Title	Altered gene regulation as a candidate mechanism by which ciliopathy gene SDCCAG8 contributes to schizophrenia and cognitive function
Author(s)	Flynn, Mairéad; Whitton, Laura; Donohoe, Gary; Morrison, Ciaran G.; Morris, Derek W.
Publication Date	2019-12-23
Publication Information	Flynn, Mairéad, Whitton, Laura, Donohoe, Gary, Morrison, Ciaran G, & Morris, Derek W. (2019). Altered gene regulation as a candidate mechanism by which ciliopathy gene SDCCAG8 contributes to schizophrenia and cognitive function. Human Molecular Genetics, 29(3), 407-417. doi:10.1093/hmg/ddz292
Publisher	Oxford University Press (OUP)
Link to publisher's version	https://doi.org/10.1093/hmg/ddz292
Item record	http://hdl.handle.net/10379/16274
DOI	http://dx.doi.org/10.1093/hmg/ddz292

Altered gene regulation as a candidate mechanism by which ciliopathy gene *SDCCAG8* contributes to schizophrenia and cognitive function

Mairéad Flynn[a,b], Laura Whitton[a], Gary Donohoe[a], Ciaran G. Morrison[b,*] and Derek W. Morris[a,*].

[a]*Cognitive Genetics and Cognitive Therapy Group, Neuroimaging and Cognitive Genomics (NICOG) Centre, School of Psychology and Discipline of Biochemistry,* [b]*Centre for Chromosome Biology, School of Natural Sciences, National University of Ireland Galway, Ireland.*

***Contributed equally to this work and joint corresponding authors**: Ciaran.Morrison@nuigalway.ie; Derek.Morris@nuigalway.ie

Abstract

Mutations in genes that encode centrosomal/ciliary proteins cause severe cognitive deficits, while common SNPs in these genes are associated with schizophrenia (SZ) and cognition in GWAS. The role of these genes in neuropsychiatric disorders is unknown. The ciliopathy gene *SDCCAG8* is associated with SZ and educational attainment (EA). Genome editing of *SDCCAG8* caused defects in primary ciliogenesis and cilium-dependent cell signalling. Transcriptomic analysis of SDCCAG8-deficient cells identified differentially expressed genes that are enriched in neurodevelopmental processes such as generation of neurons and synapse organization. These processes are enriched for genes associated with SZ, human intelligence (IQ) and EA. Phenotypic analysis of SDCCAG8-deficent neuronal cells revealed impaired migration and neuronal differentiation. These data implicate ciliary signalling in the aetiology of SZ and cognitive dysfunction. We found that centrosomal/ciliary genes are enriched for association with IQ, suggesting altered gene regulation as a general model for neurodevelopmental impacts of centrosomal/ciliary genes.

Introduction

The centrosome is the key microtubule organising centre of animal somatic cells (1). Centrosomes play a crucial role in brain development, influencing cell shape, polarity, motility, and division (2). The older of the two barrel-shaped centrioles that form the core around which the pericentriolar material assembles can serve as the base of the primary cilium, an antenna-like structure with sensory and signal transduction functions that are essential for brain development (3). Mutations that affect ciliary function cause ciliopathies, heterogeneous developmental or degenerative disorders that affect multiple organs, including the brain (4).

Perturbation of centrosomal function in cortical development may contribute to neurodevelopmental disorders such as schizophrenia (SZ) (5), where cognitive deficits (poor memory, attention, and IQ) are a key factor for explaining disability, leading to significant unemployment, homelessness and social isolation (6). SZ is highly heritable (7) and significant genetic overlap between SZ and cognitive ability (8) means that some genes contribute to both phenotypes. Genome-wide association studies (GWAS) have now identified hundreds of loci for SZ (9) and cognition phenotypes such as human intelligence (IQ; (10)) and educational attainment (EA; (11)). The molecular mechanisms by which associated genes contribute to SZ risk and cognitive function are not well understood. However, the likelihood is that genetic variation within multiple biological pathways contributes to both illness and lower cognition function. In this context, it is striking that mutations in 77 of 87 genes that encode ciliary and centrosomal proteins associated with human ciliopathies result in neurodevelopmental or cognitive deficits (12). Variable neuropsychiatric phenotypes have been reported in patients with ciliopathies (13). A study of 41 risk genes for SZ and other neuropsychiatric disorders found that for 20 of these, knockdown of gene expression resulted in a ciliary phenotype (14).

Here we explore the hypotheses that genes that function at the centrosome contribute to SZ and cognitive function and that ciliary dysfunction might impact potential neurodevelopmental mechanisms that could contribute to SZ risk or cognitive ability. We use reverse genetic analyses of serologically defined colon cancer antigen 8 (*SDCCAG8*), a ciliopathy gene where the same common genetic variants are associated with SZ and EA at genome-wide significant levels to test these potential mechanisms. SDCCAG8 deficiency causes defects in cell migration and cilium-dependent cell

signalling, leading to altered expression of genes within neurodevelopmental pathways that are enriched for genes associated with SZ, IQ and EA. The pathway analysis results prompted experiments that revealed impaired migration and neuronal differentiation in SDCCAG8-deficent cells. These observations implicate centrosomal/ciliary gene regulation as a candidate pathway in neuropsychiatric disorders. Interestingly, we find that genes that encode proteins with centrosomal functions are enriched for association with IQ, suggesting a possible broad relevance of this model.

Results and Discussion

SDCCAG8 as a candidate centrosomal/ciliary gene associated with SZ and EA

We chose the centrosomal-/ciliary-associated gene, *SDCCAG8*, as a test for our hypotheses because of *SDCCAG8* mutations in ciliopathies and reported associations with neurodevelopmental phenotypes. *SDCCAG8* is mutated in several developmental disorders, notably the ciliopathy, Bardet-Biedl syndrome (15-17), the cystic kidney and retinal disorder, Senior-Loken syndrome, and Leber congenital amaurosis (16). Patients with *SDCCAG8* mutations present with intellectual disability and sometimes seizures (16). Elements of obsessive compulsive behaviour can be seen in some patients with Bardet-Biedl syndrome, others have more severe behavioural phenotypes and develop autism or psychosis (13). SDCCAG8 has been shown to regulate the centrosomal accumulation of pericentriolar material and neuronal migration in the developing mouse cortex (18). Furthermore, analyses of *Sdccag8* gene-trap mice have implicated SDCCAG8 in the regulation of the DNA damage response, at least in kidney (19). GWAS analyses have implicated *SDCCAG8* in SZ (20, 21). SNPs surpassing genome-wide significance, a statistical threshold ($P<5x10^{-8}$) used to differentiate true positives from false positives in GWAS, have been reported at *SDCCAG8* in the most recent and largest studies of SZ (9) and EA (11). Of these SNPs, rs10803138 (associated with SZ) and rs2992632 (associated with EA) are in high linkage disequilibrium (LD; r^2=0.81), indicating that the same genetic variants may be contributing to both phenotypes. These two SNPs are both reported to be associated, although not genome-wide significant, with IQ (P<0.001; (10)). A caveat is that the association signals at *SDCCAG8* extend to the neighbouring AKT serine/threonine kinase 3 (*AKT3*) gene. The 3' ends of *SDCCAG8* and *AKT3* genes overlap on opposite DNA strands at chr1:243,488,233-243,500,091 (hg38). For SZ risk SNP rs10803138, the G allele that is associated with increased SZ risk is also associated with reduced expression of *SDCCAG8* in three of thirteen brain regions tested in the Genotype-Tissue Expression (GTEx) expression quantitative trait locus (eQTL) database (https://gtexportal.org/) (cerebellar hemisphere, P=0.0003; anterior cingulate cortex (BA24), P=0.001; caudate (basal ganglia), P=0.0014; **Supplementary Table 1**). This possibly suggests that lower SDCCAG8 expression is associated with increased SZ risk. A caveat is that these data are based on small numbers of *post mortem* brain samples and more definitive functional work would be required to determine the effect of associated SNPs on SDCCAG8 function.

Intact centriole duplication in the absence of SDCCAG8

We used cell lines to explore the roles of SDCCAG8 and how it contributes to neurodevelopmental phenotypes. We first determined the localisation of endogenous SDCCAG8 at the centrosome/cilium in cycling cells. Using immunofluorescence microscopy, we found that it consistently localised to the centrosome, on both centrioles, at all stages of the cell cycle (**Figure 1A**). However, in 76% of cells in G1 and 62% of cells in metaphase, preferential loading of SDCCAG8 onto the daughter centriole was observed.

We then used CRISPR-Cas9 genome editing to disrupt exon 3 of *SDCCAG8* in the immortalized, non-transformed hTERT-RPE1 cell line (hereafter RPE1), an excellent model for studying ciliogenesis, and the immature neuronal cell line SHSY5Y, a useful model for neuronal differentiation. Immunoblot screening of candidates yielded two RPE1 and six SHSY5Y clones that lacked detectable SDCCAG8. Genomic PCR and DNA sequencing were used to confirm that *SDCCAG8* disruption generated premature stop codons (**Supplementary Figure 1**). Western blotting confirmed the absence of SDCCAG8, and stable expression of full-length SDCCAG8 was used to obtain rescued RPE1 clones (**Figure 1B, C**). Flow cytometry analysis revealed that the *SDCCAG8*$^{-/-}$ clones generated on the RPE1 background were aneuploid, although re-expression of SDCCAG rescued ciliogenesis, cilium length and ciliary signalling. SHSY5Y clones showed a normal cell cycle distribution compared to wildtype cells. Centriolar proteins (centrin and centrobin), centriolar appendage proteins (CEP164), centrosomal linker proteins (CNAP-1) and centriolar satellites (CEP290 and PCM-1) all localized normally in the absence of SDCCAG8 (**Figure 1D**), although we noted a decline in pericentrosomal pericentrin levels to 76% of the wild-type controls, consistent with observations made in knockdown experiments in murine neurons (18). Proliferative analysis showed no significant impact on cell doubling times in the absence of SDCCAG8 (**Figure 1E**). We next tested whether DNA damage-induced centrosome amplification (CA), a cellular response that integrates DNA damage signalling and centriole duplication (22), was affected by the loss of SDCCAG8. As shown in **Figure 1F**, **1G**, CA induced by exposure to 5Gy ionising radiation was not significantly affected by the absence of SDCCAG8. Together, these data indicate that SDCCAG8 is not required for centrosome duplication or amplification.

