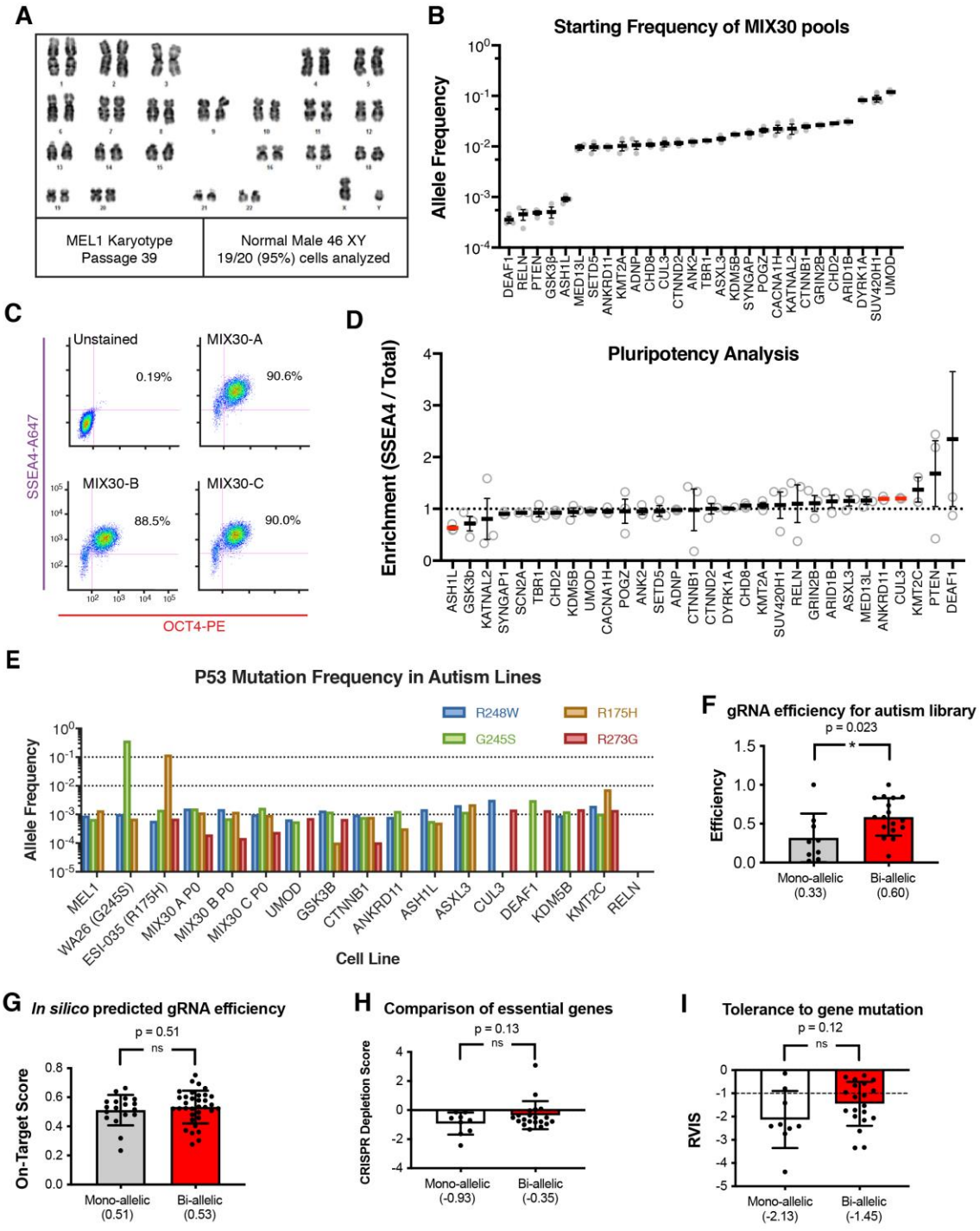
Extended Data

Title: A MULTIPLEX HUMAN PLURIPOTENT STEM CELL PLATFORM DEFINES MOLECULAR AND FUNCTIONAL SUBTYPES OF AUTISM-RELATED GENES.

