Manuscript Information

Journal name: Molecular psychiatry
NIHMS ID: NIHMS923166
Manuscript Title: New considerations for hiPSC-based models of neuropsychiatric disorders
Submitter: Nature Publishing Group (repositorynotifs@nature.com)

Manuscript Files

Type	Fig/Table #	Filename	Size	Uploaded
manuscript	1	article_1.docx	787438	2017-11-28 06:13:29
figure	1	figure_1.pdf	196063	2017-11-28 06:13:33
figure	2	figure_2.pdf	45501	2017-11-28 06:13:33
figure	3	figure_3.pdf	30931	2017-11-28 06:13:33
figure	4	figure_4.pdf	176723	2017-11-28 06:13:34

TITLE

New considerations for hiPSC-based models of neuropsychiatric disorders

AUTHORS

Gabriel E. Hoffman[1,4], Nadine Schrode[1,2,5], Erin Flaherty[2,5], Kristen J. Brennand[1,2,3,5#]

AFFILIATIONS

Department of Genetics and Genomics[1], Neuroscience[2] and Psychiatry[3], Icahn Institute of Genomics and Multiscale Biology[4], Friedman Brain Institute[5], Icahn School of Medicine at Mount Sinai, New York, NY 10029

[#] Correspondence: kristen.brennand@mssm.edu

KEYWORDS

Human induced pluripotent stem cells; psychiatric disease; complex genetic disorders; transcriptomics

ABSTRACT

The development of human induced pluripotent stem cells (hiPSCs) has made possible patient-specific modeling across the spectrum of human disease. Here we discuss recent advances in psychiatric genomics and post-mortem studies that provide critical insights concerning cell type composition and sample size that should be considered when designing hiPSC-based studies of complex genetic disease. We review recent hiPSC-based models of SZ, in light of our new understanding of critical power limitations in the design of hiPSC-based studies of complex genetic disorders. Three possible solutions are a movement towards genetically stratified cohorts of rare variant patients, application of CRISPR technologies to engineer isogenic neural cells to study the impact of common variants, and integration of advanced genetics and hiPSC-based datasets in future studies. Overall, we emphasize that to advance the reproducibility and relevance of hiPSC-based studies, stem cell biologists must contemplate statistical and biological considerations that are already well accepted in the field of genetics. We conclude with a discussion of the hypothesis of biological convergence of disease - through molecular, cellular, circuit and patient level phenotypes - and how this might emerge through hiPSC-based studies.

I. Introduction.

We are each unique, comprised of distinct genetic, epigenetic and environmental risk factors that predispose us to some diseases and confer resilience to others. As expanding genetic studies increasingly demonstrate that both rare variants of large impact and common variants of small effect contribute to a variety of neuropsychiatric disorders, it becomes increasingly critical that we unravel how these risk factors interact within and between the diverse cell types populating the brain. While mouse models are uniquely suited for demonstrating how altered function of single gene products contribute to aberrant neuronal function or behavior, genetic studies of penetrance and complex gene interactions are extremely difficult to address using inbred mouse lines. Similarly, the challenges of human post-mortem tissue, coupled with the inability to conduct *in vivo* functional validations, has to date left us with a very limited understanding of how rare and common variants impact gene expression or cellular function in patients. By

developing human induced pluripotent stem cell (hiPSC)-based models for the study of predisposition to neurological disease, stem cell scientists have established a new approach by which to systematically test the impact of causal variants in human cells [1-3].

While familial mutations in a fraction of cases result in neurological diseases from autism spectrum disorder (ASD) [4] to Alzheimer's disease (AD) [5], a large majority of the patient population is designated idiopathic, arising from unknown genetic risk factor(s). Well established for schizophrenia (SZ) [6] and increasingly accepted across the breadth of neuropsychiatric disease, is the model that a combination of rare and common variants contributes to disease with differing frequencies and penetrance, with highly penetrant rare variants being particularly relevant for diseases with major fitness consequences [7]. Although the hiPSC-based models discussed herein are specifically applied to SZ, the tools and insights acquired are being rapidly and concurrently applied to other neuropsychiatric disorders.

II. Advances in psychiatric genomics.

Genetic and environmental effects together contribute to neuropsychiatric disease risk [8] (**Figure 1**). Genetic epidemiology, including twin studies [9], provides substantial evidence that inherited and *de novo* genetic variants contribute substantially to disease liability (well-reviewed across SZ [10], bipolar disorder (BD) [11], ASD [12], intellectual disability (ID) [13], obsessive-compulsive disorder (OCD) [14], depression [15] and attention-deficit hyperactivity disorder (ADHD) [16]). The current model posits that multiple types of genetic risk influence a continuum of behavioral and developmental traits, the severe tail of which results in neuropsychiatric disease [17].

Rare variants associated with SZ currently include sixteen high confidence large copy number variations (CNVs) that are enriched for genes associated with synaptic function, and frequently shared with other neurodevelopmental disorders [18]. Unlike whole-exome sequencing studies for ASD [19, 20] and intellectual disability [21] that have identified a series of rare coding mutations enriched for synaptic genes, transcription factors and chromatin modifiers, similar studies for SZ have implied a role for functional gene sets such as voltage-gated calcium channels, ARC-associated scaffold and N-methyl-D-aspartate receptor (NMDAR) [22-24] but not yet conclusively identified specific genes. Although there is pathway level genetic convergence across neuropsychiatric diseases, it seems that the spatiotemporal activity of the precise genes involved reflects the specific phenotypes; for example, *de novo* mutations in ID are enriched for fetal cortex genes, ASD for fetal cortex, cerebellum, and striatum genes, and SZ for adolescent cortex genes [25]. Moreover, while missense mutations in neuronal development genes contribute to ID, ASD or SZ, loss-of-function mutations, particularly in chromatin genes, bias towards ID and ASD [25]. In summary, although there is substantial convergence at the pathway level of rare mutations across neuropsychiatric disease, clinical presentation reflects the precise gene(s) involved, timing of developmental expression of the affected gene(s) and type of mutation.

Common variation assessed by genome-wide association studies (GWAS) of single nucleotide polymorphisms (SNPs) have identified a growing list of risk loci significantly associated with SZ [26], BD [27] and ASD [28], which together account for most of the genetic risk for these disorders. Risk loci identified by GWAS are enriched for neuronal genes [29, 30] and show substantial overlap between these disorders [27, 31]. In SZ, these risk loci map to genes expressed in pyramidal excitatory neurons and a subset of GABAergic interneurons, substantially less to progenitor or glial cells [29, 30]. Transcriptomic profiling of

post-mortem brain tissue supports this shared molecular neuropathology, demonstrating that the degree of shared transcriptional dysregulation strongly correlates to the observed common variant overlap across psychiatric disorders [32].

While disease associated loci still explain only a small fraction of the predicted genetic liability to psychiatric disease, the "missing heritability" is now believed to largely reside amongst the common variants with effect sizes well below genome-wide statistical significance [33]. Taken to the fullest extent, a proposed ''omnigenic'' model suggests that gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes [34].

Overall, highly penetrant rare mutations tend to confer risk of neurodevelopmental disorders with earlier onset, while more common variants with much lower effect sizes contribute risk towards later onset disorders (reviewed in [35]). Different types of risk factors can interact; a higher polygenic risk score is thought to increase disease risk in carriers of rare mutations, suggesting cumulative effects between both types of variants [36, 37]. Although variants differ in penetrance and prevalence across the spectrum of neurodevelopmental disorders, there is a growing consensus that genetic risk will converge not only between psychiatric disorders, but also between common and rare variants within disorders - focused on genes expressed during fetal cortical development [38-40] and converging on common pathways such as synaptic function [18, 22, 23, 41] and epigenetic processes [41, 42]. In ASD, a shared functional signal between common and rare variants was detected after accounting for the different effect sizes of the genes involved [43]. Critically, this is not a literal overlap in risk genes or pathways; rather, functions weakly associated with common variants tend to show stronger effects in rare variants [43]. The role of effect size on these functional consequences, the degree of overlap of genetic risk and precise neural cell type(s) impacted by this growing list of disease-associated mutations remains unclear. Throughout this review we will consider the extent to which *in vitro* stem cell-derived populations can model the impact of different perturbations on these causal genes on cellular and molecular neural phenotypes. The ultimate goal of psychiatric genetics is to study the impact of patient-specific mutations, facilitating a transition into precision psychiatry [8] and personalized medicine [44].

III. Findings from recent hiPSC-based models of SZ.

Although typically diagnosed in late adolescence, SZ reflects abnormal neurodevelopmental processes that begin years before the onset of symptoms (reviewed [45]). We and others have shown that hiPSC-derived neurons most resemble fetal brain tissue [46-48], indicating that these models are most appropriate for studying aspects of disease predisposition rather than the disease state itself. (This may be changing, as recent studies have shown that long-term organoid culture can yield cells more similar to neonatal tissue [49], novel media formulations can improve the maturity [50] and genetic strategies can accelerate the aging [3, 51] of hiPSC-derived neurons, facilitating studies of adult-onset diseases). hiPSC-based models have great potential to help unravel the functional impact of the risk loci identified by genetic studies. Idiopathic cohorts capture all of the risk elements, known and unknown, that contribute to disease in any specific patient, whereas genetic cohorts (whether recruited on the basis of a shared genetic mutation (discussed in more detail in section VI) or engineered on isogenic backgrounds (discussed in section VII) provide the opportunity to test the impact of a defined genetic variant shared by a subset of patients.

Early hiPSC-based studies of SZ focused primarily on idiopathic patient collections, largely reflecting the availability of patient-derived fibroblasts during this period [2, 52, 53]. More recently, there has been a transition to more defined cohorts, in an attempt to reduce the heterogeneity between patients. Despite the obvious limitations in these early studies with respect to limited cohort size and heterogeneous neuronal populations, there was a striking convergence of findings across both idiopathic and genetic cohorts. These shared findings include aberrant migration / cell polarity (idiopathic [48], 22q11.2 deletion [54], *CNTNAP2* [55], and 15q11.2 microdeletion [56] patients), proliferation (DISC1 [57, 58]) and WNT signaling (idiopathic [59] and *DISC1* [58]) in hiPSC-derived neural progenitor cells (NPCs). Moreover, three groups, across independently reprogrammed and characterized idiopathic SZ cohorts, reported increased oxidative stress [48, 52, 53] and perturbed responses to environmental stressors [48, 60] in patient-derived NPCs. Moreover, patient-derived neurons exhibit decreased neurite outgrowth (idiopathic [2] and 22q11.2 deletion [54]), synaptic maturation (idiopathic [2, 53]; *DISC1* [61], and 15q11.2 microdeletion [56] patients) and activity (idiopathic [62]; *CNTNAP2* [63]), and *DISC1* [61]) and altered neurotransmitter release (idiopathic [64], and *DISC1* [61]). Global gene expression studies confirmed aberrant expression of synaptic genes (idiopathic [2], *DISC1* [61], 22q11.2 deletion [65], queried differential microRNA expression (idiopathic [66] and 22q11.2 deletion [67]) and demonstrated blunted activity-dependent changes of gene expression (idiopathic [68]). A limited number of studies further explored cell type specific effects, focusing specifically on hippocampal neurons [62, 69] or astrocytes [70]. Taken together, these reports mirror findings in postmortem pathological studies (reviewed in [71]) and animal models (reviewed in [72]) and provide a convincing proof-of-concept demonstration that at least some of the cellular and molecular factors underlying SZ are conserved between hiPSC cohorts, suggesting that, at least *in vitro*, the myriad genetic mechanisms contributing to SZ predisposition may manifest through a more limited number of cellular outcomes.