Ciliation and ciliary dependent signalling are defective in SDCCAG8-deficient cells

Previous data have shown that knockdown or loss of SDCCAG8 reduces ciliation frequency (23). Therefore, we examined SDCCAG8 localisation after serum starvation, which induces ciliogenesis in

RPE1 cells. We observed a daughter centriole localisation, as in asynchronous cells, but a proportion of SDCCAG8 was also found in a discrete band near the base of the primary cilium and co-localised with the established transition zone (TZ) marker, CEP290 (**Figure 2A**). In the mouse retina, SDCCAG8 has previously been shown to localise to the TZ of photoreceptor cilia, although no direct co-localisation with CEP290 was reported (16). We found that SDCCAG8-deficient RPE1 cells had a marked defect in forming primary cilia, which was rescued by re-expression of SDCCAG8 (**Figure 2B, 2C**). The *SDCCAG8*$^{-/-}$ SHSY5Y cells also demonstrated reduced ciliation capacity (**Figure 2D, 2E**). The cilia that did form in SDCCAG8-deficient cells were also shorter than those in wild-type controls, a phenotype which was rescued by restoration of SDCCAG8 expression (**Figure 2F**). Exploring potential mechanisms by which depletion of SDCCAG8 reduces ciliary capacity, we found that removal of the CP110-CEP97 cap complex from the distal end of the centriole (24-26), the assembly of tau tubulin kinase 2 (TTBK2) at the centriole (27, 28), and the localisation of the intraflagellar transport protein IFT88 (29) to the axoneme were unaffected by the loss of SDCCAG8 (**Figure 2G**). We conclude that the defect in ciliogenesis seen in SDCCAG8-deficient cells is downstream of ciliary cap removal or membrane docking of the mother centriole, consistent with phenotypes observed in mouse gene-trap and RPE1 knockdown experiments (23).

We next investigated the signalling capacity of the residual cilia formed in SDCCAG8 depleted cells. The Hedgehog (Hh) pathway plays an important role in brain development and the best-characterized ligand, Sonic Hedgehog, has been shown to regulate axonal guidance (30). After treatment with Smoothened (SMO) agonist, translocation of SMO to the axoneme can be used as a readout of Hedgehog (Hh) signalling (31, 32). We saw a marked decrease in SMO translocation to the axoneme in SDCCAG8-deficient cells, which was rescued by re-expression of SDCCAG8 (**Figure 2H, 2I**), consistent with previous data from the mouse that implicate SDCCAG8 in Hh signalling (23). All together, these data show that SDCCAG8 is required for both efficient ciliogenesis and effective ciliary signalling.

Genes that are transcriptionally dysregulated in the absence of SDCCAG8 map to GO terms that are enriched for genes associated with SZ, IQ and EA

We then analysed global gene expression changes in SDCCAG8-deficient cells using RNA-seq. We focused on those protein-coding genes most dysregulated by the absence of SDCCAG8, which we defined as genes that had an absolute $\log_2$ fold change of three or greater and a Benjamini-Hochberg

adjusted p-value of 0.05 or less. Using this threshold, we identified 585 differentially expressed genes (DEGs) in RPE1 cells under baseline conditions, and 604 DEGs after serum starvation to induce ciliogenesis (**Supplementary Tables 2 and 3**). We observed 281 and 575 DEGs in SHSY5Y cells under baseline conditions and after serum starvation, respectively (**Supplementary Figure 2** and **Supplementary Tables 4 and 5**).

We performed gene ontology (GO) enrichment analysis of the DEGs to identify the molecular functions, biological pathways and cellular components that may be disrupted as a consequence of SDCCAG8 deficiency. The most enriched GO terms for DEGs from the SDCCAG8-deficient RPE1 cells, for both baseline conditions and after serum starvation, related to processes involved in extracellular matrix composition and structure (**Supplementary Tables 6 and 7**). In SHSY5Y cells, the most enriched GO terms related to synaptic signalling as well as nervous system development (**Supplementary Table 8**), although terms related to the cell-cycle progression were also enriched in the GO analysis of DEGs from the serum-starved SHSY5Y cells (**Supplementary Table 9**).

We next investigated if the GO terms enriched for the DEGs from the RPE1 cells or the SHSY5Y cells were also enriched for genes associated with SZ, IQ or EA using MAGMA gene-set analysis of GWAS data for these phenotypes. A gene-set analysis is a statistical method for simultaneously analysing GWAS results for multiple genes to test if variants in those genes jointly affect a phenotype. As there is considerable overlap between related GO terms, we used ancestral GO charts to identify the most enriched but non-overlapping GO terms from our analyses of DEGs (five GO terms from each cell line, highlighted in bold in **Supplementary Tables 6-9**). For RPE1 cells, those enriched GO terms were *collagen-containing extracellular matrix*, *extracellular matrix organization*, *skeletal system morphogenesis*, *cell differentiation* and *regulation of multicellular organismal process*. For SHSY5Y cells, those enriched GO terms were *trans-synaptic signalling*, *dendrite*, *generation of neurons*, *synapse organization* and *sister chromatid segregation*. After correcting for multiple testing, all SHSY5Y GO terms except *sister chromatid segregation* were significantly enriched for genes associated with SZ, IQ and EA (**Figure 3**). There were no significant enrichments for any RPE1 GO terms. Neuronally-expressed genes are a major contributor to the phenotypes tested here. It is possible that the enrichments detected for SZ, IQ and EA could be due to the SHSY5Y GO terms representing a subset of neuronally-expressed genes. To control for this we also conditioned our analyses on 'SHSY5Y expressed genes', which is a gene-set containing all genes for which

expression was detected in wild-type (WT) SHSY5Y cells in our RNA-seq experiment. All enrichments remained significant in these conditional analyses (**Supplementary Tables 10-12**). As a further control, all significantly enriched SHSY5Y GO terms were tested for enrichment in "control" GWAS datasets from non-neurodevelopmental diseases (Alzheimer's disease (AD), Crohn's disease (CD), stroke (STR), type 2 diabetes (T2D), coronary artery disease (CAD) and ulcerative colitis (UC)). As no significant results were detected (**Supplementary Table 13**), we conclude that the GO term enrichments observed for SZ, IQ and EA are not a property of polygenic phenotypes in general. Overall, these analyses indicate that the specific biological pathways and cellular components that are dysregulated in immature neurons upon depletion of SDCCAG8 contain genes that contribute to SZ risk and cognitive function in the general population. Similar pathways were not affected in the non-neuronal RPE1 cells, even though they show a ciliary defect similar to that seen in the SHSY5Y cells, supporting the cell-specific nature of the GO terms identified by the gene expression analysis.

SDCCAG8-deficent cells show impaired migration and neuronal differentiation

Given that the most dysregulated pathways in the absence of SDCCCAG8 included *generation of neurons*, we hypothesised that migration would be impaired in our SHSY5Y cells. Neuronal migration is a key feature of brain development. In some migrating neurons, the centrosome is positioned ahead of the nucleus, the organelle moves into the leading process first, followed rapidly by translocation of the nucleus into the leading process (33). Knockdown of *SDCCAG8* expression has been shown previously to impair neuronal migration in the developing cortex of mice (18). We assessed migration in SHSY5Y cells using transwell inserts between serum-free and serum-containing conditions. At 16 h after exposure to serum as an attractant, most cells remained within pores and had not completed migration to the underside of the insert, although by 24 h migration was complete and WT cells regained their morphology (**Figure 4A**). Colorimetric quantification with a microplate reader revealed that the migration of SDCCAG8-deficient cells was significantly defective (**Figure 4B**), consistent with previous findings in mouse (18) and with the impact of SDCCAG8 deficiency on gene expression. Cells derived from patients with SZ have been shown to have impairments in migration in multiple studies (34-37), suggesting that cell motility may be important in pathogenesis of the disorder.

Another GO term that was highly enriched for SDCCAG8 dysregulated genes was *neuron differentiation*. We therefore hypothesised that the transcriptional dysregulation caused by SDCCAG8

depletion may affect the cells' ability to differentiate. To test this idea, both WT and SDCCAG8-deficient SHSY5Y cells were differentiated with retinoic acid for 7 days. After 7 days, phase contrast images of the cells were taken to assess morphological appearance of the cells. Undifferentiated WT and *SDCCAG8*$^{-/-}$ SHSY5Y cells had a flat phenotype with few projections, while differentiated WT cells showed extensive and elongated neuron projections (**Figure 4C**). Cells lacking SDCCAG8 failed to undergo this morphological change and remained very flat with few projections. Furthermore, the formation of neurite projections, a measure of differentiation, was greatly reduced in SDCCAG8-deficient cells (**Figure 4D**). These data demonstrate a marked deficiency in neuronal differentiation capacity in the absence of SDCCAG8. This finding is consistent with data that demonstrate the importance of the centrosome in differential neurite formation (38) and with the impact of SDCCAG8 deficiency on gene expression. Induced pluripotent stem cell-based SZ case and control studies have previously demonstrated a delayed hippocampal differentiation of patient-derived cells (39).

Centrosome/centriole/ciliary genes as contributors to neurodevelopmental phenotypes

Our findings with SDCCAG8 suggest a general impact on neuronal gene expression and behaviour induced by the dysregulation of cilia. Therefore, we explored whether genes that encode centrosome/centriole/ciliary proteins were associated with neurodevelopmental phenotypes. We generated a "centrosomal gene-set" that contained 808 genes (listed in **Supplementary Table 14**) identified as having centrosomal localisation based on available databases MiCroKiTS 4.0 (http://microkit.biocuckoo.org/index.php) and Centrosome:db (40). The centrosome gene-set was modestly enriched for genes associated with IQ (P=0.006; below our multiple test correction threshold for three tests, P<0.0167) but not for genes associated with SZ or EA using MAGMA gene-set analysis (**Figure 5; Supplementary Table 15**). It is possible that the enrichment detected for IQ could be due to the centrosome gene-set representing a subset of brain-expressed genes. However, the centrosome gene-set enrichment for IQ was robust to the inclusion in the analyses of sets of 'brain-elevated' genes (n=1,424) or 'brain-expressed' genes (n=14,243) as covariate (P=0.0048 and P=0.019 respectively). To examine if the enrichment we detect for IQ is a property of polygenic phenotypes in general, we obtained GWAS summary statistics for six other phenotypes and we tested the centrosome gene-set for enrichment in each one. These were other brain-related disorders (Alzheimer's disease (AD) and stroke (STR)) and non-brain related diseases (coronary artery disease

(CAD), Crohn's disease (CD), ulcerative colitis (UC) and type 2 diabetes (T2D)). No enrichment was detected for any of the six phenotypes (**Supplementary Table 15**).

