Feasibility and application of polygenic score analysis to the morphology of human induced pluripotent stem cells

Jonathan R. I. Coleman^{1,*}

¹Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

*Address correspondence to Jonathan Coleman, SGDP Centre, IoPPN - PO80, DeCrespigny Park, Denmark Hill, London, UK, SE5 8AF.

jonathan.coleman@kcl.ac.uk

Abstract

Genome-wide association studies have identified thousands of significant associations between genetic variants and complex traits. Inferring biological insights from these associations has been challenging. One approach attempted has been to examine the effects of individual variants in cellular models. Here, I demonstrate the feasibility of examining the aggregate effect of many variants on cellular phenotypes. I examine the effects of polygenic scores for cross-psychiatric disorder risk, schizophrenia, body mass index and height on cellular morphology, using 1.5 million induced pluripotent stem cells (iPSC) from 60 European-ancestry donors from the Human iPSC Initiative dataset. I show that measuring multiple cells per donor provides sufficient power for polygenic score analyses, and that cross-psychiatric disorder risk is associated with cell area ($p = 0.004$). Combined with emerging methods of high-throughput iPSC phenotyping, cellular polygenic scoring is a promising method for understanding potential biological effects of the polygenic component of complex traits.

1 **Introduction**

2 Genome-wide association studies (GWAS) allow associations between
3 genotypes and a trait of interest to be explored without a pre-specified idea of the
4 relevant biological mechanisms ¹. GWAS have successfully identified statistical
5 associations with common genetic variants, with over 150,000 variant-trait
6 associations described as of December 2020 ². In comparison with earlier methods
7 such as genome-wide linkage analysis, GWAS have been effective in implicating
8 variants associated with complex traits (traits affected by many factors, including
9 environmental influences and genetic variants) ³. GWAS have highlighted the
10 polygenic component of complex traits, wherein thousands of variants with very
11 small individual effects together account for a substantial proportion of the genetic
12 influences on the trait ⁴.

13 Gaining insights into the biology of complex traits has been a main motivator
14 for GWAS. However the promise of new biology has been slow to emerge ⁴. Some of
15 these limitations are inherent to GWAS, such as the difficulty of determining which of
16 the numerous correlated variants associated with a trait at a given locus contribute to
17 causality. Others reflect the limits of model systems. Animal models are useful for
18 understanding the function of conserved genes within a living organism. In contrast,
19 they are less useful when seeking to translate the effects of variants, because causal
20 variants are poorly conserved across species ⁵. Studying certain phenotypes
21 introduces further limitations. Perhaps the clearest example of this is the study of
22 behaviour, where the brain is the focus of biological interest ⁶. There are obvious
23 ethical and logistical impediments to accessing living human brain tissue, which
24 largely prevent direct functional experiments that would provide vital context for
25 understanding and validating behavioural GWAS results.

1 Efforts to address these limitations of biological context are emerging, for
2 example through the work of the PsychENCODE consortium in developing large
3 datasets of biological annotations from post-mortem human brains ⁷. A further
4 promising area is the increasing diversity in neuron types derived from human
5 induced pluripotent stem cells (iPSCs; reviewed in ⁸). Mature human cells can be
6 reprogrammed to a pluripotent state (iPSCs) by controlled exposure to specific
7 transcription factors ⁹. Self-renewing iPSCs can be maintained under standardised
8 conditions ¹⁰. Further treatment can direct the subsequent differentiation of the
9 iPSCs down specific developmental pathways. For example, highly pure populations
10 of all of the major neuron types in the human brain can now be derived from iPSCs ⁸.
11 Cells from human donors are a valuable model system when directly studying the
12 relevant tissue is challenging, as they enable the assessment of living human cell
13 types. Typically, derived cell types only approximate the real cell types they model;
14 for example, derived neurons are developmentally immature, resembling foetal
15 developmental stages ⁸. In addition, in vivo cells reflect the developmental history of
16 the organism (including the effects of intrinsic and extrinsic environments), a history
17 that is not transferred to iPSCs when they are generated. However, for the purposes
18 of modelling diseases, these approximations can be useful simplifications ⁸.

19 As such, iPSCs can be a valuable model system in studying complex
20 behavioural traits. To date, most genetic research using iPSCs in behavioural
21 phenotypes has focussed on the effects of single variants of moderate effect,
22 primarily in schizophrenia and autism spectrum disorder, where such effects have
23 been shown to have an important contribution ^{6,11}. However, both of these disorders
24 also have sizable polygenic components ⁶. Most of the genetic contribution to
25 behavioural traits is likely to be similarly polygenic. It would be valuable to be able to

1 use iPSCs to examine polygenic components. One means to assess polygenic
2 effects is through polygenic scoring, wherein multiple genetic variants are used to
3 create a single score in an individual as a weighted sum of the alleles that individual
4 carries ¹². Typically, the weights for polygenic scoring are derived from a base
5 GWAS of a trait of interest, providing a means by which to assess the shared genetic
6 variance between the trait of interest from the base GWAS and a phenotype in a
7 target genotyped cohort ¹³.

8

9 In this exploratory paper, I aim to demonstrate the feasibility of polygenic
10 scoring in iPSC datasets. I do this first through power calculations, assessing the
11 number of cell donors required to provide >80% power to detect plausible levels of
12 genetic covariance between the GWAS phenotype and the cellular phenotype. I then
13 test the association of polygenic scores from four highly powered GWAS of complex
14 traits with iPSC cell morphology. These GWAS are the most powerful available
15 studies of a single psychiatric disorder (schizophrenia ¹⁴) and of shared genetic
16 effects on psychiatric disorders ¹⁵, and of two similarly-powered GWAS of other
17 complex traits, one partially behavioural (body mass index ¹⁶) and one without a
18 behavioural aetiology (height ¹⁶).