It remains unclear to what extent hiPSC-based studies are sufficiently powered to discover the shared molecular mechanism(s) driving phenotypic differences in patient neural cells. Nonetheless, our recent discovery-based approach combining RNA, microRNA and protein analyses found microRNA-9 (miR-9) to be significantly downregulated in a subset of idiopathic SZ NPCs, a finding that was corroborated in a second hiPSC cohort derived from ten childhood-onset-SZ (COS) patients and ten unrelated controls [66]. Overexpression of miR-9 ameliorated a migration deficit in SZ NPCs, whereas knockdown partially phenocopied aberrant migration in control NPCs [66]. Concurrently, a gene-set enrichment analysis of the largest SZ GWAS to date [26] found an enrichment of predicted miR-9 targets among SZ-associated genes [73], demonstrating a remarkable convergence of independent hiPSC- and genetics-based discovery approaches.

Even within idiopathic hiPSC cohorts, there is a recent effort to focus on patients with shared clinical features such as age of onset, endophenotypes (ie. neurophysiological, biochemical, endocrinological, neuroanatomical, cognitive, or neuropsychological features), or pharmacological response, with the expectation that this may reduce inter-individual variation *in vitro*. In SZ, COS patients represent a subset of adult onset SZ patients defined by onset and severity, with no genetic or clinical differences to chronic poor outcome adult onset SZ [74]; although we have twice applied a COS cohort to replicate findings in our original idiopathic cohort [66, 75], we have no evidence of exaggerated and/or less heterogeneous phenotypes in hiPSC-derived COS neurons, relative to adult-onset SZ neurons. In ASD, two recent studies have now focused specifically on cases with increased head size (macrocephaly) and poorer clinical outcomes [76, 77]. Although their differentiation paradigms differed substantially and cohort

sizes were small (four [76] and eight [77] ASD patients), both reported perturbations in synaptogenesis (increased [76] and decreased [77]), GABAergic differentiation (increased [76] and decreased [77]) and *FOXG1* expression (increased [76, 77]). Given that these macrocephaly-ASD hiPSC studies were not directly contrasted to a more general idiopathic cohort, one cannot say whether this experimental strategy reduced patient heterogeneity or improved disease signal. This issue is better resolved in a fourth hiPSC-based cohort, this one comparing three BD patients with known clinical lithium responsiveness and three with known nonresponsiveness [78]. Here, while hiPSC neurons from all BD patients showed evidence of hyperactive neuronal firing, the phenotype was selectively reversed by lithium treatment only in neurons derived from lithium-responsive BD patients [78], consistent with a genetic mechanism underlying clinical lithium response [79]. A similar pharmacological patient stratification is frequently proposed for SZ; while it has not been fully explored to date, one study examined a pair of monozygotic twin cases with treatment-resistant SZ in which one twin responded well to clozapine treatment and the other twin did not, finding subtle gene expression differences between (untreated) twin neurons [80] but failing to ask if the patient-derived neurons differed in response to clozapine treatment. Overall, we posit that only for those endophenotypes with substantial heritability (ie. cognition [81, 82] and neuroticism [83, 84]) or treatment responses with clear pharmacogenomic evidence [85] will cohort heterogeneity be even modestly reduced through this strategy.

IV. Lessons from case-control SZ post-mortem analyses.

Understanding the natural variation in gene expression between control brains, between individuals and across development, is an important first-step before conducting case-control analyses. Towards this, the GTEx project integrates genetic and transcriptomic data across brain regions in the same subjects [86] while the BrainSpan project characterizes gene expression levels across multiple brain regions during human development [87, 88]. Large-scale multi-site studies have now established post-mortem brain collections with a comprehensive analysis of the genome, epigenome [89] and transcriptome [90-92] of hundreds of neuropsychiatric disorder patients and controls across multiple brain regions (reviewed in [93]).

Post-mortem gene expression studies (and meta-analyses) have identified subtle abnormalities in multiple brain regions and neural cell types, but the results have been inconsistent (reviewed in [94]), likely reflecting the small sample sizes involved in the primary cohorts as well as difficulties in accounting for important covariates or accurately combining datasets that were generated at different sites via divergent methods. In addition to the frequently discussed confounds of drug/alcohol abuse, anti-psychotic treatment, cause of death, post-mortem interval and RNA quality, some of this confusion also reflects a failure to account for cellular composition in postmortem brain tissue [95]; for example, any loss of neurons (or glia) associated with disease progression could result in many false positives due solely to this changing composition. Moreover, cell-type specific perturbations are frequently missed in tissue level analyses (reviewed [93]); for example, genes related to mitochondrial function and ubiquitin-proteasome functions seem to be perturbed in SZ in layer 3 and 5 pyramidal neurons, respectively [96]. Until case-control single cell datasets are available, integrating control single cell data into the analysis of case-control comparisons of heterogeneous cell populations will improve compensation for variable cell type composition [97] and resolution of cell-type specific effects [98].

With dramatically expanded sample sizes, consistency of recent results has improved (reviewed [94].) While the CommonMind Consortium (CMC) RNA sequencing (RNA-seq) evaluation of prefrontal cortex (PFC) brain tissue from 537 individuals concluded that current post-mortem studies are dramatically underpowered to detect differential expression directly attributable to SZ risk variants, others hope that this is an overly pessimistic interpretation [99]. In fact, when variation due to degradation of post-mortem RNA [99] was accounted for in the LIBR RNA-seq analysis of 495 post-mortem brains, they identified 237 genes significantly associated with SZ that implicated synaptic processes, were strongly regulated in early development, and replicated in the CMC dataset [92]. Finally, a large cross-disorder meta-analysis of 715 brain samples across five major neuropsychiatric disorders identified shared down-regulation of neuronal gene networks in ASD, SZ, and bipolar disorder (BD), and up-regulation of astrocyte networks in ASD and SZ [32].

Postmortem studies can be further refined by incorporating genetic data, both from the individual from whom the tissue was obtained as well as the GWAS for the disease being studied. With a sufficiently large cohort, common variants that regulate gene expression can be identified. These expression quantitative trait loci (eQTLs) are regions of the genome containing DNA sequence variants that influence the expression level of one or more genes; they are now understood to vary between cell types and conditions (well reviewed [100]). eQTLs can be integrated with GWAS data to identify genes whose expression levels are associated with disease. One such analysis linked *SNX19* and *NMRAL1* to SZ, noting that the genes identified were not generally the nearest annotated gene to the top associated GWAS SNP. The CMC estimated that ~20% of SZ GWAS loci have eQTLs that could regulate gene expression, and for five demonstrated that perturbing gene expression can have effects on neurodevelopment *in vivo* and/or *in vitro* [91]. A more refined analysis, considering both cellular and temporal contexts, determined that conditional eQTLs are widespread and revised the estimate of SZ GWAS loci with eQTLs to ~37%[101]. A third group, with a revised analytical strategy that incorporated additional post-mortem brain samples across development, concluded that 42.5% of SZ GWAS variants have eQTLs that converge on gene regulation [92]. A similar strategy substitutes open chromatin state for gene expression [102]; an enrichment of neuronal open chromatin regions at SZ GWAS loci was subsequently confirmed in hiPSC-derived neurons [103]. SZ GWAS risk loci are more likely to present expression differences during the fetal-postnatal developmental transition, with a specific enrichment for both dopaminergic and glutamatergic synapse pathways [92]. Given recent progress in other fields, it is likely that further integration of GWAS and eQTL data to generate a SZ transcriptional risk score will better predict disease risk that using a polygenic risk score alone [104].

V. Understanding critical power limitations in the design of hiPSC-based studies of complex genetic disorders

Given the limitations of small sample size and intra-donor variation, future hiPSC-based studies should be designed to maximize statistical power. One critical issue is the tradeoff between increasing the number total donors versus increasing the number of independent hiPSC clones per donor. In general, the best way to maximize effective sample size, while controlling the false positive rate, is to use one hiPSC line per donor and increase the number of donors, rather than using multiple replicate clones from a smaller set of donors [97, 105]. We suggest that future hiPSC-based studies of SZ (and other diseases) will achieve the greatest benefit by focusing on three strategies:

increasing the total number of individuals, reducing inter-donor heterogeneity (by focusing on patients with shared genetic variants) and optimizing neuronal protocols to decrease cellular heterogeneity (and thus decreasing intra-donor variation).

Overall, the issues of sample heterogeneity and statistical power are to a large extent shared between post-mortem and hiPSC-based studies; for example, both approaches are equally impacted by the polygenic nature of SZ and genetic heterogeneity between individuals. Unlike the confounds of post-mortem studies discussed above, hiPSC-based studies are impacted by variation in reprogramming (ie. reprogramming method and batch, technician, hiPSC culture) and neuronal differentiation (ie. reagent batch as well as more stochastic experimental effects affecting cell type composition) efficiencies between hiPSC lines derived from both the same (intra-individual) and different (inter-individual) donors. The fraction of expression variation attributable to each of these factors can be quantified using the variancePartition software [106]. Previous applications have specifically revealed impact of donor, cell type composition and technical artifacts [97, 106-108]

While it was recently estimated that ~28,500 subjects are required for an adequately powered post-mortem case/control study [91], it is unclear whether more or fewer subjects would be necessary for an equivalently powered hiPSC-based cohort. It is simply too early to state with confidence which set of post-mortem and hiPSC specific confounding variables have a greater impact on the observed donor effect in gene expression studies; however, the ability to computationally account for RNA quality [92] and cell-type composition [97] will improve our ability to resolve biological signals moving forward. Critically, recent efforts have spurred the establishment of larger NextGen Consortium hiPSC-based studies (discussed below), from which it is now possible to estimate overall sources of variance and observed donor effects in hiPSC-based studies. Both genetic [109-112] and epigenetic [113-116] errors occur during the reprogramming process. Donor cell type [117, 118] and age [119] can further influence the genome and epigenome of hiPSCs. Moreover, batch effects are seemingly unavoidable in both the reprogramming and differentiation processes. In aggregate, these processes underlie “intra-individual variation”, the subtle differences in gene expression and propensity towards neural differentiation between independent hiPSC lines generated from a given donor. Traditionally, stem cell biologists have attempted to account for these effects by including multiple hiPSCs (up to three) per donor; critically, for complex genetic disorders, we instead propose here that it is more important to power experiments to explore inter- rather than intra-individual variation. For a fixed budget, it is nearly always advisable to add additional donors rather than generate and validate additional hiPSC clones for any given individual (https://gabrielhoffman.shinyapps.io/design_ips_study) [97].

Practical limits of cohort size may mean that studies are inadequately powered at the present time for this highly polygenic disorder. Stochastic differences in the differentiation process can result in dramatic differences in cell type composition that can negatively impact analysis. Moreover, the dynamic nature of gene expression means that such differences in cell type composition, neuronal density or other culture conditions can lead to differences in neuronal activity and other physiological events that impact RNA levels. Finally, the constraints of transcriptome-wide multiple testing burden can make it difficult to identify real signals with statistical significance. It is now widely accepted that it is necessary to decrease patient heterogeneity by selecting more (genetically) homogeneous cases and controls. Hoesktra et al [120] define this objective more precisely, proposing that future studies either (i) select patients with a rare highly penetrant disease-associated genetic variant with a large effect size, or (ii) select

patients with high polygenic risk score based on common genetic variants. They further detail that for both strategies, the ideal design would be to include four groups of individuals: patients with and without the disease penetrant variant/high polygenic risk and controls with and without the disease penetrant variant variant/high polygenic risk [120]. Moreover, a more tenable focus, rather than to continue applying hiPSC-based models for the discovery of SZ risk factors, may be for the field to apply a combination of genetic stratification of patient cohorts and gene-edited isogenic hiPSCs in order to evaluate the functional effects of manipulating putative rare and common causal variants identified through genetic studies.

To a large extent, inter-individual (genetic) differences explain transcriptional variability between hiPSCs [107, 121, 122], consistent with the degree of genetic variation in human gene expression regulation [123-125]. Retaining the donor-specific signature is essential to studying case control differences. In two recent studies of hiPSCs, variance across donors explained a median of ~6% [126] and 48.8% [107] of expression variation, while we observed a much smaller (2.2%) donor effect in hiPSC neurons [97]. Donor effects in differentiated cells may be reduced due to stochastic noise in the differentiation from hiPSCs to neurons. Consistent with this, regulatory variation between individuals is lower in hiPSCs than in two differentiated cell types [121]. It remains unclear whether different hiPSC-derived cell types will retain more or less donor signal over the course of differentiation.