These data provide evidence that variations in centrosome genes contribute to cognitive ability. We next investigated which genes were individually associated with our neurodevelopmental phenotypes of interest in GWAS to identify genes associated with multiple phenotypes. Using the association test result for each gene from the MAGMA analysis (following correction for the number of genes tested), we identified five centrosome genes that were associated with SZ, IQ and EA and a further ten genes that were associated with SZ and with one of the two cognition phenotypes (IQ or EA) (**Supplementary Figure 3**). We reviewed the SNP-level association data at each gene and identified three genes where the same SNP, or SNPs in high LD with each other (r^2>0.8), were associated with both SZ and IQ or EA at genome-wide significant levels. This reduced our list of centrosome genes to three: *SDCCAG8*, of which we knew already, M-phase phosphoprotein 9 (*MPHOSPH9*) and yippee like 4 (*YPEL4*).

The link between cell migration, cell differentiation and the ciliary control of such activities through regulated gene expression that is evidenced by our analyses of *SDCCAG8* provides a model for how centrosomal genes may contribute to SZ risk and cognitive function. Thus, *MPHOSPH9* and *YPEL4* are interesting candidates for further study. Taken together, our data provide evidence that variations in centrosome genes contribute to cognitive ability.

Methods

Centrosome gene analysis

A list of genes with centrosomal localisation was compiled based on results from the MiCroKiTS 4.0 database (http://microkit.biocuckoo.org/index.php), which has collected all proteins identified to be localized in several sub-cellar regions, including the centrosome (41). To the list of genes that encode for these proteins (n=660), we added results from Centrosome:db, which contains a set of human genes encoding proteins that are localized to the centrosome, either as centrosome constituents or as centrosome visitors (n=366) (40). These genes were assembled into a final list of unique genes encoding proteins with centrosomal localisation (**Supplementary Table 15**), to which we refer as the centrosomal gene-set in this study (n=808).

GWAS data

GO terms and the centrosomal gene-set were tested for enrichment of genes associated with different neurodevelopmental phenotypes using GWAS summary statistics for SZ (9), IQ (10) and EA (11). For control purposes, we tested other phenotypes: AD (42), CD (43), STR (44), T2D (45), CAD (46) and UC (47).

Gene-set analysis

We performed gene-set analysis using MAGMA (de Leeuw et al., 2015) (http://ctg.cncr.nl/software/magma) and summary statistics from various GWAS. The analysis involved three steps. First, in the annotation step, we mapped SNPs with available GWAS results on to genes (GRCh37/hg19 start-stop coordinates +/-20kb). Second, in the gene analysis step, we computed gene P values for each GWAS dataset (**Supplementary Tables 16-18**). This gene analysis is based on a multiple linear principal components regression model that accounts for LD between SNPs. The European panel of the 1000 Genomes data was used as a reference panel for LD. Third, a competitive GSA based on the gene P values, also using a regression structure, was used to test if the genes in a gene-set were more strongly associated with either phenotype than other genes in the genome. The MHC region is strongly associated in the SZ GWAS data. This region contains high LD and the association signal has been attributed to just a small number of independent variants (48). However, MAGMA still identifies a very large number of associated genes despite

factoring in the LD information. Of 278 genes that map to chromosome 6 (25-35Mb), 130 genes were associated with SZ in our MAGMA analysis. To avoid the excessive number of associated genes biasing the MAGMA GSA, we excluded all genes within the MHC region from our GSA of SZ. MAGMA was chosen because it corrects for LD, gene size and gene density (potential confounders) and has significantly more power than other GSA tools (49). Sets of 'brain-expressed' genes (n=14,243) and 'brain-elevated' genes (n=1,424) were sourced from the Human Protein Atlas (https://www.proteinatlas.org/humanproteome/brain) and used as covariates in some of the GSA. Brain-elevated genes are those that show an elevated expression in brain compared to other tissue types.

Cell culture and transfections

hTERT-RPE1 and SHSY5Y cells were acquired from the American Type Culture Collection. Both cell lines were cultured in DMEM/F12 1:1, media were supplemented with 10% vol/vol FBS (Sigma-Aldrich) and penicillin-streptomycin (100U/ml and 100µg/ml, respectively; Sigma-Aldrich). Cell lines were cultured in a humidified 5% CO_2 atmosphere at 37°C, and mycoplasma testing was performed every 3 months. Primary cilium formation was induced by culturing cells in DMEM F-12 supplemented with 0.1% FBS and 1% penicillin-streptomycin for up to 48h.

For transient transfections, we used Lipofectamine 2000 (Thermo Fisher Scientific) complexed with DNA at a 2:1 ratio in Opti-MEM (Gibco) for 20 min.

For differentiation of SHSY5Y cells, 1.5 x 10^{-5} cells were plated in media containing 10% FBS onto 6-well dishes that had been coated with 4µg/cm^2 poly-lysine (Sigma) and 1µg/cm^2 lamin (Sigma), as previously described (50). 24h after plating, media was replaced with fresh media containing 2.5% FBS and 10µM retinoic acid (Sigma). Every 48 h, fresh media was added to the cells containing 10µM retinoic acid and the FBS concentration was lowered to 1% and then finally 0.1%. Cells were differentiated with retinoic acid for 7 days. Tracing and analysis of neuronal processes was performed using the ImageJ plugin NeuronJ (https://imagej.net/NeuronJ).

Transwell migration assay

SHSY5Y cells were first serum starved for 24 h, then trypsinised and re-suspended in media containing 0.1% FBS. 3.5 x 10^{-5} cells were added to the upper chamber of transwell inserts (6.5 mm

Transwell with 8.0 µm pore polyester (PET) insert, Corning Costar). Medium with 10% serum was added to the lower chamber. After 16 and 24 h incubation, non-migrated cells on the upper surface of the membrane were scraped off, cells on the lower surface were stained with PBS/ 0.5% crystal violet and imaged. The crystal violet was then solubilized using 10% acetic acid and absorbance was read at 595 nm using a Victor 3 1420-032 plate reader (Perkin Elmer). The percentage of migrated cells was calculated relative to a positive control signal from 3.5×10^{-5} cells plated on the bottom of the 24-well dish.

Cloning

To generate the CRISPR genome editing plasmid, pX330-SDCCAG8-Ex3, primers 5'-caccgTTGCGCCAACAAGCAGATA-3' and 5'-aaacTATCTGCTTGTTGGCGCAAC-3 were heated at 95°C for 5 min and then cooled slowly, prior to phosphorylation with T4 Polynucleotide Kinase. Annealed primers were then ligated into BbsI-digested pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (Plasmid 43330; Addgene (51)). RT-PCR amplification of full-length *SDCCAG8* was performed with the High Capacity RNA to cDNA kit (Applied Biosystems) on hTERT-RPE1 RNA that had been isolated using isolated using TRIzol (Invitrogen) using primers 5'-taagcagaattctATGGCGAAGTCCCCGGAGAACTCT-3' and 5'-tgcttaggatccttGCAATCAGATTGTGGCATGCTGGG-3'. *SDCCAG8* was cloned into pGEM-T-Easy (Promega), subcloned into pEGFP-N1 (Clontech) and then into pcDNA3.1(+) (Thermo Fisher) using NheI and BamHI. Successful cloning was verified by commercial DNA sequencing.

Stable cell line generation

For the generation of *SDCCAG8* null cells using CRISPR/Cas9 technology, 2 µg of pX330-SDCCAG8-Ex3 and 0.7µg of pLox-Neo plasmids were complexed with Lipofectamine 2000 (Invitrogen) in Opti-MEM (Gibco) for 20 min. For rescue cell lines, linearised plasmid was complexed with Lipofectamine 2000. The mixture was added to cells and incubated at 37°C for 4–6 h. After 24 h of recovery, cells were trypsinised and serial dilutions were performed into media containing G418 (Invivogen) at a final concentration of 1mg/ml and were incubated for at least 3 days, after which cells were incubated in antibiotic free growth media at 37°C until single colonies appeared after 10-14 days. Single colonies were picked and expanded using 3 mm Scienceware cloning discs (Sigma-

Aldrich) into 48-well plates (Sarstedt). Cells were further expanded and screened by immunoblotting and immunofluorescence microscopy and DNA sequencing (Source Bioscience, Waterford, Ireland).

Immunofluorescence (IF) microscopy

Cells were grown on sterile coverslips and fixed in methanol containing 5 mM EGTA at −20 °C for 10 min. To stain with antibodies against modified tubulins, cells were incubated on ice for 30 min to depolymerize microtubules. The cells were blocked in 1% BSA in 1×PBS before incubation in primary antibody for 1h followed by 45 min incubation with fluorescently labelled secondary antibodies (Jackson Immunoresearch Laboratories Inc.). DNA was stained with DAPI (Sigma-Aldrich), and slides were mounted in 80% vol/vol glycerol containing 3% N-propyl-gallate in 1×PBS. Cells were imaged using an IX81 microscope (Olympus) with a C4742-80-12AG camera (Hamamatsu) with a 100× oil objective, NA 1.35, using Volocity software (Perkin-Elmer). Images are presented as maximum intensity projections of z-stacks after deconvolution. Merges and individual channel images were exported as tagged image file formats (TIFFs) for publication and then cropped for publication using Photoshop CS6 (Adobe).

Primary antibodies used in this study were as follows: acetylated tubulin (Sigma-Aldrich; T6793, clone 6-11B-1; 1:2,000); ARL13B (ProteinTech; 17711–1-AP; 1:2,000); CP110 (ProteinTech; 12780-1-AP; 1:2,000); IFT88 (ProteinTech; 13967–1-AP; 1:800); pericentrin (Abcam; Ab28144; 1:500); Centrin (Millipore; 20H5; 1:1,000); CEP164 ((52); 1F3G10; 1:10,000); CEP135 ((53); 1420 738; 1:1,000); SDCCAG8 (ProteinTech; 13471-1-AP; 1:1,000); centrobin ((54); 6D4F4; 1:10,000); CEP97 (ProteinTech; 22050-1-AP; 1:1,000); C-NAP1 ((55); 6F2C8; 1:2); PCM1 ((56); 817; 1:10,000); Cep290 (1C3G10; 1:1,000); SMO (Abcam; Ab113438 ; 1:300); TTBK2 (ProteinTech; 15072-1-AP; 1:500).

