# Loss-of-function of neuroplasticity-related genes confers risk for human neurodevelopmental disorders

*Milo R. Smith[1-6]*, Benjamin S. Glicksberg[1,3,4,6]*, Li Li[3,6], Rong Chen[3], Hirofumi Morishita[1,2,4,5], Joel T. Dudley[3,6]*

*Department of Neuroscience[1], Departments of Psychiatry and Opthamology[2], Department of Genetics and Genomic Sciences[3], Friedman Brain Institute[4], Mindich Child Health and Development Institute[5], Institute for Next Generation Healthcare[6]*
*Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl.*
*New York City, NY 10029, USA*
** Authors contributed equally*
*Corresponding authors: joel.dudley@mssm.edu, hirofumi.morishita@mssm.edu*

High and increasing prevalence of neurodevelopmental disorders place enormous personal and economic burdens on society. Given the growing realization that the roots of neurodevelopmental disorders often lie in early childhood, there is an urgent need to identify childhood risk factors. Neurodevelopment is marked by periods of heightened experience-dependent neuroplasticity wherein neural circuitry is optimized by the environment. If these critical periods are disrupted, development of normal brain function can be permanently altered, leading to neurodevelopmental disorders. Here, we aim to systematically identify human variants in neuroplasticity-related genes that confer risk for neurodevelopmental disorders. Historically, this knowledge has been limited by a lack of techniques to identify genes related to neurodevelopmental plasticity in a high-throughput manner and a lack of methods to systematically identify mutations in these genes that confer risk for neurodevelopmental disorders. Using an integrative genomics approach, we determined loss-of-function (LOF) variants in putative plasticity genes, identified from transcriptional profiles of brain from mice with elevated plasticity, that were associated with neurodevelopmental disorders. From five shared differentially expressed genes found in two mouse models of juvenile-like elevated plasticity (juvenile wild-type or adult *Lynx1*$^{-/-}$ relative to adult wild-type) that were also genotyped in the Mount Sinai Bio*Me* Biobank we identified multiple associations between LOF genes and increased risk for neurodevelopmental disorders across 10,510 patients linked to the Mount Sinai Electronic Medical Records (EMR), including epilepsy and schizophrenia. This work demonstrates a novel approach to identify neurodevelopmental risk genes and points toward a promising avenue to discover new drug targets to address the unmet therapeutic needs of neurodevelopmental disease.

*Keywords: integrative genomics, neuroplasticity, neurodevelopment, risk genes, inflammation, drug targets*

## 1. Introduction

Neurodevelopmental disorders place enormous personal and economic burdens on society [1,2]. In addition to environmental factors, genetic factors are known to be important predictors of neurodevelopmental outcomes, and the perinatal period comprises critical windows of disease susceptibility when mutations may express their deleterious effects on neurodevelopment. Particularly important windows of susceptibility are childhood critical periods that allow brain circuits to be refined by sensory and social experiences to establish normal perception and cognition [3–6].

Disruption of these critical periods can alter the developmental trajectory and confer risk for neurodevelopmental disorders [7,8]. Previous studies have found that genetic disruptions in neurodevelopmental disorder-related genes (MeCP2, Ube3a, Fmr1) led to disruptions in critical period plasticity [9–11]. The finding that alteration in neurodevelopmental genes disrupts developmental plasticity and leads to neurodevelopmental phenotypes calls for a comprehensive search for neurodevelopmental risk genes associated with plasticity.

To systematically identify plasticity gene variants, we used one of the best-studied models of childhood plasticity, namely critical period plasticity of visual cortex [12]. In response to deprivation of light to a single eye for a few days, neural responses in the cortex diminish correspondingly. Underlying this plasticity is major circuit remodeling [8] and only occurs naturally in youth - in adults there is minimal plasticity. Moreover, perturbations during the critical period are permanent - if the eye is occluded throughout juvenile life, there will be a persistent cortex-dependent reduction in visual acuity - a condition called amblyopia. Here, we use the mouse model of critical period visual plasticity [13] as our starting point to systematically identify genes related to neuroplasticity. This model has emerged as an indispensable model system to dissect the molecular mechanisms underlying functional cortical plasticity and whose transcriptional representation is functionally predictive [14]. Importantly, to control for age, elevated plasticity can be recapitulated in adult mice by genetically manipulating genes important for critical period plasticity. Here, we took a strategy of generating two transcriptional plasticity signatures from the visual cortex, one from juvenile and the other from adult *Lynx1*$^{-/-}$ mice, the latter upon release of the Lynx1 cholinergic plasticity brake exhibit juvenile-like plasticity [15]. These signatures represent the plasticity-permissive transcriptional landscape of visual cortex and the genes shared between these signatures are high confidence plasticity-related genes referred to here as "putative plasticity genes." Identifying putative plasticity genes in a data-driven, genome-wide manner sets the stage for high-throughput detection of potential novel risk variants associated to neurodevelopmental disease.