Although not focused on neuronal cell types, many of the insights from the large collection of hiPSCs recently characterized by the NextGen Consortium [127] are very relevant to neuropsychiatric disease, as these reports quantified the genetic [107, 128] and epigenetic [129] basis of variation between hiPSC lines and differentiated progeny [130-132]. Separate work focused on the transcriptomes and chromatin accessibility of hiPSC-derived sensory neurons [108]. These large cohorts facilitated both genome-wide insights into common variants underlying disease [130-132], but also served as a platform to begin personalized (or at least genome-first) drug screening against disease mechanisms or phenotypes [133-135].

VI. Genetically stratified cohorts of rare variant patients

Despite their rarity in the genome (0.01-0.02 per generation [136]), CNVs (frequently *de novo*) were among the first genetic variants associated with SZ. There is an increased CNV burden in SZ cases compared to controls [137, 138], so one method to increase the power of hiPSC-based studies, without increasing sample size, is to focus on genetically homogenous cohorts harboring such a rare variant of large effect size.

22q11.2 deletion was the first CNV associated with SZ and remains the strongest risk factor for developing the disorder, with 25% of carriers exhibiting psychotic symptoms [139-141]. hiPSC-derived neurons from two 22q11.2 patients diagnosed with SZ demonstrated significant reductions in a number of cellular phenotypes including neurosphere size, neuronal differentiation capacity (neuron to glial ratio), neurite outgrowth and cellular migration [54]. Additionally, given that the 22q11.2 deletion encompasses *DGCR8*, a known regulator of miRNA processing, hiPSC cohorts of 22q11.2 deletion cases have been subject to both miRNA and RNA sequencing. miRNA sequencing, performed on a cohort of six patients compared to six controls, identified a significant increase in the expression of several miRNAs that overlap with differentially expressed miRNAs in post-mortem brain and peripheral cells from 22q11.2 cases [67]. In a larger cohort of eight 22q11.2 cases and seven controls, RNA-seq identified changes

in expression of nearly all genes within the deletion region, as well as 745 genes outside of the region implicating apoptosis, cell cycle and MAPK signaling [142]. These studies demonstrate that rare variant cohorts can identify cellular phenotypes and molecular pathways that are consistent with human post mortem findings.

A gene-editing approach can also be used to investigate the causal role of a gene at a SZ-associated CNV. Given that *NRXN1* deletions are only 6.4% penetrant, Pak *et al* used CRISPR/Cas9 (CRISPR-associated protein 9) technology to engineer two heterozygous deletions in a control line, to assess if these deletions were sufficient to produce neuronal deficits. Both deletions resulted in deficits in neurotransmitter release, with no changes in synapse number or neuronal differentiation capacity, a finding consistent with mouse models [143]. Isogenic comparisons are crucial for establishing a causal role for incompletely penetrant rare variants such as *NRXN1*.

While focusing on rare genetic variants with large effect size will increase the power of hiPSC-based analyses, their incomplete penetrance and pleiotropic effects may complicate the identification of disease specific phenotypes, which are likely impacted by polygenic risk. Therefore, it is imperative to consider the genetic background of common variations in each donor, even when focusing on variants of large effect size. Ideally, isogenic controls should be used whenever possible to demonstrate causal relationships between genetic variants and their phenotypic consequences.

VII. CRISPR-editing and manipulating of expression to study common variant effects

hiPSC-based models can be applied to functionally evaluate common variant risk genes identified by such analyses. For example, induced neurons (iNeurons) generated from over 20 individuals with different genotypes for a non-coding common variant in the voltage-gated calcium channel subunit gene *CACNA1C* demonstrated that the homozygous risk allele genotype correlated with increased *CACNA1C* mRNA expression and calcium channel current density [144]. While this study correlated genotype to expression, it did not directly demonstrate the role of the SNP in regulating gene expression.

A combination of CRISPR-based tools now make it possible to precisely engineer hiPSC lines with single nucleotide mutations to recapitulate common variants, finely tuned endogenous gene expression changes, as well as the multiplexing of such modifications [145-147]. A number of critical issues limit these studies and will be discussed below. First are the decisions of which hiPSC line (genotype and diagnosis) and which precise SNP(s) to engineer, and whether to edit one or both alleles. Second is how to prevent and identify off-target effects, in order to ensure the construct validity of any biological findings.

Obviously, as well established between different mouse strains, genetic background has dramatic impact on gene expression and would be predicted to significantly impact the effects of genetic manipulation [148-152]. Therefore, one important question is whether to conduct these isogenic experiments in control or SZ hiPSCs. If penetrance is to be first established in control backgrounds, there is the possibility that potential (protective) compensatory processes may limit the impact of the genetic change. If effects are queried in patients, there is the possibility that any of the other risk alleles may interact with the edited SNP to confuse the results. Likely, isogenic comparisons will need to be conducted across a number of genetic backgrounds. Establishing the polygenic risk

scores and haplotypes of different donor cell lines will be critical when choosing a cell line for genetic engineering. Close collaboration between computational geneticists and biologists is essential for the study of complex genetic disorders.

It remains unknown to what extent, and across how many putative causal loci, manipulating eQTLs in control hiPSCs will be sufficient to alter the levels and/or splicing of associated mRNAs, leading to downstream phenotypic changes in one or more neural cell types. To date this has been accomplished comparatively few times, so careful choice of the candidate variant will prove beneficial. Forrest et al. prioritized SZ-associated variants located in neuron-specific open chromatin regions in iNeurons. CRISPR correction of a common variant risk allele near the mir137 locus increased expression of mir137, dendritic complexity and synapse maturation [103]. CRISPR-editing of common variants associated with other diseases has similarly been shown to affect gene expression in hiPSC-derived disease-relevant cell types [153-155]. In a novel attempt to bridge the study of common and rare variants, Castel et al. [156] tested the potential effect of regulatory variants on the penetrance of pathogenic coding variants. By introducing a pathogenic Mendelian variant into a known eQTL haplotype they elegantly demonstrated a possible mechanism in which common variants can contribute to variable penetrance of genetic disorders

A distinct advantage of isogenic comparisons is that small effect sizes can be more readily resolved than in non-isogenic comparisons (see section VIII). However, even between isogenic cell lines, inherent heterogeneity caused by extended culture and variability during differentiation could mask subtle effects [157, 158]. Therefore, although the obvious experiment is to compare effects between homozygous risk and non-risk genotypes, an alternative is to instead compare allele-specific expression in heterozygous neurons in order to avoid this inherent transcriptional heterogeneity between cell lines. Such an approach was used to demonstrate that a common SNP risk allele associated with Parkinson's disease affects transcription factor binding efficiency and thereby alters gene expression [159].

Particularly when considering common alleles with small predicted effect sizes, it is critical to consider the possibility of off-target Cas9 nuclease activity. Although whole genome sequencing of ten CRISPR/Cas9-edited hPSC lines suggested low occurrence of off-target effects in hPSCs [160], other evidence suggests that the risk of off-target may have been underestimated [161-164]. Off-target effects can be reduced by incorporating transfection of Cas9 protein or mRNA (rather than plasmid DNA) [165], biochemically modified Cas9 varieties (ie. eSpCas9 and SpCas9-HF1) that exhibit reduced interaction between Cas9 and the target DNA [166, 167], improved gRNA design [168] and screening methods (ie. CIRCLE-seq and GUIDE-seq) [169, 170]. If off-target effects cannot be eliminated, independent engineering of multiple isogenic pairs via different gRNAs should be sufficient to rule out their impact on the biological effects observed.

A significant hindrance to the study of common disease-associated variants is their great likelihood to be in high linkage disequilibrium (LD) with other variants. While the causal SNP may be predicted via fine mapping [171, 172], it cannot always be identified unequivocally. It may be more simple to directly manipulate endogenous gene expression; in this way, hiPSC-derived neural cells were used to show that the SZ candidate gene *ZNF804A* impacts inflammatory cytokine response in differentiating neurons [173], *CYFIP1* affects neural polarity in NPCs [174] and *FURIN* alter neural migration in NPCs [91]. Not all genes are appropriate targets for traditional overexpression or knockdown experiments, particularly those that are very long or highly alternatively spliced. Fortuitously, gene expression modulation via CRISPR/Cas9 occurs at the

promoter or enhancer, and therefore is predicted to include the full range of alternative splice isoforms that are frequently overlooked by cDNA overexpression or RNAi approaches. By introducing nuclease-null mutations into Cas9 [175, 176] and fusing the catalytically inactive or dead Cas9 (dCas9) to a variety of effector protein domains, the modulation of transcription [175, 176], DNA methylation [177, 178] and histone modifications [179] have all been demonstrated. By testing dCas9-mediated transcriptional modulation using three such platforms, we recently fully evaluated the efficacy and variability of dCas9-protein fusion-based transcriptional modulation of seven different SZ-associated risk genes in three different hiPSC-derived neural cell types (NPCs, neurons and astrocytes), using hiPSCs from three unique donors [180]. While this platform has not proven equally efficacious across all neuronal genes, donors and cell types tested, it can be a fast method to achieve disease relevant changes in gene expression.

VIII. Integrating advanced genetics and hiPSCs in future approaches

As hiPSC-based studies increasingly incorporate isogenic comparisons, stem cell biologists must contemplate statistical and biological considerations that are already well accepted in the field of genetics. When planning CRISPR-mediated isogenic hiPSC-based comparisons to validate disease-associated eQTLs, care must be taken to design a well powered study. In general, the power to identify a statistically significant difference in the expression of a particular gene between two or more groups of samples depends on the magnitude of the expression differences between the groups (i.e. effect size) and the amount of expression heterogeneity within each group (i.e. variance). Standard eQTL analysis divides samples into three allelic categories based on a given SNP and determines if the number of reference alleles for that SNP is significantly associated with expression of a particular gene. For a particular SNP, the statistical power to detect a significant effect on gene expression varies by gene, reflecting both the effect size of the eQTL and the expression variance within each of the three allelic groups (**Figure 2**). Although isogenic hiPSC and post-mortem experiments are predicted to have equivalent eQTL effect sizes, we strongly believe that isogenic hiPSC-based approaches will demonstrate reduced expression variance due to the common genetic background in isogenic comparisons. Therefore, assuming even a 5-fold reduction in standard deviation (i.e. square root of the variance) for most genes, an isogenic design is well powered to identify differences in the cis-gene targets (assuming a comparison of four isogenic hiPSC replicates from a single donor with the CRISPR allele swap to the homozygous alternative allele to twelve isogenic replicates with the homozygous reference allele). While we hypothesize that using isogenic lines will reduce biological and technical variation, until the extent to which this is true is empirically evaluated, it must be modeled across a range of values of expression variability.