Immunoblotting

Total cell extracts were prepared by lysing cells in lysis buffer (50mM Tris HCl, pH 7.4, 150mM NaCl, 5% glycerol, 1mM EDTA, 0.5% sodium deoxycholate, 1% IGEPAL, protease inhibitor cocktail (Roche), and phosphatase inhibitors (Sigma-Aldrich)) for 20 min on ice. Samples were then centrifuged for 20 min at 18,000 g at 4°C, and supernatant was transferred to a fresh tube. Protein concentration was determined by Bradford assay on a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific). For loading on SDS-PAGE gel, 20–80 µg of whole cell lysate was transferred into a fresh tube and 5× sample buffer containing 20% β-mercaptoethanol was added to the samples and

boiled at 95°C for 5 min. After proteins were resolved on 10% SDS–polyacrylamide gels, they were transferred to a nitrocellulose membrane (GE Healthcare) using a semi-dry transfer unit (Hoeffer TE77) at 1 mA/cm^2 for 2 h. Blot detection was performed using ECL (GE Healthcare) after blocking and incubation in primary and secondary antibodies.

The primary antibodies used in this study were as follows: α-tubulin (B512; Sigma-Aldrich; 1:10,000), SDCCAG8 (13471-1-AP; ProteinTech; 1:2,500). HRP-labelled goat anti–mouse or anti–rabbit secondary antibodies were used at 1:10,000 (Jackson Immunoresearch Laboratories).

RNA-seq and pathway analysis

Cells were plated to be 80% confluent the next day, the following day either total RNA was extracted or cells were serum starved to induce primary cilium formation, in which RNA was extracted 48 h post serum starvation. For stranded RNA-seq, cDNA libraries were prepared with a TruSeq stranded mRNA library prep Kit (cat# RS-122-2101, Illumina, San Diego, CA , USA) as per the manufacturer's protocol. The quality of cDNA generated was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies). The resulting libraries were sequenced on the NextSeq 550 (Illumina) using a paired-end run (2 × 75 bases). A minimum of 20 M reads were generated per sample. Raw fastq files were aligned to human genome build GRCh37 using Spliced Transcripts Alignment to a Reference (STAR) to produce bam files. STAR is a high precision mapping strategy that allows for accurate alignment of high throughput RNA-seq data (57). Reads were then aggregated by gene using HTSeq-counts in order to produce raw counts per gene per sample. Only reads mapping unambiguously to a single gene are counted (58). HTSeq-count files were used for differential expression (DE) analysis. The DE analysis was performed by the R package DESeq2. *SDCCAG8* null cells were compared to wildtype cells under baseline and serum starved conditions in SHSY5Y and RPE1 cells. Briefly, DESeq2 performs internal normalisation where a geometric mean is calculated for each gene across all samples inputted. The median of all these ratios becomes the size factor for that sample and counts are then divided by this size factor. This effectively corrects for sequencing depth discrepancies and RNA composition bias. DESeq2 uses shrinkage estimators for dispersion and fold change, which leads to improved gene ranking and a highly sensitive and precise methodology that controls effectively for false positives (59). All genes with an absolute $\log_2$ fold change of greater than

three and a Benjamini-Hochberg adjusted p-value less than 0.05 were considered differentially expressed genes.

ConsensusPathDB-human (http://cpdb.molgen.mpg.de/), an open source online tool that integrates interaction networks was used to perform an over-representation analysis of Gene Ontology Terms categories 2- 5 of significant DEGs.

All RNA-seq data generated as part of this study have been submitted to the Gene Expression Omnibus repository and can be accessed using the accession number GSEXXXXXX (awaiting GEO accession number).

Data availability statement

RNA-seq data that support the findings of this study have been deposited in Gene Expression Omnibus repository with the accession code GSEXXXXXX (awaiting GEO accession number). All GWAS summary statistics used in this study are publicly available and can be accessed via the cited publications supplied in the text.

Acknowledgements

We thank Uri Frank and Bob Lahue for critical comments on the manuscript. MF received a College of Science Fellowship from NUI Galway, a Government of Ireland Postgraduate Scholarship GOIPG/2016/506 from the Irish Research Council and a Thomas Crawford Hayes Trust Research fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare no competing financial interests.

Author contributions

Conceptualization, M.F., C.G.M., D.W.M.; Methodology, M.F., L.W., G.D., C.G.M., D.W.M.; Investigation, M.F., L.W.; Writing – Original Draft, M.F., C.G.M., D.W.M.; Writing – Review & Editing, M.F., G.D., C.G.M., D.W.M.; Funding Acquisition, G.D., C.G.M., D.W.M.; Resources, C.G.M., D.W.M.; Supervision, C.G.M., D.W.M.

References

1 Conduit, P.T., Wainman, A. and Raff, J.W. (2015) Centrosome function and assembly in animal cells. *Nat Rev Mol Cell Biol*, **16**, 611-624.

2 Kuijpers, M. and Hoogenraad, C.C. (2011) Centrosomes, microtubules and neuronal development. *Mol Cell Neurosci*, **48**, 349-358.

3 Guemez-Gamboa, A., Coufal, N.G. and Gleeson, J.G. (2014) Primary cilia in the developing and mature brain. *Neuron*, **82**, 511-521.

4 Hildebrandt, F., Benzing, T. and Katsanis, N. (2011) Ciliopathies. *N Engl J Med*, **364**, 1533-1543.

5 Kamiya, A., Tan, P.L., Kubo, K., Engelhard, C., Ishizuka, K., Kubo, A., Tsukita, S., Pulver, A.E., Nakajima, K., Cascella, N.G. *et al.* (2008) Recruitment of PCM1 to the centrosome by the cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses. *Arch Gen Psychiatry*, **65**, 996-1006.

6 Green, M.F. (2006) Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. *J Clin Psychiatry*, **67**, e12.

7 Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T.M., Nordentoft, M. and Glenthoj, B. (2018) Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol Psychiatry*, **83**, 492-498.

8 Lencz, T., Knowles, E., Davies, G., Guha, S., Liewald, D.C., Starr, J.M., Djurovic, S., Melle, I., Sundet, K., Christoforou, A. *et al.* (2014) Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT). *Mol Psychiatry*, **19**, 168-174.

9 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. *Nature Genetics*, **50**, 381-389.

10 Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I. *et al.* (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, **50**, 912-919.

11 Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R. *et al.* (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, **50**, 1112-1121.

12 Guo, J., Higginbotham, H., Li, J., Nichols, J., Hirt, J., Ghukasyan, V. and Anton, E.S. (2015) Developmental disruptions underlying brain abnormalities in ciliopathies. *Nat Commun*, **6**, 7857.

13 Forsythe, E. and Beales, P.L. (2013) Bardet-Biedl syndrome. *Eur J Hum Genet*, **21**, 8-13.

14 Marley, A. and von Zastrow, M. (2012) A simple cell-based assay reveals that diverse neuropsychiatric risk genes converge on primary cilia. *PLoS One*, **7**, e46647.

15 Janssen, S., Ramaswami, G., Davis, E.E., Hurd, T., Airik, R., Kasanuki, J.M., Van Der Kraak, L., Allen, S.J., Beales, P.L., Katsanis, N. *et al.* (2011) Mutation analysis in Bardet-Biedl syndrome by DNA pooling and massively parallel resequencing in 105 individuals. *Hum Genet*, **129**, 79-90.

16 Otto, E.A., Hurd, T.W., Airik, R., Chaki, M., Zhou, W., Stoetzel, C., Patil, S.B., Levy, S., Ghosh, A.K., Murga-Zamalloa, C.A. *et al.* (2010) Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat Genet*, **42**, 840-850.

17 Schaefer, E., Zaloszyc, A., Lauer, J., Durand, M., Stutzmann, F., Perdomo-Trujillo, Y., Redin, C., Bennouna Greene, V., Toutain, A., Perrin, L. *et al.* (2011) Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly. *Mol Syndromol*, **1**, 273-281.

18 Insolera, R., Shao, W., Airik, R., Hildebrandt, F. and Shi, S.H. (2014) SDCCAG8 regulates pericentriolar material recruitment and neuronal migration in the developing cortex. *Neuron*, **83**, 805-822.

19 Airik, R., Slaats, G.G., Guo, Z., Weiss, A.C., Khan, N., Ghosh, A., Hurd, T.W., Bekker-Jensen, S., Schroder, J.M., Elledge, S.J. *et al.* (2014) Renal-retinal ciliopathy gene Sdccag8 regulates DNA damage response signaling. *J Am Soc Nephrol*, **25**, 2573-2583.

20 Chang, S., Fang, K., Zhang, K. and Wang, J. (2015) Network-Based Analysis of Schizophrenia Genome-Wide Association Data to Detect the Joint Functional Association Signals. *PLoS One*, **10**, e0133404.

21 Hamshere, M.L., Walters, J.T., Smith, R., Richards, A.L., Green, E., Grozeva, D., Jones, I., Forty, L., Jones, L., Gordon-Smith, K. *et al.* (2013) Genome-wide significant associations in

schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Mol Psychiatry*, **18**, 708-712.

22 Mullee, L.I. and Morrison, C.G. (2016) Centrosomes in the DNA damage response-the hub outside the centre. *Chromosome Res*, **24**, 35-51.

23 Airik, R., Schueler, M., Airik, M., Cho, J., Ulanowicz, K.A., Porath, J.D., Hurd, T.W., Bekker-Jensen, S., Schroder, J.M., Andersen, J.S. *et al.* (2016) SDCCAG8 Interacts with RAB Effector Proteins RABEP2 and ERC1 and Is Required for Hedgehog Signaling. *PLoS One*, **11**, e0156081.

24 Spektor, A., Tsang, W.Y., Khoo, D. and Dynlacht, B.D. (2007) Cep97 and CP110 suppress a cilia assembly program. *Cell*, **130**, 678-690.

25 Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.D. and Nigg, E.A. (2007) Plk4-induced centriole biogenesis in human cells. *Dev Cell*, **13**, 190-202.

26 Tsang, W.Y., Bossard, C., Khanna, H., Peranen, J., Swaroop, A., Malhotra, V. and Dynlacht, B.D. (2008) CP110 suppresses primary cilia formation through its interaction with CEP290, a protein deficient in human ciliary disease. *Dev Cell*, **15**, 187-197.

27 Cajanek, L. and Nigg, E.A. (2014) Cep164 triggers ciliogenesis by recruiting Tau tubulin kinase 2 to the mother centriole. *Proc Natl Acad Sci U S A*, **111**, E2841-2850.

