

Neuroinflammation — using big data to inform clinical practice

Calliope A. Dendrou¹, Gil McVean² and Lars Fugger¹

¹Oxford Centre for Neuroinflammation, Nuffield Department of Clinical Neurosciences, and MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK.

²Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, and Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.

Correspondence to L.F., lars.fugger@imm.ox.ac.uk

Abstract | Neuroinflammation is emerging as a central process in many neurological conditions, either as a causative factor or as a secondary response to nervous system insult. Understanding the causes and consequences of neuroinflammation could, therefore, provide insight that is needed to improve therapeutic interventions across many diseases. However, the complexity of the pathways involved necessitates the use of high-throughput approaches to extensively interrogate the process, and appropriate strategies to translate the data generated into clinical benefit. Use of ‘big data’ aims to generate, integrate and analyse large, heterogeneous datasets to provide in-depth insights into complex processes, and has the potential to unravel the complexities of neuroinflammation. Limitations in data analysis approaches currently prevent the full potential of big data being reached, but some aspects are already yielding results. The implementation of ‘omics’ analyses in particular is becoming routine practice in biomedical research, and neuroimaging is producing large sets of complex data. In this Review, we evaluate the impact of the ‘Big Data’ drive on neuroinflammation in disease. We describe the breadth of big data that are leading to an evolution in our understanding of this field, exemplify how these data are beginning to be of use in a clinical setting, and consider possible future directions.

The action of the immune system is fundamental to our body’s ability to fight off infection, respond to injury and maintain homeostasis and a state of health. This action requires tightly regulated orchestration of various soluble mediators and cell types that sense pathogenic insults and tissue damage, transduce these signals to

coordinate and amplify an effector response, and ultimately attenuate this response to enable tissue repair. In the CNS, microglia are the main resident innate immune cells. Under physiological conditions, microglia survey their surroundings, produce neurotrophic factors and influence tissue remodelling during fetal development, childhood and puberty by pruning synaptic connections¹. In response to infection or CNS injury, microglia become activated and switch from an anti-inflammatory to a proinflammatory phenotype that causes stimulation of other CNS cell types; these cells include astrocytes, which can also become proinflammatory and engage other components of the immune system²⁻⁴. Typically, this response is self-limiting, but if inflammation persists, it can contribute to a variety of long-term neurological diseases, including autoimmune conditions such as multiple sclerosis (MS) and neuromyelitis optica, as well as conditions not considered to be primarily inflammatory such as Alzheimer disease (AD) and Parkinson disease (PD). Neurological disorders and their sequelae are a staggering burden on global health (**Box 1**) (ref. 5), and the increasingly apparent ubiquity of neuroinflammation in these conditions suggests that it represents an important avenue for designing new strategies for disease monitoring, management and therapeutic intervention.

The complexity of the molecular and cellular pathways involved in neuroinflammation means that an understanding of exactly how it contributes to disease probably requires a similarly sophisticated approach to the investigation of pathophysiological mechanisms across different neurological disorders. Technological advances have given us the ability to generate and analyze high-throughput and in-depth data from hundreds of thousands of individuals, even down to the level of single cells, which can help to meet this challenge. The capacity to translate such 'big data' into information with practical medical benefit is, however, a fundamental challenge^{6,7} that currently prevents us from fulfilling the potential of these approaches. Nevertheless, some aspects of the big data drive are starting to yield results.

In this Review, we consider what constitutes big data, and exemplify how some elements of big data are already shaping our concept of neuroinflammation, particularly in the context of neurodegenerative diseases. We also consider how big data might start to be of pragmatic use in healthcare provision in the context of neuroinflammation. We specifically address how big data can aid the diagnosis and prognosis of neurological disorders, and whether they can facilitate the identification and prioritization of neuroinflammatory drug targets according to predictions of efficacy and adverse effects. Lastly, we discuss the potential big data offers for

appreciating heterogeneity within and between patients, and evaluate the feasibility of big-data-driven personalized medicine for neurological diseases.