19 Donor number is among the principal limitations of polygenic analyses in
20 iPSCs and their derivatives. Polygenic scores in complex traits typically capture only
21 a small amount of variance in the trait, and as such usually require hundreds of
22 participants for sufficient power ¹³. I address this in two ways. I use data from the
23 Human Pluripotent Stem Cell Initiative (HipSci), an openly-accessible, large
24 collection of iPSCs generated with a standardised pipeline, and with extensive
25 phenotypic and genetic data ¹⁷. Furthermore, I use mixed linear models to analyse

1 numerous cells from each donor. Although these technical replicates do not provide
2 as much power as new donors, they provide some increase because they limit the
3 impact of measurement error ¹⁸.

4 I focus my analyses on the morphology of iPSC cells, due to the availability of
5 sufficient data on these cells compared to their neuronal derivatives. It also is
6 feasible that the polygenic components studied may affect the morphology of iPSCs
7 directly. For example, genes involved in neuronal morphology have been implicated
8 in numerous GWAS of psychiatric disorders ^{15,19–21}. While strongest in neurons,
9 these effects may also be apparent in the morphology of cells more generally,
10 particularly in iPSCs, whose cellular fate is undetermined.

11

12 **Methods**

13 Human Induced Pluripotent Stem Cell Initiative (HipSci)

14 This paper consists of secondary analyses of existing data from HipSci. Full
15 descriptions of the generation of these data are provided elsewhere ^{17,22,23}.

16 I used data on genome-wide genotyping and cellular microscopy phenotyping
17 from the HipSci project. HipSci comprises several disease cohorts, as well as a
18 cohort of unaffected individuals. Participants in this study were unaffected
19 individuals. Broader phenotypic information (such as height, BMI or mental health
20 phenotyping, which would be relevant to the polygenic score analyses) was not
21 available on these participants. Following quality control (described below), data
22 from 103 iPSC cell lines from 60 participants was included in the final analysis
23 (Supplementary Table 1, Supplementary Figure 1). All but two cell lines were
24 included in a previous publication ²².

25

1 Cellular phenotyping

2 Cellular phenotyping data was available for all analysed participants, and has
3 been described in detail elsewhere ²². In brief, iPSCs were generated from
4 fibroblasts or peripheral blood mononuclear cells via episomal DNA or Sendai vector
5 transduction ²². iPSCs were cultured in a feeder (mouse embryonic fibroblasts) or
6 feeder-free environment, before assessment on a fibronectin adhesion assay ²².
7 From each cell line, 3000 iPSCs were plated for 24 hours on a single well coated
8 with varying concentrations of the extracellular matrix protein fibronectin: 1µg/mL
9 (sub-optimal for cell adhesion), 5µg/mL and 25µg/mL ²². For each cell line, each
10 condition was performed in triplicate to mitigate well edge and position effects ²².
11 Single iPSCs, alone or aggregated in clumps of cells, were imaged using an
12 Operetta (Perkin Elmer) high content device, and processed, quantified, and
13 normalised as previously described ²³. Three morphological phenotypes were
14 determined from cell image data: cell area, roundness, and width-to-length ratio ²².
15 Cell width-to-length ratio was defined as the length (the longest line that could be
16 drawn within the cell) divided by the width (the longest line perpendicular to the
17 length that could be drawn within the cell). Cell roundness was defined using the
18 equation below ²²:

$$Roundness = 3.544 \times \sqrt{\frac{area - (perimeter/2)}{perimeter - 0.1}}$$

19

20 Genome-wide genotype data

21 Genome-wide genotype data was available for all analysed participants from
22 the Illumina HumanCoreExome-12 v1 BeadChip, previously imputed to a combined
23 1000 Genomes and UK10K reference panel ¹⁷. I initially included data from 280 cell

1 lines (including source cells - 60 fibroblasts and 5 peripheral blood mononuclear cells
2 - and 215 iPSC lines) from 60 participants with cellular phenotype data. Data
3 underwent quality control in PLINK 1.9 and R 3.6^{24–26}. For genotype quality control
4 and the generation of polygenic risk scores, I retained a single cell-line for each
5 individual (as cell lines from the same donor were approximately genetically
6 identical; Supplementary Methods). I excluded variants with minor allele frequency
7 (MAF) < 5%, call rate < 99%, and which were outside of Hardy-Weinberg equilibrium
8 (HWE p-value < 10⁻⁵). Following this step, I limited variants to those present on the
9 genotyping array only, and re-imputed the data on the Sanger imputation server,
10 imputing to the Haplotype Reference Consortium panel release 1.1, using the PBWT
11 + Eagle pipeline^{27–29}. I retained variants from the HRC-imputed data with imputation
12 INFO > 0.9, and MAF > 5%. I plotted cell line data on principal components of
13 genome-wide genotype data from 1000 Genomes project participants³⁰. This
14 confirmed all cell lines were from individuals of European ancestries. I also
15 generated polygenic scores from the genome-wide genotype data of the participants.
16 Reported sex of donors was consistent with the heterozygosity of X chromosome
17 variants in the cell lines. All participants were unrelated (π -hat < 0.125) and well-
18 genotyped (call rate > 99%).

19

20 Polygenic scoring

21 I generated polygenic scores using PRSice v2.3.1e³¹. I obtained summary
22 statistics from GWAS of height and BMI¹⁶ from the GIANT consortium, and from
23 GWAS of schizophrenia¹⁴ and cross-psychiatric disorder analyses¹⁵ from the
24 Psychiatric Genomics Consortium. The cross-psychiatric disorder GWAS captures
25 shared genetic effects from GWAS of schizophrenia, bipolar disorder, major

1 depression, autism spectrum disorder, attention-deficit hyperactivity disorder,
2 obsessive-compulsive disorder, anorexia nervosa, and Tourette's syndrome. The
3 method used in the cross-psychiatric disorder GWAS results in equal contributions
4 from each disorder, despite differences in power between the GWAS included.

