Network Effects of the 15q13.3 Microdeletion on the Transcriptome and Epigenome in Human Induced Neurons

Supplement 1

Supplementary Methods

Cell line characterization

Fibroblasts derived from skin biopsies were obtained from 3 patients with heterozygous 15q13.3 microdeletions and 3 controls (Figure 1A, Supplementary Table S1). Samples were obtained after informed consent was given by all subjects and were subsequently de-identified. All procedures were approved by the Stanford University Institutional Review Board.

Induced pluripotent stem cells (iPSCs) were generated from fibroblasts using Sendai viral vectors Oct4, Sox2, Klf4, and c-Myc from the CytoTune-iPS 2.0 Sendai Reprogramming Kit (Life Technologies, Carlsbad, CA, USA). iPSCs were maintained on irradiated mouse embryonic fibroblasts. Individual iPSC colonies were manually selected between 3-4 weeks following Sendai virus transduction for further expansion as clonal lines. The expanded iPSCs were grown on Matrigel (Corning, Corning, NY, USA) coated plates in mTeSR1 media (Stem Cell Technologies, Vancouver, Canada) and passaged using 0.02% EDTA (Lonza, Basel, Switzerland). All cell lines were confirmed to be free of mycoplasma contamination.

For each patient and control subject, 2 clonal iPSC lines were selected for neural differentiation using the neurogenin-2 induction method (1). On day 0, iPSCs were dissociated into single cells using Accutase (Thermo Fisher Scientific, Waltham, MA, USA). On day 1, the lentiviral vectors FUW-rtTA, which expresses rtTA, and pTet-O-Ngn2-puro, which expresses Ngn2 and the puromycin resistant gene under control of the TetON promoter, were added to mTeSR1 medium containing 10 μM ROCK inhibitor (Tocris Bioscience, Bristol, United Kingdom). On day 2, the culture medium was replaced

with DMEM/F12 (Life Technologies) mixed with N2 supplement (Life Technologies), penicillin-streptomycin (Life Technologies), and doxycycline at 2 μg/ml (Sigma Aldrich, St. Louis, MO, USA) to induce TetO expression. On days 3-5, puromycin at 1-2 μg/ml (Sigma Aldrich) was added to the media to select for cells expressing Ngn2. On day 6, early stage induced neurons (iNs) were collected for downstream analysis (Figure 1B). These immature iNs are not yet electrophysiologically active but can mature into excitatory projection neurons if co-cultured with glial cells. iPSCs and iNs used for RNA-Seq, Methyl-Seq, and ATAC-Seq studies were grown independently for each assay.

To confirm successful reprogramming, iPSCs were stained with the pluripotency markers Nanog, Tra-1-60, and SSEA-4. Day 6 iNs were stained for the neuronal identity markers TUJ1, VGLUT1, and MAP2. For both iPSCs and iNs, cells were fixed with 4% paraformaldehyde for 15 minutes, permeabilized with 0.1% Triton X-100 for 10 minutes, blocked with 5% goat serum for 1 hour and labeled with primary antibody overnight at 4°C. The samples were incubated for 1 hour with the secondary antibody, followed by DAPI staining for 10 minutes. For a list of antibodies used, see Supplementary Table S17. All primary antibodies were used at 1:100 dilution and all secondary antibodies were used at 1:1000 dilution. Cells were imaged using the Leica AF 6000 microscope at 20x magnification using the same acquisition settings for all samples and analyzed using ImageJ.

CRISPR-Cas9 knockout of 15q13.3 microdeletion genes

Gene knockouts targeting CHRNA7, KLF13, FAN1, OTUD7A, and MTMR10 were generated in 4 control iPSC lines (Supplementary Table S2). Two gRNAs were designed for each gene knockout. Control lines were generated using a safe-targeting gRNA and a non-targeting gRNA (Supplementary Table S3). For each gene knockout, a ribonucleoprotein (RNP) complex was prepared by mixing 40 pmol Cas9 and 300 pmol sgRNA (Synthego, Menlo Park, CA, USA), followed by a 10 minute room temperature incubation. iPSCs were dissociated into single cells with Accutase (Thermo Fisher