[H1] Inflammation in neurological disorders

The involvement of neuroinflammation in the development and progression of neurological disease is particularly evident in CNS-targeted autoimmune conditions, such as MS⁸. The CNS is normally an immune-privileged site, relatively protected from the influence of the peripheral immune system^{3,9}, but MS is characterized by infiltration of immune cells from the blood stream (**Fig. 1a**), predominantly macrophages and activated microglia in early disease, followed by adaptive immune cells — mainly CD8⁺ T cells and, to a lesser extent, CD4⁺ T cells, B cells and antibody-secreting plasma cells¹⁰. The resulting inflammation damages the axonal myelin sheath and myelin-producing oligodendrocytes, causing demyelination in the vicinity of the lesion. As the disease develops, infiltrates become more widespread, glial cell activation becomes more pronounced, and demyelination and neuroaxonal injury become more widespread¹¹. Eventually, infiltration wanes and fewer invading cells are present within lesions in late disease, but microglia, macrophage and astrocyte activation is sustained. These cells continue to produce neurotoxic inflammatory mediators, including cytokines, reactive oxygen species (ROS) and reactive nitrogen species (RNS). As a consequence, neurodegeneration progresses¹².

Beyond autoimmune diseases, neuroinflammation is increasingly recognized as an important component of primary neurodegenerative diseases, including AD (**Fig. 1b**), PD, Huntington disease (HD) and amyotrophic lateral sclerosis (ALS), and other neurological conditions, such as schizophrenia. This importance is despite the fact that peripheral immune cell infiltration of the CNS is not a typical feature of these disorders^{2,13,14}; chronic microglia and astrocyte activation is usually prominent instead. In primary neurodegenerative diseases, this activation is thought to be triggered by accumulation of peptide or protein aggregates — amyloid- β (A β), α -synuclein, mutant huntingtin and mutant superoxide dismutase 1 (SOD1) in AD, PD, HD and ALS, respectively — leading to a vicious cycle of inflammation, neurotoxicity and neuronal cell death. The mechanistic parallels between these conditions suggest the possibility that neuroinflammatory responses across the diseases are similar, and cytokines such as IL-1 β , IL-6 and tumour necrosis factor (TNF), and neurotoxic factors such as ROS and RNS, are widely implicated in the pathogenesis of the diseases^{2,13,14}.

Emerging evidence indicates that the exact nature and role of the neuroinflammatory response in neurological diseases is much more complex than previously anticipated. For example, genome-wide association studies (GWAS) have shown that genetic variation in *CD33* is associated with AD^{15,16} but not with other neurodegenerative diseases. Evidence indicates that this gene is involved in regulating clearance of debris by microglia, suggesting that microglial dysfunction in AD does not only lead to chronic production of inflammatory mediators, but also to impaired neuroprotection owing to insufficient phagocytosis of A β ^{16–18}. In schizophrenia, evidence that the *complement C4A* gene is associated with the disease indicates that increased complement activity in the CNS promotes disease by driving excessive synaptic pruning, probably via microglia¹⁹. The idea that aberrant synaptic pruning mediated by crosstalk between the complement system and microglia is involved in AD is also gaining momentum, supported by genetic associations¹⁵ and functional follow-up investigations²⁰. Similarly, an intriguing study using a mouse model of ALS, in which pathology is driven by expression of the human disease-associated *SOD1* variant, has found that astrocytes can have pleiotropic effects, mediating cytotoxicity through a function similar to that of natural killer cells when levels of major histocompatibility complex class I are reduced on motor neurons²¹. The transpiring diversity of inflammatory responses in neurological disease demonstrates the need for investigative approaches that can interrogate the breadth and depth of the processes involved in neuroinflammatory mechanisms.