5 For each base GWAS, I calculated polygenic scores without performing
6 regression, limiting to autosomal variants in common between each set of summary
7 statistics, the HipSci data, and the non-Finnish European participants in 1000
8 Genomes³⁰. I removed variants in linkage disequilibrium ($r^2 < 0.1$) with a variant with
9 a lower p-value within 250 kilobases. The HipSci cohort is small for polygenic score
10 analysis, and so I estimated the linkage disequilibrium structure from the non-Finnish
11 European participants in 1000 Genomes^{13,30}. I generated polygenic scores at the
12 threshold capturing the most variance in leave-one-sample-out polygenic scoring
13 from the initial publications ($p < 0.05$ for schizophrenia, $p < 0.001$ for height and
14 BMI). I also generated a polygenic score at $p < 1$, to capture the effects of all variants
15 (at the expense of including noise from variants not truly associated with the trait).
16 For the cross-psychiatric disorder analysis, I used two sets of summary statistics,
17 including and excluding data from 23andMe. Data including 23andMe is limited (to
18 protect participant privacy) to 10,000 variants in linkage equilibrium, with $p < 0.001$.
19 Leave-one-sample-out polygenic scoring was not reported in the original cross-
20 psychiatric disorder publication. Accordingly, I generated three polygenic scores for
21 the cross-psychiatric disorder data: 1) including data from 23andME, limited to
22 $p < 0.001$; 2) excluding 23andME and limiting to $p < 0.001$ for comparison; and 3)
23 excluding 23andME and including all variants ($p < 1$). In total, I analysed nine
24 polygenic scores. Polygenic scores are referenced in the format $\text{Trait}_{\text{Threshold}}$, for
25 example Schizophrenia_{0.05} (Table 1).

1 Power analyses

2 To assess the current power for polygenic score analyses in iPSCs, I
 3 performed a series of power analyses in R version 3.6, using the AVENGEME
 4 package^{26,32}. I performed power analyses for each of the full polygenic scores used
 5 in this paper, assessing the power of analyses assuming iPSCs from 60 donors. The
 6 first set of analyses (N=60) do not take account of the multiple phenotypic
 7 measurements made for each donor in the analyses presented, and so are
 8 conservative. Estimates of the effective sample size for the cohort analysed in this
 9 paper (taking into account the multiple phenotypic measurements) give a range of
 10 estimates between 850 and 2435 respectively (Supplementary Methods;
 11 Supplementary Table 2). Accordingly, I ran additional power analyses using these
 12 two values.

13 I defined parameters for power analysis from the GWAS from which each
 14 polygenic score originated (Supplementary Table 3). I estimated the power of
 15 polygenic score analysis for a hypothetical cellular phenotype, varying the
 16 covariance between the polygenic score trait and the phenotype. Covariance can be
 17 described as a function of the common genetic contribution to the polygenic score
 18 trait ($vg1$, which is fixed), the common genetic contribution to the phenotype ($vg2$,
 19 which I varied), and the genetic correlation between the polygenic score trait and the
 20 phenotype (which I varied):

$$Covariance = Correlation * \sqrt{vg1 \times vg2}$$

21 Statistical analysis

22 I ran all analysis in R version 3.6²⁶. I restricted the data for analyses to 103
 23 cell lines (from 60 donors) for which cell morphological data and polygenic scores
 24 were available (Supplementary Table 1, Supplementary Figure 1). Cell

1 morphological phenotype data were available for 1,543,624 individual cells in 2484
2 plate wells. In addition to polygenic scores, I also considered the effect of the
3 fibronectin concentration (coded as an ordinal variable of 1, 5, 25) of the wells on
4 which cells were plated as a variable of interest, as this has been shown to
5 contribute importantly to cell morphology²². I controlled for continuous variables of
6 four genomic principal components (to control for population stratification), and the
7 number of cells in each clump, as well as factors assessing the origin cell for the
8 iPSC (fibroblasts or peripheral blood mononuclear cells), the method of iPSC
9 reprogramming (episomal DNA or Sendai vector), and whether or not the origin cell
10 was maintained on a feeder. All of these variables may have effects on cell
11 morphology.

12 Data were drawn from different, nested levels of analysis (Figure 1).
13 Polygenic scores were assessed at the level of individual donors, whereas
14 phenotype data were assessed at the individual cell level. Covariates were assessed
15 at the donor (genomic principal components), well (fibronectin concentration), cell
16 line (origin cell, method of reprogramming, feeder status), and cellular (number of
17 cells in each clump) levels. Accordingly, I calculated associations between polygenic
18 scores and cell shape phenotypes using mixed linear models from the *lmerTest*
19 package³³. Specifically, I included all variables described above as fixed effects,
20 alongside a random intercept term of plate well nested within donor. As each well
21 contains only cells from a single cell line, the random effect of well also captures
22 variance between cell lines from the same donor. This allows for all observations to
23 be included while controlling for pseudoreplication resulting from measuring multiple
24 cells from each donor.

1 I performed analyses with and without assessing the interaction between the
2 variables of interest (polygenic score and fibronectin concentration). In analyses
3 including this interaction, I included additional interaction terms between each of the
4 variables of interest and all covariates^{34,35}. I fit all mixed linear models using REML,
5 and the significance of coefficients was calculated using Satterthwaite's
6 approximation for degrees of freedom. I visualised interaction models using the R
7 packages *ggplot2* and *interactions*^{36,37}.