Past work to identify neurodevelopmental risk variants has traditionally focused on genome wide association studies (GWAS), family-based, or hereditary (e.g. twin, adoption) studies. While all are successfully used to identify risk variants, these approaches are inherently disease-centric rather than function-centric. By limiting discovery to a specific disease (e.g. microcephaly), discovery of cross-disease functional factors are missed. Our approach begins with functional plasticity-related genes and identifies any associated disease or phenotypic risk genes. This allows for greater biological insight downstream while increasing sensitivity by shrinking the search space to identify real associations between neurodevelopmental genes and disease. Moreover, by deriving putative plasticity genes using a genome-wide transcriptional approach across multiple models of elevated plasticity and coupling it with an integrative genomics methodology to identify risk genes across many diseases, we propose a highly systematic approach to identifying neurodevelopment risk genes, which does not depend on prior knowledge of either the specific functional role of the plasticity genes nor specific diseases.

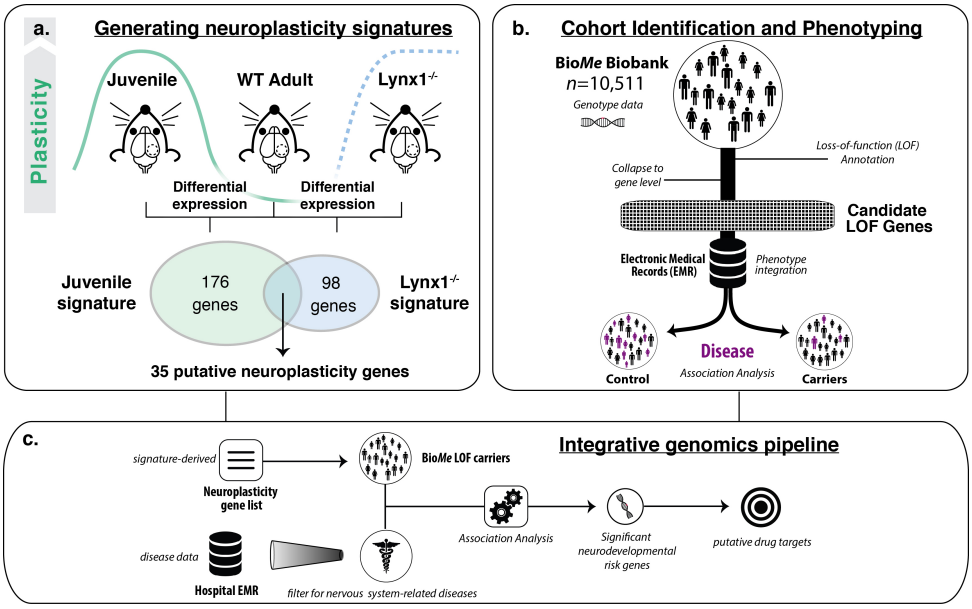


**Figure 1. An integrative genomics approach to validate a role for putative neuroplasticity genes in human neurodevelopmental disorder.** (**a**) We generated transcriptional neuroplasticity signatures from two mouse models of elevated, neurodevelopmental plasticity (juvenile and Lynx1-/-) to identify 35 shared putative plasticity genes used for downstream analysis. (**b**) We derived 2117 putative loss-of-function (LOF) variants in 1665 genes in a population of 10,510 patients from the Mount Sinai BioMe BioBank coupled with disease diagnosis data from EMR. (**c**) We applied an integrative genomics pipeline to identify associations between LOF of genotyped putative plasticity genes and nervous system diseases by logistic regression controlling demographic covariates to provide human-level evidence for multiple neurodevelopmental risk genes.

In order to assess the relationship between putative plasticity genes and neurodevelopment outcomes, we utilized a biobank of individuals with genetic data and longitudinal phenotype information from a large hospital system. To identify associations of large effect within this human dataset, we focused on the impact of loss-of-function (LOF) mutations on nervous system disease susceptibility. The study of LOF mutations in the human genome has played an important role in understanding etiologies of human disease, as these natural human knockouts shed light on gene function in the context of disease [16,17]. In a landmark study, MacArthur et al. identified LOF variants within protein coding genes using whole genome data from the 1000 Genomes Project [18,19]. They estimated that the typical human genome contains around 100 LOF variants and identified rare LOF variants that likely confer risk for disease. This work has been extended to elucidate the role and function of genes through LOF mutations in a variety of diseases: *ABCA1* with pancreatic β-cell dysfunction in Type 2 Diabetes [20]; *SETD5* with intellectual disability [21]; *APOC3* with reduced risk of both ischemic vascular disease and coronary disease [22,23]; *SLC30A8* with protection from Type 2 Diabetes [24, p. 30], among others. These findings have direct applications for identifying molecular targets to guide and accelerate drug discovery [25]. Using this strategy, Graham et al. found that antisense oligonucleotides targeting *ANGPTL3* transcripts reduced levels of atherogenic lipoproteins in humans [26]. In the current study, human findings may reveal novel drug targets relevant to neuroplasticity and neurodevelopment. Genes identified may be appropriate to directly target with small molecules. In addition, molecular editing of these targets in mouse could reveal novel molecular machinery important for disease phenotypes seen in human and lead to novel rescue therapeutics. In fact, Diamantopoulou *et al.* used

such a strategy to identify a LOF mutation in *Mirta22* that rescues schizophrenia-related phenotypes in a mouse model of 22q11.2 deletion [27].