Scientific) and 5×10^5 iPSCs were resuspended in 20 µl P3 nucleofection buffer and mixed with 5 µl of previously prepared RNP complex. The nucleofection was carried out with the Lonza P3 primary cell 4D nucleofection kit (Lonza) using the "CA137" program. After nucleofection, cells were placed on Matrigel (Corning) coated plates in mTeSR Plus media (Stem Cell Technologies) containing 10 µM ROCK inhibitor (Tocris Bioscience). To generate single clones with homozygous gene knockouts, iPSCs were replated as single cells on 96-well plates using limiting dilution. iPSCs were grown in CloneR supplement and mTeSR Plus (Stem Cell Technologies) for the first 3 days, then grown in mTeSR Plus alone. Genomic DNA was extracted from the clones using the Quick-DNA Universal Kit (Zymo Research) and successful clones were confirmed with Sanger sequencing.

Whole-genome DNA sequencing

Genomic DNA was isolated from fibroblasts and iPSCs using the Quick-DNA Universal Kit (Zymo Research, Irvine, CA, USA). DNA sequencing libraries were generated using the TruSeq DNA Nano Library kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina HiSeq X instrument using 2×150 bp paired-end runs. Reads were mapped to GRCh37 using BWA (2). The location of the 15q13.3 microdeletion was confirmed using read depth coverage with BEDtools (3) and CNVnator (4).

To detect additional CNVs in iPSCs compared to parental fibroblast lines, CNVnator (4) calls were filtered as follows: (1) pval1 < 0.05; (2) q0 < 0.5; (3) <= 70% of CNV overlap with repetitive and low-complexity regions as previously defined (5); (4) CNV size larger than 0.1 Mb; (5) CNV found in iPSC line and not in parental fibroblast line; (6) CNV not found in another clonal line from same parental fibroblast line.

RNA sequencing (RNA-Seq)

Total RNA was extracted from samples using the Direct-zol RNA MiniPrep Kit (Zymo Research, Irvine, CA, USA) and mRNA was isolated using the Dynabeads mRNA Purification Kit

(Thermo Fisher Scientific). Libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) and sequenced on the Illumina NextSeq 500 and Hiseq X using 2x150 bp paired-end runs with 1% PhiX spike-in. The sequencing reads were mapped to GRCh38 using Tophat (6). Principal component analysis (PCA) and differential gene expression analysis were performed using DESeq2 (7), with significant genes defined as having padj <= 0.05, using the Benjamini-Hochberg (BH) adjustment method. WebGestalt (8) was used to perform gene set enrichment analysis (GSEA), with ranking based on signed fold change * -log10pvalue. Ingenuity Pathway Analysis (Qiagen, Hilden, Germany) was used for over-representation analysis.

The RNA-Seq data was also used to analyze 1) sample identity, 2) Sendai virus expression, 3) cell type composition, and 4) contributions of expression variance.

1. Sample identity was confirmed in iPSCs and iNs by performing variant calling on mapped RNA-Seq reads using samtools mpileup. The number of overlapping heterozygous single nucleotide variants (SNVs), defined as reference allele percentage > 0.3 and < 0.7, were compared between all pairwise samples.

2. Sendai viral clearance was assessed in iPSCs by using BWA (2) to align reads that did not map to the human genome (GRCh38) to the Sendai virus genome sequence (GenBank: AB855655.1). The number of reads mapping to the Sendai virus along with the total number of RNA-Seq reads was calculated for each sample.

3. To examine the cell type composition of iPSC and iN samples, cell type composition scores were generated using Cibersort (9). The reference panel was based on single-cell RNA-Seq data from the following cell types: iPSCs, neural precursor cells, fetal replicating neurons, fetal quiescent neurons, oligodendrocyte progenitor cells, adult neurons, astrocytes, oligodendrocytes, microglia, and endothelial cells (10–12).

4. The contribution of different sources of variation (cell type, donor, genotype) to gene expression was calculated using variancePartition (13).