Figure 1

Figure 1: Source of variables for analysis, and sample size at each level of analysis.

8 I standardised all continuous variables (polygenic scores, genomic principal
9 components, number of cells in each clump, and the phenotypes) within-cohort prior

1 to analysis. I included all other variables as binary factors. As such, the effect size
2 (B) for polygenic scores can be interpreted as the change in standard deviations in
3 the phenotype for one standard deviation change in the polygenic score. The effect
4 size for fibronectin concentration can be interpreted as the change in standard
5 deviations in the phenotype when comparing the tested conditions (either 5 or
6 25µg/mL concentration) to the baseline condition (1 µg/mL).

8 Sensitivity analyses

9 I performed several sensitivity analyses to assess the robustness of the
10 model. First, in addition to the method described above, I assessed the significance
11 of the polygenic score effects as a likelihood ratio test comparing the mixed linear
12 models (fit with maximum likelihood) with and without the polygenic score. For the
13 models including polygenic score-by-fibronectin concentration interaction terms, I
14 compared models with and without these interaction terms (but including the main
15 effects of polygenic score and fibronectin concentration). Second, to assess the
16 contribution of individual donors to the observed associations, I ran leave-one-donor-
17 out models for all models without interaction terms. Finally, I assessed the
18 importance of the analytical choices made in coding and including certain covariates
19 on the statistically significant finding reported. Specifically, I varied the number of
20 genomic principal components in the model (comparing models with 2 principal
21 components and 10 principal components), and I separately recoded the number of
22 cells in each clump as an ordinal variable (1 cell [single, reference], 2 or 3 cells
23 [multiple, no cells surrounded by other cells], 4 or more cells [multiple, cells
24 surrounded by other cells]).

25

1 Multiple testing correction

2 In total, I assessed the association of nine non-independent polygenic scores
3 with three non-independent phenotypes. To assess the effective number of tests
4 incurred, I performed principal component analysis on the pairwise correlation
5 matrices of the polygenic scores and the phenotypes separately. I defined the
6 effective number of tests incurred as the number of principal components required to
7 account for 99.5% of the variance in the correlation matrices. This resulted in 12
8 effective tests in total (6 effectively independent polygenic scores, 2 effectively
9 independent phenotypes; Supplementary Table 4). As such, statistical significance
10 was set at $p < 4.16 \times 10^{-3}$ (i.e., $0.05/12$).

11

12 Results

13 Power analyses

14 Power analyses examined the power of different numbers of donors to detect
15 significant associations between a hypothetical cellular phenotype and the polygenic
16 scores used in this analysis. Results for the cross-psychiatric polygenic score using
17 all variants with $p < 1$ (Cross-psychiatric₁) are described here (Figure 2,
18 Supplementary Table 3). Results for other polygenic scores are described in the
19 Supplementary Results (Supplementary Table 3, Supplementary Figures 2-4).

20 At the current donor number ($n = 60$), analyses only have >80% power when
21 the genetic covariance between Cross-psychiatric₁ and the cellular phenotype was
22 0.26 or greater. This could correspond to a genetic correlation of 0.7 when the
23 cellular phenotype has a SNP-based heritability of 0.55 (Figure 2a, Supplementary
24 Table 3).

Figure 2

Figure 2: Power (y axis) to detect a genetic relationship between Cross-psychiatric₁ and a cellular phenotype with a common genetic component of varying size (coloured lines) at different values of genetic correlation (x-axis), for differing values of N:
a) 60 [not accounting for multiple measurements],
b) 850 [lower estimate of effective N],
c) 2435 [higher estimate of effective N].

1 However, this does not take into account the potential power gain from
2 measuring multiple cells from the same donor. Taking this into account yields
3 estimates of effective n ranging 850-2435. At these sample sizes, analyses have
4 >80% power when the genetic covariance between Cross-psychiatric₁ and the
5 cellular phenotype was 0.04-0.07 or greater. This corresponds to (for example) a
6 genetic correlation of 0.15-0.2 when the cellular phenotype has a SNP-based
7 heritability of 0.55, or a genetic correlation of 0.3-0.5 when the SNP-based heritability
8 is 0.1 (Figures 2b, 2c, Supplementary Table 3).

9

10 Analyses without interactions

11 I assessed the association of each polygenic score with morphological
12 variability between cells. Across 27 analyses, one polygenic score was significantly
13 associated ($p < 4.16 \times 10^{-3}$) with a cell morphological phenotype (Cross-psychiatric₁
14 associated with cell area, $B = 0.085$, $p = 3.59 \times 10^{-3}$; Table 1; full models in
15 Supplementary Table 5). This association persisted in all sensitivity analyses
16 (Supplementary Results; Supplementary Table 5). As expected, fibronectin
17 concentration had a strong effect on all cell morphological phenotypes (absolute B
18 0.124-0.744, $p < 10^{-10}$; Supplementary Table 5). P-values obtained from the
19 likelihood ratio test were consistent with those using Satterthwaite's approximation
20 (Supplementary Table 5).