Statistical considerations for CRISPR validation studies

Designing a successful study requires adapting the experiment, question or analysis so that it is experimentally tractable, biologically plausible and statistically well powered. Here, although we consider the power to detect a significant expression change following a targeted CRISPR allele swap of a SNP affecting gene expression, where the eQTL association was identified in postmortem brain, the same reasoning applies to testing case/control expression differences. The statistical power to detect expression differences between two sets of samples depends on four major factors discussed below: magnitude of expression difference, expression variance, sample size and significance level.

i) *Magnitude of the expression difference.* For a given SNP-gene pair, a linear regression model can be fit to estimate the effect size (β) of increasing the number of minor alleles and the variance of the expression residuals (σ^2) after the SNP effect is considered. Let μ_0 and μ_2 be the average expression levels of the two homozygote classes. Since β is the difference in expression from changing a single allele, then $\mu_2 - \mu_0$ = 2β. Power calculations depend on the term $2\beta/\sigma$ and generally assume that the effect size remains the same in the discovery and validation datasets. But, when isogenic hiPSC studies are used to validate findings identified in whole brain tissue, this variation in cell type composition can reduce the observed magnitude of a signal present in only one cell type. Therefore, it is obviously desirable to conduct hiPSC-based isogenic experiments in the appropriate neural cell type where the eQTL is most active, increasing the statistical power by increasing the effect size.

ii) *Expression variance.* Gene expression levels measured in postmortem brain can be highly variable due to donor genetics, variation in cell type composition, environmental effects experienced by the donor, and technical variation in sample processing. Increasing the fraction of expression variation explained by the SNP of interest by reducing the variation attributable to extraneous variables will improve the statistical power to validate an initial finding. Using isogenic cell culture model as a validation system should remove some of this variation and increase statistical power.

iii) *Sample size*. Collecting a sufficient number of samples is essential to conducting a well-powered study, especially when effect sizes are small. Of course, findings should always report the number of hiPSC lines (and individuals represented), experimental replicates and technical replicates conducted for each analysis. Observations should be presented by individual, rather than as averaged group effects, so as to most transparently reflect variation within these isogenic comparisons. Interpreting expression variation within each phenotype class is essential to interpreting the biological relevance of a finding.

Obviously, cost constraints are often the major factor limiting sample size. Therefore, one solution is to conduct all isogenic comparisons in a few common hiPSC line(s), allowing all engineered lines to serve as additional isogenic controls for the other pairs. The effects of biological heterogeneity can be further reduced by conducting a paired statistical analysis whereby each perturbed sample is compared to a donor-specific baseline (**Figure 3**).

iv) *Significance level*. The central goal of most gene expression studies is to identify expression differences that are larger than expected by chance. Findings are generally evaluated based on passing statistical significance, by convention often $p<0.05$. When multiple genes are tested in an analysis, the p-value cutoff to attain significance must be adjusted in order to account for the number of statistical hypothesis tests. This is known as the 'multiple testing problem'. Analyses focusing on a single gene can use a nominal p-value cutoff of 0.05. Genome-wide analyses must overcome a higher multiple testing burden in order to control the false positive rate. The Bonferroni correction is widely used and intuitive so that a nominal p-value cutoff of 0.05 corresponds to 0.05/k when k genes are tested. When 20,000 genes are tested the Bonferroni cutoff becomes $p<2.5e-6$ and is much more stringent that the nominal cutoff.

The scope of the biological question and the statistical analysis has major impact on the multiple testing burden and the power of the study. Thus decreasing the scope of an analysis by only considering genes near a SNP of interest can dramatically increase

statistical power compared to a genome-wide analysis. Overall, focused validation experiments will always be better powered than genome-wide discovery.

Biological considerations for CRISPR validation studies

While the effect of modifying an eQTL SNP in an hiPSC-based study of gene expression can be estimated from post mortem data, considering the effect of a SNP identified by GWAS of a psychiatric trait is more challenging and open-ended. Whereas genetic variants have a large effect on direct phenotypes (such as gene expression), the effect on higher-level phenotypes (such as neuronal function or diagnosis) will be attenuated by buffering and/or environmental effects. Thus, the effect size will likely reflect how far the assayed phenotype is from genetics [181, 182] and assumes that one has selected the correct intermediate phenotype to study. For a given eQTL SNP, the sorting of effect sizes might look like: gene expression > protein > neuronal function > psychosis > SZ. A particular SNP might not act through gene expression, or not under the specific conditions of the experiment, or not in a particular cell type.

Moreover, the question remains as to how to link eQTL associations to the complex behavioral phenotypes associated with SZ, ranging from delusions, hallucinations, negative affect and impaired cognition. Obviously, once the constraints of these isogenic studies are better understood, future studies may wish to improve the complexity and maturity of the neural cells being queried, advancing towards circuit-level complexity. Moving forward, it will be increasingly straightforward to incorporate three-dimensional culture techniques and/or generate defined neuronal circuits comprised of specific neuronal cell types, synapsed in a defined orientation, together with oligodendrocytes to provide myelination, and astrocytes and microglia to incorporate critical aspects of inflammation and synaptic pruning. Moreover, either by neuronal stimulation or treatment with physiologically-relevant exposures of stress hormones and/or other environmental factors, it may one day be possible to further unravel the causality of environmental risk factors such as neuroinflammation, stress and drug exposure [183] using isogenic hiPSC-based comparisons. While models, by definition, will always lack the intricacies of human disease, our goal must always be to strive towards the complexity of the human brain.

IX. A perspective on the hypothesis of biological convergence of disease

Much of the current work to improve our understanding of disease etiology and develop novel treatment strategies is implicitly predicated on an assumption of biological convergence. Under this model, patients who share high-level psychiatric symptoms are clinically diagnosed with a particular disease and are hypothesized to share a disruption of some lower-level biological function. As discussed in section II, this convergence model has yielded themes of disruption in ion channels and neuronal genes in SZ. Recent psychiatric research has pursued models of biological convergence at multiple levels (**Figure 4**), including genetic [26], epigenetic [89, 184] gene expression [91, 92], gene module [32, 91, 92], proteome [185, 186], braining imaging [187, 188], drug response [85], psychiatric endophenotypes and disease subtypes [189-191] and high-level disease phenotypes [26].

The best level to study this common hypothesis of functional convergence is unclear; it may be too low level or underpowered to detect at the level of gene expression and more likely to be identified at the level of modules/pathways/gene-sets/subnetworks.. For example, while there may be hundreds of genetic perturbations that converge at the level of synaptic function, the majority of these likely have a distinct expression signature

[34]. Therefore, detecting a convergence at the molecular level is dependent on having the proper module/pathway/gene-set/subnetwork to test. Whereas these pathways show clear convergence at the higher behavioral/psychiatric level, and deconstructing this convergence to lower level phenotypes can reveal shared molecular etiology and potential therapies, which level to focus on is not clear. The amount of convergence increases with complexity (i.e. toward psychiatry) since the diagnosis is based on is psychiatric symptoms, but so does the cost per patient to study. Going down a level (towards DNA) allows increased resolution, sample size and prospects for understanding molecular mechanism, but also dramatically increases multiple testing burden and is susceptible to a lack of biological convergence. Obviously, each has its own technical and logistical challenges.

A major challenge in psychiatric genomics is how best to align a patient cohort, biological assay and analytical approach with a hypothesis of convergence that is both biologically feasible and statistically well powered. While in ASD there is sufficient convergence at the genetic and pathway levels that exome and genome sequencing have already identified multiple genetic disruptions of the same genes and pathways, such clear convergence is not seen in SZ. Since SZ risk is conferred by genetic variants of weaker effect sizes, there is weaker convergence at the genetic and expression levels in SZ, and so convergence has only been observed at the pathway (i.e. ion channel) level. Moving forward, this can be addressed by increasing sample size, developing focused cohorts to increase statistical power and applying integrative statistical methods to detect convergence at new levels [34, 43]. The open questions in psychiatric genetics are therefore, given a set of patients, with a given amount of heterogeneity in genetics, phenotype and technical noise, as well as cost limitations: what level of biological convergence can be expected, is the assay able to detect this level of convergence and are we statistically well-powered to detect a signal of a reasonable effect size? By thoughtfully integrating GWAS, post-mortem and hiPSC-based approaches, we hope that the molecular convergence underling SZ and other complex genetic disorders will become better resolved, revealing novel points of therapeutic intervention.

ACKNOWLEDGEMENTS

Kristen J Brennand is a New York Stem Cell Foundation - Robertson Investigator. The Brennand Laboratory is supported by a Brain and Behavior Young Investigator Grant, National Institute of Health (NIH) grants R01MH101454 and R01MH106056 and the New York Stem Cell Foundation. Erin Flaherty is supported by F31MH112285. The authors declare no conflict of interest exists.

FIGURES

Figure 1: Genetic contributions to neuropsychiatric disorders. A) Current state of psychiatric genomics with the assays and analyses linking genetics, gene expression and disease biology. **B)** Genetic liability to neuropsychiatric disease. Common variants constitute the majority of the genetic contribution to neuropsychiatric disorders. Graph adapted from Gandal et al. [8]. **C/D)** Representative plots illustrating the polygenetic nature of neuropsychiatric diseases. The liability threshold plot shows the normal distribution of disease liability among the population. Genetic and environmental factors combined may lead to crossing of the threshold into disease. The dose-response plot visualizes the same additive effect of genetic liability to disease risk.

Figure 2: Genetics of gene expression and statistical power. A) Statistics from eQTL analysis of 3 representative genes from post mortem RNA-seq data from the CommonMind Consortium [91]. The reference SNP ID (rsid) of the SNP with the smallest p-value for each gene is shown along with the corresponding p-value, effect size (β), residual variance (σ^2, but shown on the scale of standard deviation (σ)), and the percentage of expression variation explained by the SNP. **B)** Observed expression of the 3 genes from (A) in post mortem brains stratified by the corresponding eQTL SNP. **C)** Hypothetical gene expression in isogenic hiPSC-derived cell lines substantially reduces the amount of expression variation due to multiple genetic and non-genetic factors. **D)** Statistical power of targeted CRISPR allele swap from one homozygote to the other as a function of the residual variance, where effect size and variance of each gene were computed from CommonMind Consortium data of 534 post mortem brains. For each gene the power is computed as a function of the residual variance for a study of 4 hiPSC lines that are homozygous for the alternative allele at the relevant SNP compared to 12 hiPSC lines that are homozygous for the reference allele. Power is evaluated at $p<0.05$ corresponding to a focused analysis testing only a single gene. Power is shown as σ decreases by up to a factor of 10. We note that statistical power varies substantially across genes.

Figure 3: Statistical tests for independent and paired samples. **A)** Testing expression differences between two genotypes in independent samples involves comparing the mean expression between the two genotypes. The statistical analysis considers whether the difference in the mean expression in the homozygous reference genotype (i.e. Ref/Ref) is statistically different from the mean expression in the homozygous alternative genotype (i.e. Alt/Alt). This corresponds to testing if the difference in means is statistically different from zero. Here the color of the observation indicates the donor, but the fact that each donor is measured twice is not used. **B)** When a study design involves paired samples as in a perturbation experiment with treated and control samples from the same individual, an individual-specific baseline can be used to compute the expression change between the two genotypes within each individual. In this case, the statistical test considers the difference between the treated and control samples for each individual, and then tests if the mean of the differences is statistically different from zero. The colors of the observations indicate the donor and the statistical model explicitly considers the fact that the paired observations are from the same donor. In R, the paired model is used by including the donor in the regression formula.

Figure 4: Hypothesis of multiple levels of biological convergence. A) Multiple genetic variants directly affect the same risk gene. **B)** Risk variants have an indirect

effect on the expression of a downstream gene. **C)** Risk variants indirectly affect a higher level molecular process by disrupting the function of a biological module of genes. **D)** Risk variants disrupt intermediate genes to indirectly affect a psychiatric phenotype. **E)** Biological variation is traced from low level DNA variants to high level disease phenotype through intermediate levels of increasing complexity. The possibility of performing biological assays at each level is indicated for hiPSCs, post mortem brains and living patients. Gradients indicate the level of biological convergence expected for idiopathic versus a genetically selected cohort. The feasible sample size is largest for studies of DNA and disease phenotype (i.e. GWAS), but is much lower for intermediate phenotypes.

REFERENCES

1. Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y *et al.* A model for neural development and treatment of rett syndrome using human induced pluripotent stem cells. *Cell* 2010; **143**(4)**:** 527-539.

2. Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* 2011; **473**(7346)**:** 221-225.

3. Miller JD, Ganat YM, Kishinevsky S, Bowman RL, Liu B, Tu EY *et al.* Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* 2013; **13**(6)**:** 691-705.