28 Goetz, S.C., Liem, K.F., Jr. and Anderson, K.V. (2012) The spinocerebellar ataxia-associated gene Tau tubulin kinase 2 controls the initiation of ciliogenesis. *Cell*, **151**, 847-858.

29 Pazour, G.J., Dickert, B.L., Vucica, Y., Seeley, E.S., Rosenbaum, J.L., Witman, G.B. and Cole, D.G. (2000) Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene tg737, are required for assembly of cilia and flagella. *J Cell Biol*, **151**, 709-718.

30 Hor, C.H. and Tang, B.L. (2010) Sonic hedgehog as a chemoattractant for adult NPCs. *Cell Adh Migr*, **4**, 1-3.

31 Chen, J.K., Taipale, J., Young, K.E., Maiti, T. and Beachy, P.A. (2002) Small molecule modulation of Smoothened activity. *Proc Natl Acad Sci U S A*, **99**, 14071-14076.

32 Rohatgi, R., Milenkovic, L. and Scott, M.P. (2007) Patched1 regulates hedgehog signaling at the primary cilium. *Science*, **317**, 372-376.

33 Solecki, D.J., Model, L., Gaetz, J., Kapoor, T.M. and Hatten, M.E. (2004) Par6alpha signaling controls glial-guided neuronal migration. *Nat Neurosci*, **7**, 1195-1203.

34 Tee, J.Y., Sutharsan, R., Fan, Y. and Mackay-Sim, A. (2016) Schizophrenia patient-derived olfactory neurosphere-derived cells do not respond to extracellular reelin. *NPJ Schizophr*, **2**, 16027.

35 Tee, J.Y., Sutharsan, R., Fan, Y. and Mackay-Sim, A. (2017) Cell migration in schizophrenia: Patient-derived cells do not regulate motility in response to extracellular matrix. *Mol Cell Neurosci*, **80**, 111-122.

36 Sei, Y., Ren-Patterson, R., Li, Z., Tunbridge, E.M., Egan, M.F., Kolachana, B.S. and Weinberger, D.R. (2007) Neuregulin1-induced cell migration is impaired in schizophrenia: association with neuregulin1 and catechol-o-methyltransferase gene polymorphisms. *Mol Psychiatry*, **12**, 946-957.

37 Fan, Z., Wu, Y., Shen, J., Ji, T. and Zhan, R. (2013) Schizophrenia and the risk of cardiovascular diseases: a meta-analysis of thirteen cohort studies. *J Psychiatr Res*, **47**, 1549-1556.

38 Andersen, E.F. and Halloran, M.C. (2012) Centrosome movements in vivo correlate with specific neurite formation downstream of LIM homeodomain transcription factor activity. *Development*, **139**, 3590-3599.

39 Ahmad, R., Sportelli, V., Ziller, M., Spengler, D. and Hoffmann, A. (2018) Tracing Early Neurodevelopment in Schizophrenia with Induced Pluripotent Stem Cells. *Cells*, **7**.

40 Alves-Cruzeiro, J.M., Nogales-Cadenas, R. and Pascual-Montano, A.D. (2014) CentrosomeDB: a new generation of the centrosomal proteins database for Human and Drosophila melanogaster. *Nucleic Acids Res*, **42**, D430-436.

41 Huang, Z., Ma, L., Wang, Y., Pan, Z., Ren, J., Liu, Z. and Xue, Y. (2015) MiCroKiTS 4.0: a database of midbody, centrosome, kinetochore, telomere and spindle. *Nucleic Acids Res*, **43**, D328-334.

42 Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W. *et al.* (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*, **45**, 1452.

43 Franke, A., McGovern, D.P.B., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R. *et al.* (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics*, **42**, 1118-1125.

44 Traylor, M., Farrall, M., Holliday, E.G., Sudlow, C., Hopewell, J.C., Cheng, Y.-C., Fornage, M., Ikram, M.A., Malik, R., Bevan, S. *et al.* (2012) Genetic risk factors for ischaemic stroke and its

subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *The Lancet Neurology*, **11**, 951-962.

45 Replication, D.I.G., Meta-analysis, C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., South Asian Type 2 Diabetes, C., Mexican American Type 2 Diabetes, C., Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples, C., Mahajan, A., Go, M.J., Zhang, W., Below, J.E. *et al.* (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*, **46**, 234-244.

46 Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F.R., Barbalic, M., Gieger, C. *et al.* (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*, **43**, 333.

47 Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A. *et al.* (2011) Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature Genetics*, **43**, 246.

48 Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V. *et al.* (2016) Schizophrenia risk from complex variation of complement component 4. *Nature*, **530**, 177-183.

49 de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology*, **11**, e1004219.

50 Dwane, S., Durack, E. and Kiely, P.A. (2013) Optimising parameters for the differentiation of SH-SY5Y cells to study cell adhesion and cell migration. *BMC Res Notes*, **6**, 366.

51 Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.

52 Daly, O.M., Gaboriau, D., Karakaya, K., King, S., Dantas, T.J., Lalor, P., Dockery, P., Krämer, A. and Morrison, C.G. (2016) CEP164-null cells generated by genome editing show a ciliation defect with intact DNA repair capacity. *Journal of Cell Science*, **129**, 1769.

53 Bird, A.W. and Hyman, A.A. (2008) Building a spindle of the correct length in human cells requires the interaction between TPX2 and Aurora A. *The Journal of Cell Biology*, **182**, 289.

54 Ogungbenro, Y.A., Tena, T.C., Gaboriau, D., Lalor, P., Dockery, P., Philipp, M. and Morrison, C.G. (2018) Centrobin controls primary ciliogenesis in vertebrates. *The Journal of Cell Biology*, **217**, 1205.

55 Flanagan, A.-M., Stavenschi, E., Basavaraju, S., Gaboriau, D., Hoey, D.A. and Morrison, C.G. (2017) Centriole splitting caused by loss of the centrosomal linker protein C-NAP1 reduces centriolar satellite density and impedes centrosome amplification. *Molecular biology of the cell*, **28**, 736-745.

56 Dammermann, A. and Merdes, A. (2002) Assembly of centrosomal proteins and microtubule organization depends on PCM-1. *J Cell Biol*, **159**, 255-266.

57 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, **29**, 15-21.

58 Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, **31**, 166-169.

59 Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, **15**, 550-550.

Figure legends

Figure 1. Normal centriolar structures in *SDCCAG8* null cells

A. Localisation of SDCCAG8 in hTERT-RPE1 cells during the indicated cell cycle stages. Asynchronous cells were fixed and stained for SDCCAG8 (green), CEP164 (red) and DNA (blue). Scale bar, 5 µm and inset scale bar, 2.5 µm.

B. Western blot confirming the loss of SDCCAG8 expression in genome-edited hTERT-RPE1 and SHSY5Y clones (KO) compared to their respective wild-type (WT) cells. The SDCCAG8 protein can be visualized at its expected molecular weight of 83kDa (arrowhead); the lower band (arrowed) appears to be isoform 4 of the protein, which is also targeted in the KO clones generated. A background band is indicated with an asterisk. Actin is shown as a loading control. Molecular mass markers are at left in kDa.

C. Generation of rescue clones. SDCCAG8 null RPE1 cells were transfected with expression construct for SDCCAG8 and immunoblot analysis was used to screen for clones with expression restored to levels equivalent to those seen in wild-type cells. SDCCAG8 is indicated with an arrowhead; a background band is asterisked. Rescue clone 1 was used in this study. Molecular mass markers are at left in kDa. Ponceau staining of the membrane was used as a loading control.

D. Growth curves show mean ± SEM of three independent experiments. No significant difference was observed between WT and SDCCAG8-deficient cells at any time point.

E. IF microscopy of the indicated PCM, centriole, and centriolar satellite makers in asynchronous WT and SDCCAG8 null (KO) cells. Scale bars: 5 µm; (inset) 1 µm.

F. Representative micrographs of imaging used to determine centriole number in RPE1 cells. Centrioles were visualized by centrin (red) and CEP135 (green) staining before and 48 hours after cells were treated with 5Gy IR. Scale bar, 1 µm.

G. Quantitation of DNA damage-induced centrosome amplification in cells of the indicated genotype. Bar graph shows mean ± SEM from 3 experiments in which at least 100 cells were counted.

Figure 2. Defective primary ciliogenesis in *SDCCAG8* null cells

A. Localisation of SDCCAG8 (green) in 48-h serum starved hTERT-RPE1 cells. Ac-Tub, Acetylated tubulin. Scale bar, 5 µm, inset scale bar, 2.5 µm.

B. Representative images of the ciliary marker, acetylated tubulin (red) with the centriolar marker CEP135 (green) in hTERT-RPE1 cells of the indicated genotype after 48h serum starvation. Wild-type, KO, SDCCAG8 knockout. Scale bar, 5 µm, inset scale bar, 2.5 µm.

C. Quantification of ciliation frequency, based on acetylated tubulin staining. Histogram depicts the mean ± SEM of 4 independent repeats of which at least 100 cells were counted. ***, P ≤ 0.001; ns, non-significant in comparison to RPE1 WT cells by unpaired t-test.

D. Representative images of the ciliary marker, ARL13B (green) with the centriolar marker centrin (red) in SHSY5Y cells of the indicated genotype after 48h serum starvation. Scale bar, 5 µm, inset scale bar, 2.5 µm.

E. Quantification of SHSY5Y ciliation frequency. Histogram depicts the mean ± SEM of 3 independent repeats of which at least 100 cells were counted. **, P ≤ 0.01 in comparison to WT SHSY5Y cells by unpaired t-test.

F. Quantitation of cilium length in RPE1 cells of the indicated genotype. Individual data points are shown, as is the mean cilium length ± SEM from at least 88 ciliated cells quantitated in 3 separate experiments. **, P ≤ 0.01; ***, P ≤ 0.001; ns, non-significant in comparison to WT RPE1 cells by unpaired t-test.

G. IF microscopy of indicated regulators of ciliation in hTERT-RPE1 WT and SDCCAG8 null cells 48-h serum starved. Scale bar, 5 µm, inset scale bar, 2.5 µm.

H. Localisation of Smoothened (SMO, green) in hTERT-RPE1 cells of the indicated genotype after 48h serum starvation and 4 h treatment with 100 nM SAG. Scale bar, 5 µm, inset scale bar, 2.5 µm.

