King's Research Portal

DOI:
10.1038/s41467-017-00422-w

Document Version
Publisher's PDF, also known as Version of record

Link to publication record in King's Research Portal

Citation for published version (APA):
Krishnan, M. L., Van Steenwinckel, J., Schang, A., Yan, J., Arnadottir, J., Le Charpentier, T., ... Gressens, P. (2017). Integrative genomics of microglia implicates DLG4 (PSD95) in the white matter development of preterm infants. Nature Communications, 8(1). https://doi.org/10.1038/s41467-017-00422-w

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 11. Mar. 2020
Integrative genomics of microglia implicates DLG4 (PSD95) in the white matter development of preterm infants

Michelle L. Krishnan1, Juliette Van Steenwinckel2,3, Anne-Laure Schang2,3, Jun Yan2,3, Johanna Arnadottir2,3, Tifenn Le Charpentier2,3, Zsolt Csaba2,3, Pascal Dournaud2,3, Sara Cipriani2,3, Constance Auvenet4, Luigi Titomanlio2, Julien Pansiot2,3, Gareth Ball 1, James P. Boardman5, Andrew J. Walley6, Alka Saxena7, Ghazala Mirza8,9, Bobbi Fleiss1,2,3, A. David Edwards1, Enrico Petretto10 & Pierre Gressens1,2,3

Preterm birth places infants in an adverse environment that leads to abnormal brain development and cerebral injury through a poorly understood mechanism known to involve neuroinflammation. In this study, we integrate human and mouse molecular and neuroimaging data to investigate the role of microglia in preterm white matter damage. Using a mouse model where encephalopathy of prematurity is induced by systemic interleukin-1β administration, we undertake gene network analysis of the microglial transcriptomic response to injury, extend this by analysis of protein-protein interactions, transcription factors and human brain gene expression, and translate findings to living infants using imaging genomics. We show that DLG4 (PSD95) protein is synthesised by microglia in immature mouse and human, developmentally regulated, and modulated by inflammation; DLG4 is a hub protein in the microglial inflammatory response; and genetic variation in DLG4 is associated with structural differences in the preterm infant brain. DLG4 is thus apparently involved in brain development and impacts inter-individual susceptibility to injury after preterm birth.
The majority of preterm infants develop encephalopathy of prematurity characterised by oligodendrocyte maturation arrest, hypomyelination and reduced brain growth\(^1\)-\(^3\). This is associated in over 30% of infants with long-term neurocognitive problems\(^4\), \(^5\), and has characteristic correlates on magnetic resonance imaging (MRI) and diffusion MRI (d-MRI)\(^6\), \(^7\). It is strongly associated with systemic and cerebral inflammation, with the prominent involvement of microglia\(^2\), \(^8\)-\(^10\).

Microglia are found preferentially within the developing white matter\(^1\)-\(^11\) and have both pro-inflammatory and restorative functions that are essential for normal brain development\(^15\), \(^13\). However, the microglial response to preterm birth is poorly understood and a better understanding of microglial function in this context could allow therapeutic modulation to mitigate the brain injury of prematurity and other cerebral inflammatory conditions.

We have used integrative genomics to investigate the role of microglia in preterm brain development, employing a clinically relevant mouse model of interleukin-1β (IL1B)-induced systemic inflammation that recapitulates the essential features of the encephalopathy of prematurity\(^14\), integrating microglial-specific data from this model with ex vivo and in vitro experiments, analysis of human microglia and imaging-genomic data from preterm infants.

We now report: (a) endogenous expression of DLG4 (PSD95) by microglia in early development, which is modulated by developmental stage and inflammation; (b) a role for DLG4 as a hub protein in the microglial inflammatory response; and (c) an association between genetic variability in DLG4 and white matter structure in the preterm neonatal brain.

**Results**

**Global effects of IL1B on mouse microglial transcriptome.** In a previously validated mouse model of encephalopathy of prematurity in which IL1B is administered intra-peritoneally from postnatal days 1–5\(^14\) (Fig. 1a), we verified that microglia were the predominant myeloid cell in the brain (Supplementary Fig. 1a, b) and that the blood-brain barrier remained appreciably intact, with increased expression of tight junction and adherens genes (Supplementary Fig. 2a–d). P5 is comparable with human gestational age around 32 weeks\(^15\). CD11B+ cells were then isolated by magnetic-activated cell sorting (MACS) at postnatal days 1, 5, 10 and 45, and fluorescence-activated cell sorting (FACS) and qPCR analysis used to demonstrate these cells to be >95% pure microglia (Supplementary Fig. 1c–e). These isolated ex vivo microglia were then used for microarray gene expression analysis.

We examined the effects of systemic IL1B on the microglial transcriptome by: (a) comparing systemic IL1B exposure with control conditions (“IL1B”); (b) examining transcriptional changes over time (“Development”) and; (c) assessing whether there is a different transcriptional response to IL1B as a function of time (“Interaction”). Development captured development of microglia themselves as well as systemic maturation, and the interaction assessed whether IL1B effects were time dependent.

After accounting for multiple testing, we found thousands of genes with altered expression in each of the three responses (Supplementary Data 1). Functional enrichment analysis of the differentially expressed genes implicated numerous biological processes and pathways (Supplementary Fig. 3 and Supplementary Data 2 and 3). IL1B and Development showed some interdependency, with over-representation of cytokine-cytokine receptor interaction, transmembrane signalling and cell adhesion molecules. Using REVIGO\(^16\), we defined a set of representative gene ontology (GO) processes common to IL1B and Development, including immune system processes, cell adhesion, morphogenesis, regulation of multicellular organismal process and cell surface receptor signalling (Supplementary Data 4). IL1B exposure during development thus has a pervasive effect on the microglial transcriptome, engaging pathways relevant to both immune function and growth.