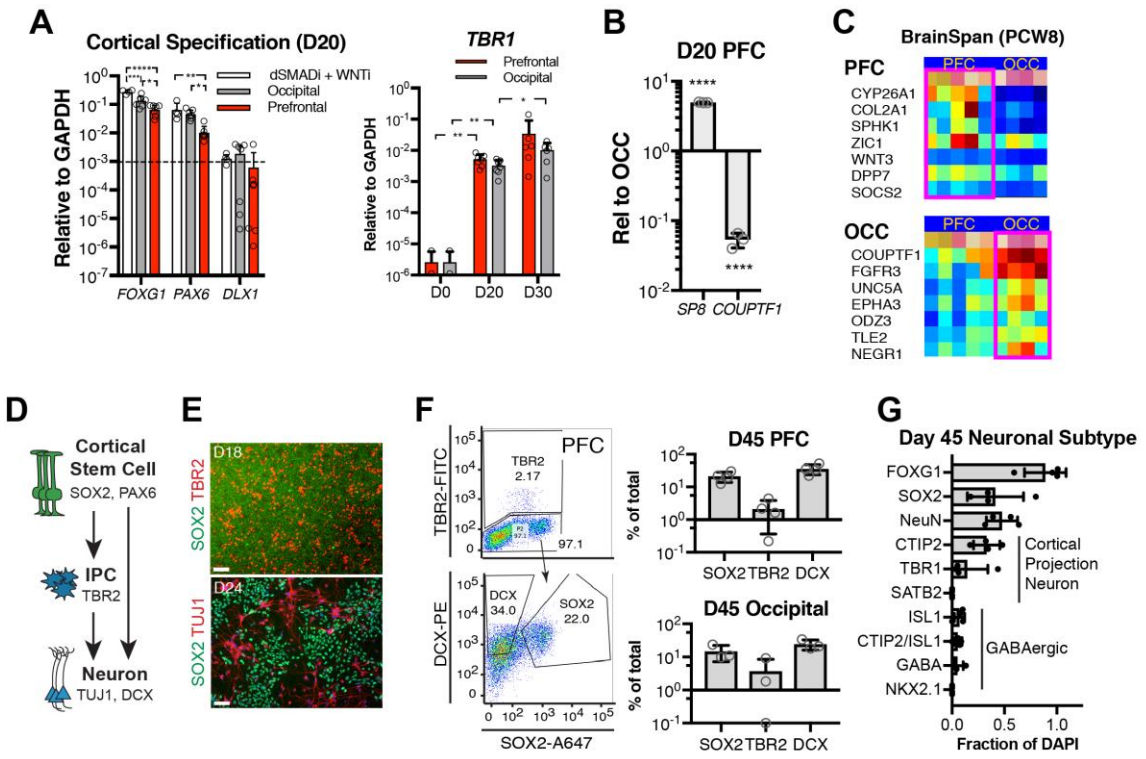
Authors: Gustav Y. Cederquist[1,2], Jason Tchieu[1], Scott J. Callahan[1,3,4], Kiran Ramnarine[1], Sean Ryan[1], Chao Zhang[5,6], Chelsea Rittenhouse[1], Nadja Zeltner[1,7], Sun Young Chung[1], Ting Zhou[8], Shuibing Chen[8,9], Doron Betel[5,6], Richard M. White[3], Mark Tomishima[1], Lorenz Studer*[1,10]



Supplemental Figure 1: Characterization of MIX30 pool, related to Figure 1.

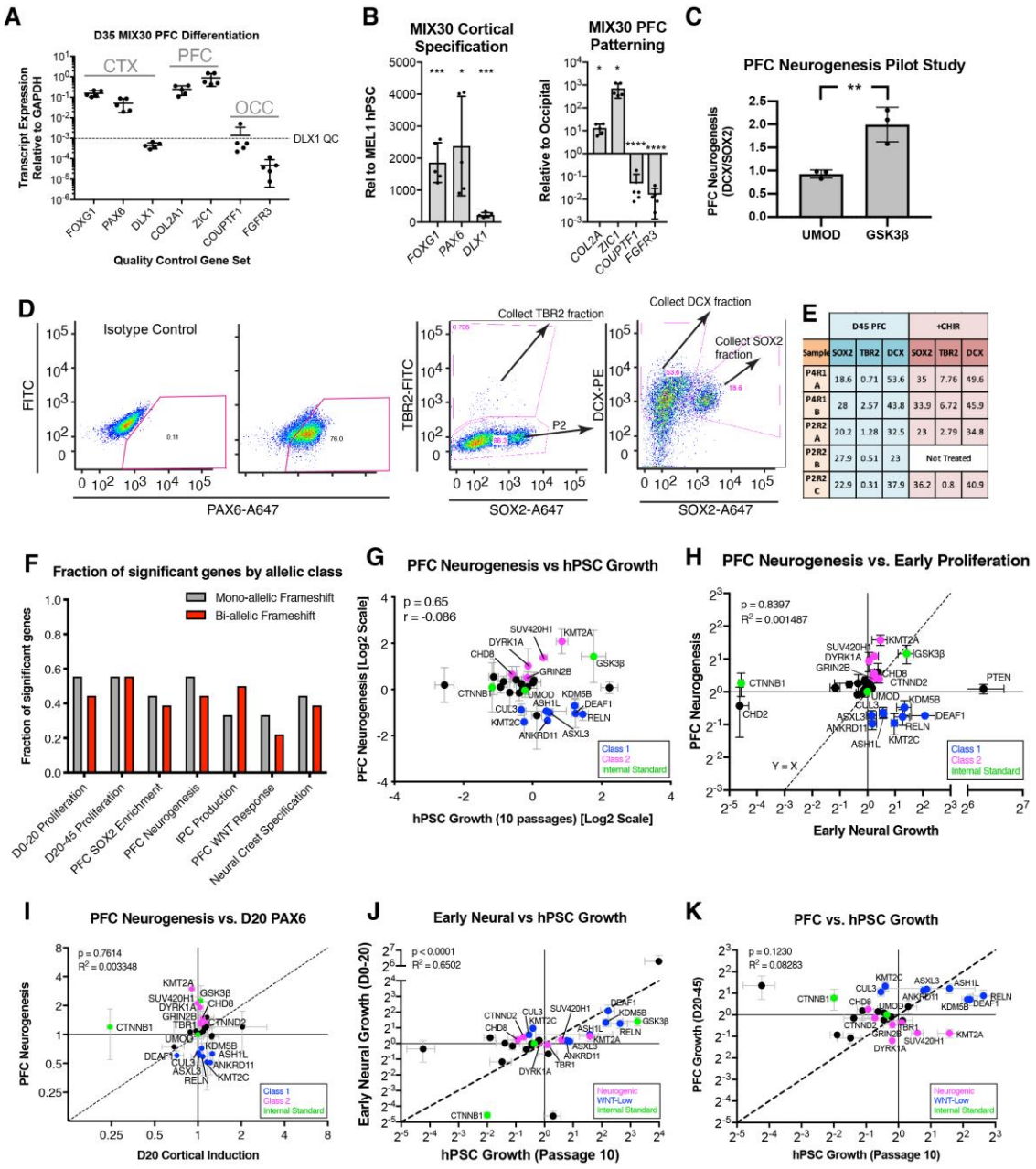
(A) Karyotype analysis of 46XY MEL1 founder line demonstrates normal karyotype in 19/20 (95%) of cells. **(B)** Average starting frequency of each line in three MIX30 pools. Some lines that showed increased fitness in preliminary studies were mixed at low frequencies while the negative control, UMOD, and some lines that showed decreased