I. Quantification of SMO localisation to cilia. Histogram depicts the mean ± SEM of 3 independent repeats of which at least 30 ciliated cells were counted. . ***, P ≤ 0.001; **, P ≤ 0.01; in comparison to WT RPE1 cells (black) or to KO cells (red) by unpaired t-test.

Figure 3. Gene expression pathways impacted by SDCCAG8 deficiency affect IQ, SZ and EA

Gene-set analysis of the SHSY5Y and RPE1 pathways in IQ, SZ and EA. GO terms are listed on the y-axis. P-values are shown above each data point, which represent beta-values (regression coefficients) plotted on the x-axis. Horizontal bars indicate standard error. Significant p-values after multiple test correction are indicated in bold.

Figure 4. Migration and differentiation defects in *SDCCAG8* null SHSY5Y cells

A. Representative images of wild-type (WT) and *SDCCAG8* null (KO) SHSY5Y cells that were allowed to migrate for 16 and 24 h on transwell inserts. Cells were stained with crystal violet. Scale bar, 250 µm.

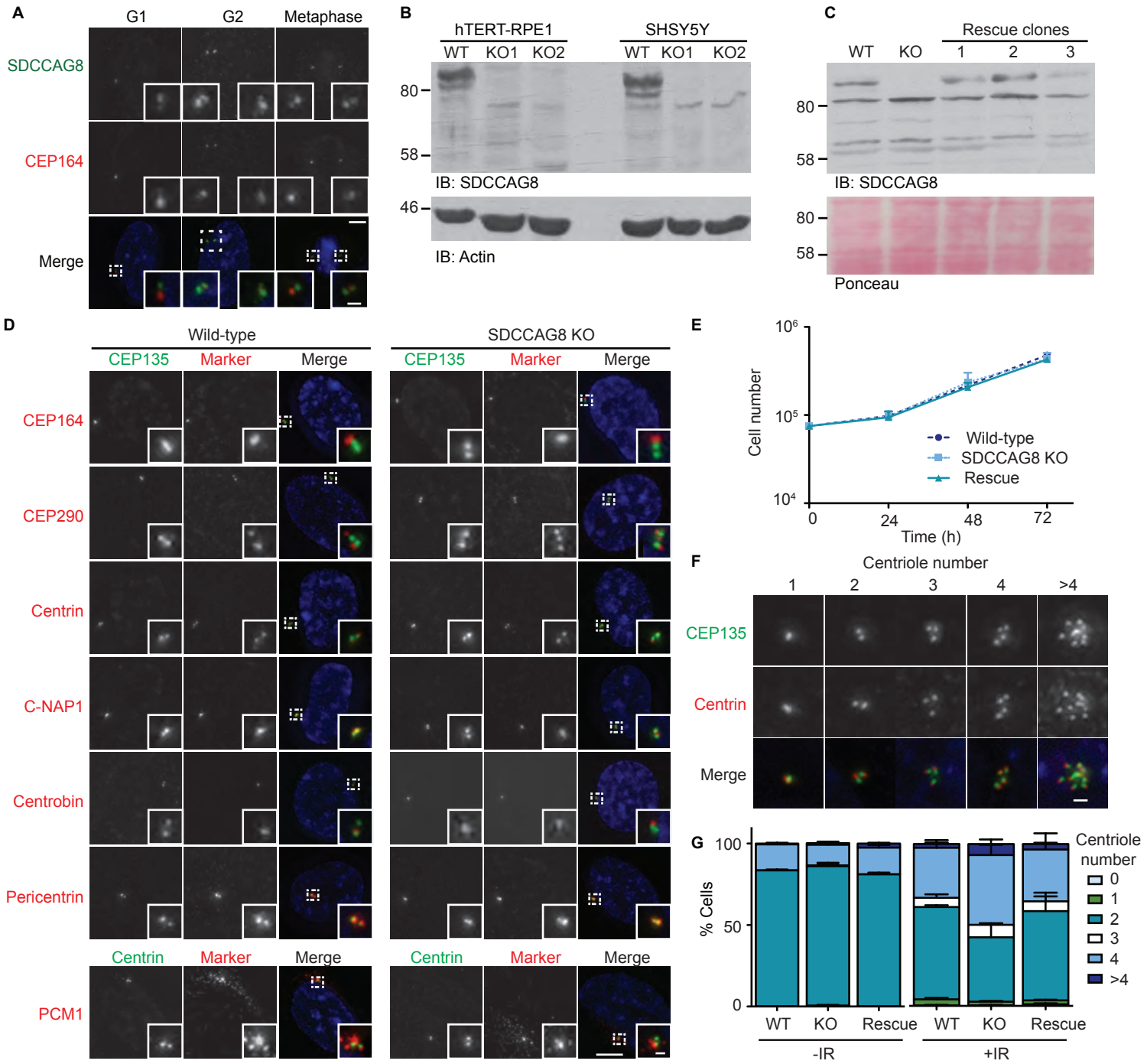
B. Quantification of cell migration based on absorbance of crystal violet staining relative to the positive control of plating all cells on the opposite side of the transwell insert. Data show means and individual data points of percentage absorbances from two separate experiments.

C. Representative images of differentiation in SHSY5Y cells of the indicated genotype following treatment with vehicle or RA for 7 days. Neuritic processes can be seen as darker, linear extensions from the cell body. Scale bar, 250 µm.

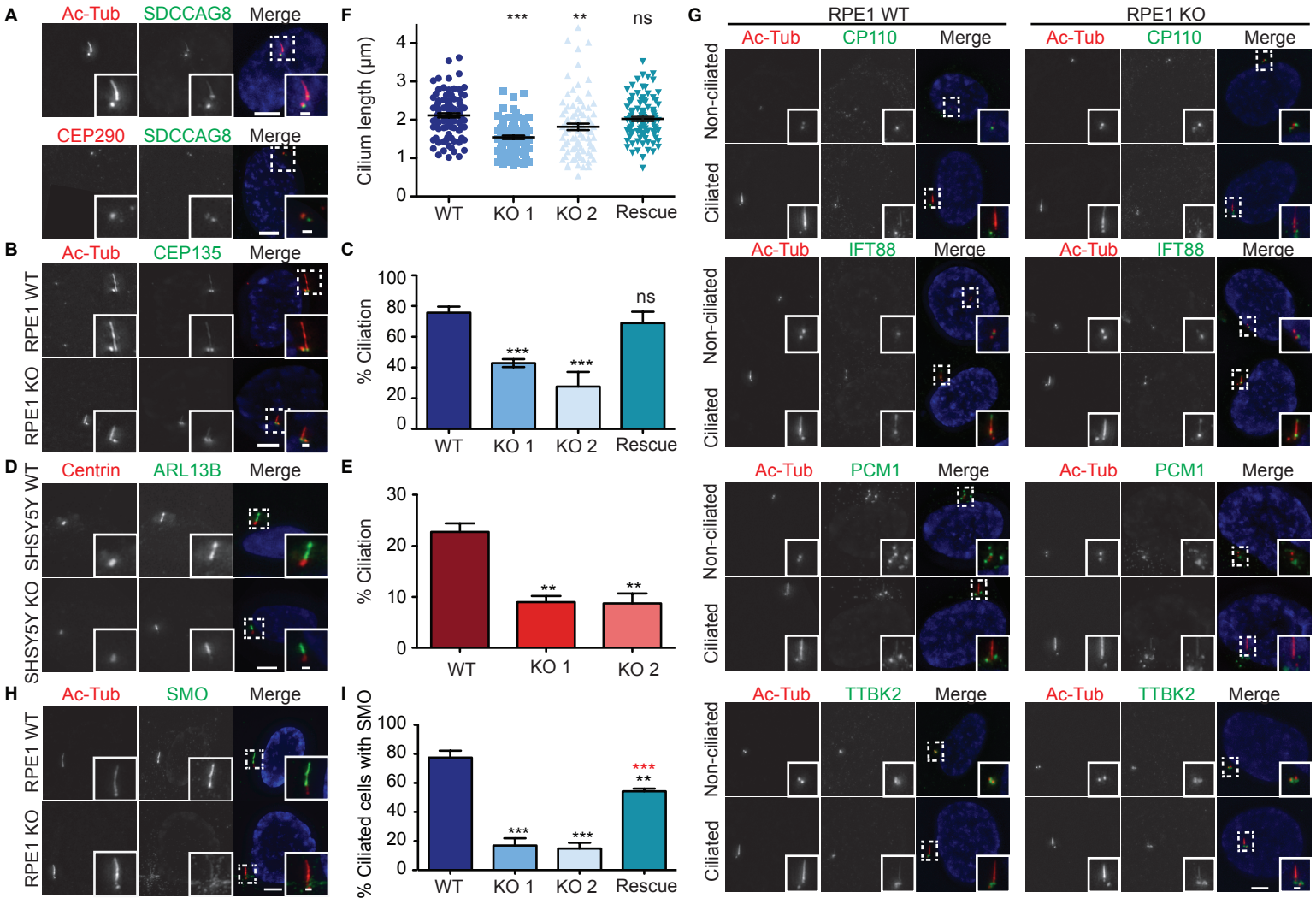
D. Quantitation of neuritic projections as a measure of differentiation in SHSY5Y cells of the indicated genotype. Data show mean length $\pm$ SEM of N=292 (WT) and N=244 (KO) projections analysed in 2 separate experiments. ***, $P \leq 0.0001$; in comparison to WT SHSY5Y cells by unpaired t-test.