**Patterns of microglia gene expression response to IL1B.** When expression profiles were clustered by their response to IL1B at each time point, several gene clusters became apparent (Fig. 1b and Supplementary Fig. 4), with the most prominent clusters observed at P1 and P5 (Z-score above 1 or below −1, difference between IL1B and control \(p < 0.05\), false discovery rate (FDR) = 10%, two-sample Welch \(t\)-statistics, adjusted with Benjamini and Hochberg FDR controlling procedure). Functional annotation (Fig. 1c and Supplementary Data 5) revealed inflammatory genes upregulated at P1 and P5 in two distinct waves: an immediate response at P1 and a subsequent early response at P5. In contrast, processes related to anatomical development and DNA replication were downregulated at P1, and cell structure and binding-related processes were downregulated at P5. This transcriptional pattern was neutralised by P10 with suggestion of a reversal by P45, which was not quantified further. Systemic IL1B thus induces a rapid transcriptional response in microglia that prioritises inflammatory functions over growth.

**Gene co-expression network analysis.** Gene and protein network-based analysis uncovers processes involved in disease\(^17\), with topological measures being informative of underlying biology.\(^18\) The functionally coherent gene clusters suggested gene co-regulation, and an analysis using graphical Gaussian models\(^19\) revealed characteristic co-expression networks emerging in response to IL1B (IL1B), over time (Development), or with a differential response to IL1B over time (Interaction) (Fig. 2).

The three networks showed distinct gene membership and function, with only 22 genes in common (Supplementary Table 1), and notable topological differences (Supplementary Table 2). The Development network had small world topology with a high clustering coefficient and degree exponent close to 2. In contrast, the IL1B network was bigger and more homogenous, and although there were genes with high degree (i.e., hub genes with many other genes connected to them) this did not result in the formation of obvious subclusters. The Interaction network has a topological structure intermediate to the other two networks (Fig. 2, lower panel, and Supplementary Movie 1). This suggests that IL1B disrupts the normal small world topology, leading to a transcriptional response akin to the genomic storm previously noted in human leukocytes following severe inflammatory stress\(^20\).

Functional enrichment analysis showed that differences in network topology were paralleled by different biological processes and pathways (Supplementary Fig. 5 and Supplementary Data 6), with no overlap between the annotation categories (GO terms or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways) for IL1B and Development networks. The IL1B network was predominantly enriched for GO terms related to defence response, transmembrane signalling and channel activity. The Development and Interaction networks encompassed broad functional categories, spreading the networks’ genes among many different categories, so that enrichment results rarely survive multiple testing correction; nominally significant functional enrichment terms are therefore included for the annotation of the Development and Interaction networks (Fig. 2, lower panel).
**Fig. 1** Overview of in vivo mouse model, and clustering of expression responses to IL1B by time point, with functional annotation summary. a IL1B mouse model experimental set-up. b Clustering of expression profiles in response to IL1B, showing up- and downregulated clusters at each time point (Student’s t test, \( p < 0.05 \), FDR 10%). c Summary functional annotation of four main clusters identified in b. Red = upregulated in IL1B vs. PBS, Blue = downregulated in IL1B vs. PBS. Mouse illustration: “Drawing of a grey mouse”, Author Jan Gillbank, licence CC-BY-3.0. Syringe icon made by Freepik from www.flaticon.com

**Fig. 2** Gene co-expression networks for three responses. Development, IL1B and interaction. Top: summary of functional annotation. Hypergeometric test, * = nominal \( p < 0.05 \), ** = adjusted \( p < 0.05 \). Middle: Gene co-expression networks; nodes = genes, edges = partial correlations, local FDR = \( 1 \times 10^{-13} \). Bottom: Hiveplots for each network illustrating differences in topology between responses, where nodes are arranged on axes according to their degree; axes range clockwise from top: degree < 30, 30 \( \leq \) degree \( \leq \) 80, degree \( x > 80 \)
This annotation implies that differences between gene networks are reflected at the functional level, with IL1B activating immune responses, while Development and Interaction have broader biological functions.

**Protein-protein interactions and neuropsychiatric genes.** We asked whether these co-expression relationships are conserved at the protein level and relevant to neuropsychiatric disorders linked to brain development and prematurity. The genes from all three co-expression networks were combined and their shared interactome investigated by searching for known protein-protein interactions (PPI) to highlight functional interactions, an approach previously shown to be useful for disease gene prioritisation. We used a curated data set to build a robust PPI network and then carried out a power graph analysis (PGA) to identify simplified coherent PPI networks. Power graphs are lossless representations of graphs based on power nodes (sets of nodes brought together) and power edges (connecting two power nodes, so that all the nodes in the first power node are connected to all the nodes in the second power node). To further rationalise the power graph structures, we introduce the term super-power node (SPN) to refer to a set of power nodes that form a connected graph. When applied to all genes present in the three co-expression networks, we identified high-confidence connections between 96 proteins. The PGA revealed that 71 of these proteins belong to either one of two main SPNs: SPN1 or SPN2 (Fig. 3 and Supplementary Fig. 6). To corroborate the identification of these two SPNs independently, we used a separate approach (DAPPLE), which interrogates a large database of experimentally derived PPI to assess the physical connections among proteins encoded by the genes of interest. In this case, the significance of the PPI derived from the input gene list is assessed empirically. This analysis replicated the identification of 23/46 edges from SPN1 and 36/41 edges from SPN2, providing empirical support for the significance of these PPI (permutation-based p value <0.001, Supplementary Fig. 7).

PPIs have rarely been measured in the context of distinct cell types, tissues or in specific disease conditions, making it challenging to model and understand context-related phenotypes, but it has been shown that heterogeneous genomic data contain functional information of protein-DNA, protein-RNA, protein-protein and metabolite-protein interactions.

We therefore sought additional external validation of cell-type specificity for SPN1 and SPN2 by querying the GIANT database of tissue-specific gene networks that includes mapping to tissue and cell-lineage-specific functional contexts. Glia-specific gene interactions within SPN1 and SPN2 were reconstructed with high confidence from this prior experimental data (Supplementary Fig. 8a), further supporting the consistency and specificity of our findings.