Here, applying an integrative genomics approach we identified 35 putative plasticity genes across two mouse models, including those important in inflammatory processes. After identifying putative plasticity genes, we systematically identified LOF variants within these genes across the Charles Bronfman Institute of Personalized Medicine Mount Sinai Bio*Me* BioBank and linked Electronic Medical Record (EMR) cohort of 10,510 patients. We then assessed associations between putative plasticity genes and various neurodevelopment-related diseases using logistic regression that controls for demographic covariates (see **Figure 1** for the research design). This approach revealed potential risk variants in multiple putative plasticity genes for neurodevelopmental disorders, including epilepsy and schizophrenia. These findings provide human evidence for a role of plasticity-related genes in neurodevelopment and establish a novel approach to identify human neurodevelopment risk variants. Using model-derived putative plasticity genes as seeds to identify neurodevelopmental risk genes in human immediately sets the stage for pre-clinical studies to determine the mechanisms by which these novel risk genes disrupt neurodevelopment, and provides novel targets for therapeutic discovery.

## 2. Methods

All data processing and statistical analyses were conducted in R v 3.2.2 and Python v 2.7.10.

### 2.1 Neuroplasticity signatures

To identify putative neuroplasticity genes, we compared primary visual cortex transcriptomes of juvenile wild-type mice or adult *Lynx1*$^{-/-}$ compared to adult wild-type (n = 3 all groups). We used Limma [28] to quantile normalized raw microrray probe-level data and RankProd [29] to compute rank-based differential expression of mouse genes, which we mapped to orthologous human genes using the Mouse Genome Informatics homology reference to yield 176 and 98 gene signatures (juvenile wild-type and adult *Lynx1*$^{-/-}$ respectively), 35 of which were shared (Fisher Exact Test: OR=37.1, 95% CI = 23.8–58.0, $p < 2.2 \times 10^{-16}$, replication of comparison found in [14]) **(Figure 1a).** Both juvenile and *Lynx1*$^{-/-}$ mice have elevated experience-dependent plasticity, whereas adult wild-type mice have reduced plasticity. Transcriptional data was derived from publicly available data (GSE89757 [14]). We used the well-established gene set enrichment approach from Enrichr [30] to determine known Gene Ontology Biological Processes relevant to the 35 putative plasticity genes (using a FDR < 0.05) and further assessed relevance of individual genes that mapped to genotyped variants using a literature-based approach.

### 2.2 Hospital and biobank cohort

The Mount Sinai Hospital, located in Upper Manhattan, NY, has EMR that are de-identified and stored within the Mount Sinai Data Warehouse. These records contain clinical (e.g. disease diagnoses) and demographic data for over four million patients as of February 2015. The Charles

Bronfman Institute of Personalized Medicine Bio*Me* biobank (http://icahn.mssm.edu/research/ipm) within the Icahn School of Medicine at Mount Sinai has collected genetic data for over 30,000 patients with linked EMR as of 2016. For the current analysis, we utilized a subset of Bio*Me*, consisting of over 11,000 individuals that were genotyped using the Illumina Human Omni Express Exome Bead-8 BeadChip v1.1 array. This cohort consists of 61.2% females and 38.8% males and the self-reported racial breakdown is as follows: 46.3% Hispanic/Latino, 33.6% African American, 18.6% Caucasian, and 1.5% Other (merged from several smaller racial group categories). To account for relatedness within this cohort, we used PLINK v1.9 [31] to identify pairs of directly related individuals (PI-HAT scores > 0.25). From these pairs, we randomly selected one from each to exclude ($n$=612), resulting in 10,510 individuals used for the analyses.

## 2.3 Variant annotation

We adapted the protocol used by Glicksberg et al. to annotate genotyped variants as LOF [32]. Briefly, we ran 906,917 genotyped variants through three different public annotation tools, namely Variant Annotation Tool (VAT) [33], ANNOVAR (v. 2015Apr14) [34], and SnpEff (v. 3.6) [35]. Following established procedures [19,36], we restricted output from these annotators to "High" effect and relevant types: stop gain, frameshift/indel, and splice site. We performed further quality control by excluding variants that were in the final exon of the transcript and those with >2% alternate allele frequency. To enhance confidence of these annotations, we only included variants that passed these criteria in at least two out of three of the annotators for at least one overlapping transcript. Following these steps, we derived 2,117 putative LOF variants in 1,665 genes. For the purposes of this study, we collapsed variants to the gene level. When intersecting with the 35 neurodevelopmental genes of interest, there were five (*IL33, INMT, MAP9, LCN2, LRG1*) overlapping with at least one LOF variant (**Table 1**) used for subsequent analyses.

Table 1. Loss-of-function variants mapped to five of 35 putative neuroplasticity genes

| Gene | Chr | Position | RSID | Ref | Alt | VAT effect type | snpEff effect type | ANNOVAR effect type | BioMe aaf | ExAC aaf |
|---|---|---|---|---|---|---|---|---|---|---|
| IL33 | 9 | 6253575 | rs145735086 | G | T | . | STOP_GAINED | stopgain | 0.000143 | 0.000025 |
| INMT | 7 | 30791800 | rs190694809 | C | T | stop_gained | STOP_GAINED | stopgain | 0.000190 | 0.000033 |
| MAP9 | 4 | 156294610 | rs149881598 | T | G | splice_site | SPLICE_SITE_ACCEPTOR | . | 0.000476 | 0.000190 |
| LCN2 | 9 | 130913998 | rs139329518 | T | C | splice_site | SPLICE_SITE_DONOR | . | 0.001332 | 0.000289 |
| LRG1 | 19 | 4540004 | rs116733978 | G | A | stop_gained | . | stopgain | 0.000999 | 0.000189 |