fitness were mixed at high frequencies. Graph shows mean±S.E.M. n = 3 MIX30 pools. **(C)** Flow cytometry analysis for expression of pluripotency markers OCT4 and SSEA4. n = 3 MIX30 pools. **(D)** Average cell line enrichment in MIX30 SSEA4 sorted fractions relative to unsorted MIX30 fractions. Red bars indicate cell lines with significant increases or decreases in enrichment score compared to UMOD, FDR < 0.05. n = 3 MIX30 pools. Dots represent independent MIX30 pools. Graph show mean±S.E.M. **(E)** Bar plots summarizing results of ddPCR analysis for p53 DNA binding domain mutations that have been reported to arise spontaneously in pluripotent stem cells (Merkle et al., 2017) on control lines (MEL1 founder, UMOD, CTNNB1, and GSK3β) and Class 1 lines (ANKRD11, ASH1L, ASXL3, CUL3, DEAF1, KDM5B, KMT2C, and RELN), which showed competitive growth advantages (see **Figure 3**). Typically, minor allele frequencies were estimated to be ~0.1% or less, though the R175H mutation may approach 1% frequency in the KMT2C line. Graph depicts mean allele frequencies. ddPCR, droplet digital PCR. **(F)** Average efficiency in generating indels for homozygous and heterozygous gene classes. The efficiency in generating bi-allelic indels is higher than generating mono-allelic indels. Graph shows mean±S.D, dots represent gene targeted. Student two-tailed t-test. * P = 0.023. Mono-allelic n = 9 genes; Bi-allelic n = 18 genes. **(G)** *In silico* prediction of guide efficiency using Broad GPP Portal (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) predicts equivalent efficiency between guides that generate mono-allelic and bi-allelic mutations. Graph shows mean±S.D, dots represent individual gRNAs. Student two-tailed t-test. P = 0.51. Monoallelic n = 18 gRNAs; bi-allelic n = 36 gRNAs. **(H)** Requirement of autism gene function in pluripotency assessed using data from genome-wide screen for essential genes in hPSCs (Yilmaz et al., 2018). No significant difference between mono-allelic and bi-allelic groups. Graph shows mean±S.D, dots represent gene targeted. Student two-tailed t-test. P = 0.13. mono-allelic n = 9 genes; bi-allelic n = 18 genes. **(I)** Tolerance of autism genes to sequence variation in the human population, measured by residual variant intolerance score (RVIS)(Petrovski et al., 2013). No significant difference between mono-allelic and bi-allelic groups. Graph shows mean±S.D., dots represent gene targeted. Student two-tailed t-test. P = 0.12. mono-allelic n = 9 genes; bi-allelic n = 18 genes.



Supplemental Figure 2: Additional characterization of hPSC-derived PFC, related to Figure 2.