[H1] Big data in neuroinflammatory research

The concept of big data is constantly evolving and expanding, such that all data could be included. However, a commonly used definition of big data takes into account volume (the amount of data in a given data set), variety (heterogeneity of data owing to integration of information from different types of sources) and velocity (the speed of data integration and analysis)²². In biomedicine, potential sources of big data include the laboratory (genomics and phenomics research), the clinic (biometrics, medical imaging, patient-reported health surveys, clinical registries and electronic medical records), administrative databases (which provide information about health-associated care delivery, insurance and service reimbursement) and the Internet (through, for example, social media and web-based applications) (**Fig. 2**). The use of big data typically requires non-traditional, non-hypothesis-driven analytical methodologies, such as machine learning, cluster analyses and Bayesian

approaches, to find meaningful patterns amidst multidimensional, multiparametric data²². However, these methods are not without challenges, and the types of big data currently available are not without limitations²³ (**Box 2**); until these challenges and limitations are overcome, an important, albeit relatively constrained, application of big data is their integration, mining and filtering to effectively, quickly and accurately address specific hypotheses and research questions.

Neuroinflammation is defined by complex molecular and cellular processes — as a result, much of the big data that has been generated and applied in this context has come from use of ‘omics’ technologies. It is anticipated that as other types of big data, including imaging, epidemiological, lifestyle and behavioural, clinical and para-clinical information, are amassed and integrated these too will increasingly contribute to the understanding of neuroinflammatory diseases. To date, with the advent of high-throughput genomics²⁴, establishing the genetic basis of neurological disorders — from common, complex diseases, studies of which require genotypic screening of thousands of patients and controls^{15,25–27}, to rare, monogenic conditions that necessitate complete sequencing of the entire genome or exome^{28–30} — has become a mainstream practice in biomedical research. Deciphering the functional consequences of the identified disease-associated genetic variants has necessitated development of similarly high-throughput phenotyping technologies³¹. The challenge now is to translate the outputs of these high-throughput approaches into clinical practice and meaningful benefits for patients.

[H2] Genomics

Genetic associations with common neurodegenerative diseases, such as MS and AD, and identified by GWAS have confirmed the relevance of inflammatory responses in these diseases, and indicate that neuroinflammation is not merely a secondary response in many of these conditions, but has a central role in pathogenesis^{15,25}. Over 100 MS-associated loci have been identified, and the genes thought to drive these associations are predominantly immunological²⁵. *Human leukocyte antigen (HLA)* genes are the strongest genetic determinants, emphasizing the importance of interactions between antigen-presenting cells and CD4⁺ and CD8⁺ T cells in MS pathophysiology³². Many non-*HLA* candidate genes are also likely to be involved in innate and adaptive immune cell function, and efforts to organize these genes into networks implicate the IL-2, nuclear factor- κ B, and TNF superfamily signalling pathways, amongst many others^{33,34}.

In MS, associated genomic regions indicate a greater overlap of the disease with other autoimmune diseases^{35,36} than with other neurodegenerative conditions, consistent with the hypothesis that infiltration of peripheral immune cells drives demyelination and neuroaxonal degeneration. In AD, by contrast, many associated loci also contain immunological candidate genes, but these genes are predominantly implicated in myeloid cell function, suggesting that microglia are central to pathogenesis. In addition to the role of *CD33* discussed above, candidate genes such as *INPP5D*, *PILRA* and *ABCA7* are thought to influence microglial and macrophage signalling, phagocytosis, proliferation and survival¹⁵. Similarly, genetic variation in the *Triggering receptor expressed on myeloid cells 2 (TREM2)* gene, which is involved in microglial activity, is associated with AD³⁷⁻⁴¹ and other neurological disorders, including Nasu–Hakola disease⁴², frontotemporal dementia^{43,44}, PD⁴⁴ and ALS⁴⁵. As for MS, variation in the *HLA* region has been associated with AD¹⁵, frontotemporal dementia⁴⁶ and PD²⁶. Given the absence of evident adaptive immune cell involvement in these disorders, however, the interpretation of these associations is not obvious; notably, in schizophrenia, associations with the *HLA* region²⁷ are driven not by *HLA* alleles but by complement gene haplotypes in the same region¹⁹.