## 2.4 Neurodevelopmental disease phenotyping

Disease diagnoses are encoded in the Mount Sinai Hospital de-identified EMR as International Classification of Diseases (ICD)-9 codes. In order to increase power for our analyses, we mapped these codes to the Clinical Classification Software (CCS; https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp) for ICD-9-CM single level categories. In total, there are 283 single level categories. As our focus is neurodevelopment and the nervous system, we restricted this list to 38 disease categories where the nervous system is considered the primary affected organ, specifically: "Meningitis (except that caused by tuberculosis or sexually transmitted disease)",

"Inflammation; infection of eye (except that caused by tuberculosis or sexually transmitted disease)", "Other CNS infection and poliomyelitis", "Otitis media and related conditions", "Cancer of brain and nervous system", "Delirium, dementia, and amnestic and other cognitive disorders", "Alcohol-related disorders", "Substance-related disorders", "Schizophrenia and other psychotic disorders", "Mood disorders", "Anxiety disorders", "Personality disorders", "Screening and history of mental health and substance abuse codes", "Developmental disorders", "Adjustment disorders", "Attention-deficit, conduct, and disruptive behavior disorders", "Impulse control disorders, NEC", "Other nervous system disorders", "Other hereditary and degenerative nervous system conditions", "Parkinson`s disease", "Headache; including migraine", "Multiple sclerosis", "Paralysis", "Epilepsy; convulsions", "Acute cerebrovascular disease", "Coma; stupor; and brain damage", "Spinal cord injury", "Other eye disorders", "Retinal detachments; defects; vascular occlusion; and retinopathy", "Glaucoma", "Cataract", "Blindness and vision defects", "Other ear and sense organ disorders", "Conditions associated with dizziness or vertigo", "Transient cerebral ischemia", "Nervous system congenital anomalies", "Poisoning by psychotropic agents", "Suicide and intentional self-inflicted injury".

### 2.5 LOF gene and disease association analysis

With the genotype and disease data processed, we assessed associations between LOF in these five putative plasticity genes and the 38 nervous system-related disease categories of interest (**Figure 1b**). Specifically, we performed a logistic regression for all gene-disease combinations for which there were at least three carriers of the gene afflicted with the disease. We also controlled for demography in the form of age, self-reported sex, and genetic ancestry using Principal Component Analysis (PCA) in the form of the first five Principal Components, which constituted the majority of variance explained (Eq. 1). The use of PCA on genetic data for determining and controlling for genetic ancestry in association studies is well established [37]. We focused on the significance of the gene term and magnitude and direction of the associated $\beta_l$ value, which represents effect size after controlling for other covariates (positive values indicate increased risk and vice versa).

$$P(\text{disease} \mid \beta_0 + \beta_l \cdot \text{gene} + \beta_g \cdot \text{sex} + \beta_a \cdot \text{age} + \beta_{pc1} \cdot \text{PC1} \ldots + \ldots \beta_{pc5} \cdot \text{PC5}) \quad (1)$$

where disease is a binary Yes/No outcome, gene is binary Yes/No indicating presence of LoF mutation, age is a continuous constant per year, sex is binary piecewise Female/Male, and PC# is continuous.

## 3. Results

### 3.1 Identifying putative neuroplasticity genes

To identify putative plasticity genes, we generated transcriptional signatures of plasticity by comparing primary visual cortex transcriptomes of juvenile wild-type or adult *Lynx1*$^{-/-}$ compared to adult wild-type mice yielding 176 and 98 differentially expressed genes (**Figure 1a**). *Lynx1*$^{-/-}$ mice have elevated, juvenile-like plasticity [15] and were used to control for non-plasticity aspects of the

juvenile signature. We defined putative plasticity genes as the 35 shared between the two signatures (Fisher Exact Test: OR=37.1, 95% CI = 23.8–58.0, $p < 2.2 \times 10^{-16}$, this statistic reproduced as in [14]). Interestingly, using gene set enrichment we found that these genes are predominantly enriched for immune processes, including gene sets related to neutrophil degranulation, defense response to fungus, immune cell chemotaxis, apoptotic pathways, and cytokine production (**Table 2**).

Table 2. Enrichment of biological pathways across 35 putative neuroplasticity genes reveals inflammatory pathways

| Term | P-value | Z-score | Comb. Score | Genes |
|---|---|---|---|---|
| neutrophil degranulation | 7.8E-04 | -4.99 | 22.6 | LRG1;LCN2;PDAP1;S100A9;S100A8 |
| defense response to fungus | 1.0E-04 | -3.45 | 20.3 | S100A9;S100A8 |
| antimicrobial humoral response | 6.7E-05 | -2.68 | 16.1 | LCN2;S100A9;S100A8 |
| negative regulation of transcription from RNA polymerase II promoter | 5.9E-03 | -5.03 | 15.6 | MTF2;ARID5B;TBL1X;CPEB3 |
| activation of cysteine-type endopeptidase activity involved in apoptotic process | 1.3E-04 | -2.46 | 14.5 | ACER2;S100A9;S100A8 |
| cytokine production | 4.9E-05 | -2.01 | 12.0 | S100A9;S100A8 |
| regulation of cytoskeleton organization | 3.9E-04 | -2.34 | 11.5 | S100A9;S100A8 |
| leukocyte migration involved in inflammatory response | 3.5E-05 | -1.87 | 11.2 | S100A9;S100A8 |
| positive regulation of intrinsic apoptotic signaling pathway | 6.3E-04 | -2.34 | 10.8 | S100A9;S100A8 |