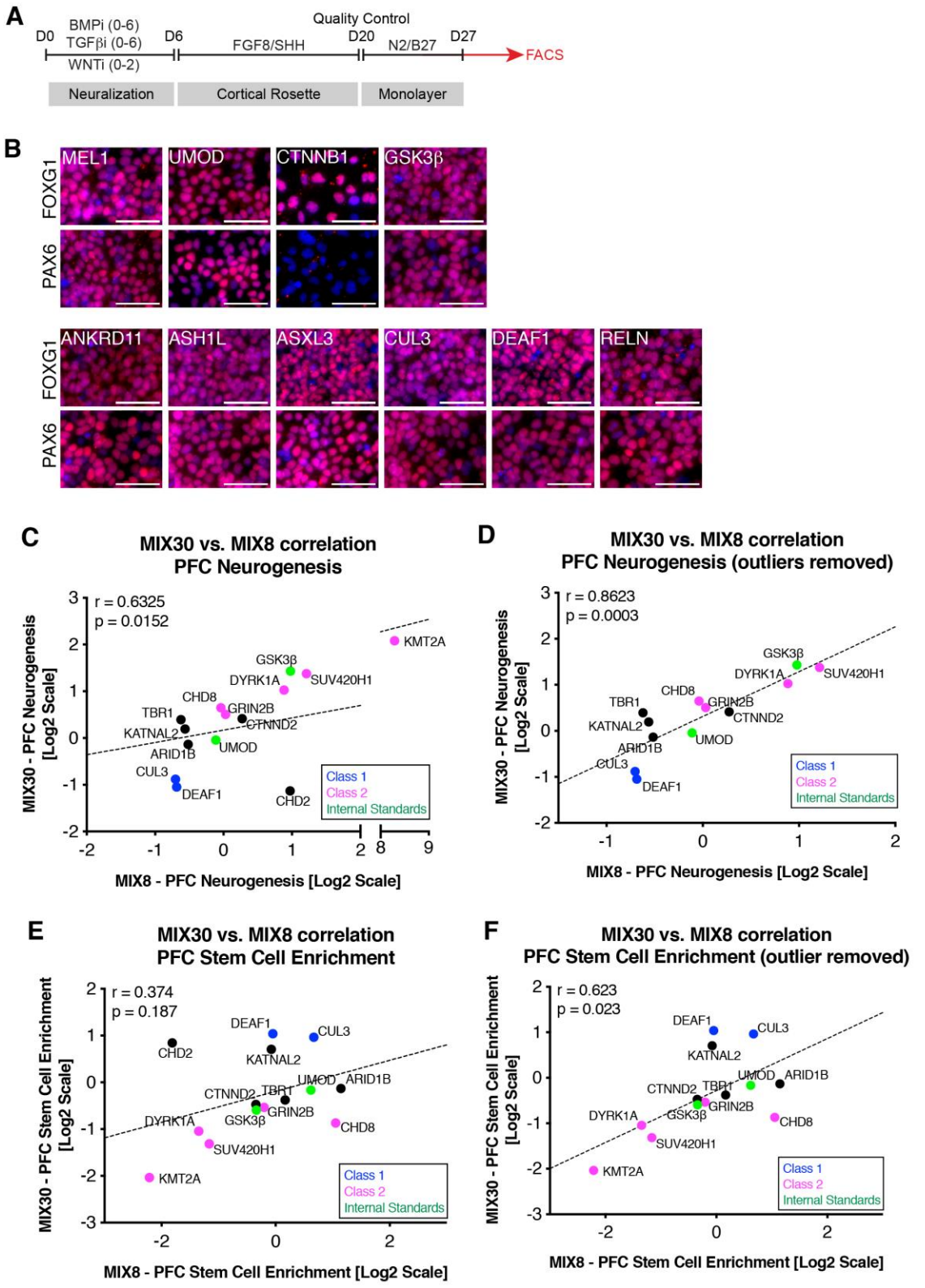
(A) qRT-PCR analysis of cortical identity genes at day 20 PFC and OCC differentiations are compared to standard dual SMAD inhibition protocol (dSMADi+WNTi), which is known to produce dorsal pallium/cortical neurons. Expression of *FOXG1*, *PAX6*, and DLX1 was comparable between differentiation protocols. (FOXG1 ****p<0.0001, ***p=0.0001, *p=0.016; PAX6 **p=0.003, *p=0.024). Day 20 *DLX1* expression > 0.098% of *GAPDH* expression was used as a quality control cut value to identify ventralized cultures. *TBR1*, which is expressed in deep layer cortex, shows induction by Day 20 in PFC and OCC diffs (**p<0.01, *p=0.024). Graphs show mean±S.D., dots represent individual differentiations. Day 20 n = 8 differentiations, day 30 n = 5 differentiations. **(B)** qRT-PCR analysis of differential gene expression between day 20 PFC and OCC cultures for *SP8* (4.972 ± 0.045, p < 0.0001) and *COUPTF1* (0.05339 ± 0.013, p < 0.0001). Graphs depict mean±S.D, student two-tailed t-test. n = 3 differentiations. **(C)** Identification of genes that are differentially expressed between PFC and OCC in PCW 8 human fetal tissue. Heatmap generated using Brainspan.org. **(D)** Schematic illustration of cortical neurogenesis. **(E)** Representative immunocytochemistry for SOX2, TBR2, and TUJ1 demonstrates appropriate neurogenic capacity during PFC differentiation. n = 3 differentiations. **(F)** Day 45 fixed intracellular flow cytometry analysis and percent of total quantification for SOX2, TBR2, and DCX populations in PFC (SOX2: 21.18 ± 7.36, TBR2: 2.12 ± 1.76, DCX; 35.88 ± 12.34). and OCC (SOX2: 14.86 ± 7.68, TBR2: 3.71 ± 4.84, DCX: 24.37 ± 8.28) differentiations. Graphs show mean±S.D., dots represent individual differentiations. n = 4 differentiations. **(G)** Immunocytochemical quantification of cortical projection neuron and GABAergic neuronal subtypes in day 45 PFC differentiations. Graphs show mean±S.D., dots represent individual differentiations. n = 4. PCW, post-conception week; PFC, prefrontal cortex; OCC, occipital cortex; QC, quality control. Scale bars = 50 μm.