Genome-wide targeted-capture DNA methylation sequencing (Methyl-Seq)

Genomic DNA was extracted from samples using the Quick-DNA Universal Kit (Zymo Research). Bisulfite converted libraries were prepared using the SeqCap Epi CpGiant System (Roche, Basel, Switzerland), which targets 80.5 Mb containing over 5.5 million CpGs. Libraries were sequenced on the Illumina HiSeq 4000 using 2x150 bp paired-end runs and 30% PhiX spike-in. Sequencing reads were trimmed using Cutadapt and mapped to GRCh38 using Bowtie 2 (14). DNA methylation ratios for CpGs, defined as methylated reads divided by total reads, were called by Bismark (15). PCA was performed using Methylkit (16). Differentially methylated regions (DMRs) were identified using Metilene (17) using a minimum cutoff of 0.1 mean methylation difference between deletion and control samples and padj <= 0.05, with Bonferroni correction. For association with nearby genes, DMRs were assigned first to promoter regions (defined as the 2 kb region upstream of the transcriptional start site), then to gene bodies, and finally to the nearest gene (intergenic). For genes with multiple significant DMRs, the DMR with the most significant p-value was chosen for downstream gene-based analyses. GSEA was performed on DMR-associated genes using WebGestalt, with ranking based on signed mean methylation difference * -log10pvalue (8).

Assay for transposase-accessible chromatin sequencing (ATAC-Seq)

For each sample, 50,000 cells were lysed, transposed, and PCR amplified as previously described (18), using the Illumina Nextera DNA kit. PCR products were purified using the Qiagen MinElute Kit (Qiagen), followed by a 1.8x AMPure XP bead selection (Beckman Coulter, Brea, CA, USA) to remove primer dimers. Libraries were sequenced on the Illumina NextSeq 500 using 2x75 bp paired-end runs and 1% PhiX spike-in. Reads were mapped to GRCh38 with Bowtie 2 (14) and peaks

were called using MACS2 (19). Peak reads from each sample were normalized to the total number of aligned reads and PCA was performed using Diffbind (20). For peak annotation to nearby genes, differential peak analysis by Diffbind was rerun with MACS2 peaks in the 15q13.3 microdeletion region removed. To generate a list of differentially accessible genes, significant peaks (pval <= 0.05, Benjamini-Hochberg (BH) corrected) were assigned first to promoters (2 kb region upstream of the transcriptional start site), then to gene bodies, and finally to the nearest gene (intergenic). For genes with multiple significant peaks, the peak with the most significant p-value was chosen for downstream gene-based analyses.

Homer motif enrichment analysis (21) was performed using a list of differentially open promoters and a custom motif file. The openness of each promoter, defined as the 2 kb region upstream of the transcriptional start site, was calculated from mapped reads using the surrounding 1 Mbp as background as previously described (22). Reads in the 15q13.3 microdeletion region were doubled prior to calculating openness. A list of promoters with differential openness was generated using the following filters: 1) average openness in control samples > 2 and average openness in deletion samples < 2, or vice versa, and 2) absolute value difference in average openness in control samples versus average openness in deletion samples >= 0.5. A custom list of TFs with known motifs was collected from existing databases and previous publications including Homer, Jaspar, ENCODE, and Taipale. Motif results were filtered for best match score >= 0.6 and p-value <= 10^{-12}.

Supplementary Results

Cell line characterization

To study the molecular effects of the 15q13.3 microdeletion, we obtained fibroblasts from three patients with the 15q13.3 microdeletion and three sex-matched controls (Supplementary Table S1). Heterozygous 15q13.3 microdeletions in all three patient fibroblasts were located between breakpoints 3 and 4 (Figure 1C), which is the most common combination of breakpoints for deletion CNVs in the 15q13.1-q13.3 region (23).

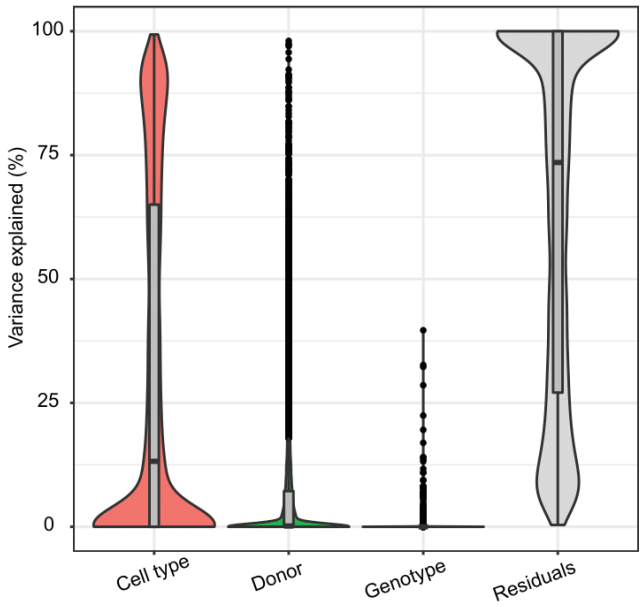
iPSCs generated from fibroblasts using Sendai viral reprogramming stained positive for the pluripotency markers SSEA4, TRA-1-60, and Nanog (Figure 1D). Whole genome sequencing of iPSCs identified no novel CNVs above the detection limit of traditional karyotyping (>5 Mb). Between 0-4 smaller additional CNVs were detected in each iPSC line relative to its parental fibroblast line (Supplementary Table S4), in line with expectations (24,25). iPSCs were checked for Sendai viral clearance, with one line showing residual expression above one read per million mapped reads, but no apparent expression changes of the reprogramming transcription factors (TFs) MYC, SOX2, KLF4, and OCT4 (Supplementary Table S1).