Genes coding proteins in the SPNs came from all three gene networks with minimal overlap (Supplementary Fig. 9), suggesting different transcriptional networks (Fig. 2) might converge to less redundant structures at the protein level (Fig. 3), probably reflecting the known modular architecture of PPI networks reflected in the identification of SPN1 and SPN2. Functional annotation of the two SPNs indicated that these are distinct in function, with SPN1 being significantly enriched for proteins involved in nervous system processes and SPN2 for cell signalling and transcriptional regulation (Fig. 3 and Supplementary Table 3). Consistent with this, SPN1 showed tissue expression specific to the brain, whereas SPN2 had a broader tissue distribution (Supplementary Data 7 and 8).
mice. Scale bars (sensorimotor cortex (TEX). IBA1 and DLG4 were exclusively co-localised in mouse microglia in vivo. Double labelling of MACS-isolated mouse microglia showed that SPN1 is significantly and specifically enriched for genes in SPN1 with any of STAT1, STAT3 or STAT5 indicated STAT binding at the promoters of 8/35 (22%) of the enriched gene sets (containing genes that share a common TFBS) and the genes in SPN1/2. This showed a significant association of members of SPN1 (Med13, Kcnj16 and Ifit3) with STAT1 and STAT2 TFs (FDR = 2.8%), further supporting a link between SPN2 and SPN1 via the STAT family of TFs.

To seek additional experimental support for the predicted interaction between SPN1 members and the STATs TFs, we investigated the genomic distribution of binding sites of the family of signal transducers and activators of transcription (STAT1, STAT3 and STAT5) in a data set of lipopolysaccharide (LPS)-stimulated primary microglial cultures. Chromatin immunoprecipitation-promoter microarray data (ChIP-Chip) investigated the genomic distribution of binding sites of the transcriptional targets of STAT3 (signal transducer and activator of transcription 3), as well as other members of the signal transducers and activators of transcription (STAT) family of TFs (STAT6 and STAT1-alpha), and are significantly predicted to bind the promoters of 22/36 (61%) members of SPN1 (hypergeometric test, $p < 0.05$, Supplementary Tables 5–7). The link between the STAT family of TFs and genes in SPN1 was also supported by an independent analysis using the gene set enrichment analysis tool in the molecular signatures database (Broad Institute) to assess the overlap between 615 gene sets (containing genes that share a common TFBS) and the genes in SPN1/2. This showed a significant association of members of SPN1 (Med13, Kcnj16 and Ifit3) with STAT1 and STAT2 TFs (FDR = 2.8%), further supporting a link between SPN2 and SPN1 via the STAT family of TFs.

Transcriptional regulation of SPNs. We examined the potential regulation of SPN1 and SPN2 by transcription factors (TFs) and searched for potential regulatory relationships mediated by TFs, which can determine coordinated expression of several target genes. Analysis of transcription factor-binding site (TFBS) motifs using the PASTA algorithm indicated that an SPN2 member, STAT3 (signal transducer and activator of transcription 3), as well as other members of the signal transducers and activators of transcription (STAT) family of TFs (STAT6 and STAT1-alpha), are significantly predicted to bind the promoters of 22/36 (61%) members of SPN1 (hypergeometric test, $p < 0.05$).

Confirmation of effects of STAT3 TF on SPN members. Using CD11B+ cells MACS-isolated at P1 and exposed to a vehicle solution or IL1B + IFN gamma (chosen because they are highly regulated in the brains of IL1B exposed mice (Gressens, unpublished data) and cause a moderate but consistent inflammatory reaction in vitro), we tested the transcriptional role of STAT3 in the interaction between SPN1 and SPN2 by STAT3 pharmacological inhibition followed by RT-qPCR. MACS-isolated microglia were stimulated by IL1B + IFNg and exposed to vehicle or a small molecule inhibitor of STAT3 (BP-1-102) that binds to the three subpockets of STAT3 SH2 domain and blocks STAT3 phosphorylation, dimerization and DNA-binding activity. Gene expression for a set of genes representing general microglia function markers and a subset of genes from SPN1 and SPN2 was measured by qPCR (Supplementary Table 13). The profiles and features of the markers were previously characterised by us with in vitro microglia. Given the high purity of the MACS-isolated CD11B+ cells (>95% microglia, Supplementary Fig. 1), we employ them as markers of general microglia activity states and a means of assessing the potential role of the cell rather than for the
purposes of cell-type specificity. We used 11 markers broadly grouped as markers of classic pro-inflammatory actions (Pigs2, Cd32, Cd86 and Nos2), immunomodulatory markers (Il1rn, Il1ra, Socs3 and Sphk1) and regenerative function markers (Lgs3, Igf1 and Cd206)39. IL1B + IFNg induced expression of several pro-inflammatory and immunomodulatory markers and repressed expression of regenerative markers as expected. STAT3 inhibition decreased expression of three of the four pro-inflammatory markers (Cd32, Cd86 and Nos2), but had no significant effect on the repression of regenerative markers. By contrast, IL1B + IFNg reduced expression of immunoregulatory marker Socs3, while Sphk1 expression was increased by STAT3 inhibition (Supplementary Fig. 11).

Regarding the analysis of genes from SPN1 (Dlg4 and Notch1), we found no effect on Dlg4 of inflammation or STAT3 inhibition. This lack of response of Dlg4 gene expression to IL1B is supported by the microarray measurements at P1, which show no significant difference between IL1B and control (PBS) (Supplementary Fig. 10, Student’s t test, p = 0.32). However, expression of DLG4 protein in neurons is strongly regulated by post-translational modifications, and these data may not provide reliable information on changes at the protein level. Notch1 gene expression was also unaffected by either inflammation or STAT3 inhibition. For the analysis of SPN2 genes, we observed that IL1B + IFNg exposure significantly increased expression of Stat3, and STAT3 inhibition significantly decreased Arnt2 and Hif1a (Supplementary Fig. 11).