Supplemental Figure 3: Additional characterization of multiplex neurogenesis assay, related to Figure 3.

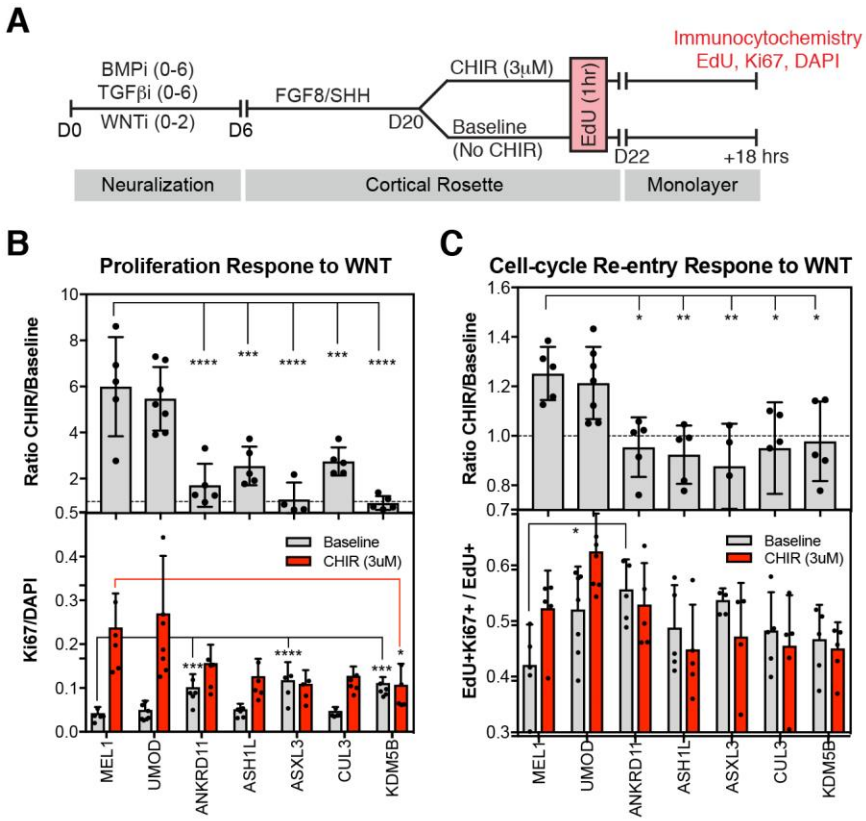
(A) *GAPDH* normalized qRT-PCR analysis of D35 MIX30 PFC differentiation shows appropriate PFC patterning. Graphs depict mean±S.D. n = 5 differentiations from three MIX30 pools. **(B)** Values from panel C plotted as relative to historical controls from Supplemental Figure 2A (left panel) and Figure 2C (right panel)). Graphs depict mean±S.D., One-sample t-test versus hypothetical value of 1. *** *FOXG1* p = 0.0026, * *PAX6* p = 0.0267, *** *DLX1* p = 0.0024, * *COL2A1* p = 0.0131, * *ZIC1* p = 0.0231, **** *COUPTF1* p < 0.0001, **** *FGFR3* p < 0.0001. PFC, prefrontal cortex; FACS, fluorescent activated cell sorting. **(C)** Pilot study testing whether multiplex approach can detect a known neurogenesis phenotype using pools of 8 lines. The GSK3β positive control line exhibited an expected increase in PFC neurogenesis compared to the