Induced neurons (iNs) generated from iPSCs exhibited neurite morphology and stained positive for the neuronal markers TUJ1, VGLUT1, and MAP2 (Figure 1E). Gene set enrichment analysis (GSEA) showed that genes involved in neuronal processes were up-regulated in iNs compared with iPSCs (Figure 1F). All cell lines passed sample identity verification and were assessed for cell type composition. The iPSCs had the strongest cell type composition score for the iPSC reference category and the day 6 iNs had the strongest cell type composition score for the fetal quiescent neuron category (Supplementary Table S5). Cell lines were also assessed for contributors of expression variance (Supplementary Figure S1). Cell type had the highest genome-wide contribution (13.2% median

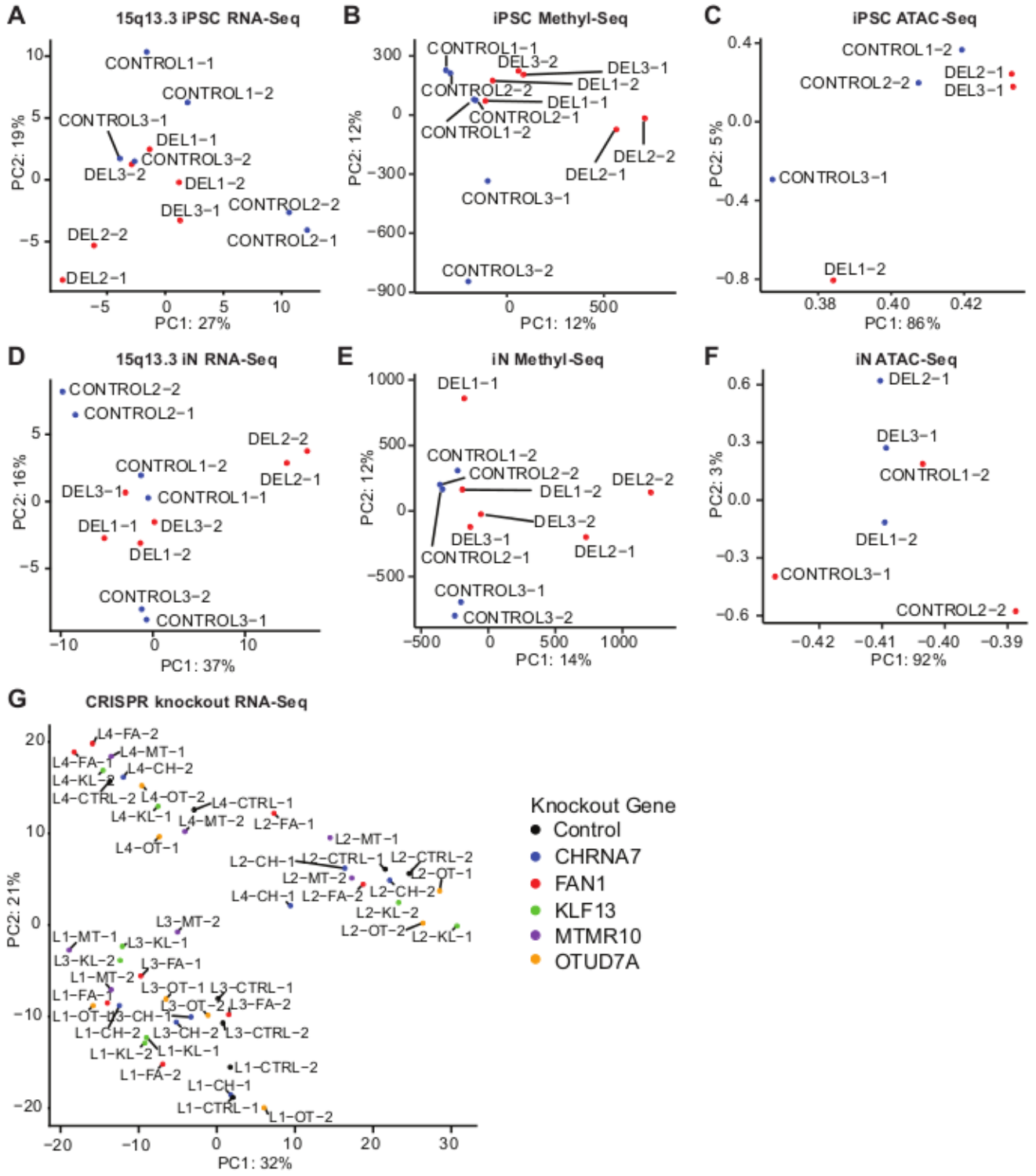
contribution), followed by donor (<1%) and genotype (<1%), which was in line with previous data for iPSC-derived neural precursor cells and neurons derived from schizophrenia patients (26).