Taken together, these data confirm that IL1B + IFNg exposure induces both pro-inflammatory and immunomodulatory microglial responses, and that STAT3 presence is required for the pro-inflammatory response and typical modulatory component to occur. The transcriptional relationship between SPN1 and SPN2 in the context of inflammation is less clear, and possibly post-transcriptional.

**DLG4 is central to SPN1 and has a novel role in microglia.**

To clarify how the systemic IL1B inflammatory stimulus leads to specific neuronal and neurological effects, we focused on DLG4, a member of the membrane-associated guanylate kinase family. DLG4 has not previously been known to have a function in microglia, but is the hub protein of SPN1 (Fig. 3), and has an established neurodevelopmental role in synaptic plasticity in neurons41. Our data suggest the hypothesis that DLG4 plays a role in the brain’s response to inflammation and may be a potential link between inflammation and both preterm brain injury and neuropsychiatric disease.

**DLG4 at the microglial membrane is dynamically modulated.**

We investigated whether DLG4 protein could be observed in mouse microglia, and whether this is modulated by IL1B and development. We assessed immunofluorescence both in MACS-isolated CD11B+ cells cultured from P1 mice (Fig. 4 and Supplementary Fig. 12), and in tissue sections from pups exposed to IL1B from P1 to P3 (Supplementary Fig. 13). Under control conditions at P1, microglia produce DLG4 protein that is localised to the cell membrane (>95% of IBA1+ cells stained for DLG4) (Supplementary Movie 2). This DLG4 staining disappears by P3 (Supplementary Fig. 13, postnatal day 1 and 3 PBS overlay), and the protein is still absent from IBA1+ cells at P45. Following exposure to IL1B, DLG4 was detected at the microglial membrane at both P1 and P3 (Supplementary Fig. 13, postnatal day 1 and 3 IL1B overlay) but not P45, suggesting a delay rather than a permanent change in development, consistent with current hypotheses42.

**Microglia synthesise endogenous DLG4 in early development.**

To verify that DLG4 protein detected in microglia is endogenously synthesised, rather than the phagocytosed synaptic debris seen later in development43, we stained for DLG4 and lysosomal-associated membrane protein 1 (LAMP1), a marker of lysozymes. No co-localisation of these markers was demonstrable in IBA1+ cells, with LAMP1 immunofluorescence mainly confined to intracellular vesicles and DLG4 consistently and clearly at the microglial cell membrane (Supplementary Fig. 14). A 3D reconstruction of DLG4 and IBA1 immunofluorescence showed evident membranous staining (Supplementary Movie 2). In addition, in the developing cortex at P1 and P3, DLG4 protein staining was absent from cells not expressing IBA1. At P5, we observed DLG4 protein in a limited number of putative cortical neurons, in keeping with the reported normal developmental expression in the cortex in mice and comparative time points in humans44. We also found DLG4 protein at the membrane of MACS-isolated primary mouse microglia maintained ex vivo for 96 h, in the absence of neurons (Supplementary Fig. 12).

**DLG4 is expressed in developing and adult human brain tissue.**

Interrogating the BrainCloud resource45, we found the DLG4 gene expressed in human cortex from early fetal development (14+ gestational weeks (GW)) throughout life, with expression increasing rapidly during the first year (Supplementary Fig. 15a)46. The BrainSpan Atlas of the developing human brain47 additionally suggests a relatively robust expression from early gestation with an apparent decrease in DLG4 expression at around 21 GW, then increasing values from 37 GW (Supplementary Fig. 15b). The Allen Brain Atlas Brain Explorer indicates widespread cortical expression in adulthood (Supplementary Fig. 15c) and the UK Brain Expression Consortium (UKBEC) database48 records DLG4 widely expressed in both grey and white matter in adults (Supplementary Fig. 8c). However, DLG4 mRNA is targeted for degradation via specific post-translational modifications and may not be strictly predictive of protein levels49.

**DLG4 is expressed by microglia in the developing human brain.**

We then assessed cell-type specific expression of DLG4 protein from microglia of the developing human brain in isolated CD11B+ cells and in tissue sections. In MACS-isolated human fetal microglia at 19 and 21 GW (roughly equivalent to P1 in mouse45) DLG4 protein co-localised with IBA1 (Supplementary Fig. 16); this staining appeared to increase with activation of microglia to a pro-inflammatory state using LPS. Further, we stained human fetal brain sections through the dorsal cortex. At 20 GW, there was co-localisation of IBA1 and DLG4 protein (Fig. 5) limited to cells in the proliferative zone, and the absence of any other DLG4-positive cells. At 26 GW, IBA1 and DLG4 again co-localised, but DLG4 was also detected in sparse cells, putative neurons, in the cortex. At 30 GW, we noted very few IBA1 and DLG4 co-localised cells, but many DLG4-positive cells in the cortex.

**DLG4 is thus widely expressed in the human brain during development with synthesis of DLG4 protein by human microglia at a time when there appears to be minimal expression by neurons.**

**DLG4 gene variants and matter structure in preterm infants.**

To assess whether the DLG4 gene might be important in infants surviving preterm birth, we undertook an imaging-genomic analysis. We focused on white matter, which has a well-established correlation with outcome6, 7. 51, and a high
microglia content particularly at this stage of development. d-MRI brain images for two separate cohorts of preterm infants (n = 70 and 271, Supplementary Table 9) were acquired at term-equivalent age, and a measure of white matter microstructure known to be affected by prematurity (fractional anisotropy, FA) was extracted using tract-based spatial statistics (TBSS) (Supplementary Methods). DNA extracted from saliva was genotyped on the Illumina HumanOmniExpress-12 array. Seven single nucleotide polymorphisms (SNPs) on this array mapped to DLG4 (Supplementary Table 10), including rs17203281, which has previously been associated with neuropsychiatric disease, significantly predicting schizophrenia risk and changes in cortical regional volume in patients with Williams’ syndrome, a well-characterised genetic syndrome overlapping with autism.