Figure 5. Gene-set analysis of the centrosomal gene-set in neurodevelopmental phenotypes

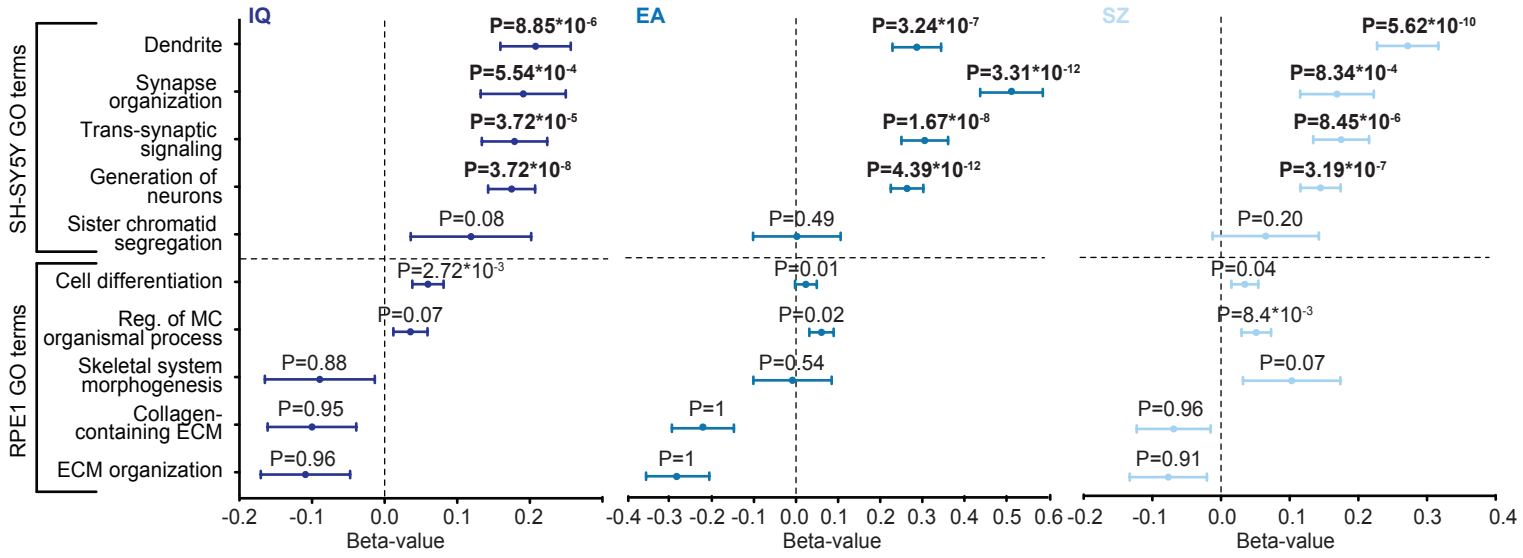
Gene-set analysis of the centrosomal gene-set in IQ, SZ, and EA, plus *post-hoc* analysis of six other GWAS datasets (two brain-related phenotypes (Alzheimer's disease, AD; stroke, STR) and four non-brain diseases (coronary artery disease, CAD; Crohn's disease, CD; type 2 diabetes, T2D; ulcerative colitis, UC)). Phenotypes are listed on the y-axis. P-values are shown above each data point, which represent beta values (regression coefficients) plotted on the x-axis. Horizontal bars indicate standard error. Significant p-values, after multiple test correction, are indicated in bold.

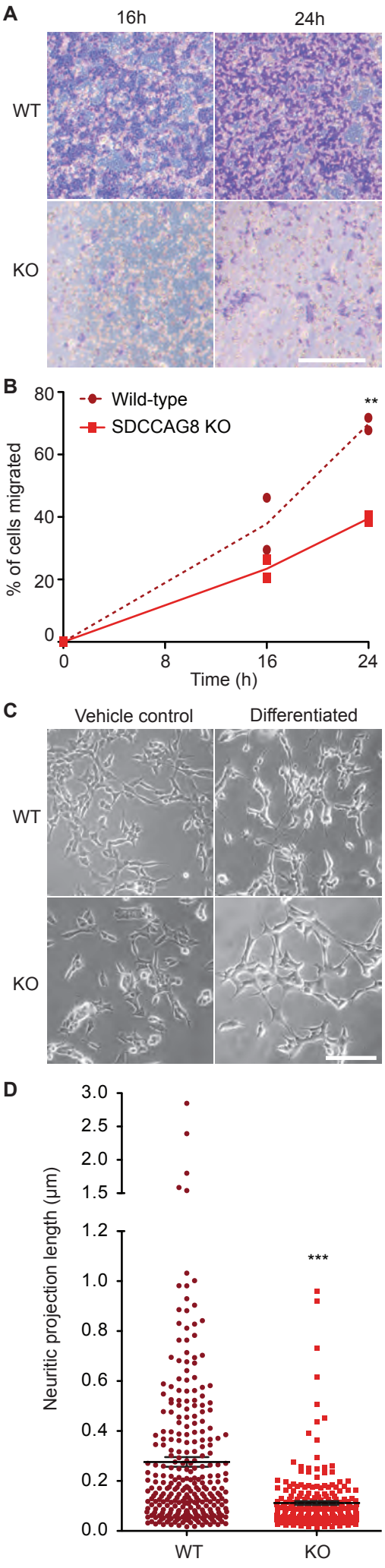

A
G1
G2
Metaphase
SDCCAG8
CEP164
Merge
B
hTERT-RPE1
SHSY5Y
WT
KO1
KO2
WT
KO1
KO2
80
58
IB: SDCCAG8
46
IB: Actin
C
Rescue clones
WT
KO
1
2
3
80
58
IB: SDCCAG8
80
58
Ponceau
D
Wild-type
SDCCAG8 KO
CEP135
Marker
Merge
CEP164
CEP290
Centrin
C-NAP1
Centrobin
Pericentrin
PCM1
E
Cell number
Wild-type
SDCCAG8 KO
Rescue
Time (h)
F
Centriole number
1
2
3
4
>4
CEP135
Centrin
Merge
G
% Cells
Centriole number
0
1
2
3
4
>4
WT
KO
Rescue
-IR
+IR

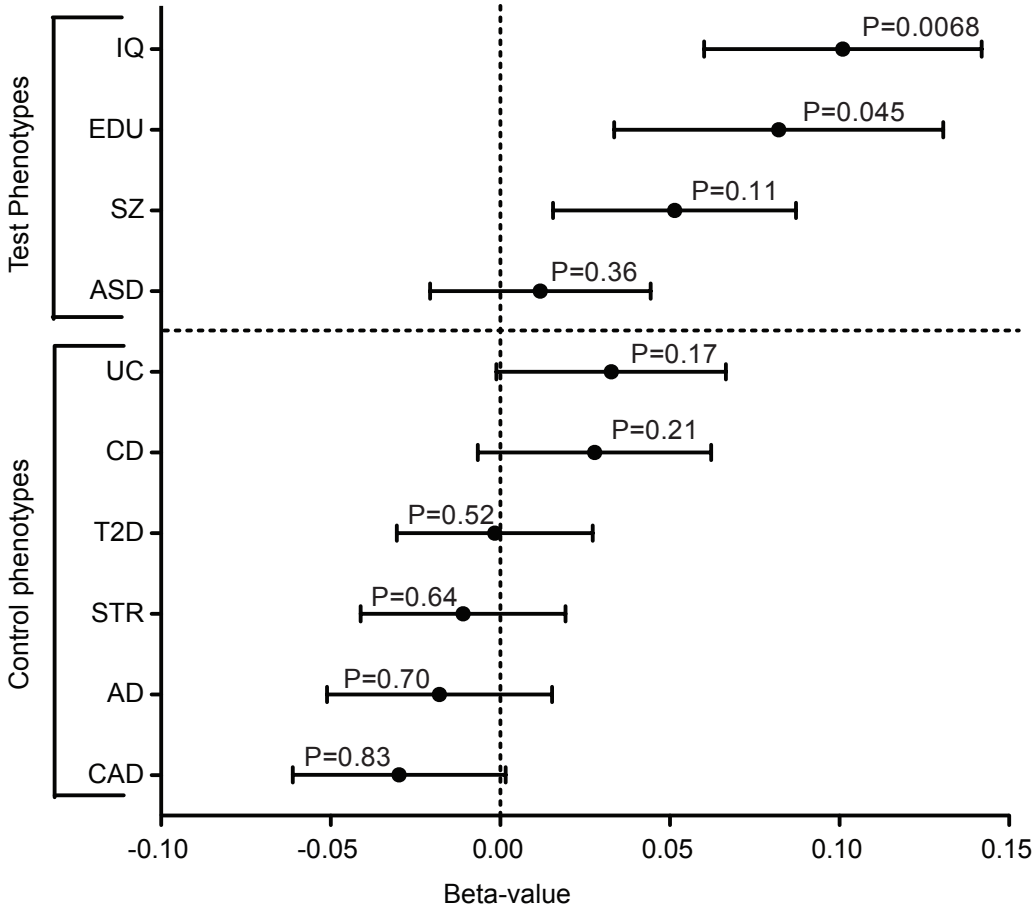
Flynn Figure 2



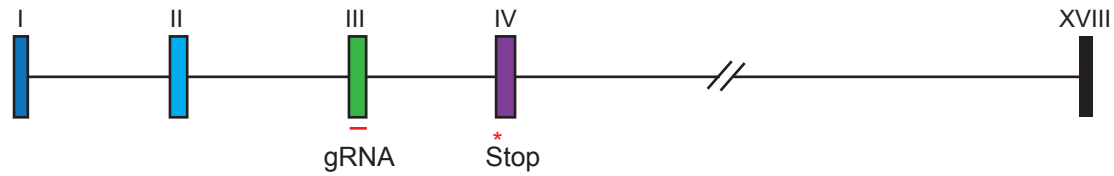
Flynn Figure 3




A
16h
24h
WT
KO
B
Wild-type
SDCCAG8 KO
**
% of cells migrated
0
20
40
60
80
0
8
16
24
Time (h)
C
Vehicle control
Differentiated
WT
KO
D
Neuritic projection length (µm)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.5
2.0
2.5
3.0
***
WT
KO


Test Phenotypes
IQ
EDU
SZ
ASD
P=0.0068
P=0.045
P=0.11
P=0.36
Control phenotypes
UC
CD
T2D
STR
AD
CAD
P=0.17
P=0.21
P=0.52
P=0.64
P=0.70
P=0.83
-0.10
-0.05
0.00
0.05
0.10
0.15
Beta-value