### 3.2 *LOF variants in putative plasticity genes confer risk for neurodevelopmental and nervous system-related disorders*

Applying an integrative genomics approach (**Figure 1c**), we determined that five of 35 putative plasticity genes (*IL33, INMT, MAP9, LCN2, LRG1*) contained a LOF variant (**Table 1**) that had been genotyped in the Bio*Me* biobank and were included in subsequent analyses. Using a disease carrier minimum frequency of three, we were able to perform 27 association tests for three genes (*MAP9, LCN2, LRG1*) across 15 nervous system-related diseases. We found that two genes, *LRG1* and *LCN2,* conferred risk for five nervous system diseases (**Table 3**). Strikingly, two of these diseases, schizophrenia and epilepsy, have putative etiologies based in perinatal and childhood neurodevelopment (*LRG1* - schizophrenia: $\beta = 1.27$, $p = 0.04$; *LCN2* - epilepsy: $\beta = 1.22$, $p = 0.03$). Additionally, we identified a trending association between *MAP9* and blindness and vision defects *(*$\beta = 1.15$, $p = 0.08$).

Table 3. Putative neuroplasticity genes confer risk for neurodevelopmental and brain-related diseases

| Disease | Gene | Disease Carriers | Carriers | Disease | P-value | Beta |
|---|---|---|---|---|---|---|
| Schizophrenia and other psychotic disorders | LRG1 | 3 | 21 | 407 | 0.042 | 1.27 |
| Epilepsy; convulsions | LCN2 | 4 | 28 | 455 | 0.025 | 1.22 |
| Conditions associated with dizziness or vertigo | LCN2 | 10 | 28 | 1646 | 0.013 | 1.01 |
| Blindness and vision defects | LRG1 | 8 | 21 | 1745 | 0.042 | 0.93 |
| Anxiety disorders | LCN2 | 8 | 28 | 1672 | 0.049 | 0.83 |

## 4. Discussion

We demonstrate an innovative use of genes relevant to neuroplasticity to identify potential human neurodevelopmental risk genes. By applying an integrative genomics approach, we identified *LRG1* and *LCN2* as putative plasticity genes associated with the neurodevelopmental diseases epilepsy and schizophrenia in a human population. These genes are correlated to experience-dependent neural plasticity across two mouse models, suggesting that LOF in human may confer risk for neurodevelopmental disorders by disrupting plasticity. Moreover, these genes are regulated by inflammation via lipopolysaccharide, which also disrupts experience-dependent plasticity in juvenile mouse [14], suggesting LOF in these genes may disrupt a component of neural-immune interaction to confer risk for human neurodevelopment. Consistent with that perspective, schizophrenia and epilepsy have numerous aberrations in immune function [38,39] and neural plasticity [40–42] and this work suggests that a nexus of these aberrations may be juvenile experience-dependent plasticity, which is increasingly postulated as an important locus of neurodevelopmental risk [7,8].

*LRG1* has been previously identified as dysregulated in the choroid plexus of individual's with schizophrenia [43] and marks early granulocyte maturation [44], consistent with gene set enrichments indicating the 35 putative plasticity genes are enriched for granulocyte function (see Table 2). In contrast, antipsychotics appear to induce an immature granulocytic phenotype [45]. This has generally been considered a side-effect (and is separate from potentially fatal agranulocytosis, as induced by clozapine [46]), but neutrophils in drug-free individuals with schizophrenia generate elevated reactive oxygen species (ROS) [47,48] and ROS levels can be normalized by antipsychotics [49,50]. It should be noted, however, that one study found antipsychotics did not decrease ROS in patients with schizophrenia [51]. In animal models and postmortem brains of individuals with schizophrenia, there is evidence of elevated oxidative stress associated with the parvalbumin interneuron cell type [52]. Moreover, genetically reducing the antioxidant glutathione specifically in parvalbumin cells (which elevates ROS) leads to dysregulated critical period plasticity [53]. Therefore, we speculate that neutrophils may be a source of oxidative stress (i.e. ROS) in schizophrenia and that the suppressive effective of antipsychotics on neutrophil function may in fact be a therapeutic phenomenon. Together this suggests neutrophils and *LRG1* as previously unrecognized components of schizophrenia pathophysiology and as putative therapeutic targets that should be explored further.

*LCN2* is an important cell-autonomous marker of astrocyte activation - a phenotype that shifts astrocytes away from their resting-state role in maintaining neural circuit homeostasis to an active watchfulness against cellular damage and other forms of danger. In epilepsy, abberations in astrocytic regulation of neurotransmitters (i.e. glutamate and GABA) and ions (i.e. $K^+$) likely contribute to excitotoxicity and reduced threshold for induction of seizure [54]. Therefore, we hypothesize that LOF mutations in *LCN2* could cause astrocytes to exit their normal resting-state wherein they homeostatically support neural equilibrium, leading to chronic neurotransmitter and ionic dysregulations. Moreover, Lcn2 is an exogenous activator of microglia [55] and microglia are critically important to juvenile experience-dependent plasticity *per se* [56]. Together, this suggests mutations in *LCN2* may confer risk for epilepsy via dysregulation of multiple glial types to produce a multi-faceted disruption across neurodevelopment and suggests glia may be a promising therapeutic target at the intersection of inflammation and plasticity in epilepsy.

