinical psychiatry* **71**, 902-9 (2010).

270. Le-Niculescu, H. *et al.* Convergent functional genomic studies of omega-3 fatty acids in stress reactivity, bipolar disorder and alcoholism. *Translational Psychiatry* **1**, e4 (2011).

271. McGrath, C.L. *et al.* Evidence for genetic association of RORB with bipolar disorder. *BMC Psychiatry* **9**, 70 (2009).

272. Bunney, B.G. & Bunney, W.E. Mechanisms of Rapid Antidepressant Effects of Sleep Deprivation Therapy: Clock Genes and Circadian Rhythms. *Biol Psychiatry* (2012).

273. McCall, W.V. *et al.* Nightmares and Dysfunctional Beliefs about Sleep Mediate the Effect of Insomnia Symptoms on Suicidal Ideation. *J Clin Sleep Med* **9**, 135-40 (2013).

274. Meltzer, H.Y. *et al.* Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Arch Gen Psychiatry* **60**, 82-91 (2003).

275. Niwa, M. *et al.* Adolescent stress-induced epigenetic control of dopaminergic neurons via glucocorticoids. *Science* **339**, 335-9 (2013).

276. Kelleher, I. *et al.* Psychotic Symptoms in Adolescence Index Risk for Suicidal Behavior: Findings From 2 Population-Based Case-Control Clinical Interview Studies. *Arch Gen Psychiatry*, 1-7 (2012).

277. Niculescu, A.B. *et al.* PhenoChipping of psychotic disorders: a novel approach for deconstructing and quantitating psychiatric phenotypes. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **141B**, 653-62 (2006).

278. Schipper, R.G., Penning, L.C. & Verhofstad, A.A. Involvement of polyamines in apoptosis. Facts and controversies: effectors or protectors? *Semin Cancer Biol* **10**, 55-68 (2000).

279. Pietila, M. *et al.* Activation of polyamine catabolism profoundly alters tissue polyamine pools and affects hair growth and female fertility in transgenic

mice overexpressing spermidine/spermine N1-acetyltransferase. *The Journal of biological chemistry* **272**, 18746-51 (1997).

280. Min, S.H. *et al.* Altered levels of growth-related and novel gene transcripts in reproductive and other tissues of female mice overexpressing spermidine/spermine N1-acetyltransferase (SSAT). *The Journal of biological chemistry* **277**, 3647-57 (2002).

281. Fiori, L.M. *et al.* Global gene expression profiling of the polyamine system in suicide completers. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum* **14**, 595-605 (2011).

282. Duckworth, C.A., Clyde, D. & Pritchard, D.M. CD24 is expressed in gastric parietal cells and regulates apoptosis and the response to Helicobacter felis infection in the murine stomach. *Am J Physiol Gastrointest Liver Physiol* **303**, G915-26 (2012).

283. Malkesman, O. *et al.* Targeting the BH3-interacting domain death agonist to develop mechanistically unique antidepressants. *Mol Psychiatry* **17**, 770-80 (2012).

284. Tondo, L. & Baldessarini, R.J. Long-term lithium treatment in the prevention of suicidal behavior in bipolar disorder patients. *Epidemiol Psichiatr Soc* **18**, 179-83 (2009).

285. Lowthert, L. *et al.* Increased ratio of anti-apoptotic to pro-apoptotic Bcl2 gene-family members in lithium-responders one month after treatment initiation. *Biol Mood Anxiety Disord* **2**, 15 (2012).

286. Benedetti, F. *et al.* Opposite effects of suicidality and lithium on gray matter volumes in bipolar depression. *J Affect Disord* **135**, 139-47 (2011).

287. Schork, N.J. Personalized medicine: Time for one-person trials. *Nature* **520**, 609-11 (2015).

288. Niculescu, A.B. *et al.* Psychiatric blood biomarkers: avoiding jumping to premature negative or positive conclusions. *Mol Psychiatry* **20**, 286-8 (2015).

289. Borges, G., Angst, J., Nock, M.K., Ruscio, A.M. & Kessler, R.C. Risk factors for the incidence and persistence of suicide-related outcomes: a 10-year follow-up study using the National Comorbidity Surveys. *J Affect Disord* **105**, 25-33 (2008).

290. Nock, M.K. Future directions for the study of suicide and self-injury. *J Clin Child Adolesc Psychol* **41**, 255-9 (2012).