UMOD negative control line, p = 0.0084, unpaired two-tailed t-test. Error bars are s.d. n = 3 differentiations. **(D)** Representative FACS plots showing isolation Day 20 sort of PAX6+ fraction (left panels) and Day 45 sort of $TBR2^+$ fraction, followed by isolation of $SOX2^+$ and DCX^+ fractions (right panels). CHIR experiments were performed using the same sorting strategy. **(E)** Percent of total values for each sorted fraction for all five replicates of multiplex neurogenesis assay (see **Figure 3**) and all four replicates of multiplex PFC WNT response assay (see **Figure 5**). *ddPCR failed. **(F)** Mono-allelic and bi-allelic mutations equally likely to give phenotypes during this study. **(G)** No correlation between cell line growth at the pluripotent stage (**Figure 1E**) and PFC neurogenesis phenotype (**Figure 3C**). Correlation r = -0.086, p = 0.65. **(H)** No correlation between early neural growth (**Figure 3E**) and PFC neurogenesis. Correlation r = 0.044, p = 0.817. **(I)** No correlation between cortical induction (**Figure 3B**) and PFC neurogenesis. Correlation r = 0.0315, p = 0.869. **(J)** Positive correlation between early neural growth (**Figure 3E**) and hPSC growth phenotypes (**Figure 1E**). r = 0.593, p = 0.0005. **(K)** No correlation between PFC growth phenotype (**Figure 3E**) and hPSC growth phenotypes (**Figure 1E**). r = 0.176, p = 0.351. Dotted line is plotted at Y = X. Error bars are s.e.m.



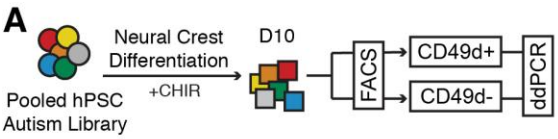
Supplemental Figure 4: Additional data related to validation of multiplex neurogenesis assay, related to Figure 4.

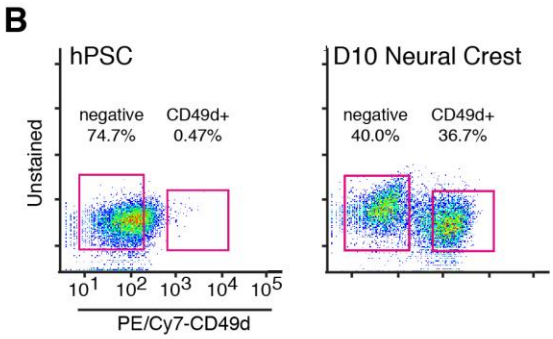
(A) Schematic illustration of single genotype neurogenesis assay. Quality control by immunocytochemistry was performed at day 20 to ensure appropriate cortical patterning for each line. Lines were passaged at day 20 to a saturating density of 10^6 cells / cm^2 and allowed to undergo neurogenesis for 7 days. A negative control line, UMOD, was included in each differentiation batch for normalization. **(B)** Day 20 immunocytochemistry for FOXG1 and PAX6 to confirm appropriate cortical patterning. **(C-F)** Correlation between MIX30 experiments and pilot studies that used mini-pools of only 8 lines (MIX8) demonstrates that multiplex assay phenotypes are stable to changes in pool number and reproducible over time. **(C)** PFC neurogenesis phenotypes showed positive correlation (r = 0.6325, p = 0.0152). **(D)** PFC neurogenesis correlation repeated with outliers removed (r = 0.8623, p = 0.0003). The CHD2 allele frequency approached zero and thus exhibited high technical variability. KMT2A was a statistical outlier. **(E)** PFC stem cell enrichment did not show a positive correlation (r = 0.374, p = 0.187), though this was influenced by a CHD2 outlier. CHD2 absolute allele frequency was near the limit of detection (< 1:13,000 in SOX2 fraction), precluding reliable analysis. **(F)** PFC stem cell enrichment shows a positive correlation when CHD2 outlier is removed (r = 0.623, p = 0.023). Time point for MIX30 data is day 45, and time point for MIX8 data is day 55. MIX30 data is from **Figure 3**. n = 1 differentiation for MIX8 data.



Supplemental Figure 5: Proliferation and cell-cycle re-entry response to WNT signaling, related to Figure 5.

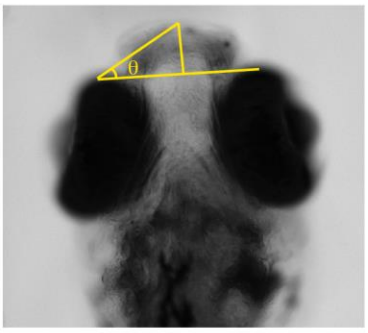
(A) Experimental design for single genotype analysis of WNT responsiveness during PFC development. **(B)** Baseline proliferation rates (Ki67/DAPI) and proliferation response to CHIR (CHIR/Baseline). 3/5 class 1 lines show increased proliferation at baseline compared to MEL1 (ANKRD11 ***p = 0.0009, ASXL3 ****p < 0.0001, KDM5B ***p = 0.0001), and 5/5 class 1 lines have a blunted proliferation response to CHIR compared to MEL1 (ANKRD11 ****p < 0.0001, ASH1L ***p = 0.0004, ASXL3 ****p < 0.0001, CUL3 ***p = 0.0008, KDM5B ****p < 0.0001). **(C)** rates of cell-cycle re-entry (Ki67+EdU+/EdU+) at baseline and in response to CHIR. Control lines show increased rates of cell-cycle re-entry in response to CHIR, while 5/5 Class 1 lines showed a blunted response compared to MEL1 (ANKRD11 *p = 0.015, ASH1L **p = 0.0067, ASXL3 **p = 0.0033, CUL3 *p = 0.0139, KDM5B *p = 0.0284). Graph shows mean± s.d., dots represent individual differentiations, One-way ANOVA with Dunnett Test. n = 7 differentiations for UMOD, n = 5 differentiations for MEL1, ANKRD11, ASH1L, CUL3, and KDM5B, n = 4 differentiations for ASXL3.