Table 1

Polygenic Score	Area Beta	Area SE	Area P	Roundness Beta	Roundness SE	Roundness P	Width:Length Beta	Width:Length SE	Width:Length P
Cross-psychiatric _{0.001}	0.0068	0.0281	0.810	-0.0261	0.0345	0.454	-0.0210	0.0212	0.327
Cross-psychiatric ₁	0.0845	0.0277	0.00359	-0.1026	0.0345	0.00441	-0.0425	0.0223	0.0617
Cross-psychiatric _{23andMe}	0.0100	0.0297	0.737	-0.0305	0.0365	0.407	-0.0218	0.0224	0.335
Schizophrenia _{0.05}	-0.0131	0.0295	0.660	0.0171	0.0364	0.640	-0.0128	0.0224	0.572
Schizophrenia ₁	-0.0008	0.0263	0.975	0.0099	0.0323	0.761	-0.0140	0.0199	0.486
BMI _{0.001}	-0.0186	0.0330	0.577	0.0033	0.0408	0.935	-0.00198	0.0252	0.938
BMI ₁	-0.0222	0.0299	0.461	-0.0331	0.0367	0.372	-0.0377	0.0223	0.0972
Height _{0.001}	0.0644	0.0301	0.0371	-0.0592	0.0379	0.124	-0.0323	0.0236	0.176
Height ₁	0.0467	0.0293	0.116	-0.0585	0.0361	0.111	-0.0185	0.0227	0.419

Table 1: Main effects of polygenic scores on cellular phenotypes. Polygenic scores are referenced as Trait_{Threshold}. Betas are on the scale of standard deviation changes in phenotype for one standard deviation change in polygenic score. Significant associations shown in bold ($p < 0.00416$). All values drawn from multiple linear regressions, full models in Supplementary Table X.

Analyses with interactions between polygenic scores and fibronectin concentration

I then assessed how each polygenic score altered the known effect of fibronectin concentration on cellular morphology. Four significant interaction terms were observed between fibronectin concentration and polygenic scores (Figure 3; Table 2; Supplementary Table 6). Of most interest is the interaction between Cross-psychiatric₁ and fibronectin concentration of 5µg/mL on cell area ($B = 0.052$, $p = 5.61 \times 10^{-4}$; Figure 3a). This suggests the effect of Cross-psychiatric₁ should be interpreted in the context of differing fibronectin concentrations. Specifically, the effect of plating cells on 5µg/mL fibronectin (compared to a suboptimal concentration of 1µg/mL) on increased cell area is greater in cells with higher Cross-psychiatric₁ polygenic scores. Further significant interactions were seen in the analysis of cell width-to-length ratio, between a fibronectin concentration of 25µg/mL and BMI₁ ($B = -0.037$, $p = 2.70 \times 10^{-3}$; Figure 3b), Height_{0.01} ($B = 0.057$, $p = 5.65 \times 10^{-6}$; Figure 3c), and Height₁ ($B = 0.050$, $p = 3.70 \times 10^{-5}$; Figure 3d) respectively. In the absence of a main effect of these polygenic scores on cell width-to-length ratio, and the weaker effect of fibronectin on cell width-to-length ratio compared to cell area, these results are harder to interpret than the interaction described above. P-values obtained from the likelihood ratio test were consistent with those using Satterthwaite's approximation (Supplementary Table 6).

Figure 3

Figure 3: Significant interactions between polygenic scores and fibronectin concentrations.

Lines reflect the relationship between each polygenic score and the phenotype from the relevant model. Points reflect the fitted value of the phenotype from the relevant model, averaged for each value of the polygenic score (that is, for all cells from a given donor).

- a) Cross-psychiatric₁ and fibronectin concentration of 5 $\mu\text{g/mL}$ on cell area,
- b) BMI₁ and fibronectin concentration of 25 $\mu\text{g/mL}$ on cell width-to-length ratio,
- c) Height_{0.01} and fibronectin concentration of 25 $\mu\text{g/mL}$ on cell width-to-length ratio,
- d) Height₁ and fibronectin concentration of 25 $\mu\text{g/mL}$ on cell width-to-length ratio.

Abbreviations: PGS = polygenic score, BMI = body mass index.

Table 2

Polygenic Score	Fibronectin conc	Area Beta	Area SE	Area P	Roundness Beta	Roundness SE	Roundness P	Width:Length Beta	Width:Length SE	Width:Length P
Cross-psychiatric _{0.001}	5	0.0213	0.0130	0.103	-0.0176	0.0144	0.221	-0.0136	0.0104	0.190
Cross-psychiatric _{0.001}	25	0.0235	0.0130	0.071	-0.0124	0.0144	0.390	-0.0075	0.0104	0.469
Cross-psychiatric ₁	5	0.0519	0.0150	5.61x10⁻⁴	-0.0138	0.0166	0.406	0.0023	0.0119	0.847
Cross-psychiatric ₁	25	0.0218	0.0150	0.147	0.0073	0.0166	0.659	0.0220	0.0119	0.0645
Cross-psychiatric _{23andMe}	5	0.0162	0.0135	0.230	-0.0064	0.0149	0.666	0.0007	0.0108	0.945
Cross-psychiatric _{23andMe}	25	0.0246	0.0135	0.0691	-0.0014	0.0149	0.923	0.0075	0.0108	0.488
Schizophrenia _{0.05}	5	0.0082	0.0127	0.518	-0.0092	0.0141	0.513	-0.0182	0.0101	0.0724
Schizophrenia _{0.05}	25	0.0009	0.0127	0.944	-0.0053	0.0141	0.706	-0.0182	0.0101	0.0722
Schizophrenia ₁	5	0.0166	0.0130	0.201	-0.0103	0.0144	0.476	-0.0132	0.0104	0.203
Schizophrenia ₁	25	0.0074	0.0130	0.569	-0.0184	0.0144	0.706	-0.0097	0.0103	0.351
BMI _{0.001}	5	-0.0140	0.0157	0.374	0.0002	0.0173	0.992	-0.0134	0.0124	0.280
BMI _{0.001}	25	-0.0050	0.0157	0.753	-0.0157	0.0173	0.365	-0.0259	0.0124	0.0370
BMI ₁	5	-0.0299	0.0155	0.0535	0.0006	0.0171	0.973	-0.0166	0.0122	0.174
BMI ₁	25	-0.0304	0.0155	0.0501	-0.0203	0.0171	0.234	-0.0366	0.0122	0.00270
Height _{0.001}	5	0.0336	0.0157	0.0327	0.0052	0.0172	0.764	0.0353	0.0125	0.00458
Height _{0.001}	25	0.0236	0.0157	0.133	0.0268	0.0173	0.121	0.0566	0.0125	5.65x10⁻⁶
Height ₁	5	0.0267	0.0151	0.0782	-0.0017	0.0166	0.918	0.0278	0.0120	0.0203
Height ₁	25	0.0156	0.0152	0.304	0.0199	0.0167	0.231	0.0495	0.0120	3.70x10⁻⁵