291. Pandey, G.N. *et al.* Proinflammatory cytokines in the prefrontal cortex of teenage suicide victims. *Journal of psychiatric research* **46**, 57-63 (2012).

292. Hoyo-Becerra, C. *et al.* Concomitant interferon alpha stimulation and TLR3 activation induces neuronal expression of depression-related genes that are elevated in the brain of suicidal persons. *PloS one* **8**, e83149 (2013).

293. Lindqvist, D. *et al.* CSF biomarkers in suicide attempters--a principal component analysis. *Acta psychiatrica Scandinavica* **124**, 52-61 (2011).

294. Bay-Richter, C. *et al.* A role for inflammatory metabolites as modulators of the glutamate N-methyl-D-aspartate receptor in depression and suicidality. *Brain, behavior, and immunity* **43**, 110-7 (2015).

295. Kim, Y.K. *et al.* Differences in cytokines between non-suicidal patients and suicidal patients in major depression. *Progress in neuro-psychopharmacology & biological psychiatry* **32**, 356-61 (2008).

296. Sokolowski, M., Ben-Efraim, Y.J., Wasserman, J. & Wasserman, D. Glutamatergic GRIN2B and polyaminergic ODC1 genes in suicide attempts: associations and gene-environment interactions with childhood/adolescent physical assault. *Mol Psychiatry* **18**, 985-92 (2013).

297. Kim, S., Choi, K.H., Baykiz, A.F. & Gershenfeld, H.K. Suicide candidate genes associated with bipolar disorder and schizophrenia: An exploratory gene expression profiling analysis of post-mortem prefrontal cortex. *BMC Genomics* **8**, 413 (2007).

298. Ballesteros, J., Maeztu, A.I., Callado, L.F., Meana, J.J. & Gutierrez, M. Specific binding of [3H]Ro 19-6327 (lazabemide) to monoamine oxidase B is increased in frontal cortex of suicide victims after controlling for age at death. *Eur Neuropsychopharmacol* **18**, 55-61 (2008).

299. Roggenbach, J. *et al.* Peripheral serotonergic markers in acutely suicidal patients. 1. Comparison of serotonergic platelet measures between suicidal

individuals, nonsuicidal patients with major depression and healthy subjects. *J Neural Transm* **114**, 479-87 (2007).

300. Labonte, B. *et al.* Genome-wide methylation changes in the brains of suicide completers. *Am J Psychiatry* **170**, 511-20 (2013).

301. Pandey, G.N. *et al.* Altered expression and phosphorylation of myristoylated alanine-rich C kinase substrate (MARCKS) in postmortem brain of suicide victims with or without depression. *Journal of psychiatric research* **37**, 421-32 (2003).

302. Karege, F. *et al.* Alterations in phosphatidylinositol 3-kinase activity and PTEN phosphatase in the prefrontal cortex of depressed suicide victims. *Neuropsychobiology* **63**, 224-31 (2011).

303. Darlington, T.M. *et al.* Identifying rare variants for genetic risk through a combined pedigree and phenotype approach: application to suicide and asthma. *Transl Psychiatry* **4**, e471 (2014).

304. Hishimoto, A. *et al.* A functional polymorphism of the micro-opioid receptor gene is associated with completed suicides. *J Neural Transm* **115**, 531-6 (2008).

305. Gabilondo, A.M., Meana, J.J. & Garcia-Sevilla, J.A. Increased density of mu-opioid receptors in the postmortem brain of suicide victims. *Brain Res* **682**, 245-50 (1995).

306. Mann, J.J. & Currier, D. Medication in Suicide Prevention Insights from Neurobiology of Suicidal Behavior. in *The Neurobiological Basis of Suicide* (ed. Dwivedi, Y.) (Boca Raton (FL), 2012).

307. Oquendo, M.A. *et al.* Toward a biosignature for suicide. *Am J Psychiatry* **171**, 1259-77 (2014).

308. Kelleher, I. *et al.* Psychotic symptoms in adolescence index risk for suicidal behavior: findings from 2 population-based case-control clinical interview studies. *Arch Gen Psychiatry* **69**, 1277-83 (2012).

309. Schoenbaum, M. *et al.* Predictors of suicide and accident death in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS): results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry* **71**, 493-503 (2014).

310. Posner, K. *et al.* The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry* **168**, 1266-77 (2011).

311. Wemmie, J.A., Taugher, R.J. & Kreple, C.J. Acid-sensing ion channels in pain and disease. *Nat Rev Neurosci* **14**, 461-71 (2013).