Comparison with other neuropsychiatric associated CNVs

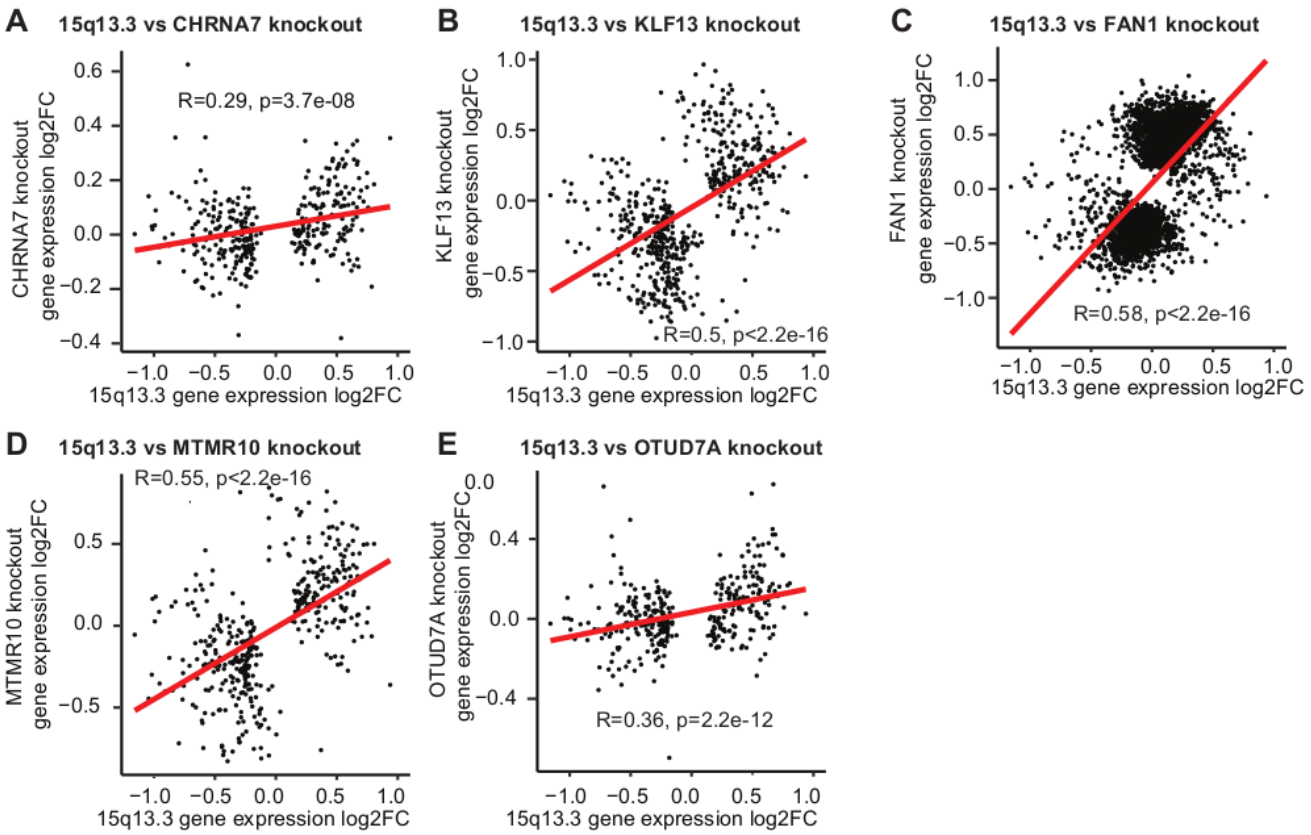
We compared our results for the 15q13.3 CNV with previous findings in other large neuropsychiatric-associated CNVs. Currently the available functional genomics data in neuronal cell types is relatively limited. We reanalyzed published RNA-seq data from 22q11.2 neurons and did not observe an overlap in gene ontology terms from 22q11.2 neurons, which primarily showed an enrichment of leukocyte and immune-related terms (27). We also compared the findings from our 15q13.3 iNs with those in 16p11.2 iPSC-derived iNs generated using similar approaches and observed a shared enrichment in a gene ontology term related to ribosome biogenesis (GO:0003735 structural constituent of ribosome). In addition, differential DNA methylation of the protocadherin gene family was observed in the 16p11.2 iPSC-derived iNs (28). Similar to our findings in the 15q13.3 CNV, both the 22q11.2 and 16p11.2 analyses also reported that the expression of genes within the CNV boundaries typically followed the direction of the copy-number change. Finally, the disruption to Wnt signaling, which we observed in multiple omics layers in the 15q13.3 CNV, was also documented in the gene expression analysis of 7q11.23 CNV/Williams syndrome neural precursor cells (29).