Table 2: Effects of polygenic score-by-fibronectin concentration interaction on cellular phenotypes. Polygenic scores are referenced as Trait_{Threshold}. Fibronectin concentrations are on the scale of µg/mL, and are compared to a suboptimal concentration of 1µg/mL. Betas are on the scale of standard deviation changes in phenotype for one standard deviation change in polygenic score at each fibronectin concentration, above and beyond the effect of the polygenic score at the reference fibronectin concentration. Significant associations shown in bold (p < 0.00416). All values drawn from multiple linear regressions, full models in Supplementary Table Y.

1 **Discussion**

2

3 The genetic component of population-level variance in complex disorders is
4 primarily polygenic⁴. Understanding the biological effects of this polygenic
5 component will be a challenging but necessary step in understanding the aetiology of
6 complex disorders. Human induced pluripotent stem cells offer a potentially valuable
7 model for studying such effects. In this paper, I have shown that applying polygenic
8 scores to databases of iPSCs is feasible but requires overcoming a number of
9 technical challenges. I discuss each of these challenges below.

10 The primary challenge facing polygenic score analyses in iPSCs is donor
11 number. The power analyses I present show that measuring multiple cells from the
12 same donor provides sufficient power for informative polygenic analyses, despite the
13 relatively small number of donors assessed. The power gain will vary between
14 different datasets and analysed phenotypes, dependent on the average correlation
15 between multiple measurements from the same individual³⁸. In these analyses I
16 estimate that measuring approximately 25,000 cells per donor on average increased
17 the power of polygenic score analyses by a factor of 14-40 relative to the donor
18 number alone. This increase in power is sufficient to provide 80% power for plausible
19 values of both the heritability of the cellular phenotype, and of the genetic correlation
20 with complex traits. As such, these analyses show that currently available iPSC
21 datasets are powered for the polygenic exploration of complex trait phenotypes.

22 The power analyses presented are generalisable to polygenic score analyses
23 examining any repeated measure from the same individual, not just measures of cell
24 morphology. Furthermore, they are generalisable to other model systems, including
25 iPSC-derived neurons, which are better models for behavioural phenotypes. The
26 genetic covariance between behavioural polygenic scores and neuronal phenotypes

1 is likely to be small. Nonetheless, the power analyses suggest that a dataset of
2 derived neuron lines from 10s-100s of donors, with 1000s of neurons per line, would
3 be a practical target for analysing polygenic scores derived from GWAS of
4 behavioural phenotypes.

5 The creation of the HipSci dataset was a collaborative effort across multiple
6 laboratories, taking a considerable amount of time and investment, and intended as
7 a community resource ³⁹. Creating similarly sized and measured datasets will be a
8 major undertaking. However, high-throughput methods for the rapid assessment of
9 cellular phenotypes are being developed. For example, the Census-seq cell village
10 approach allows cells from multiple donors to be plated and assayed together on a
11 single dish ⁴⁰. In addition to enabling large-scale cell-level analyses, individual donor
12 genetic associations with cellular phenotypes can be inferred computationally using
13 this method, enabling polygenic score analyses. Building large scale cell datasets is
14 therefore technically feasible, and further technical developments may reduce the
15 time and costs required.

16

17 One of the exploratory polygenic score analyses presented in this paper was
18 significant, namely the association of higher Cross-psychiatric₁ polygenic scores with
19 greater cell area (the association of higher Cross-psychiatric₁ with cell roundness
20 was also sizable, and close to statistical significance). This association was robust to
21 sensitivity analyses, suggesting it was not an artefact of the analytical method. The
22 interaction between Cross-psychiatric₁ and the plate fibronectin concentration was
23 also significantly associated with cell area. Fibronectin is an extracellular matrix
24 protein, the concentration of which influences how well the cell can adhere to the
25 plate surface, and therefore affects the cell area ²². The effect of plating on an

1 optimal concentration of fibronectin (compared to a suboptimal concentration) was
 2 greater in cells with a high polygenic score than in cells with a low polygenic score.
 3 As such, the polygenic score may (in part) be capturing variability in cell-extracellular
 4 matrix adhesion. Cell-extracellular matrix adhesion is fundamental to cell crawling
 5 and axon guidance^{41–43}, which has also been implicated in psychiatric disorders
 6^{15,19–21}. The composition of the extracellular matrix, and increased adhesion of cells
 7 to the matrix, has been shown to increase neurite extension in neurons^{42,43}, to
 8 stimulate gyrification in the developing human neocortex⁴⁴, and to promote
 9 branching and migration of neurons^{42,45}, among other roles⁴².