For less common neurological conditions in which rare variants are primarily implicated and genome-wide screening is inadequate, next-generation sequencing has been of particular use for identifying associated genes. For example, exome sequencing has revealed a role of the *TBK1* gene in ALS^{28,29}. This gene encodes the non-canonical I κ B kinase TANK-binding kinase 1, which phosphorylates proteins involved in the innate immune response and autophagy, including optineurin, which has also been genetically implicated in ALS⁴⁷. Rare variants can also contribute to common neurological diseases, as suggested in schizophrenia by a higher load of *de novo* mutations⁴⁸⁻⁵¹. In this case, however, attributing the variant load to specific risk genes has been problematic, partly because there are ~100 loss-of-function variants and >10 complete gene inactivations per genome, suggesting considerable redundancy in the human genome⁵². A recent analysis of thousands of patients and tens of thousands of controls, however, provides strong evidence for an association of rare loss-of-function variants in the *SETD1A* gene (which is involved in the histone methylation pathway) with schizophrenia and other neurodevelopmental disorders⁵³.

Whether disease-associated genetic variants are rare or common, determining their biological consequences requires investigation of their impact on phenotype at the

molecular and cellular levels. The advent of CRISPR–Cas9 genome engineering⁵⁴ used in conjunction with induced pluripotent stem cell-derived CNS cell types^{55,56} is likely to greatly facilitate functional studies, particularly to assess the biological effect of rare variants. To date, however, analysis of common variants has mainly relied on interrogation of genotype-to-phenotype correlations with non-hypothesis-driven, high-throughput methodologies. These approaches are needed to study the many tens of loci that are associated with each common neurological disorder, and to avoid the need for generating accurate hypotheses about which variants affect which genes. Formulating such hypotheses is difficult because even if an associated genomic region includes genes with seemingly disease-relevant functions, the majority of common variants implicated in disease are non-coding and could regulate genes that are hundreds of kilobases away⁵⁷.

[H2] Phenomics

[H3] Epigenomics. Epigenomic data, which collectively describe the DNA modifications that indicate gene activation states, can provide information about how neuroinflammation is involved in neurological conditions in several ways. Co-localization of disease-associated genetic variants with specific epigenetic marks in specific cell types can provide insight into disorder-specific genetic effects. One such example is the distinct localization of an MS-associated polymorphism at the *IL2RA* locus within a super-enhancer specific to CD4⁺ T cells⁵⁸. Such findings can be added to by probing the physical interactions between regulatory variants and the promoters of the genes they regulate with chromatin conformation capture technologies^{59,60}.

In addition, epigenomic information can aid characterization of the spatiotemporal landscape of gene regulation within the CNS. DNA methylation patterns in post-mortem brain tissue suggest that, in MS, pathology-free regions of the CNS have an altered sensitivity to damage⁶¹. In AD, early alterations in brain DNA methylation indicate that changes in modulation of gene expression — including genes implicated in microglial function — have a role in disease onset⁶².

[H3] Transcriptomics. Transcriptomics are also proving to be a powerful tool for investigating neuroinflammatory signatures in disease. For example, analysis of expression quantitative trait loci in human myeloid cells indicates that gene expression differences driven by AD-associated genetic variants are enriched in these cells as opposed to other immune cell types⁶³. Mouse studies have shown how