312. Wasserman, D. *et al.* School-based suicide prevention programmes: the SEYLE cluster-randomised, controlled trial. *Lancet* (2015).

313. Li, N. *et al.* mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science* **329**, 959-64 (2010).

314. Ballard, E.D. *et al.* Improvement in suicidal ideation after ketamine infusion: relationship to reductions in depression and anxiety. *J Psychiatr Res* **58**, 161-6 (2014).

315. Canetto, S.S. & Sakinofsky, I. The gender paradox in suicide. *Suicide Life Threat Behav* **28**, 1-23 (1998).

316. Moller-Leimkuhler, A.M. The gender gap in suicide and premature death or: why are men so vulnerable? *Eur Arch Psychiatry Clin Neurosci* **253**, 1-8 (2003).

317. Hawton, K. Sex and suicide. Gender differences in suicidal behaviour. *Br J Psychiatry* **177**, 484-5 (2000).

318. Zhang, E.E. *et al.* A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell* **139**, 199-210 (2009).

319. McCarthy, M.J. & Welsh, D.K. Cellular circadian clocks in mood disorders. *J Biol Rhythms* **27**, 339-52 (2012).

320. Niculescu, A.B. *et al.* Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach. *Mol Psychiatry* (2015).

321. Smalheiser, N.R. *et al.* MicroRNA expression is down-regulated and reorganized in prefrontal cortex of depressed suicide subjects. *PLoS One* **7**, e33201 (2012).

322. Dick, D.M. *et al.* Evidence for genes on chromosome 2 contributing to alcohol dependence with conduct disorder and suicide attempts. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 1179-88 (2010).

323. Hesselbrock, V. *et al.* The search for genetic risk factors associated with suicidal behavior. *Alcohol Clin Exp Res* **28**, 70S-76S (2004).

324. Falcone, T. *et al.* Serum S100B: a potential biomarker for suicidality in adolescents? *PLoS One* **5**, e11089 (2010).

325. Hennings, J.M. *et al.* RNA expression profiling in depressed patients suggests retinoid-related orphan receptor alpha as a biomarker for antidepressant response. *Transl Psychiatry* **5**, e538 (2015).

326. Martins-de-Souza, D. *et al.* Blood mononuclear cell proteome suggests integrin and Ras signaling as critical pathways for antidepressant treatment response. *Biol Psychiatry* **76**, e15-7 (2014).

327. Goodwin, R.D. & Marusic, A. Association between short sleep and suicidal ideation and suicide attempt among adults in the general population. *Sleep* **31**, 1097-101 (2008).

328. Gaiteri, C., Guilloux, J.P., Lewis, D.A. & Sibille, E. Altered gene synchrony suggests a combined hormone-mediated dysregulated state in major depression. *PLoS One* **5**, e9970 (2010).

329. Sainz, J. *et al.* Inflammatory and immune response genes have significantly altered expression in schizophrenia. *Mol Psychiatry* **18**, 1056-7 (2013).

330. Sokolov, B.P., Jiang, L., Trivedi, N.S. & Aston, C. Transcription profiling reveals mitochondrial, ubiquitin and signaling systems abnormalities in postmortem brains from subjects with a history of alcohol abuse or dependence. *J Neurosci Res* **72**, 756-67 (2003).

331. Yu, Z. *et al.* Four mood stabilizers commonly induce FEZ1 expression in human astrocytes. *Bipolar Disord* **13**, 486-99 (2011).

332. Sood, S. *et al.* A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol* **16**, 185 (2015).

333. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929-35 (2006).

334. Lillie, E.O. *et al.* The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Per Med* **8**, 161-173 (2011).

Curriculum vitae

Daniel F. Levey

Education

Indiana University, Indianapolis, IN

Ph.D. in Medical Neuroscience, April 2017

Indiana State University, Terre Haute, IN

B.S. in Psychology, May 2003

Awards

Larry Kays Fellowship Award for Scientific Achievement, 2016

First Place Poster Award in Translational Research, 6th Annual Veteran Affairs Scientific Symposium, Indianapolis, IN, 2014

University Fellowship, Indiana University School of Medicine, IBMG Program 2014

Dean's List Highest Honors, Indiana University-Purdue University Indianapolis, 2009 - 2010

Dean's List, Indiana State University, 2000 - 2003

Sigma XI Grant-in-Aid of Research: Mechanisms of Attention in an Animal Model of Alzheimer's Disease, 2001

Publications

Rangaraju S*, Levey DF*, Nho N*, Jain N, Andrews KD, Le-Niculescu H, Salomon DR, Saykin AJ, Petrascheck M, and Niculescu AB.