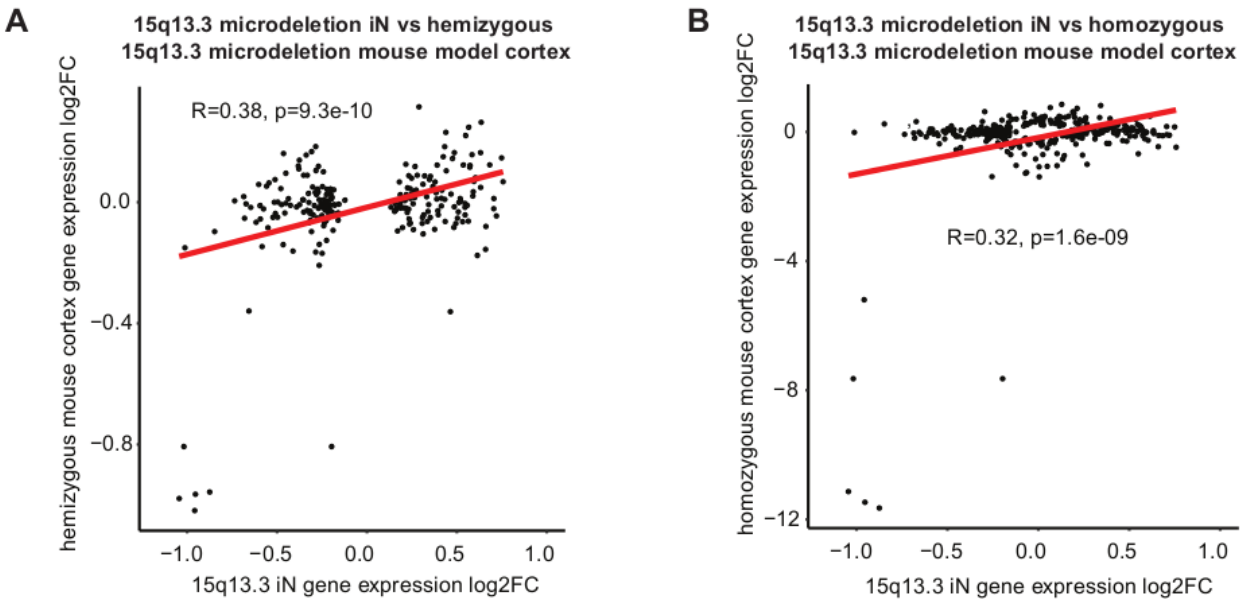
Supplementary Figure S1. Expression variance. Percentage of variance explained by cell type (iPSC, iN), donor, and genotype (15q13.3 microdeletion, control) over all expressed genes.