Given that neurodevelopmental disease is highly polygenic, it may be unsurprising that in addition to epilepsy, *Lcn2* is dysregulated in the *Disc1*-L100P mouse model of schizophrenia [57]. There is a growing but unclear role of both microglia and astrocytes in schizophrenia [58]. Functionally, activated microglia go on to secrete soluble inflammatory cytokines C1q, Tnf, and Il1-α to activate astrocytes, which then secrete Lcn2 and an unknown toxic substance that inhibits synaptic efficacy and kills neurons [59]. We speculate that a microglia-astrocyte-neural circuit may be involved in plasticity aberrations in schizophrenia and future work should explore this possibility. Consistent with this hypothesis, *Gfap* expression is elevated in the *Disc1*-L100P model, indicating a reactive astrocyte phenotype [57]. Moreover, sodium valproate normalized *Gfap* and *Lcn2* levels, as well as functional correlates of schizophrenia, indicating *Lcn2* may be a novel drug target or biomarker of successful treatment in schizophrenia. More generally, astrocytes and microglia may be an inflammatory hub in epilepsy and schizophrenia that could be targeted for therapeutic intervention.

We provide here a highly systematic and high-throughput integrative genomics approach to identify neurodevelopmental risk genes. This approach is complementary to existing approaches including GWAS, family-based, and hereditary (e.g. twin, adoption) studies. Those approaches have been extremely useful for identifying risk variants in a disease-focused manner; our integrative genomics approach extends on these by liberating from disease-centric constraints to orient the analysis on a function-based approach to identify relevant risk genes across multiple diseases. Implementing this approach here, we find that two genes implicated in neural plasticity, *LRG1* and *LCN2*, are associated with the neurodevelopmental diseases epilepsy and schizophrenia and may play a pathophysiological role at the nexus of immune-brain function. As such, we believe these genes may be biomarkers for such neurodevelopmental-related diseases and candidates for drug targets. On the other hand, there are a few caveats and limitations to our integrative genomics approach. We used two models of plasticity (juvenile and *Lynx1*$^{-/-}$), but transcriptional changes in other models could further contribute to the identification of neurodevelopmental risk genes in human. Limiting to the models used here could exclude genes relevant to neuroplasticity (i.e. false negatives). Additionally, though we used a strict FDR threshold to identify putative plasticity genes, the possibility of including genes that are not directly relevant to plasticity (i.e. false positives) is possible given the variable nature of gene expression profiling. In addition, the specific molecular function of these genes in plasticity is not yet established, making interpretation of their role in developmental neuroplasticity *per se* more challenging. Moreover since these genes were identified using differential expression analysis, making interpretations of LOF in a given gene challenging and robust experimental work should follow. Given the input set of plasticity genes used in this study, we were limited by the number of genes that were genotyped and the number of LOF variants in these genes. Separately, there are known issues surrounding the accuracy of defining a disease by ICD codes. While robust, multimodal electronic phenotyping algorithms exist for many diseases (e.g. PheKB; https://phekb.org/), we utilized ICD-based definitions for diseases (via CCS) because there are not many algorithms that exist for our disease domain of interest (nervous-system and neurodevelopment). Finally, this study used only a single cohort (Mount Sinai Bio*Me*) and given the relatively low sample sizes for the diseases for which we identified LOF variants in putative plasticity genes (see Table 3) we considered a nominal *p* value threshold of 0.05 as appropriate for discovery. Follow up studies in larger, independent cohorts using a multiple test correction approach, as well as

functional experiments to elucidate the specific neurobiological relevance, is critical to validate these findings. We further discuss potential approaches to address these issues in follow-up studies within the next section.

## 5. Conclusions and Future Directions

This study provides significant impact to the field by identifying unrecognized neurodevelopment risk genes for schizophrenia and epilepsy through a novel systematic approach leveraging Mount Sinai's BioMe BioBank and linked Mount Sinai Hospital's Electronic Medical Record (EMR) data. This integrative genomics approach facilitates high-throughput identification of LOF risk variants that may have a deleterious impact on neurodevelopment and the findings set the groundwork for functional studies to determine the mechanisms by which these novel risk genes disrupt neurodevelopment and to investigate their utility for therapeutic discovery. Using putative plasticity genes as the seed genes to identify neurodevelopmental risk genes immediately sets the stage to rigorously test the hypothesis that these genes play a role in childhood neurodevelopmental. Using the ocular dominance animal model of developmental neuroplasticity [13] from which the plasticity genes were derived allows investigators to rapidly return to the mouse to test the effect of gene perturbation in neurodevelopment and neuroplasticity.