Infants in both cohorts were categorised by presence or absence of the minor allele (A) for DLG4 at rs17203281 to test for the effect of minor allele load (minor allele frequency (MAF) in cohort 1 (pilot) = 0.22; MAF in cohort 2 (replication) = 0.28), and a general linear model was used to test for a correlation between genotype and white matter FA. This analysis showed a significant difference (p < 0.05, family-wise error (FWE)-corrected by threshold-free cluster enhancement, FDR=10%) in FA between infants with or without the minor allele (A) for SNP rs17203281 in both cohorts (Fig. 6). There were no significant differences in other clinical features (gestational age at birth, age at scan, days of ventilation, incidence of chorioamnionitis, bacterial sepsis or necrotising enterocolitis, Supplementary Table 15) between infants with or without the minor allele. None of the other SNPs at the DLG4 locus had a replicable effect in both cohorts.

To investigate whether the observed effect of common genetic variation in DLG4 is specific to this gene or could be extended to SNPs in other genes as a group, we used a list of all genes in SPN1 as a set of interest and carried out association testing with the phenotype using joint association of genetic variants in the larger cohort of 271 infants. This tests the null hypothesis of there being no more evidence for association of common genetic variants (SNPs) in genes from SPN1 with the phenotype under study than any other random set of an equal effective number of SNPs (see “Methods” section). The SPN1-gene set was tested for association with the real phenotype (query data), and this was repeated with 1000 permutations of the phenotype (self-contained test) to estimate empirical p values. This analysis showed no significant effect of the genes in SPN1 on FA, supporting the specific association of DLG4 with white matter features in the developing brain rather than a wider and more general effect of the larger SPN1 network.

**Fig. 5** Co-localisation of DLG4 protein and microglial marker IBA1 in the developing human brain. Representative photomicrographs from the subplate (SP) and subventricular zone (SVZ) through the dorsal cortex of a 20 GW (upper two rows) and 26 GW (lower row) human brain. No IBA1 or DLG4 co-localisation was observed at 30GW. In green (IBA1 co-localisation) and in red (DLG4 (PSD95) protein) shows co-localisation of IBA1 and DLG4, together with DAPI nuclear staining. Arrows in top rows point to area of distinct co-localisation at ×20 and a higher magnification is shown of a single double-stained cell in the lower panel. Scale bar two top rows = 40 µm and lower row = 10 µm.

**Fig. 6** 3T d-MRI brain images for two cohorts of preterm infants acquired at term-equivalent age (Pilot: cohort 1, n = 70; Replication: cohort 2, n = 271). Images from replication analysis (cohort 2) shown. Views = sagittal, coronal, axial (left-right). Coloured voxels have significantly different diffusion features (fractional anisotropy, FA) between infants with or without the minor allele for DLG4 (n = 271, threshold-free cluster enhancement (TFCE) p value <0.05, FDR=10%). Table inset: replication of findings in two independent cohorts of preterm infants.

eQTL effect of DLG4 in human brain. We then queried whether rs17203281 in the DLG4 region might be regulating DLG4 mRNA expression, and searched for a possible expression quantitative trait locus (eQTL) effect. We analysed DLG4 mRNA expression within individual tissues in the brain, treating the expression levels of the gene as a quantitative trait, so that variations in gene expression that are highly correlated with genetic variation can be identified as eQTLs. A SNP located within or in close proximity to a gene that is significantly associated with the gene’s mRNA variation defines a cis-acting eQTL. We returned to the UKBEC database and extracted genotypes for the rs17203281 plus DLG4 gene expression in the white matter, and compared gene expression levels to genotypes. This analysis suggested that normal individuals homozygote for the rs17203281 minor allele (AA) genotype have a significantly higher expression of DLG4 in white matter than those with the alternate genotypes (AG or GG) (Student’s t test, p < 0.05), suggesting a cis-eQTL effect for rs17203281 (Supplementary Fig. 17).

To cross-validate this result, we queried the GTEx portal, which provides a searchable resource of multiple different human tissues with genotyping, gene expression profiling, whole-genome sequencing and RNA sequencing data. This revealed that DLG4 is preferentially expressed in the brain (Supplementary Fig. 8b),
and has multiple SNPs regulating its expression levels in cis (Supplementary Table 11), of which one SNP (rs3826408) is in high linkage disequilibrium (i.e., is a proxy SNP) with rs17203281 ($r^2 = 0.667, D = 1$). Together, these analyses suggest cis-acting genetic regulation of DLG4 mRNA expression in the brain by rs17203281.

Overlaps between SPNs and neuropsychiatric gene modules. Rare genetic variants in DLG4 have previously been associated with autism, schizophrenia and epilepsy. This identified DLG4 within a small module of 24 genes involved with synaptic function, in which de novo and more severe missense mutations were more likely in individuals with significantly higher intellectual impairment. We investigated the overlaps between the main gene modules reported in ref. 30 and genes in SPN1 and SPN2 using a formal test for intersecting gene lists (Supplementary Fig. 18 and Supplementary Table 12). We found that there were significant overlaps between SPN1/2 and the reported disease-associated modules. In particular, DLG4 was the most frequently occurring gene in these overlaps (11/51 occurrences) followed by another gene from SPN1, CAMK2A (4/51). Notably, SPN1/2 were broadly captured by the autism modules 1/2 reported in ref. 30.

Discussion

The primary finding of this study is that DLG4 is expressed in microglia during early development in a developmentally regulated expression pattern that is altered by neuroinflammation associated with brain damage in preterm born infants. This is a novel role for DLG4, previously considered an archetypal marker of the neuronal post-synaptic density.