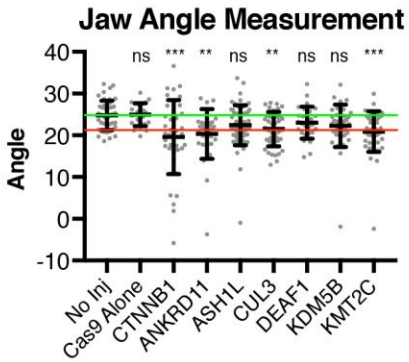


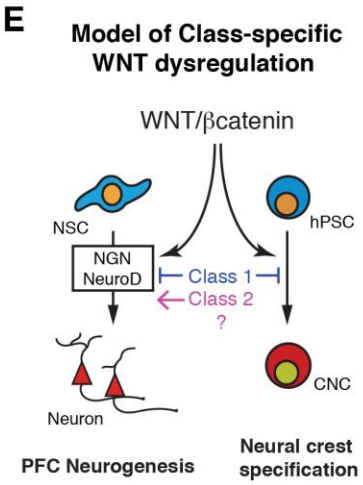


C

Sample	Negative	CD49d+
MIX30 A	40	36.7
MIX30 B	40.6	36.1
MIX30 C	42.8	34.9



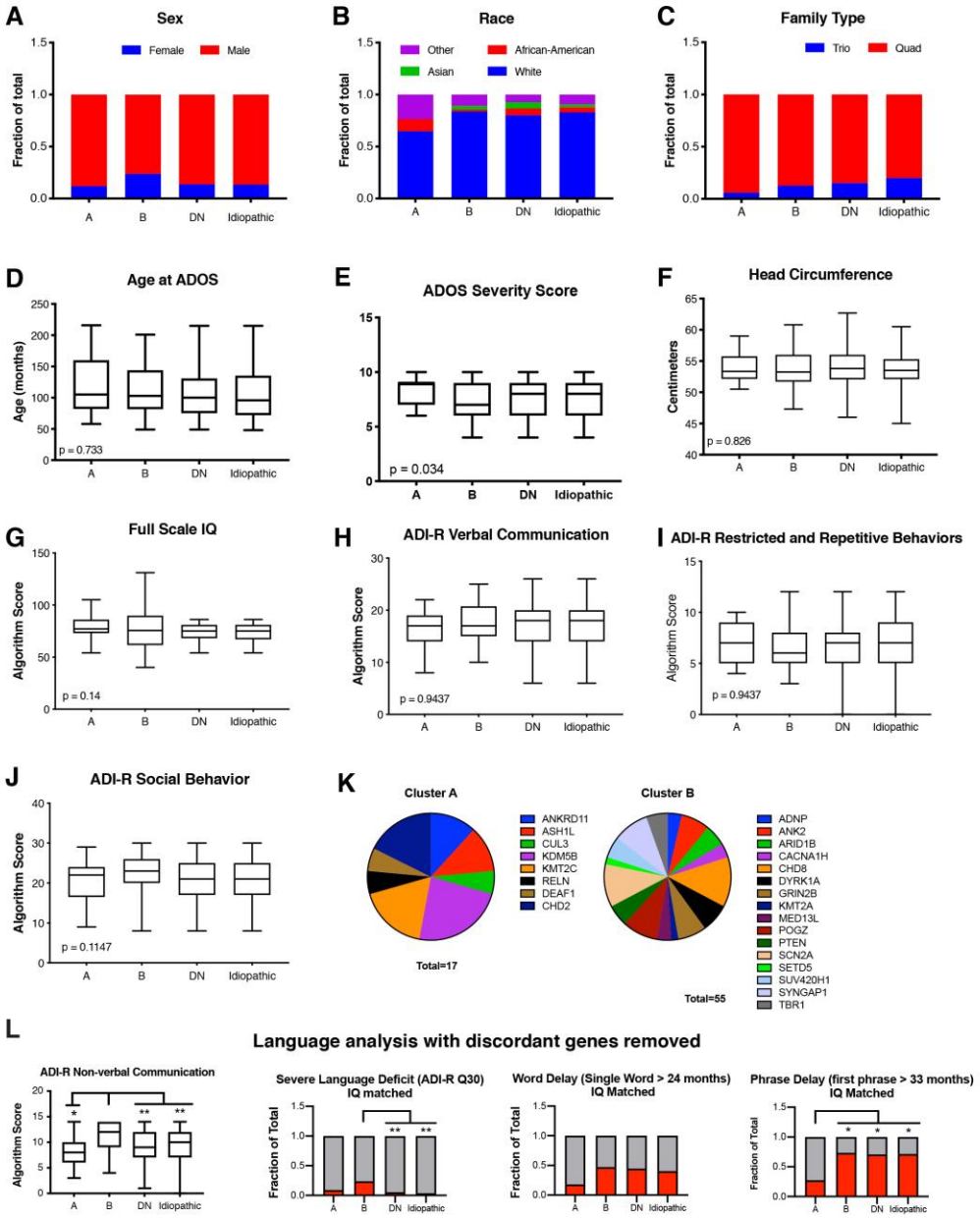




Supplemental Figure 6: Additional data related to multiplex neural crest assay, related to Figure 6.

(A) Schematic illustration of multiplex strategy to test autism lines for cranial neural crest (CNC) phenotypes. At day 10, MIX30 pools are separated into CNC ($CD49^+$) and non-CNC ($CD49d^-$) fractions using FACS. Allele frequencies are calculated for each fraction using ddPCR. **(B)** Representative FACS plots showing isolation of $CD49^+$ and $CD49d^-$ fractions during CNC differentiation. **(C)** Percent of total values for sorted fractions for all three replicates of multiplex neural crest assay (**See Figure 6**). **(D)** Schematic of quantification for zebrafish jaw development at 7 dpf. A straight line is drawn between the top of the eyes, and second line is then drawn to the middle of the upper jaw. The angle between the two lines (θ) is measured using ImageJ. A cut-off value of -1 s.d. (red line) from the mean (green line) of the no injection group was used to define hypomorphic animals. Graphs show raw data that underlie the hypomorph quantification in **Figure 6C**. $p < 0.0001$ for group differences calculated by non-parametric Kruskal-Wallis test, followed by Dunn's test for multiple comparisons. CTNNB1 $p = 0.0007$,

ANKRD11 p = 0.0013, CUL3 p = 0.0073, KMT2C p = 0.0007. **(E)** Potential model of WNT-dysregulation for class 1 and 2 autism mutations. WNT/βCatenin promotes neural crest specification and regulates cortical neurogenesis through transcriptional activation of proneural factors NGN and NeuroD. Class 1 mutations block WNT-dependent responses during neural crest specification and neurogenesis. dpf, Days post fertilization; CNC, cranial neural crest; ddPCR, droplet digital PCR; FACS, fluorescent activated cell sorting; IPC, intermediate progenitor cell; PFC, prefrontal cortex.