4. Talkowski ME, Minikel EV, Gusella JF. Autism spectrum disorder genetics: diverse genes with diverse clinical outcomes. *Harv Rev Psychiatry* 2014; **22**(2)**:** 65-75.

5. Pimenova AA, Raj T, Goate AM. Untangling Genetic Risk for Alzheimer's Disease. *Biol Psychiatry* 2017.

6. Owen MJ, Sawa A, Mortensen PB. Schizophrenia. *Lancet* 2016; **388**(10039)**:** 86-97.

7. Simons YB, Sella G. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Current opinion in genetics & development* 2016; **41:** 150-158.

8. Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH. The road to precision psychiatry: translating genetics into disease mechanisms. *Nat Neurosci* 2016; **19**(11)**:** 1397-1407.

9. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015; **47**(7)**:** 702-709.

10. Corvin A, Sullivan PF. What Next in Schizophrenia Genetics for the Psychiatric Genomics Consortium? *Schizophr Bull* 2016; **42**(3)**:** 538-541.

11. Neale BM, Sklar P. Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation. *Curr Opin Neurobiol* 2015; **30:** 131-138.

12. De Rubeis S, Buxbaum JD. Genetics and genomics of autism spectrum disorder: embracing complexity. *Hum Mol Genet* 2015; **24**(R1)**:** R24-31.

13. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet* 2016; **17**(1)**:** 9-18.

14. Browne HA, Gair SL, Scharf JM, Grice DE. Genetics of obsessive-compulsive disorder and related disorders. *Psychiatr Clin North Am* 2014; **37**(3)**:** 319-335.

15. Flint J, Kendler KS. The genetics of major depression. *Neuron* 2014; **81**(3)**:** 484-503.

16. Hawi Z, Cummins TD, Tong J, Johnson B, Lau R, Samarrai W *et al.* The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol Psychiatry* 2015; **20**(3)**:** 289-297.

17. Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet* 2016; **48**(5)**:** 552-555.

18. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 2017; **49**(1)**:** 27-35.

19. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**(7526)**:** 216-221.

20. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**(7526)**:** 209-215.

21. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014; **511**(7509)**:** 344-347.

22. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**(7487)**:** 185-190.

23. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**(7487)**:** 179-184.

24. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landen M *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci* 2016; **19**(11)**:** 1433-1441.

25. Shohat S, Ben-David E, Shifman S. Varying Intolerance of Gene Pathways to Mutational Classes Explain Genetic Convergence across Neuropsychiatric Disorders. *Cell reports* 2017; **18**(9)**:** 2217-2227.

26. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**(7510)**:** 421-427.

27. Cross-Disorder Group of the Psychiatric Genomics C, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**(9)**:** 984-994.

28. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* 2014; **46**(8)**:** 881-885.

29. Finucane H, Reshef Y, Anttila V, Slowikowski K, Gusev A, Byrnes A *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *bioRxiv* 2017.

30. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar H *et al.* Genetic Identification Of Brain Cell Types Underlying Schizophrenia. *bioRxiv* 2017.

31. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**(7256)**:** 748-752.

32. Gandal MJ, Haney J, Parikshak N, Leppa V, Horvath S, Geschwind DH. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *bioRxiv* 2016.

33. Shi H, Kichaev G, Pasaniuc B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet* 2016; **99**(1)**:** 139-153.

34. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 2017; **169**(7)**:** 1177-1186.

35. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* 2017; **101**(1)**:** 5-22.

36. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet* 2017; **49**(7)**:** 978-985.

37. Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Moran JL *et al.* Common alleles contribute to schizophrenia in CNV carriers. *Mol Psychiatry* 2016; **21**(8)**:** 1153.

38. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A *et al.* Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 2012; **149**(3)**:** 525-537.

39. Loohuis LM, Vorstman JA, Ori AP, Staats KA, Wang T, Richards AL *et al.* Genome-wide burden of deleterious coding variants increased in schizophrenia. *Nat Commun* 2015; **6:** 7501.

40. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 2013; **154**(3)**:** 518-529.

41. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 2015; **87**(6)**:** 1215-1233.

42. Network, Pathway Analysis Subgroup of the Psychiatric Genomics C, International Inflammatory Bowel Disease Genetics C, International Inflammatory Bowel Disease Genetics Consortium I. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 2015; **18**(2)**:** 199-209.

43. Ballouz S, Gillis J. Strength of functional signature correlates with effect size in autism. *Genome Med* 2017; **9**(1)**:** 64.

44. McMahon FJ, Insel TR. Pharmacogenomics and personalized medicine in neuropsychiatry. *Neuron* 2012; **74**(5)**:** 773-776.

45. Millan MJ, Andrieux A, Bartzokis G, Cadenhead K, Dazzan P, Fusar-Poli P *et al.* Altering the course of schizophrenia: progress and perspectives. *Nature reviews* 2016; **15**(7)**:** 485-515.

46. Nicholas CR, Chen J, Tang Y, Southwell DG, Chalmers N, Vogt D *et al.* Functional maturation of hPSC-derived forebrain interneurons requires an extended timeline and mimics human neural development. *Cell Stem Cell* 2013; **12**(5)**:** 573-586.

47. Mariani J, Simonini MV, Palejev D, Tomasini L, Coppola G, Szekely AM *et al.* Modeling human cortical development in vitro using induced pluripotent stem cells. *Proc Natl Acad Sci U S A* 2012; **109**(31)**:** 12770-12775.

48. Brennand K, Savas JN, Kim Y, Tran N, Simone A, Hashimoto-Torii K *et al.* Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia. *Mol Psychiatry* 2015; **20**(3)**:** 361-368.

49. Sloan SA, Darmanis S, Huber N, Khan TA, Birey F, Caneda C *et al.* Human Astrocyte Maturation Captured in 3D Cerebral Cortical Spheroids Derived from Pluripotent Stem Cells. *Neuron* 2017; **95**(4)**:** 779-790 e776.

50. Bardy C, van den Hurk M, Eames T, Marchand C, Hernandez RV, Kellogg M *et al.* Neuronal medium that supports basic synaptic functions and activity of human neurons in vitro. *Proc Natl Acad Sci U S A* 2015; **112**(20)**:** E2725-2734.

51. Vera E, Bosco N, Studer L. Generating Late-Onset Human iPSC-Based Disease Models by Inducing Neuronal Age-Related Phenotypes through Telomerase Manipulation. *Cell reports* 2016; **17**(4)**:** 1184-1192.

52. Paulsen Bda S, de Moraes Maciel R, Galina A, Souza da Silveira M, dos Santos Souza C, Drummond H *et al.* Altered oxygen metabolism associated to neurogenesis of induced pluripotent stem cells derived from a schizophrenic patient. *Cell Transplant* 2012; **21**(7)**:** 1547-1559.

53. Robicsek O, Karry R, Petit I, Salman-Kesner N, Muller FJ, Klein E *et al.* Abnormal neuronal differentiation and mitochondrial dysfunction in hair follicle-derived induced pluripotent stem cells of schizophrenia patients. *Mol Psychiatry* 2013; **18**(10)**:** 1067-1076.

54. Toyoshima M, Akamatsu W, Okada Y, Ohnishi T, Balan S, Hisano Y *et al.* Analysis of induced pluripotent stem cells carrying 22q11.2 deletion. *Translational psychiatry* 2016; **6**(11)**:** e934.

55. Lee IS, Carvalho CMB, Douvaras P, Ho SM, Hartley BJ, Zuccherato LW *et al.* Characterization of molecular and cellular phenotypes associated with a heterozygous CNTNAP2 deletion using patient-derived hiPSC neural cells. *NPJ Schizophrenia* 2015; **1:** 15019.

56. Yoon KJ, Nguyen HN, Ursini G, Zhang F, Kim NS, Wen Z *et al.* Modeling a genetic risk for schizophrenia in iPSCs and mice reveals neural stem cell deficits associated with adherens junctions and polarity. *Cell Stem Cell* 2014; **15**(1)**:** 79-91.

57. Murai K, Sun G, Ye P, Tian E, Yang S, Cui Q *et al.* The TLX-miR-219 cascade regulates neural stem cell proliferation in neurodevelopment and schizophrenia iPSC model. *Nat Commun* 2016; **7:** 10965.

58. Srikanth P, Han K, Callahan DG, Makovkina E, Muratore CR, Lalli MA *et al.* Genomic DISC1 Disruption in hiPSCs Alters Wnt Signaling and Neural Cell Fate. *Cell reports* 2015; **12**(9)**:** 1414-1429.

59. Topol A, Zhu S, Tran N, Simone A, Fang G, Brennand KJ. Altered WNT Signaling in Human Induced Pluripotent Stem Cell Neural Progenitor Cells Derived from Four Schizophrenia Patients. *Biol Psychiatry* 2015; **78**(6)**:** e29-34.

60. Hashimoto-Torii K, Torii M, Fujimoto M, Nakai A, El Fatimy R, Mezger V *et al.* Roles of heat shock factor 1 in neuronal response to fetal environmental risks and its relevance to brain disorders. *Neuron* 2014; **82**(3)**:** 560-572.

61. Wen Z, Nguyen HN, Guo Z, Lalli MA, Wang X, Su Y *et al.* Synaptic dysregulation in a human iPS cell model of mental disorders. *Nature* 2014; **515**(7527)**:** 414-418.

62. Yu DX, Di Giorgio FP, Yao J, Marchetto MC, Brennand K, Wright R *et al.* Modeling hippocampal neurogenesis using human pluripotent stem cells. *Stem Cell Reports* 2014; **2**(3)**:** 295-310.

63. Flaherty E, Deranieh RM, Artimovich E, Lee IS, Siegel AJ, Levy DL *et al.* Patient-derived hiPSC neurons with heterozygous CNTNAP2 deletions display altered neuronal gene expression and network activity. *NPJ Schizophr* 2017; **3**(1)**:** 35.

64. Hook V, Brennand KJ, Kim Y, Toneff T, Funkelstein L, Lee KC *et al.* Human iPSC neurons display activity-dependent neurotransmitter secretion: aberrant catecholamine levels in schizophrenia neurons. *Stem Cell Reports* 2014; **3**(4)**:** 531-538.

65. Lin M, Pedrosa E, Shah A, Hrabovsky A, Maqbool S, Zheng D *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One* 2011; **6**(9)**:** e23356.

66. Topol A, Zhu S, Hartley BJ, English J, Hauberg ME, Tran N *et al.* Dysregulation of miRNA-9 in a Subset of Schizophrenia Patient-Derived Neural Progenitor Cells. *Cell reports* 2016; **15**(5)**:** 1024-1036.

67. Zhao D, Lin M, Chen J, Pedrosa E, Hrabovsky A, Fourcade HM *et al.* MicroRNA Profiling of Neurons Generated Using Induced Pluripotent Stem Cells Derived from Patients with Schizophrenia and Schizoaffective Disorder, and 22q11.2 Del. *PLoS One* 2015; **10**(7)**:** e0132387.

68. Roussos P, Guennewig B, Kaczorowski DC, Barry G, Brennand KJ. Activity-Dependent Changes in Gene Expression in Schizophrenia Human-Induced Pluripotent Stem Cell Neurons. *JAMA Psychiatry* 2016; **73**(11)**:** 1180-1188.

69. Han J, Kim HJ, Schafer ST, Paquola A, Clemenson GD, Toda T *et al.* Functional Implications of miR-19 in the Migration of Newborn Neurons in the Adult Brain. *Neuron* 2016; **91**(1)**:** 79-89.

70. Windrem MS, Osipovitch M, Liu Z, Bates J, Chandler-Militello D, Zou L *et al.* Human iPSC Glial Mouse Chimeras Reveal Glial Contributions to Schizophrenia. *Cell Stem Cell* 2017; **21**(2)**:** 195-208 e196.

71. Lewis DA, Glantz LA, Pierri JN, Sweet RA. Altered cortical glutamate neurotransmission in schizophrenia: evidence from morphological studies of pyramidal neurons. *Ann N Y Acad Sci* 2003; **1003:** 102-112.