There are several future directions we will pursue to extend and further assess the implications of our findings. The relatively low sample sizes of nervous-system related diseases in our cohort (for example, see Table 3) coupled with the rare nature of these LOF mutations, limits power to detect associations. As such, we plan to perform a cross-validation experiment using genotype and clinical data for the 500,000 individuals in the UKBioBank (http://www.ukbiobank.ac.uk/). Additionally, in the hopes of exploring associations for the entire original set of 35 putative neuroplasticity genes, we will leverage the UK10K (http://www.uk10k.org) whole exome sequencing data to identify putative LOF variants for these genes within the neurodevelopmental cohort ($N$=3,000). In addition, we will increase our collection of genes related to neuroplasticity using other models, such as calorie restriction-induced plasticity [60], exercise-induced plasticity [61], drug-induced plasticity [62], as well as other plasticity-enhancing gene perturbation models, depending on available transcriptional data. Relatedly, we aim to extend these analyses to confirmed plasticity genes whose molecular mechanisms in plasticity are well-established, to yield a hypothesis-driven iteration of our approach. While it is important to increase the number of starting plasticity genes and use larger quantities of human data, it would be additionally valuable to reassess the associations made here using PheKB algorithms for nervous-system related diseases to address the limitations of ICD-code based phenotyping. Finally, we expect this integrative genomics approach will be generalizable to identify risk genes and facilitate focused biological inquiry in other disease contexts to enable drug target and biomarker identification.

## 6. Acknowledgments

We would like to sincerely thank the Mount Sinai Data Warehouse for facilitating data accessibility and the Mount Sinai Scientific Computing team for infrastructural support. We would

also like to thank the Charles Bronfman Institute of Personalized Medicine for collecting and housing the biobank genetic data as well as Eli Stahl and Douglas Ruderfer for performing quality control on the genotyping array.

This work was funded by a Traineeship, National Institute of Child Health and Human Development - Interdisciplinary Training in Systems and Developmental Biology and Birth Defects Grant T32H-D0-75735 (to M.R.S.); the Mindich Child Health and Development Institute Pilot Fund (to J.T.D. and H.M.); the Knights Templar Eye Foundation (to H.M.); the March of Dimes (to H.M.); the Whitehall Foundation (to H.M.); the Harris Center for Precision Wellness (to J.T.D); and National Institutes of Health Grants P30-ES-023515 (to J.T.D. and H.M.), R01-DK-098242 (to J.T.D.), and R01-EY- 024918, R01-EY-026053, and R21 MH106919 (to H.M.).

## References

1. C. A. Boyle, S. Boulet, L. A. Schieve, R. A. Cohen, et al., *Pediatrics*, peds.2010-2989 (2011).
2. P. Grandjean and P. J. Landrigan, *Lancet Neurol.* **13**, 330–338 (2014).
3. J. S. Johnson and E. L. Newport, *Cognit. Psychol.* **21**, 60–99 (1989).
4. T. P. Nikolopoulos, G. M. O'Donoghue, and S. Archbold, *The Laryngoscope* **109**, 595–599 (1999).
5. T. L. Lewis and D. Maurer, *Dev. Psychobiol.* **46**, 163–183 (2005).
6. C. A. Nelson, C. H. Zeanah, N. A. Fox, P. J. Marshall, et al., *Science* **318**, 1937–1940 (2007).
7. J. J. LeBlanc and M. Fagiolini, *Neural Plast.* **2011** (2011).
8. A. E. Takesian and T. K. Hensch, in *Prog. Brain Res.* **207**, M. N. and T. M. V. V. Michael M. Merzenich, Ed. (Elsevier, 2013).
9. D. Tropea, E. Giacometti, N. R. Wilson, C. Beard, et al., *Proc. Natl. Acad. Sci.* **106**, 2029–2034 (2009).
10. K. Yashiro, T. T. Riday, K. H. Condon, A. C. Roberts, et al., *Nat. Neurosci.* **12**, 777–783 (2009).
11. E. G. Harlow, S. M. Till, T. A. Russell, L. S. Wijetunge, et al., *Neuron* **65**, 385–398 (2010).
12. T. N. Wiesel and D. H. Hubel, *J. Neurophysiol.* **26**, 1003–1017 (1963).
13. J. A. Gordon and M. P. Stryker, *J. Neurosci.* **16**, 3274–3286 (1996).
14. M. R. Smith, P. Burman, M. Sadahiro, B. A. Kidd, et al., *eNeuro* **3**, ENEURO.0240-16.2016 (2016).
15. H. Morishita, J. M. Miwa, N. Heintz, and T. K. Hensch, *Science* **330**, 1238–1240 (2010).
16. F. S. Alkuraya, *Genome Med.* **7** (2015).
17. F. E. Dewey, M. F. Murray, J. D. Overton, L. Habegger, et al., *Science* **354** (2016).
18. 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, et al., *Nature* **467**, 1061–1073 (2010).
19. D. G. MacArthur, S. Balasubramanian, A. Frankish, N. Huang, et al., *Science* **335**, 823–828 (2012).
20. M. Vergeer, L. R. Brunham, J. Koetsveld, J. K. Kruit, et al., *Diabetes Care* **33**, 869–874 (2010).
21. D. Grozeva, K. Carss, O. Spasic-Boskovic, M. J. Parker, et al., *Am. J. Hum. Genet.* **94**, 618–624 (2014).
22. A. B. Jørgensen, R. Frikke-Schmidt, B. G. Nordestgaard, and A. Tybjærg-Hansen, *N. Engl. J. Med.* **371**, 32–41 (2014).
23. *N. Engl. J. Med.* **371**, 22–31 (2014).
24. J. Flannick, G. Thorleifsson, N. L. Beer, S. B. R. Jacobs, et al., *Nat. Genet.* **46**, 357–363 (2014).
25. N. O. Stitziel and S. Kathiresan, *Trends Cardiovasc. Med.* **27**, 352–359 (2017).
26. M. J. Graham, R. G. Lee, T. A. Brandt, L.-J. Tai, et al., *N. Engl. J. Med.* **377**, 222–232 (2017).
27. A. Diamantopoulou, Z. Sun, J. Mukai, B. Xu, et al., *Proc. Natl. Acad. Sci.* **114**, E6127–E6136 (2017).
28. G. K. Smyth, in *Bioinforma. Comput. Biol. Solut. Using R Bioconductor* (Springer, New York, 2005).
29. F. Hong, R. Breitling, C. W. McEntee, B. S. Wittner, et al., *Bioinformatics* **22**, 2825–2827 (2006).