10 Variability in cell-extracellular matrix adhesion affecting neuron migration may
 11 be one among many general mechanisms influencing the neurobiology of psychiatric
 12 disorders. This might explain why there was no significant association between
 13 schizophrenia polygenic scores and cell area. There may have been no association
 14 with the schizophrenia polygenic score because it is dominated by other (potentially
 15 schizophrenia specific) mechanisms. This fits with the results of the latest
 16 schizophrenia GWAS from the Psychiatric Genomics Consortium, where genes
 17 associated with schizophrenia were not enriched for a role in neuron migration,
 18 unlike genes associated with cross-psychiatric disorder risk in the cross-psychiatric
 19 disorder GWAS^{15,20}. The association with cell area was seen only with the full
 20 polygenic score ($p < 1$) for cross-psychiatric disorder risk, not the score limited to
 21 variants with stronger evidence for association ($p < 0.001$). One explanation for this
 22 may be that individual genetic associations with cross-psychiatric disorder risk that
 23 act through effects on cell-extracellular matrix adhesion are likely to be very small.
 24 Accordingly, they may not be estimated accurately enough to be enriched in the
 25 more limited score from the cross-psychiatric disorder GWAS, and so the association

1 between cross-psychiatric disorder risk and cell area is only observed in the full
2 score ^{6,15}.

3 Even when only considering the baseline fibronectin concentration (1 µg/mL),
4 there was still an effect of the cross-psychiatric disorder polygenic score on cell area,
5 suggesting the effect cannot be purely driven by cell-extracellular matrix interactions.
6 Other possibilities may include an effect on cell morphogenic processes within the
7 cell, such as altering the action or composition of the cytoskeleton ^{46,47}. However,
8 given that the association between the cross-psychiatric disorder polygenic score
9 and cell area is close to the threshold for statistical significance, I cannot conclude
10 strongly it is stable and generalisable. As sufficiently sized iPSC datasets emerge, it
11 would be of interest to replicate this finding.

12

13 Certain limitations of this work should be taken into account. Broad
14 phenotypic information on the donors was not available. As such, I cannot exclude
15 that donors may have been phenotypic outliers for the polygenic score traits. For
16 example, if a donor had schizophrenia, they may have a higher polygenic score for
17 psychiatric disorders than the population average. Similarly, because the participants
18 are anonymous, it is not known whether they contributed to the GWAS from which
19 the polygenic scores were derived. Sample overlap would severely bias the results
20 of the study ¹³. However, the results were stable in leave-one-donor out analyses,
21 suggesting no single donor (such as one that was a phenotypic outlier, or who might
22 have participated in one of the GWAS) strongly influenced the results. An
23 unmitigated limitation is that the study was restricted only to individuals of European
24 ancestries, due to the availability of cell data. Results from polygenic scores are
25 often poorly translated across different ancestries, which limits the broader

1 generalisability of these findings⁴⁸. Including donors from diverse ancestries should
2 be a key consideration in generating new cellular datasets.

3 In summary, I have demonstrated that polygenic score analysis is feasible in
4 existing datasets of iPSCs, and that this holds considerable promise for examining
5 the polygenic effects of complex disorders in cellular models. Emergent cell biology
6 techniques such as the Census-seq approach provide a mechanism to achieve the
7 sample sizes needed for the analyses described in this paper. Coupled with further
8 advances in the derivation of diverse cell types from iPSCs, polygenic scoring in cell
9 lines has the potential to be a powerful technique for the assessment of cellular
10 proxies of complex traits.

11

12 **Data accessibility**

13 All data analysed in this project is available via the Human Induced Pluripotent Stem
14 Cell Initiative project website <http://www.hipsci.org/>. Imaging data is available open
15 access from the HipSci website. Genotype data are available under controlled
16 access (accessed under a data transfer agreement with the Wellcome Sanger
17 Institute) and open access conditions. Controlled access data is hosted at the
18 European Genome-phenome Archive at the European Bioinformatics Institute under
19 accession number EGAD00010001147. Open-access genotypes are hosted at the
20 European Nucleotide Archive at the European Bioinformatics Institute under project
21 ID PRJEB11750. Summary statistics were downloaded from the GIANT consortium
22 ([https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_d](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files)
23 [ata_files](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files)), the Walters group (<https://walters.psychm.cf.ac.uk/>), and the Psychiatric
24 Genomics Consortium (<https://www.med.unc.edu/pgc/download-results/cd/>).

25

1 Code underlying the analyses presented will be made available on publication at
2 <https://github.com/JoniColeman/hipsciPGS>.

3

4 **Acknowledgements**

5 I thank the investigators of the Human Induced Pluripotent Stem Cell Consortium for
6 their work in creating the dataset analysed in this publication. I am indebted to
7 Davide Danovi, Alessandra Vigilante, Ewan Carr and Laura Blackie for their advice
8 and discussion of this work.

9

10 **Funding and Disclosures**

11 I declare no funding or conflicts of interest associated with this work.\

References

1. Hirschhorn, J.N., and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* 6, 95–108.
2. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012.
3. Cannon, T.D., and Keller, M.C. (2006). Endophenotypes in the genetic analyses of mental disorders. *Annu. Rev. Clin. Psychol.* 2, 267–290.
4. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22.
5. Flint, J., and Mackay, T.F.C. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* 19, 723–733.
6. Sullivan, P.F., and Geschwind, D.H. (2019). Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* 177, 162–183.