microglial transcriptional signatures change with age⁶⁴ and how they compare to signatures of other myeloid cells, thereby revealing a unique profile dependent on transforming growth factor- β ⁶⁵. Similar studies in the context of disease have revealed that microglial transcriptional signatures are skewed at the onset of MS model disease experimental autoimmune encephalomyelitis (EAE) in a way that suppresses cellular metabolism⁶⁶, and skewed in the presence of mutant huntingtin protein in a way that promotes a proinflammatory phenotype¹⁴. Transcriptomic analyses of human brain samples are also helping to uncover a role for neuroinflammation in neurological disease. For example, in patients with autism spectrum disorder (ASD), genes co-expressed in patient but not control brain tissue include those with immune and glial function^{67,68}. Similarly, an integrated systems approach to the characterization of RNA-level molecular signatures in post-mortem tissue from patients with late-onset AD has highlighted the differential expression of immune-specific and microglia-specific gene modules⁶⁹.

Transcriptional analysis can also reveal spatial variation in gene expression across the CNS, an additional factor in disease development. For example, profiling of murine astrocytes suggests that these cells display region-related heterogeneity, although the implications of this heterogeneity on pathophysiological mechanisms are yet to be elucidated⁷⁰. Specialized transcriptional analyses, such as B cell receptor sequencing, have been used to map the clonal relationship of B cells in MS, and indicate that these cells traffic in and out of the CNS, an observation that supports their peripheral targeting for therapeutic benefit^{71,72}. A considerable methodological development that is set to transform our understanding of cellular heterogeneity is the ability to perform highly parallel, single-cell transcriptomics^{73,74}; for instance, application of this technology to mouse retinal tissue has defined 39 transcriptionally distinct cell populations, including potentially novel cell subsets⁷⁴.

[H3] Proteomics. Use of protein-level analyses in the context of neuroinflammation has largely been limited to confirmation of RNA-level findings, assessment of selected cerebrospinal fluid biomarkers, and the addressing of very specific research questions. Advances in mass spectrometry-based approaches, including mass cytometry⁷⁵, and proteomics of human^{76,77} and murine⁷⁸ tissues is likely to further increase our understanding of molecular and cellular diversity in the CNS and in neurological disease⁷⁹, especially as some analyses suggest that transcriptional signatures do not perfectly correspond to cell states defined by protein expression^{78,80}.

[H3] Lipidomics. Lipidomic investigations using lipid antigen microarrays and lipid mass spectrometry have revealed that naturally occurring phosphatidylserine and oxidized phosphatidylcholine derivatives have an anti-inflammatory effect in EAE, and might have therapeutic potential in MS⁸¹. Conversely, specific glycolipids, such as lactosylceramide, can promote astrocyte and, subsequently, microglial activation in EAE⁸². In a mouse model of AD, lipid bilayer-delimited exosomal vesicles that are shed from microglia contribute to tau protein propagation, and hence tauopathy progression⁸³.

[H2] Environmental ‘omics’

Despite considerable progress in understanding the role of neuroinflammation in neurological disorders through the identification of genetic associations and follow-up phenotyping analyses, the importance of non-genetic contributions is increasingly recognized. High-throughput genomic technologies have been extended to enable comprehensive characterization of the virome⁸⁴ and microbiome⁸⁵; initial studies in mice demonstrate that commensal microbiota can regulate the maturation and function of microglia within the CNS⁸⁶, and can contribute to demyelination in spontaneous EAE⁸⁷. A study of human gut microbiome diversity based on 16S rRNA sequencing has demonstrated that the microbiome differs between patients with MS and healthy controls, and between treated and untreated patients, although further investigations are required to determine whether these differences contribute to or are biomarkers of MS pathogenesis⁸⁸. Epidemiological meta-analyses that systematically integrate and re-examine existing data can help to substantiate the relevance of environmental factors, such as Epstein–Barr virus, in MS⁸⁹.

Studies of gene–environment interactions also suggest that exogenous factors have a role in modulating neuroinflammation. For example, one study has shown that air pollution and MET receptor tyrosine kinase genotype interact in ASD⁹⁰. Evidence suggests that MET function is relevant to several immunological settings, including maternal immune suppression in which it affects the level of maternal anti-fetal brain protein antibodies, which are a risk factor for ASD in the child⁹¹.