The IL1B administration model used here captures essential features of the encephalopathy of prematurity including hypomyelination linked to oligodendrocyte maturation arrest, microglial activation, cognitive deficits, decreased fractional anisotropy on MRI and axonopathy. Systemic IL1B generated two waves of gene expression in microglia, with upregulation of inflammatory genes and downregulation of genes involved in growth, creating a genonic storm that altered the normal gene and protein network topology. DLG4 was the hub of the SPN1 protein network that is highly enriched with genes implicated in neurocognitive disorders. SPN1 has a star configuration, which appears to maximise network efficiency at the expense of robustness, since perturbing the central hub broadly disrupts the system. Hub proteins are important for cellular growth, under tight regulation, continuously evolving, and more intrinsically disordered than proteins with fewer relations. DLG4 is the only member of either SPN1 or SPN2 that qualifies as a hub, and is significantly enriched for disorder promoting amino acids (Student’s t test, $p < 0.02$).

SPN1 interacted with SPN2 via the transcriptional regulation by STAT3, a TF previously linked to microglial activation and neuroprotection pathways. The combined PPI and TF analysis revealed a possible functional relationship between two protein interaction sub-networks. The same relationship was not observed between the STAT3’s TFs and genes in SPN2.

Microglia DLG4 protein synthesis at P1 disappears by P3, unless this developmental pattern is disrupted by inflammation leading to the temporary persistence of DLG4 at the cell membrane, which resolves by P15. The disappearance of DLG4 protein between P1 and P3 could be due to post-transcriptional mechanisms involving micro-RNAs, mRNA degradation and post-translational mechanisms affecting protein localisation as previously observed in neurons (review in ref. 30). Inflammation-induced persistence of DLG4 protein may be a previously unappreciated mechanism of inflammatory brain injury.

In human infants, we found developmentally regulated DLG4 gene expression throughout brain tissue. In mouse and human microglia, DLG4 protein localised to cell membranes and was not found within cytosolic vesicles: DLG4 detection was thus not due to the participation of microglia in synaptic pruning seen later in development, reflecting the well-known role of DLG4 in synapse structure and development. Indeed this is consistent with previous data showing that DLG4 mRNA was not expressed in neurons before P5-10.

Common genetic variation in DLG4 was associated with structural white matter changes in two independent infant cohorts, possibly due to measured differences in expression of DLG4 mediated through a posited cis-eQTL action of rs17203281, suggesting that inter-individual genetic variability in DLG4 gene could affect the response to perinatal inflammation. Preterm infants have an increased risk of developing autism spectrum disorders (ASD) and other neuropsychiatric disorders, and DLG4 has been consistently associated with neuropsychiatric diseases including ASD and schizophrenia. Common genetic variation in DLG4 (rs17203281) was associated with structural white matter changes in two independent infant cohorts, possibly due to measured differences in expression of DLG4 mediated through a posited cis-eQTL action of rs17203281, suggesting that inter-individual genetic variability in DLG4 gene could affect the response to perinatal inflammation. Preterm infants have an increased risk of developing autism spectrum disorders (ASD) and other neuropsychiatric disorders, and DLG4 has been consistently associated with neuropsychiatric diseases including ASD and schizophrenia.

The observed changes in d-MRI are characteristic of encephalopathy of prematurity. However, d-MRI provides no information on the cell types involved in these microstructural abnormalities, and the relationship between DLG4 variability and brain structure cannot be attributed specifically to microglia. Nevertheless, white matter is highly populated with microglia in the perinatal period in both humans and mouse, and even in adults microglia are typically much more numerous in white matter than in neocortex. TBSS provides a powerful approach to d-MRI data, providing group analyses that overcome problems of partial volume effects. However, TBSS is highly dependent on the quality of the image registrations, and as this is particularly challenging in developing brain, we have created an optimised neonatal pipeline and characterised the sensitivity of TBSS in this population. In addition, although TBSS is a powerful technique for identifying associations with white matter microstructure, it is less reliable for precise spatial localisation within white matter.

The mechanism of DLG4 action in microglia remains unclear. In vitro, inhibition of the interaction between DLG4 protein and the N-methyl-D-aspartate receptor N2B subunit (which has a moderate but significant effect on microglial activation state and response to perinatal injury) had no effect on microglia phenotype. We conjecture that the role of DLG4 in microglia may be in the domain of cell-cell communications, such as cross-talk between oligodendrocytes, astrocytes and microglia, or regulation of glutamatergic and GABAergic signalling. Of interest, DLG4 may be involved with the clustering and activity of inwardly rectifying potassium channels (Kir) predominantly expressed in glial cells. Microglial Kir have functional effects on microglial activity and could have therapeutic applications in Alzheimer’s and Parkinson’s diseases. DLG4 has also been observed in oligodendrocytes and may mediate Kir-related myelination.

The inflammatory reflex involving the vagus nerve could attenuate the IL1B response in vivo, and microglia activity may be modulated via microglial α7 nicotinic acetylcholine receptors (α7AChRs) that limit microglial activation, possibly signalling via NRG/ErbB4 interactions with DLG4 in microglia or via endothelial COX2. This might be a fruitful area for further discovery. It is also unclear how much the microglial response is conditioned by cellular development and context, and whether...
this might differ between the in vivo and in vitro settings, as well as between mouse and human. The topic of microglial development including how to best characterise it in relation to function requires further work.

Methods
Animal model. Experimental protocols were approved by the institutional guidelines of the Institut National de la Santé et de la Recherche Scientifique (Inserm, France), and met the guidelines for the United States Public Health Service’s Policy on Humane Care and Use of Laboratory Animals (NIH, Bethesda, MD, USA). The experimental set-up for inducing inflammation-induced white matter injury in the mouse has previously been described in detail. In brief, a 5 µl volume of phosphate-buffered saline (PBS) containing 10 µg/kg injection of recombinant mouse IL1B or PBS alone (control) was injected intra-peritoneally twice a day (morning and evening) on days postnatal P1-P4 and once in the morning on day P5. Animals were sacrificed 4 h after the morning injection of IL1B at P1, P5, P10 and P45. Brains from animals at P1, P3 and P10 were also used for immunofluorescence with antibodies to detect DLG4, ((66G6-1C9) (Product# MA1-045), Thermo Scientific; 1:500), ((IBA1) (ab5076, Abcam; 1:400) and LAMP1 (L1418, Sigma; 1:200). Details in Supplementary Methods.