Mood, Stress and Longevity: Convergence on ANK3. Molecular Psychiatry Epub May 2016.

Levey DF*, Niculescu EM*, Le-Niculescu H, Dainton HD, Phalen PL, Ladd TB, Weber H, Belanger E, Graham DL, Khan FN, Vanipenta NP, Stage EC, Ballew A, Yard M, Gelbart T, Shekhar A, Schork NJ, Kurian SM, Sandusky GE, Salomon DR, Niculescu AB.
Towards understanding and predicting suicidality in women: biomarkers and clinical risk assessment. Molecular Psychiatry Epub April 2016.

Niculescu AB , Levey DF, Phalen PL, Le-Niculescu H, Dainton HD ,Jain N, Belanger E, James A, George S, Weber H, Graham DL, Schweitzer R, Ladd TB, Learman R, Niculescu EM, Vanipenta NP, Khan FN, Mullen J, Shankar G, Cook S, Humbert C, Ballew A, Yard M, Gelbart T, Shekhar A, Schork NJ, Kurian SM, Sandusky GE, Salomon DR.
Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach. Molecular Psychiatry Epub August 2015.

Niculescu AB, Levey D.F., Le-Niculescu H, Niculescu E, Kurian SM, Salomon, D.
Psychiatric blood biomarkers: avoiding jumping to premature negative or positive conclusions. Molecular Psychiatry Epub January 2015.

D.F. Levey*, H. Le-Niculescu*, J. Frank, M. Ayalew, N. Jain, B. Kirlin, R. Learman, E. Winiger, Z.A. Rodd, A. Shekhar, N. J. Schork, F, Kiefe, N. Wodarz, B. Müller-Myhsok, N. Dahmer, GESGA Consortium, M. Nöthen, R. Sherva, L. Farrer, A. H. Smith, H. Kranzler, M. Rietschel, J. Gelernter, A. B. Niculescu.

Genetic Risk Prediction and Neurobiological Understanding of Alcoholism. Translational Psychiatry 2014. Epub May 2014.

Le-Niculescu H*, Levey DF*, Ayalew M, Palmer L, Gavrin LM, Jain N, Bhosrekar S, Shankar G, Belanger E, Olesek K, Duckworth H, Vergo J, Schweitzer R, Radel M, Yard M, Ballew A, Shekhar A, Sandusky G, Schork NJ, Kurian SM, Salomon DR, Niculescu AB.

Discovery and Validation of Blood Biomarkers for Suicidality. Molecular Psychiatry 2013 Dec;18(12):1249-64. doi: 10.1038/mp.2013.95. Epub August 2013.

Ayalew M*, Le-Niculescu H*, Levey DF, Jain N, Changala B, Patel SD, Winiger E, Breier A, Shekhar A, Amdur R, Koller D, Nurnberger JI, Corvin A, Geyer M, Tsuang MT, Salomon DR, Schork NJ, Fanous A, O'Donovan M, Niculescu AB.

Convergent Functional Genomics of Schizophrenia: From Comprehensive Understanding to Genetic Risk Prediction. Molecular Psychiatry 2012 Sep; 17(9):887-905. Epub May 2012.

*Authors contributed equally to this work

Research Experience

Indiana University School of Medicine, Indianapolis, IN

Graduate Research Assistant, 2012 - 2017

PI: Dr. Alexander Niculescu

Graduate Research Rotations, 2011 - 2012

PIs: Dr. Alexander Niculescu, Dr. Andrew Saykin, and Dr. Eri Hashino

Indiana University-Purdue University Indianapolis, IN

Undergraduate Research Assistant, 2010-2011

PI: Dr. Charles Goodlett

Indiana University School of Medicine, Terre Haute, IN

Undergraduate Research Assistant, 2002

PI: Dr. Peter Duong

Indiana State University, Terre Haute, IN

Undergraduate Research Assistant, 2001

PI: Dr. Carol Yoder

Undergraduate Research Assistant, 2000 – 2002

PI: Dr. Allen Butt