Supplementary Figure S2. Principal component analysis. (A, D) RNA-Seq in 15q13.3 iPSCs and iNs. (B, E) Methyl-Seq in 15q13.3 iPSCs and iNs. (C, F) ATAC-Seq in 15q13.3 iPSCs and iNs. (G) RNA-Seq in CRISPR knockout iNs.



Supplementary Figure S3. Correlation analysis between 15q13.3 iNs and CRISPR knockout iNs. Correlation of gene expression log2 fold change between 15q13.3 iNs and knockout of (A) CHRNA7, (B) KLF13, (C) FAN1, (D) MTMR10, and (E) OTUD7A.



Supplementary Figure S4. Correlation analysis between 15q13.3 microdeletion iNs and 15q13.3 microdeletion mouse model cortex. Correlation of gene expression log2 fold change between (A) 15q13.3 iNs and hemizygous 15q13.3 mouse model cortex and (B) 15q13.3 iNs and homozygous 15q13.3 mouse model cortex.

Supplementary References

1. Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Marro S, et al. Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells. 2013;78(5):785–798.

2. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–595.

3. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842.

4. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011 Jun;21(6):974–84.

5. Trost B, Walker S, Wang Z, Thiruvahindrapuram B, MacDonald JR, Sung WWL, et al. A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. The American Journal of Human Genetics. 2018 Jan 4;102(1):142–55.

6. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–1111.

7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):1–21.

8. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017 : a more comprehensive , powerful , flexible and interactive gene set enrichment analysis toolkit. 2017;45(May):130–137.

9. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015 May;12(5):453–7.

10. Song Y, Botvinnik OB, Lovci MT, Kakaradov B, Liu P, Xu JL, et al. Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. Mol Cell. 2017 Jul 6;67(1):148-161.e5.

11. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci U S A. 2015 Jun 9;112(23):7285–90.

12. Burke EE, Chenoweth JG, Shin JH, Collado-Torres L, Kim S-K, Micali N, et al. Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. Nature Communications. 2020 Jan 23;11(1):1–14.

13. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics. 2016 Nov 25;17(1):483.

14. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357–9.

15. Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571–1572.

16. Li S, Figueroa ME, Kormaksson M, Melnick A, Mason CE, Garrett-Bakelman FE, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology. 2012;13(10):R87.

17. Jühling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S. Metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. Genome Research. 2016;26(2):256–262.

18. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods. 2013 Dec;10(12):1213–1218.

19. Bernstein BE, Brown M, Johnson DS, Liu XS, Nussbaum C, Myers RM, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biology. 2008;9(9):R137.

20. Stark R, Brown G. DiffBind: differential binding analysis of ChIP-Seq peak data [Internet]. 2011. Available from: http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf

21. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576–89.

22. Duren Z, Chen X, Jiang R, Wang Y, Wong WH. Modeling gene regulation from paired expression and chromatin accessibility data. Proceedings of the National Academy of Sciences. 2017;114(25):E4914–E4923.

23. Lowther C, Costain G, Stavropoulos DJ, Melvin R, Silversides CK, Andrade DM, et al. Delineating the 15q13.3 microdeletion phenotype: a case series and comprehensive review of the literature. Genetics in medicine : official journal of the American College of Medical Genetics. 2015;17(February):1–9.

24. Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, Tomasini L, et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature. 2012 Dec 20;492(7429):438–42.

25. Kang X, Yu Q, Huang Y, Song B, Chen Y, Gao X, et al. Effects of Integrating and Non-Integrating Reprogramming Methods on Copy Number Variation and Genomic Stability of Human Induced Pluripotent Stem Cells. PLoS ONE. 2015;10(7):e0131128.

26. Hoffman GE, Hartley BJ, Flaherty E, Ladran I, Gochman P, Ruderfer DM, et al. Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. Nat Commun. 2017 20;8(1):2225.

27. 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 Syst Biol. 2016 15;10(1):105.

28. Ward TR, Zhang X, Leung LC, Zhou B, Muench K, Roth JG, et al. Genome-wide molecular effects of the neuropsychiatric 16p11 CNVs in an iPSC-to-iN neuronal model. bioRxiv. 2020 Feb 10;2020.02.09.940965.

29. Chailangkarn T, Trujillo CA, Freitas BC, Hrvoj-Mihic B, Herai RH, Yu DX, et al. A human neurodevelopmental model for Williams syndrome. Nature. 2016 Aug 18;536(7616):338–43.