CD11B+ microglia MACs in mouse. Brains were collected from mice for cell dissociation, and microglia were isolated by magnetic antibody-based cell sorting (MACS) using CD11B antibody (Miltenyi Biotech, Bergisch Gladbach, Germany, dilution 1:10) (details in Supplementary Methods) according to the manufacturer’s protocol using all recommended reagents and equipment. The purity of MACSed CD11B+ fractions was validated using FACS analysis of CD11B fluorescence, and the purity was further validated with RT-qPCR of the positive and negative CD11B cell fractions. Details in Supplementary Methods.

Gene co-expression network reconstruction. Mouse microglial RNA was extracted and hybridised to Agilent Whole Mouse Genome Oligo Microarrays (8 × 60 K). Three separate gene co-expression networks were reconstructed from these data, showing a significant response to IL1B, Development and Interaction effect (Fig. 2). Details in Supplementary Methods.

PPis and transcriptional regulation. The nodes of all three gene networks (IL1B, Development and Interaction) were aggregated into one list and used to investigate protein interactions with a PGA3, and two super-power nodes (SPNs) were aggregated into one list and used to investigate protein networks with power graph analysis.

Mouse cell culture studies. MACS-isolated microglia were stimulated by IL1B + IFNγ and exposed to one of two vehicle; a small molecule inhibitor of STAT3 (BP-1-102; Merck Millipore) or TAT-N-dimer (a inhibitor of the DLG4 protein NMDA receptor interaction; Merck Millipore). At the end of the treatment period, cells were harvested and used for immunofluorescent labelling; mRNA was extracted for gene expression analysis, and supernatant was collected for cytokines/nitric oxides. Measurements. Antibodies used were a mouse monoclonal antibody to detect DLG4, ((66G6-1C9) (Product# MA1-045), Thermo Scientific; 1:500), a goat polyclonal antibody to detect IBA1 Ionized calcium binding adaptor molecule 1 ((IBA1) (ab5076, Abcam; 1:400)), and a rabbit polyclonal antibody to detect Lysoosomal-associated membrane protein 1 ((LAMP-1) (L1418, Sigma; 1:200)). Details in Supplementary Methods.

Developing human brain post-mortem studies. All human post-mortem tissue (cells and tissues) was acquired with ethical approval at The French Agency of Biomedicine (Agence de Biomédecine; approval PFS12-0011). Written informed consent was received prior to donation of foetal tissue. For the collection of human microglia, post-mortem tissue without any neuropathological alterations was acquired from two individuals at 19 and 21 GW. MACS-isolated microglia were obtained and used for treatment with DMEM (control) or LPS, as well as pyrosequencing with antibodies to detect DLG4, ((66G6-1C9) (Product# MA1-045), Thermo Scientific; 1:500), a goat polyclonal antibody to detect Lysosomal-associated membrane protein 1 ((LAMP-1) (L1418, Sigma; 1:200)). Details in Supplementary Methods.

Infant MRI studies. Research was carried out in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), with approval from the NHS National Research Ethics Service (NRES), to the standard of the associated granting agencies. Two independent cohorts of preterm infants (cohort 1: n = 70; cohort 2: n = 271) were imaged at term-equivalent age on Philips 3-Tesla system (Philips Medical Systems, the Netherlands), with acquisition of structural (T2-weighted fast-spin echo MRI and 3D-MPRAGE) and diffusion (single-shot echo-planar diffusion tensor imaging (EPI DTI)) sequences. Details in Supplementary Methods.

Infant genotyping. Saliva samples for both infant cohorts above were genotyped on the Illumina HumanOmniExpress-12 array, as previously described (Supplementary Methods). The genotype matrix was recoded in terms of minor allele counts, including only SNPs with MAF ≥5% and ≥99% genotyping rate. Details in Supplementary Methods.

Data availability. All new data are available from the authors on request, subject to ethical restrictions. Expression profiling of human dorsolateral prefrontal data is accessible in the NCBI GEO database under accession code GSE30272.