72. Jaaro-Peled H, Ayhan Y, Pletnikov MV, Sawa A. Review of pathological hallmarks of schizophrenia: comparison of genetic models with patients and nongenetic models. *Schizophr Bull* 2010; **36**(2)**:** 301-313.

73. Hauberg ME, Roussos P, Grove J, Borglum AD, Mattheisen M, Schizophrenia Working Group of the Psychiatric Genomics C. Analyzing the Role of MicroRNAs in Schizophrenia in the Context of Common Genetic Risk Variants. *JAMA Psychiatry* 2016; **73**(4)**:** 369-377.

74. Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: update 2012. *Mol Psychiatry* 2012; **17**(12)**:** 1228-1238.

75. Xu J, Hartley BJ, Kurup P, Phillips A, Topol A, Xu M *et al.* Inhibition of STEP61 ameliorates deficits in mouse and hiPSC-based schizophrenia models. *Mol Psychiatry* 2016.

76. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* 2015; **162**(2)**:** 375-390.

77. Marchetto MC, Belinson H, Tian Y, Freitas BC, Fu C, Vadodaria KC *et al.* Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry* 2017; **22**(6)**:** 820-835.

78. Mertens J, Wang QW, Kim Y, Yu DX, Pham S, Yang B *et al.* Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature* 2015; **527**(7576)**:** 95-99.

79. Hou L, Heilbronner U, Degenhardt F, Adli M, Akiyama K, Akula N *et al.* Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. *Lancet* 2016; **387**(10023)**:** 1085-1093.

80. Nakazawa T, Kikuchi M, Ishikawa M, Yamamori H, Nagayasu K, Matsumoto T *et al.* Differential gene expression profiles in neurons generated from lymphoblastoid B-cell line-derived iPS cells from monozygotic twin cases with treatment-resistant schizophrenia and discordant responses to clozapine. *Schizophr Res* 2017; **181:** 75-82.

81. Trampush JW, Yang ML, Yu J, Knowles E, Davies G, Liewald DC *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry* 2017; **22**(3)**:** 336-345.

82. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**(7604)**:** 539-542.

83. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* 2016; **48**(6)**:** 624-633.

84. Lo MT, Hinds DA, Tung JY, Franz C, Fan CC, Wang Y *et al.* Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat Genet* 2017; **49**(1)**:** 152-156.

85. Ruderfer DM, Charney AW, Readhead B, Kidd BA, Kahler AK, Kenny PJ *et al.* Polygenic overlap between schizophrenia risk and antipsychotic response: a genomic medicine approach. *Lancet Psychiatry* 2016; **3**(4)**:** 350-357.

86. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235)**:** 648-660.

87. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A *et al.* Transcriptional landscape of the prenatal human brain. *Nature* 2014; **508**(7495)**:** 199-206.

88. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012; **489**(7416)**:** 391-399.

89. Psych EC, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ *et al.* The PsychENCODE project. *Nat Neurosci* 2015; **18**(12)**:** 1707-1712.

90. BrainSeq AHBGC. BrainSeq: Neurogenomics to Drive Novel Target Discovery for Neuropsychiatric Disorders. *Neuron* 2015; **88**(6)**:** 1078-1083.

91. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 2016; **19**(11)**:** 1442-1453.

92. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado Torres L *et al.* Developmental And Genetic Regulation Of The Human Cortex Transcriptome In Schizophrenia. *bioRxiv* 2017.

93. Jaffe AE. Postmortem human brain genomics in neuropsychiatric disorders--how far can we go? *Curr Opin Neurobiol* 2016; **36:** 107-111.

94. Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci* 2017.

95. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* 2013; **10**(10)**:** 949-955.

96. Arion D, Corradi JP, Tang S, Datta D, Boothe F, He A *et al.* Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder. *Mol Psychiatry* 2015; **20**(11)**:** 1397-1405.

97. Hoffman GE, Hartley BJ, Flaherty E, Ladran I, Gochman P, Ruderfer D *et al.* Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with signatures from post mortem adult brains. *bioRxiv* 2017.

98. Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM *et al.* Cell type-specific gene expression differences in complex tissues. *Nat Methods* 2010; **7**(4)**:** 287-289.

99. Jaffe AE, Tao R, Norris A, Kealhofer M, Nellore A, Jia Y *et al.* A framework for RNA quality correction in differential expression analysis2016. Report no. 0000000000000.

100. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 2015; **16**(4)**:** 197-212.

101. Dobbyn A, Huckins LM, Boocock J, Sloofman LG, Glicksberg BS, Giambartolomei C *et al.* Co-localization of Conditional eQTL and GWAS Signatures in Schizophrenia. *bioRxiv* 2017.

102. Bryois J, Garrett ME, Song L, Safi A, Giusti-Rodriguez P, Johnson GD *et al.* Evaluation Of Chromatin Accessibility In Prefrontal Cortex Of Schizophrenia Cases And Controls. *bioRxiv* 2017.

103. Forrest MP, Zhang H, Moy W, McGowan H, Leites C, Dionisio LE *et al.* Open Chromatin Profiling in hiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci. *Cell Stem Cell* 2017.

104. Marigorta UM, Denson LA, Hyams JS, Mondal K, Prince J, Walters TD *et al.* Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet* 2017.

105. Germain PL, Testa G. Taming Human Genetic Variability: Transcriptomic Meta-Analysis Guides the Experimental Design and Interpretation of iPSC-Based Disease Modeling. *Stem Cell Reports* 2017; **8**(6)**:** 1784-1796.

106. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC bioinformatics* 2016; **17**(1)**:** 483.

107. Carcamo-Orive I, Hoffman GE, Cundiff P, Beckmann ND, D'Souza SL, Knowles JW *et al.* Analysis of Transcriptional Variability in a Large Human iPSC Library Reveals Genetic and Non-genetic Determinants of Heterogeneity. *Cell Stem Cell* 2016; **20**(4)**:** 518-531.e510.

108. Schwartzentruber J, Foskolou S, Kilpinen H, Rodrigues J, Alasoo K, Knights A *et al.* Molecular and functional variation in iPSC-derived sensory neurons. *Nature genetics* 2017.

109. Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R *et al.* Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* 2011; **8**(1)**:** 106-118.

110. Liu P, Kaplan A, Yuan B, Hanna JH, Lupski JR, Reiner O. Passage number is a major contributor to genomic structural variations in mouse iPSCs. *Stem Cells* 2014; **32**(10)**:** 2657-2667.

111. Lu J, Li H, Hu M, Sasaki T, Baccei A, Gilbert DM *et al.* The distribution of genomic variations in human iPSCs is related to replication-timing reorganization during reprogramming. *Cell reports* 2014; **7**(1)**:** 70-78.

112. Gore A, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* 2011; **471**(7336)**:** 63-67.

113. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 2011; **471**(7336)**:** 68-73.

114. Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD *et al.* Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* 2014; **511**(7508)**:** 177-183.

115. Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I *et al.* Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* 2012; **10**(5)**:** 620-634.

116. Mekhoubad S, Bock C, de Boer AS, Kiskinis E, Meissner A, Eggan K. Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* 2012; **10**(5)**:** 595-609.

117. Kim K, Zhao R, Doi A, Ng K, Unternaehrer J, Cahan P *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat Biotechnol* 2011; **29**(12)**:** 1117-1119.

118. Bar-Nur O, Russ HA, Efrat S, Benvenisty N. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell Stem Cell* 2011; **9**(1)**:** 17-23.

119. Lo Sardo V, Ferguson W, Erikson GA, Topol EJ, Baldwin KK, Torkamani A. Influence of donor age on induced pluripotent stem cells. *Nat Biotechnol* 2017; **35**(1)**:** 69-74.

120. Hoekstra SD, Stringer S, Heine VM, Posthuma D. Genetically-Informed Patient Selection for iPSC Studies of Complex Diseases May Aid in Reducing Cellular Heterogeneity. *Frontiers in cellular neuroscience* 2017; **11:** 164.

121. Banovich NE, Li YI, Raj A, Ward MC, Greenside P, Calderon D *et al.* Impact of regulatory variation across human iPSCs and differentiated cells. *bioRxiv* 2016.

122. Rouhani F, Kumasaka N, de Brito MC, Bradley A, Vallier L, Gaffney D. Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLoS genetics* 2014; **10**(6)**:** e1004432.

123. Barrera LA, Vedenko A, Kurland JV, Rogers JM, Gisselbrecht SS, Rossin EJ *et al.* Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science* 2016; **351**(6280)**:** 1450-1454.

124. Burrows CK, Banovich NE, Pavlovic BJ, Patterson K, Gallego Romero I, Pritchard JK *et al.* Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs. *PLoS genetics* 2016; **12**(1)**:** e1005793.

125. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y *et al.* Extensive variation in chromatin states across humans. *Science* 2013; **342**(6159)**:** 750-752.

126. Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S *et al.* Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* 2017; **546**(7658)**:** 370-375.

127. Warren CR, Jaquish CE, Cowan CA. The NextGen Genetic Association Studies Consortium: A Foray into In Vitro Population Genetics. *Cell Stem Cell* 2017; **20**(4)**:** 431-433.

128. DeBoever C, Li H, Jakubosky D, Benaglio P, Reyna J, Olson KM *et al.* Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. *Cell Stem Cell* 2017; **20**(4)**:** 533-546 e537.

129. Panopoulos AD, Smith EN, Arias AD, Shepard PJ, Hishida Y, Modesto V *et al.* Aberrant DNA Methylation in Human iPSCs Associates with MYC-Binding Motifs in a Clone-Specific Manner Independent of Genetics. *Cell Stem Cell* 2017; **20**(4)**:** 505-517 e506.

130. Warren CR, O'Sullivan JF, Friesen M, Becker CE, Zhang X, Liu P *et al.* Induced Pluripotent Stem Cell Differentiation Enables Functional Validation of GWAS Variants in Metabolic Disease. *Cell Stem Cell* 2017; **20**(4)**:** 547-557 e547.

131. Pashos EE, Park Y, Wang X, Raghavan A, Yang W, Abbey D *et al.* Large, Diverse Population Cohorts of hiPSCs and Derived Hepatocyte-like Cells

Reveal Functional Genetic Variation at Blood Lipid-Associated Loci. *Cell Stem Cell* 2017; **20**(4)**:** 558-570 e510.

132. Panopoulos AD, D'Antonio M, Benaglio P, Williams R, Hashem SI, Schuldt BM *et al.* iPSCORE: A Resource of 222 iPSC Lines Enabling Functional Characterization of Genetic Variation across a Variety of Cell Types. *Stem Cell Reports* 2017; **8**(4)**:** 1086-1100.

133. Cayo MA, Mallanna SK, Di Furio F, Jing R, Tolliver LB, Bures M *et al.* A Drug Screen using Human iPSC-Derived Hepatocyte-like Cells Reveals Cardiac Glycosides as a Potential Treatment for Hypercholesterolemia. *Cell Stem Cell* 2017; **20**(4)**:** 478-489 e475.

134. Bi X, Pashos EE, Cuchel M, Lyssenko NN, Hernandez M, Picataggi A *et al.* ATP-Binding Cassette Transporter A1 Deficiency in Human Induced Pluripotent Stem Cell-Derived Hepatocytes Abrogates HDL Biogenesis and Enhances Triglyceride Secretion. *EBioMedicine* 2017; **18:** 139-145.

135. Gu M, Shao NY, Sa S, Li D, Termglinchan V, Ameen M *et al.* Patient-Specific iPSC-Derived Endothelial Cells Uncover Pathways that Protect against Pulmonary Hypertension in BMPR2 Mutation Carriers. *Cell Stem Cell* 2017; **20**(4)**:** 490-504 e495.

136. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* 2011; **470**(7332)**:** 59-65.

137. Rees E, Kirov G, O'Donovan MC, Owen MJ. De novo mutation in schizophrenia. *Schizophr Bull* 2012; **38**(3)**:** 377-381.

138. Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM *et al.* Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* 2014; **19**(7)**:** 762-773.

139. Karayiorgou M, Simon TJ, Gogos JA. 22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. *Nat Rev Neurosci* 2010; **11**(6)**:** 402-416.

140. Schneider M, Debbane M, Bassett AS, Chow EW, Fung WL, van den Bree M *et al.* Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J Psychiatry* 2014; **171**(6)**:** 627-639.

141. Warnica W, Merico D, Costain G, Alfred SE, Wei J, Marshall CR *et al.* Copy number variable microRNAs in schizophrenia and their neurodevelopmental gene targets. *Biol Psychiatry* 2015; **77**(2)**:** 158-166.

142. Lin M, Pedrosa E, Hrabovsky A, Chen J, Puliafito BR, Gilbert SR *et al.* Integrative transcriptome network analysis of iPSC-derived neurons from schizophrenia and schizoaffective disorder patients with 22q11.2 deletion. *BMC systems biology* 2016; **10**(1)**:** 105.

143. Pak C, Danko T, Zhang Y, Aoto J, Anderson G, Maxeiner S *et al.* Human Neuropsychiatric Disease Modeling using Conditional Deletion Reveals Synaptic Transmission Defects Caused by Heterozygous Mutations in NRXN1. *Cell Stem Cell* 2015; **17**(3)**:** 316-328.

144. Yoshimizu T, Pan JQ, Mungenast AE, Madison JM, Su S, Ketterman J *et al.* Functional implications of a psychiatric risk variant within CACNA1C in induced human neurons. *Mol Psychiatry* 2015; **20**(2)**:** 162-169.

145. Ding Q, Lee YK, Schaefer EA, Peters DT, Veres A, Kim K *et al.* A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 2013; **12**(2)**:** 238-251.

146. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 2011; **29**(8)**:** 731-734.

147. Soldner F, Laganiere J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* 2011; **146**(2)**:** 318-331.

148. Montagutelli X. Effect of the genetic background on the phenotype of mouse mutations. *Journal of the American Society of Nephrology : JASN* 2000; **11 Suppl 16:** S101-105.

149. Opsahl ML, McClenaghan M, Springbett A, Reid S, Lathe R, Colman A *et al.* Multiple effects of genetic background on variegated transgene expression in mice. *Genetics* 2002; **160**(3)**:** 1107-1112.

150. Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, Pajkic D *et al.* Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. *Cell* 2015; **162**(2)**:** 391-402.

151. Chandler CH, Chari S, Dworkin I. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends Genet* 2013; **29**(6)**:** 358-366.

152. Williams RB, Chan EK, Cowley MJ, Little PF. The influence of genetic variation on gene expression. *Genome Res* 2007; **17**(12)**:** 1707-1716.

153. Gupta RM, Hadaya J, Trehan A, Zekavat SM, Roselli C, Klarin D *et al.* A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell* 2017; **170**(3)**:** 522-533 e515.

154. Jin HJ, Jung S, DebRoy AR, Davuluri RV. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. *Oncotarget* 2016; **7**(34)**:** 54616-54626.

155. Ulirsch JC, Nandakumar SK, Wang L, Giani FC, Zhang X, Rogov P *et al.* Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* 2016; **165**(6)**:** 1530-1545.

156. Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A *et al.* Modified penetrance of coding variants by cis-regulatory variation shapes human traits. *bioRxiv* 2017.

157. Stockholm D, Benchaouir R, Picot J, Rameau P, Neildez TM, Landini G *et al.* The origin of phenotypic heterogeneity in a clonal cell population in vitro. *PLoS One* 2007; **2**(4)**:** e394.

158. Wu J, Tzanakakis ES. Deconstructing stem cell population heterogeneity: single-cell analysis and modeling approaches. *Biotechnol Adv* 2013; **31**(7)**:** 1047-1062.

159. Soldner F, Stelzer Y, Shivalila CS, Abraham BJ, Latourelle JC, Barrasa MI *et al.* Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* 2016; **533**(7601)**:** 95-99.

160. Veres A, Gosis BS, Ding Q, Collins R, Ragavendran A, Brand H *et al.* Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* 2014; **15**(1)**:** 27-30.

161. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013; **31**(9)**:** 822-826.

162. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013; **31**(9)**:** 827-832.

163. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 2014; **24**(1)**:** 132-141.

164. Schaefer KA, Wu WH, Colgan DF, Tsang SH, Bassuk AG, Mahajan VB. Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nat Methods* 2017; **14**(6)**:** 547-548.

165. Liang X, Potter J, Kumar S, Zou Y, Quintanilla R, Sridharan M *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* 2015; **208:** 44-53.

166. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016; **351**(6268)**:** 84-88.

167. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016; **529**(7587)**:** 490-495.

168. Perez AR, Pritykin Y, Vidigal JA, Chhangawala S, Zamparo L, Leslie CS *et al.* GuideScan software for improved single and paired CRISPR guide RNA design. *Nat Biotechnol* 2017; **35**(4)**:** 347-349.

169. Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* 2017; **14**(6)**:** 607-614.

170. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* 2015; **33**(2)**:** 187-197.

171. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Magi R, Reschen ME *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* 2015; **47**(12)**:** 1415-1425.

172. Hormozdiari F, Kichaev G, Yang WY, Pasaniuc B, Eskin E. Identification of causal genes for complex traits. *Bioinformatics* 2015; **31**(12)**:** i206-213.

173. Chen J, Lin M, Hrabovsky A, Pedrosa E, Dean J, Jain S *et al.* ZNF804A Transcriptional Networks in Differentiating Neurons Derived from Induced Pluripotent Stem Cells of Human Origin. *PLoS One* 2015; **10**(4)**:** e0124597.

174. Nebel RA, Zhao D, Pedrosa E, Kirschen J, Lachman HM, Zheng D *et al.* Reduced CYFIP1 in Human Neural Progenitors Results in Dysregulation

of Schizophrenia and Epilepsy Gene Networks. *PLoS One* 2016; **11**(1)**:** e0148039.

175. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013; **152**(5)**:** 1173-1183.

176. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013; **154**(2)**:** 442-451.

177. Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* 2016; **44**(12)**:** 5615-5628.

178. McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R *et al.* Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biol Open* 2016; **5**(6)**:** 866-874.

179. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 2015; **33**(5)**:** 510-517.

180. Ho SM, Hartley BJ, Flaherty E, Rajarajan P, Abdelaal R, Obiorah I *et al.* Evaluating Synthetic Activation and Repression of Neuropsychiatric-Related Genes in hiPSC-Derived NPCs, Neurons, and Astrocytes. *Stem Cell Reports* 2017.

181. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* 2016; **352**(6285)**:** 600-604.

182. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK *et al.* Genomic variation. Impact of regulatory variation from RNA to protein. *Science* 2015; **347**(6222)**:** 664-667.

183. Howes OD, McDonald C, Cannon M, Arseneault L, Boydell J, Murray RM. Pathways to schizophrenia: the impact of environmental factors. *Int J Neuropsychopharmacol* 2004; **7 Suppl 1:** S7-S13.

184. Sun W, Poschmann J, Rosario RC-h, Mill J, Geschwind DH, Sun W *et al.* Histone Acetylome-wide Association Study of Resource Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* 2016; **167**(5)**:** 1385-1397.e1311.

185. Nascimento JM, Martins-de-Souza D. The proteome of schizophrenia. *NPJ Schizophr* 2015; **1:** 14003.

186. Focking M, Lopez LM, English JA, Dicker P, Wolff A, Brindley E *et al.* Proteomic and genomic evidence implicates the postsynaptic density in schizophrenia. *Mol Psychiatry* 2014.

187. Egerton A, Modinos G, Ferrera D, McGuire P. Neuroimaging studies of GABA in schizophrenia: a systematic review with meta-analysis. *Translational psychiatry* 2017; **7**(6)**:** e1147.

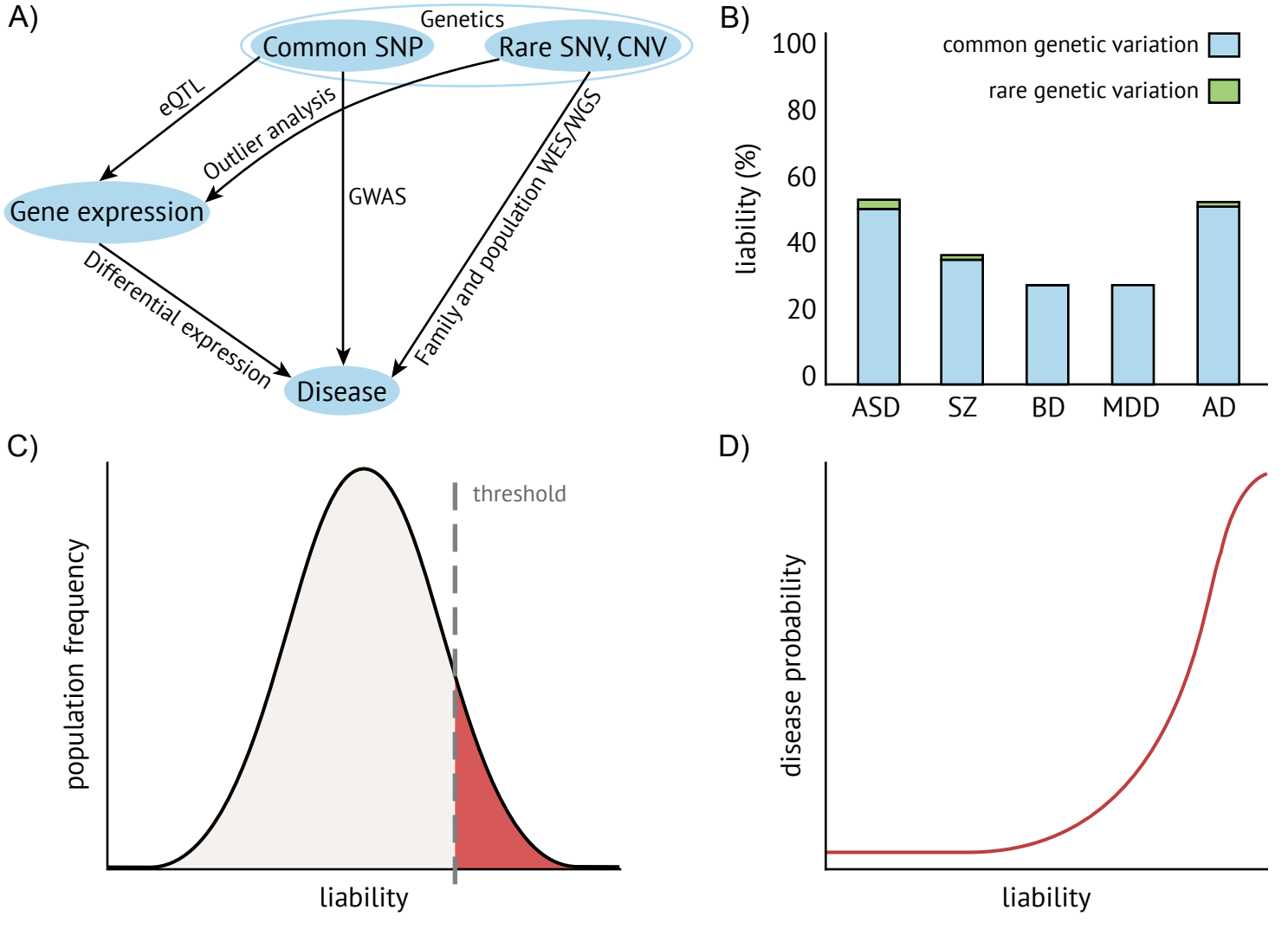
188. Howes OD, McCutcheon R, Owen MJ, Murray RM. The Role of Genes, Stress, and Dopamine in the Development of Schizophrenia. *Biol Psychiatry* 2017; **81**(1)**:** 9-20.

189. Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Schizophrenia Working Group of Psychiatric Genomics C *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* 2014; **19**(9)**:** 1017-1024.

190. Stahl EA, Forstner AJ, McQuillin A, Ripke S. Genome-wide association study identifies 30 Loci Associated with Bipolar Disorder. *biorxiv* 2017.

191. Ruderfer DM, Ripke S, Mcquillin A, Boocock J, Stahl EA, Whitehead JM. Genomic dissection of bipolar disorder and schizophrenia including 28 subphenotypes. *biorxiv* 2017.

Type of file: figure
Label: 1
Filename: figure_1.pdf

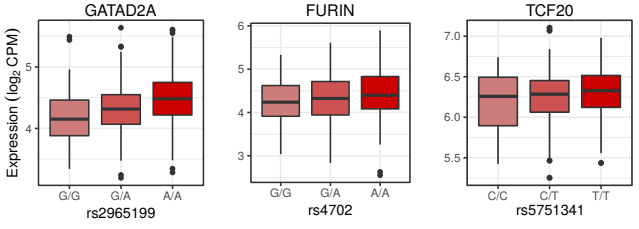

A)
Genetics
Common SNP
Rare SNV, CNV
eQTL
Outlier analysis
GWAS
Family and population WES/WGS
Gene expression
Differential expression
Disease
B)
common genetic variation
rare genetic variation
liability (%)
100
80
60
40
20
0
ASD
SZ
BD
MDD
AD
C)
threshold
population frequency
liability
D)
disease probability
liability

Type of file: figure
Label: 2
Filename: figure_2.pdf

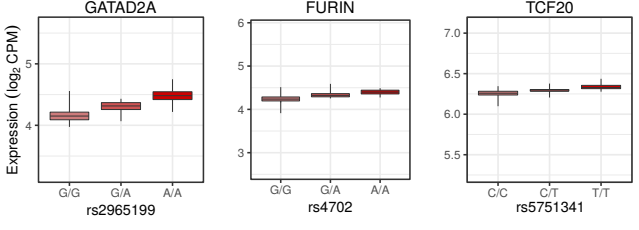
A) eQTL statistics

gene	rsid	p_{eQTL}	β	σ	% var
GATAD2A	rs2965199	5.52e−27	0.1103	0.1334	6.08
FURIN	rs4702	1.65e−12	0.0749	0.1345	1.74
TCF20	rs5751341	2.24e−02	0.0163	0.0901	1.83

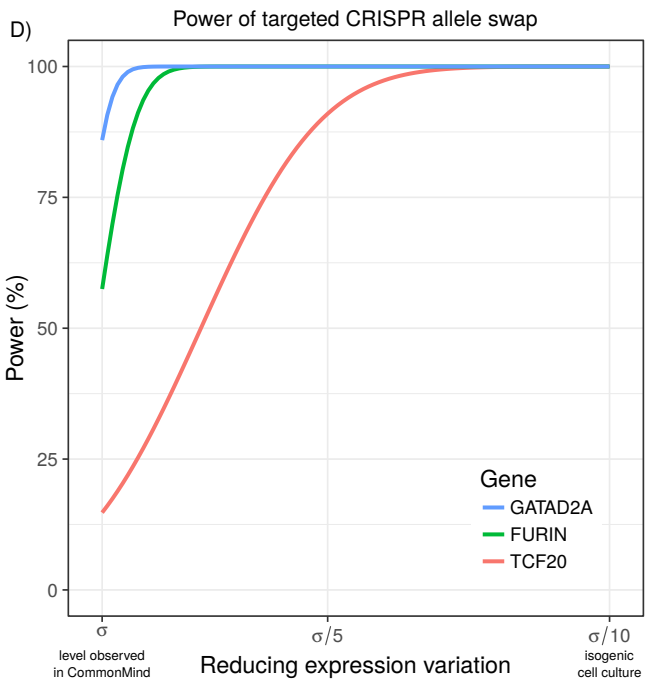
B) Observed in post-mortem brains



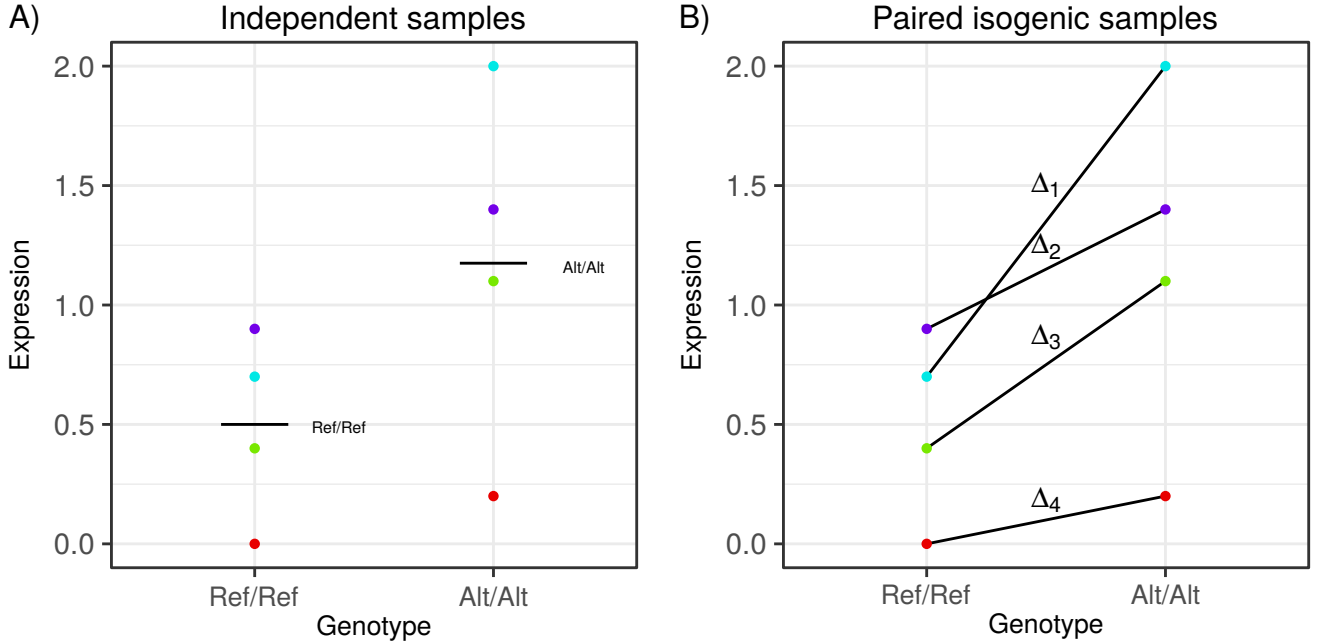
C) Hypothetical expression in isogenic hiPSC comparisons



D)



Type of file: figure
Label: 3
Filename: figure_3.pdf



Statistical test

$\bar{}_{\text{Alt/Alt}} - \bar{}_{\text{Ref/Ref}} \neq 0$

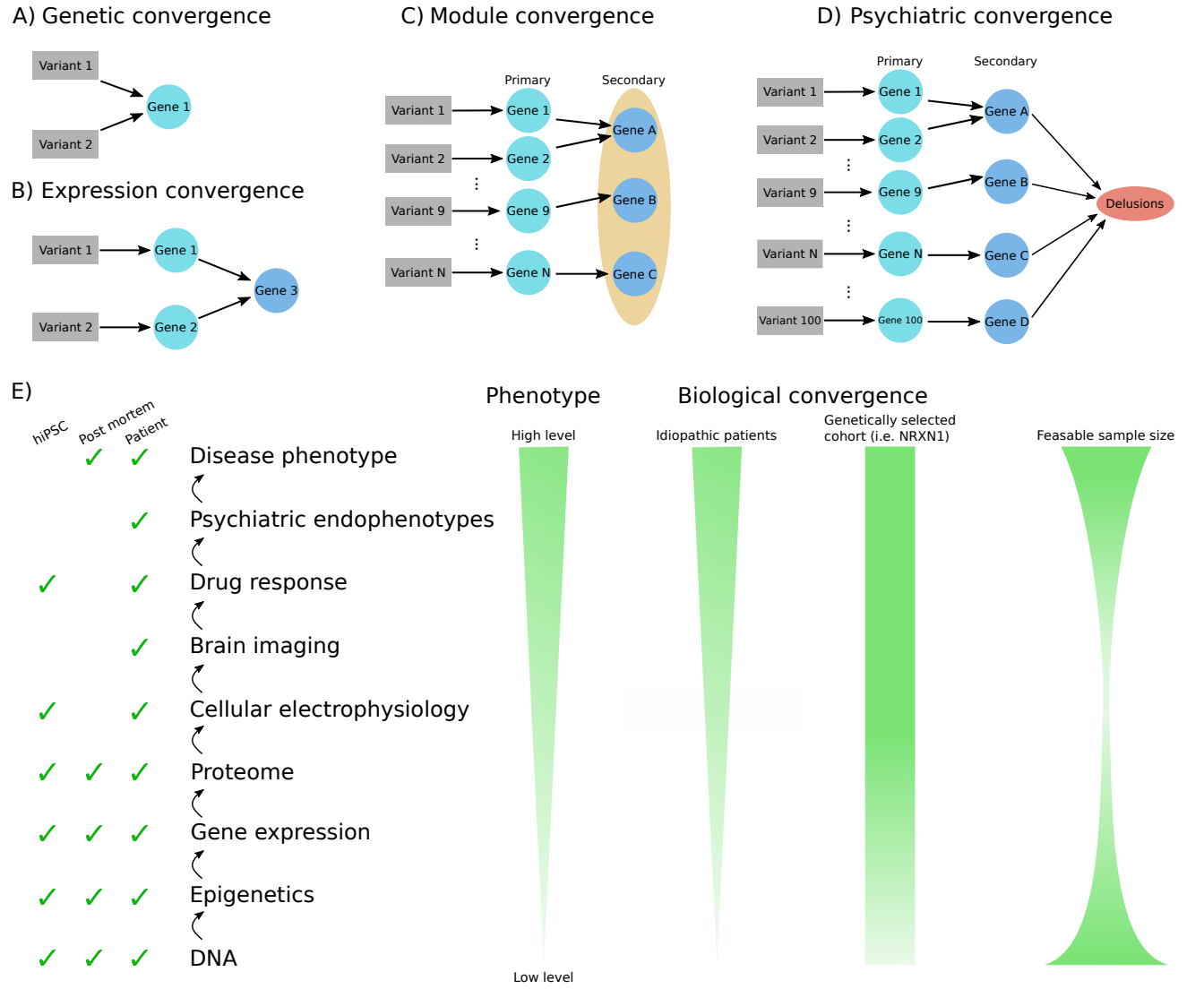
$\bar{\Delta} \neq 0$

R regression formula

```
expression ~ genotype
```

```
expression ~ genotype + donor
```

Type of file: figure
Label: 4
Filename: figure_4.pdf


A) Genetic convergence
Variant 1
Variant 2
Gene 1
B) Expression convergence
Variant 1
Variant 2
Gene 1
Gene 2
Gene 3
C) Module convergence
Primary
Secondary
Variant 1
Variant 2
Variant 9
Variant N
Gene 1
Gene 2
Gene 9
Gene N
Gene A
Gene B
Gene C
D) Psychiatric convergence
Primary
Secondary
Variant 1
Variant 2
Variant 9
Variant N
Variant 100
Gene 1
Gene 2
Gene 9
Gene N
Gene 100
Gene A
Gene B
Gene C
Gene D
Delusions
E)
hiPSC
Post mortem
Patient
Disease phenotype
Psychiatric endophenotypes
Drug response
Brain imaging
Cellular electrophysiology
Proteome
Gene expression
Epigenetics
DNA
Phenotype
High level
Low level
Biological convergence
Idiopathic patients
Genetically selected cohort (i.e. NRXN1)
Feasable sample size