A



B

hTERT-RPE1

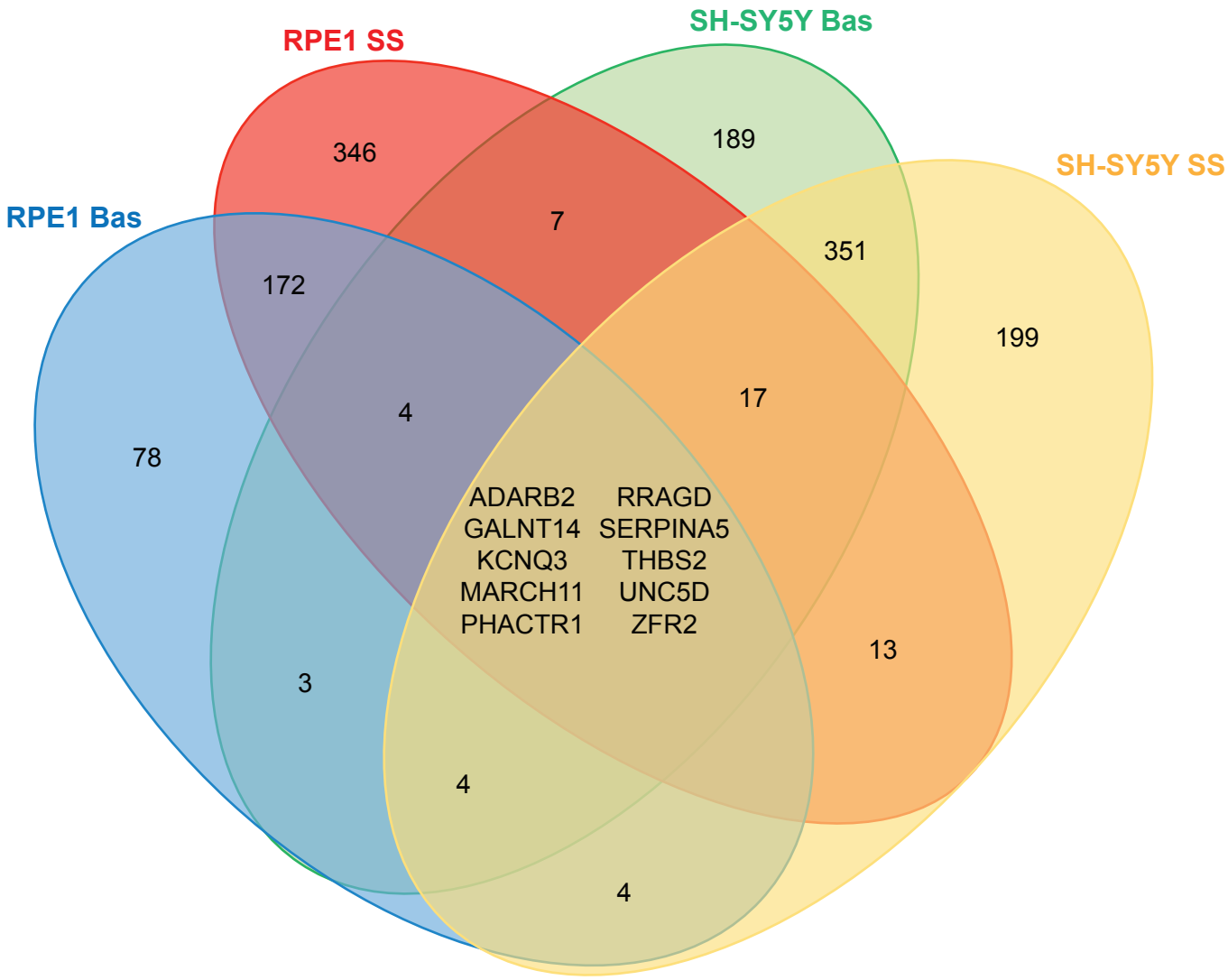
```
Wild-type     ttgttgcgccaacaagcagataaggaaagtgaagtatctccgtcaagaagaagaaaaatgtccccttgaggtcattagaa
               L  L  R  Q  Q  A  D  K  E  S  E  V  S  P  S  R  R  R  K  M  S  P  L  R  S  L  E

KO 1/ KO 2    ttgttgcgccaacaagca-ataaggaaagtgaagtatctccgtcaagaagaagaaaaatgtccccttgaggtcattag
(1 bp del)     L  L  R  Q  Q  A   I  R  K  V  K  Y  L  R  Q  E  E  E  K  C  P  P  *
```

C

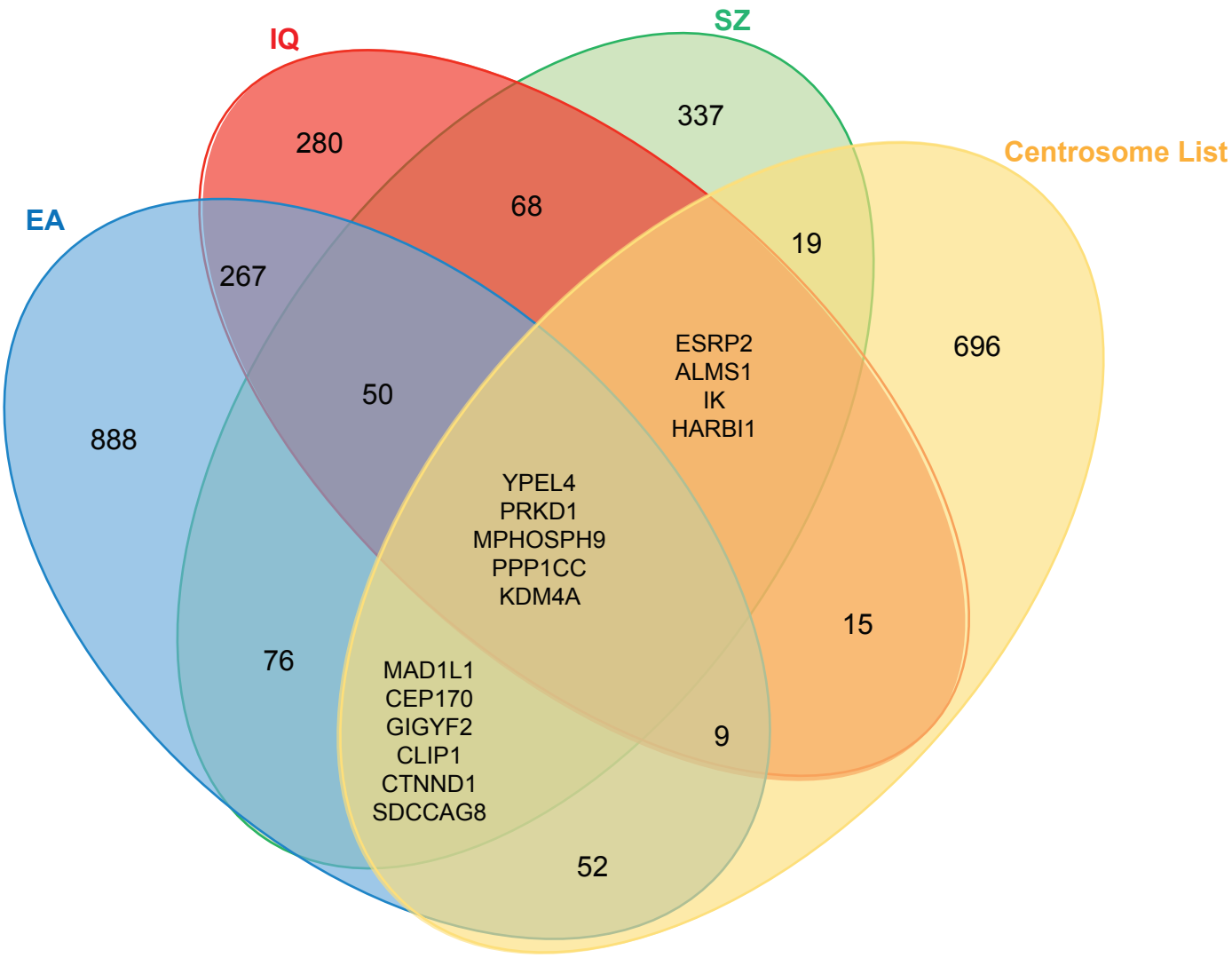
SHSY5Y

```
Wild-type     ttgttgcgccaacaagcagataaggaaagtgaagtatctccgtcaagaagaagaaaaatgtccccttgaggtcattagaa
               L  L  R  Q  Q  A  D  K  E  S  E  V  S  P  S  R  R  R  K  M  S  P  L  R  S  L  E
KO 1 A1       tt----------------gataaggaaagtgaagtatctccgtcaagaagaagaaaaatgtccccttgaggtcattag
(16 bp del)    L                 I  R  K  V  K  Y  L  R  Q  E  E  E  K  C  P  P  *
KO 1 A2       ttgttgcgccaacaa----ataaggaaagtgaagtatctccgtcaagaagaagaaaaatgtccccttgaggtcattag
(4bp del)      L  L  R  Q  Q      I  R  K  V  K  Y  L  R  Q  E  E  E  K  C  P  P  *
KO 2          ttgttgcgccaacaagcagagataaggaaagtgaagtatctccgtcaagaagaagaaaaatgtccccttgaggtcattag
(2bp ins)      L  L  R  Q  Q  A  E  I  R  K  V  K  Y  L  R  Q  E  E  E  K  C  P  P  *
```

Supplementary Figure 1. Confirmation of SDCCAG8 disruption by sequencing
A. Schematic representation of *SDCCAG8* gene on chromosome 1. The region to which the gRNA was designed is underlined below the exon, highlighting the targeted exon, and the resulting stop codon by an asterisk.
B. Nucleotide sequences and predicted protein sequence of wild-type and KO hTERT-RPE1 clones. Sequences are coloured according to their respective exons. Both clones have a biallelic deletion of 1 bp.
C. Nucleotide sequences and predicted protein sequence of wild-type and KO SHSY5Y clones. Sequences are coloured according to their respective exons. SH-SY5Y KO 1 has a unique mutation on each allele, allele 1 (A1) was a 16 bp deletion, while allele 2 was a 4 bp deletion. SH-SY5Y KO 2 has a 2 bp insertion on both alleles. All mutations lead to a frameshift and the generation of the same premature stop codon (*) at the beginning of exon 4.



Supplementary Figure 2. Overlap of DEGs upon *SDCCAG8* disruption
Venn diagram demonstrating the overlap of DEGs upon *SDCCAG8* disruption across the four conditions analysed in this study; RPE1 cells at baseline conditions ('Bas'), RPE1 cells after 48h serum starvation ('SS'), SHSY5Y cells at baseline and SH-SY5Y cells after 48h serum starvation. All DEGs had an absolute $\log_2$ fold change of 3 or greater and Benjamini-Hochberg adjusted p-value of less than 0.05. Listed are the gene symbols for the ten genes that are differentially expressed under all four conditions analysed.



Supplementary Figure 3. Overlap of EA-, IQ- and SZ-associated genes with centrosomal genes

Venn diagram demonstrating the overlap between EA, IQ and SZ associated genes and genes that encode proteins with centrosomal localisation. To generate phenotype-associated gene lists, a gene-level analysis was performed in MAGMA. Genes with significant P-values after correcting for the number of genes tested were considered associated with EA, IQ or SZ.