Supplemental Figure 7: Additional data related to clinical language analysis, related to Figure 7.

(A-C) Demographic sex, race, and family data for four IQ-matched autism cohorts, plotted as fraction of total. **(D)** Autism cohorts were not significantly different in age (ANOVA p = 0.733). **(E)** There may be differences in ADOS calibrated severity score

between groups (Kruskal-Wallis p = 0.034), however post-hoc Dunn's comparison did not identify specific group differences. **(F)** Autism cohorts were not significantly different in head circumference. **(G)** IQ-matched cohorts were not different in average IQ. **(H-J)** No significant between group differences on ADI-R verbal communication scores (Kruskal-Wallis p = 0.9437), restricted and repetitive behavior scores (Kruskal-Wallis p = 0.9437), or social behavior scores (Kruskal-Wallis p = 0.1147). p values corrected for multiple comparisons of behavioral domains using Holm-Sidak method. Error bars are s.d. **(K)** genotype composition of cohort A and B. **(L)** Re-analysis of clinical data omitting genotypes that were discordant in validation study. Cluster B shows an increased severity of non-verbal communication deficits versus other cohorts. Kruskal-Wallis p = 0.0018. Post-hoc Dunn's tests for multiple comparisons A vs. B p = 0.036, B vs. DN p = 0.0045, B vs. idiopathic p = 0.0018. Cluster B shows an increased proportion of severe language deficit compared to control cohorts. Chi-square p < 0.0001, Fisher's Exact test with correction for multiple testing B vs. DN p = 0.0018, B vs. idiopathic p = 0.0018. Cluster A shows a reduced proportion of phrase delay than other cohorts. Chi-square p = 0.034, Fisher's Exact test with correction for multiple testing A vs. B p = 0.039, A vs. DN p = 0.019, A vs. idiopathic p = 0.019. ADI-R, Autism Diagnostic Interview – Revised; ADOS, Autism Diagnostic Observation Schedule; DN, *de novo* control cohort.