References
1. Ball, G. et al. The effect of preterm birth on thalamic and cortical development. Cereb. Cortex 22, 1016–1024 (2012).
2. Verney, C. et al. Microglial reaction in axonal crossroads is a hallmark of noncytotic periventricular white matter injury in very preterm infants. J. Neuropathol. Exp. Neurol. 71, 251–264 (2012).
3. Biliards, S. S. et al. Myelin abnormalities without oligodendrocyte loss in periventricular leukomalacia. Brain Pathol. 18, 153–163 (2008).
4. Johnson, S. & Wolke, D. Behavioural outcomes and psychopathology during adolescence. Early Hum. Dev. 89, 199–207 (2013).
5. Nosarti, C. et al. Perinatal birth and psychiatric disorders in young adult life. Arch. Gen. Psychiatry 69, E1–E8 (2012).
6. Krishnan, M. L. et al. Relationship between white matter apparent diffusion coefficients in preterm infants at term-equivalent age and developmental outcome at 2 years. Pediatrics 120, e604–e609 (2007).
7. Ball, G. et al. Thalamocortical connectivity predicts cognition in children born preterm. Cereb. Cortex 25, 4310–4318 (2015).
8. Nelson, K. B., Dambrosia, J. M., Grether, J. K. & Phillips, T. M. Neonatal cytokines and coagulation factors in children with cerebral palsy. Ann. Neurol. 44, 665–675 (1998).
9. Leviton, A. & Greisen, P. Neuronal damage accompanies perinatal white-matter damage. Trends Neurosci. 30, 473–478 (2007).
10. Dammann, O. & Leviton, A. Intermittent or sustained systemic inflammation and the preterm brain. Pediatr. Res. 75, 376–380 (2014).
11. Perry, V. H., Hume, D. A. & Gordon, S. Immunohistochemical localization of macrophages and microglia in the adult and developing mouse brain. Neuroscience 15, 313–326 (1985).
12. Crottì, A. & Ransohoff, R. M. Microglial physiology and pathophysiology: insights from genome-wide transcriptional profiling. Immunity 44, 505–515 (2016).
13. Tremblay, M. E. The role of microglia at synapses in the healthy CNS: novel insights from recent imaging studies. Neuron Glia Biol. 7, 67–76 (2011).
14. Favrais, G. et al. Systemic inflammation disrupts the developmental program of white matter. Ann. Neurol. 70, 550–565 (2011).
15. Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haesaüelin, L. J. Brain development in rodents and humans: identifying benchmarks of maturation and vulnerability to injury across species. Prog. Neurobiol. 106–107, 1–16 (2013).
16. Supek, F., Bosniak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE 6, e21800 (2011).
17. Greene, C. S. et al. Understanding multicellular function and disease with human tissue-specific networks. Nat. Genet. 47, 569–576 (2015).
18. Sun, K., Goncalves, J. P., Larrinio, C. & Pruzil, N. Predicting disease associations via biological network analysis. BMC Bioinformatics 15, 304 (2014).
19. Oppen-Rhein, R. & Strimmer, K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Syst. Biol. 1, 37 (2007).
20. Xiao, W. et al. A genomic storm in critically injured humans. J. Exp. Med. 208, 2581–2590 (2011).
21. Chen, J., Aronow, B. J. & Jegga, A. G. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics 10, 73 (2009).
22. Wang, Y., Thilmony, R. & Gu, Y. Q. NetVenn: an integrated network analysis web platform for gene lists. Nucleic Acids Res. 42, W161–W166 (2014).
23. Royer, L., Reimann, M., Androunouls, B. & Schroeder, M. Unraveling protein networks with power graph analysis. PLoS Comput. Biol. 4, e1000108 (2008).
24. Rossin, E. J. et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 7, e1001273 (2011).
25. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat. Rev. Genet. 16, 441–458 (2015).

26. Johnson, M. R. et al. Systems biology and gene networks in human embryonic and fetal cerebral cortex. Nature 477, 623–628 (2011).

27. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. Neuroscience 251, 754–769 (2013).

28. Robison, A. J. Emerging role of CaMKII in neuropsychiatric disease. Trends Neurosci. 35, 442–450 (2012).

29. Chhor, V. et al. Characterization of phenotype markers and neuronotoxic potential of polarised primary microglia in vitro. Brain Behav. Immun. 32, 25–36 (2013).

30. Verderio, C. et al. Myeloid microvesicles are a marker and therapeutic target for microglia activation. Nat. Rev. Immunol. 12, 259–272 (2012).

31. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

32. Lassman, P. L. et al. Microglia and the central nervous system in health and disease. J. Neuroinflammation 10, 142 (2013).

33. Klinkert, K. et al. Microglia as a mediator of early brain injury in newborns. J. Neuroinflammation 11, 58 (2014).

34. Tsirka, S. E. et al. Microglia as a mediator of early brain injury in newborns. J. Neuroinflammation 11, 58 (2014).

35. Ekdahl, C. T. et al. Microglia as a mediator of early brain injury in newborns. J. Neuroinflammation 11, 58 (2014).

36. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

37. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

38. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

39. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

40. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

41. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

42. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

43. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

44. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

45. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

46. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

47. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

48. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

49. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

50. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

51. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

52. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

53. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

54. Lassman, P. L. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

55. Heneka, M. T. et al. Microglia in neurodegenerative diseases. Acta Neuropathol. 127, 759–782 (2014).

Acknowledgements

Our thanks to the children and families who participated in the study, and the nurses, doctors and scientists who supported the project. This work was supported by grants from Inserm, Université Paris Diderot, Université Sorbonne Paris-Cité, Investissement d’Avenir (ANR-11-IDBS-0011, NeurATRIS), ERA-ETN Network (Micromet), EU FP7 (Inflammation, FP7-22284, FP7-22284), The Swedish Research Council (2015-02493), Wellcome Trust (WSCR P32674) and The Swedish Research Council Foundation for the Advancement of Research in Cerebral Palsy (SPLP 43281) and the Swedish Research Council (2015-02493). MRI
scans of preterm infants were in part obtained in an independent programme of research funded by the National Institute for Health Research (NIHR) Programme Grants for Applied Research Programme (RP-PG-0707-10154) and in the future a compendium report of the programme will be published in the NIHR Journal. The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, MRIC, CCF, NETSCC, the Programme Grants for Applied Research programme or the Department of Health. In addition, the authors acknowledge financial support from the Department of Health via the NIHR comprehensive Biomedical Research Centre award to Guy’s & St Thomas’ NHS Foundation Trust in partnership with King’s College London and King’s College Hospital NHS Foundation Trust, as well as support from the Medical Research Council (MRC) through Strategic Grant to A.D.E. and Clinical Training Fellowship to M.L.K. (Grant ref: MR/L001578/1), Duke-NUS Medical School and Singapore Ministry of Health to E.P.

Author contributions
M.L.K., E.P., P.G. and A.D.E designed the research; M.L.K., J.V.S., A.-L.S., J.Y., J.A., T.L.C., Z.C., P.D., S.C., C.A., L.T., J.P., G.B., J.P.B., A.J.W., B.F. and P.G. performed the research; A.S. and G.M. contributed genotyping of the replication cohort; M.L.K., J.V.S., A.-L.S., J.Y., J.A., T.L.C., Z.C., P.D., S.C., C.A., L.T., G.B., B.F., P.G. and E.P. analysed the data; M.L.K., J.V.S., J.A., B.F., A.D.E., E.P. and P.G. wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00422-w.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017