7. PsychENCODE Consortium (2018). Revealing the brain's molecular architecture. *Science* 362, 1262–1263.
8. Fernando, M.B., Ahfeldt, T., and Brennand, K.J. (2020). Modeling the complex genetic architectures of brain disease. *Nat. Genet.* 52, 363–369.
9. Ebben, J.D., Zorniak, M., Clark, P.A., and Kuo, J.S. (2011). Introduction to induced pluripotent stem cells: advancing the potential for personalized medicine. *World Neurosurg.* 76, 270–275.
10. Efthymiou, A.G., Chen, G., Rao, M., Chen, G., and Boehm, M. (2014). Self-renewal and cell lineage differentiation strategies in human embryonic stem cells and induced pluripotent stem cells. *Expert Opin. Biol. Ther.* 14, 1333–1344.
11. Vadodaria, K.C., Jones, J.R., Linker, S., and Gage, F.H. (2020). Modeling brain disorders using induced pluripotent stem cells. *Cold Spring Harb. Perspect. Biol.* 12,.
12. International Schizophrenia Consortium, Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752.
13. Choi, S.W., Mak, T.S.-H., and O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* 15, 2759–2772.
14. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* 50, 381–389.
15. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, P.H., Anttila, V., Won, H., Feng, Y.-C.A., Rosenthal, J., Zhu, Z., Tucker-Drob, E.M., Nivard, M.G., Grotzinger, A.D., et al. (2019). Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. 528117.
16. Yengo, L., Sidorenko, J., Kempner, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* 27, 3641–3649.
17. Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F.P., Culley, O.J., et al. (2017). Common genetic

- variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375.
18. Blainey, P., Krzywinski, M., and Altman, N. (2014). Points of significance: replication. *Nat. Methods* **11**, 879–880.
19. Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Børte, S., Bryois, J., et al. (2020). Genome-wide association study of over 40,000 bipolar disorder cases provides novel biological insights. *MedRxiv*.
20. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke, S., Walters, J.T., and O'Donovan, M.C. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*.
21. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681.
22. Vigilante, A., Laddach, A., Moens, N., Meleckyte, R., Leha, A., Ghahramani, A., Culley, O.J., Kathuria, A., Hurling, C., Vickers, A., et al. (2019). Identifying Extrinsic versus Intrinsic Drivers of Variation in Cell Behavior in Human iPSC Lines from Healthy Donors. *Cell Rep.* **26**, 2078–2087.e3.
23. Leha, A., Moens, N., Meleckyte, R., Culley, O.J., Gervasio, M.K., Kerz, M., Reimer, A., Cain, S.A., Streeter, I., Folarin, A., et al. (2016). A high-content platform to characterise human induced pluripotent stem cell lines. *Methods* **96**, 85–96.
24. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7.
25. Coleman, J.R.I., Euesden, J., Patel, H., Folarin, A.A., Newhouse, S., and Breen, G. (2016). Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Brief. Funct. Genomics* **15**, 298–304.
26. Core Team, R., and Others (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available.
27. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283.

28. Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448.
29. Durbin, R. (2014). Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* **30**, 1266–1272.
30. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. *Nature* **526**, 68–74.
31. Choi, S.W., and O'Reilly, P.F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**,.
32. Palla, L., and Dudbridge, F. (2015). A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. *Am. J. Hum. Genet.* **97**, 250–259.
33. Kuznetsova, A., Brockhoff, P., and Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, Articles* **82**, 1–26.
34. Keller, M.C. (2014). Gene $\times$ environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol. Psychiatry* **75**, 18–24.
35. Yzerbyt, V.Y., Muller, D., and Judd, C.M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *J. Exp. Soc. Psychol.* **40**, 424–431.
36. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer).
37. Long, J.A. (2019). *interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*.
38. Faes, C., Molenberghs, G., Aerts, M., Verbeke, G., and Kenward, M.G. (2009). The Effective Sample Size and an Alternative Small-Sample Degrees-of-Freedom Method. *Am. Stat.* **63**, 389–399.
39. Streeter, I., Harrison, P.W., Faulconbridge, A., The HipSci Consortium, Flicek, P., Parkinson, H., and Clarke, L. (2017). The human-induced pluripotent stem cell initiative-data resources for cellular genetics. *Nucleic Acids Res.* **45**, D691–D697.

40. Mitchell, J.M., Nemesh, J., Ghosh, S., Handsaker, R.E., Mello, C., Meyer, D., Raghunathan, K., de Rivera, H., Tegtmeier, M., Hawes, D., et al. (2020). Mapping genetic effects on cellular phenotypes with “cell villages”. 2020.06.29.174383.
41. Walsh, F.S., and Doherty, P. (1997). Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. *Annu. Rev. Cell Dev. Biol.* 13, 425–456.
42. Long, K.R., and Huttner, W.B. (2019). How the extracellular matrix shapes neural development. *Open Biol* 9, 180216.
43. Letourneau, P.C., Condic, M.L., and Snow, D.M. (1992). Extracellular matrix and neurite outgrowth. *Curr. Opin. Genet. Dev.* 2, 625–634.
44. Long, K.R., Newland, B., Florio, M., Kalebic, N., Langen, B., Kolterer, A., Wimberger, P., and Huttner, W.B. (2018). Extracellular Matrix Components HAPLN1, Lumican, and Collagen I Cause Hyaluronic Acid-Dependent Folding of the Developing Human Neocortex. *Neuron* 99, 702–719.e6.
45. Chai, X., Fan, L., Shao, H., Lu, X., Zhang, W., Li, J., Wang, J., Chen, S., Frotscher, M., and Zhao, S. (2015). Reelin Induces Branching of Neurons and Radial Glial Cells during Corticogenesis. *Cereb. Cortex* 25, 3640–3653.
46. Courtot, A.-M., Magniez, A., Oudrhiri, N., Féraud, O., Bacci, J., Gobbo, E., Proust, S., Turhan, A.G., and Bennaceur-Griscelli, A. (2014). Morphological analysis of human induced pluripotent stem cells during induced differentiation and reverse programming. *Biores. Open Access* 3, 206–216.
47. Närvä, E., Stubb, A., Guzmán, C., Blomqvist, M., Balboa, D., Lerche, M., Saari, M., Otonkoski, T., and Ivaska, J. (2017). A strong contractile actin fence and large adhesions direct human pluripotent colony morphology and adhesion. *Stem Cell Rep.* 9, 67–76.
48. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591.