30. E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, et al., *BMC Bioinformatics* **14**, 128 (2013).
31. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, et al., *GigaScience* **4** (2015).
32. B. S. Glicksberg, L. Amadori, N. K. Akers, K. Sukhavasi, et al., *Rev.*
33. G. T. Wang, B. Peng, and S. M. Leal, *Am. J. Hum. Genet.* **94**, 770–783 (2014).
34. K. Wang, M. Li, and H. Hakonarson, *Nucleic Acids Res.* **38**, e164 (2010).
35. P. Cingolani, A. Platts, L. L. Wang, M. Coon, et al., *Fly (Austin)* **6**, 80–92 (2012).
36. A. H. Li, A. C. Morrison, C. Kovar, L. A. Cupples, et al., *Nat. Genet.* **47**, 640–642 (2015).
37. A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, et al., *Nat. Genet.* **38**, 904–909 (2006).
38. J. C. Leza, B. García-Bueno, M. Bioque, C. Arango, et al., *Neurosci. Biobehav. Rev.* **55**, 612–626 (2015).
39. A. Vezzani, J. French, T. Bartfai, and T. Z. Baram, *Nat. Rev. Neurol.* **7**, 31–40 (2011).
40. D. Ben-Shachar and D. Laifenfeld, *Int. Rev. Neurobiol.* **59**, 273–296 (2004).
41. H. E. Scharfman, *The Neuroscientist* **8**, 154–173 (2002).
42. J. W. Swann and J. J. Hablitz, *Ment. Retard. Dev. Disabil. Res. Rev.* **6**, 258–267 (2000).
43. S. Kim, Y. Hwang, D. Lee, and M. J. Webster, *Transl. Psychiatry* **6**, e964 (2016).
44. L. C. O'Donnell, L. J. Druhan, and B. R. Avalos, *J. Leukoc. Biol.* **72**, 478–485 (2002).
45. J. M. Delieu, M. Badawoud, M. A. Williams, R. W. Horobin, et al., *J. Psychopharmacol. (Oxf.)* **15**, 191–194 (2001).
46. J. M. J. Alvir, J. A. Lieberman, A. Z. Safferman, J. L. Schwimmer, et al., *N. Engl. J. Med.* **329**, 162–167 (1993).
47. P. Sirota, R. Gavrieli, and B. Wolach, *Psychiatry Res.* **121**, 123–132 (2003).
48. Y. Melamed, P. Sirota, D. R. Dicker, and P. Fishman, *Psychiatry Res.* **77**, 29–34 (1998).
49. F. Péters, T. Franck, M. Pequito, G. De La REBIÈRE, et al., *J. Vet. Pharmacol. Ther.* **32**, 541–547 (2009).
50. F. Vargas, V. Chávez, and K. Pérez, *Rev. Colomb. Cienc. Quím. - Farm.* **38**, 5–18 (2009).
51. M. Cosentino, A. Fietta, E. Caldiroli, F. Marino, et al., *Prog. Neuropsychopharmacol. Biol. Psychiatry* **20**, 1117–1129 (1996).
52. J.-H. Cabungcal, P. Steullet, H. Morishita, R. Kraftsik, et al., *Proc. Natl. Acad. Sci.* **110**, 9130–9135 (2013).
53. H. Morishita, J.-H. Cabungcal, Y. Chen, K. Q. Do, et al., *Biol. Psychiatry* **78**, 396–402 (2015).
54. N. C. de Lanerolle, T.-S. Lee, and D. D. Spencer, *Neurotherapeutics* **7**, 424–438 (2010).
55. E. Jang, S. Lee, J.-H. Kim, J.-H. Kim, et al., *FASEB J.* **27**, 1176–1190 (2013).
56. G. O. Sipe, undefined R. L. Lowery, M.-È. Tremblay, E. A. Kelly, et al., *Nat. Commun.* **7**, 10905 (2016).
57. T. V. Lipina, F. N. Haque, A. McGirr, P. C. Boutros, et al., *PLOS ONE* **7**, e51562 (2012).
58. H.-G. Bernstein, J. Steiner, and B. Bogerts, *Expert Rev. Neurother.* **9**, 1059–1071 (2009).
59. S. A. Liddelow, K. A. Guttenplan, L. E. Clarke, F. C. Bennett, et al., *Nature* **541**, 481–487 (2017).
60. M. Spolidoro, L. Baroncelli, E. Putignano, J. F. Maya-Vetencourt, et al., *Nat. Commun.* **2**, 320 (2011).
61. E. Kalogeraki, F. Greifzu, F. Haack, and S. Löwel, *J. Neurosci.* **34**, 15476–15481 (2014).
62. D. Silingardi, M. Scali, G. Belluomini, and T. Pizzorusso, *Eur. J. Neurosci.* **31**, 2185–2192 (2010).