[H2] Neuroimaging

In addition to omics data, neuroimaging, particularly MRI, is now a substantial source of big data⁹². Large imaging initiatives, such as the Human Connectome Project (**Box 3**), are fundamentally improving our understanding of connectivity and activity in the

CNS. MRI enables the assessment of functional characteristics and the measurement of microstructural and macrostructural features, thereby generating data of different scales. The complexity of the data produced and increases in the volume of data as a result of collaborative initiatives have posed a substantial computational challenge, especially data sharing⁹³ and the development of appropriate analytical tools to process digitized images⁹⁴. Nevertheless, multimodal MRI data from the Human Connectome Project have recently been used to shed light on the structural and functional organization of the human cerebral cortex⁹⁵. This work has led to the identification of previously unrecognized subdivisions with the cortex⁹⁵, insight that is likely to improve our understanding of the impact that spatial variation has in neuroinflammatory processes across neurological diseases. As neuroimaging big data generation and sharing progresses, this technology will likely have a fundamental impact on clinical decision-making for patient stratification and improved healthcare provision of patients with neuroinflammatory diseases.

[H2] Contribution to small-scale studies

In addition to spurring the development of new and improved omics technologies and large-scale genetic, phenotypic, epidemiological and imaging studies, the big data drive is also feeding into and facilitating smaller-scale, in-depth mechanistic research, which in turn provides further insight into the complexities of neuroinflammatory processes in neurological disorders. For example, genetic studies have implicated the anti-inflammatory cytokine IL-10 in AD^{15,96}, and subsequent investigations conducted on the basis of this finding have unexpectedly demonstrated that in transgenic mouse models of AD, IL-10 actually worsens disease pathology by suppressing the phagocytic capacity of microglia to clear A β ^{97,98}. Similarly, the functional relevance of TREM2 in AD is being dissected in more detail⁹⁹. *In vitro* studies of microglia have demonstrated that overexpression of TREM2 leads to more-efficient phagocytosis-mediated clearance of apoptotic neurons, suggesting that the protein is a phagocytic receptor¹⁰⁰. Studies in animal models have shown that the presence of TREM2 reduces proinflammatory cytokine secretion by myeloid cells¹⁰⁰⁻¹⁰², indicating that TREM2 is neuroprotective in AD. Moreover, microglia that express the AD-associated R47H variant of TREM2 have a diminished capacity to bind to phospholipids, suggesting that TREM2 can mediate neuroprotection by sensing alterations in the lipid microenvironment that result from A β accumulation and neuron degradation, and consequently stimulate the phagocytic function of microglia⁴¹. In the APPPS₁₋₂₁ murine model of AD, however, expression of TREM2 on peripheral blood monocyte-derived macrophages that cluster around

amyloid plaques has been linked with greater A β accumulation, suggesting that the role of TREM2 in disease differs between cell types³⁹. Such work has also led to investigations of TREM2 in other neuroinflammatory contexts, such as stroke¹⁰³, thereby paving the way for informative comparisons across neurological conditions. Collectively, these examples illustrate how big data may help to drive smaller-scale studies to address very specific medical questions, but how data derived from such studies may be easily merged and incorporated with big data sets remains to be determined.

[H1] Translation of big data to the clinic

The technical, informatics and statistical difficulties associated with collating and analysing big data are becoming less of a bottleneck in research, but a new challenge is the integration and interpretation of information generated by large-scale and small-scale studies for clinical benefit¹⁰⁴. Given the need for advances in healthcare for neurological diseases, big data must be accessed, mined and filtered to enable specific medical questions to be answered in a timely manner.

Use of big data to meet a diverse range of clinical needs requires maintained and strengthened crosstalk between the scientific and medical disciplines, and equal progress in both fields that enables convergence of clinically targeted science and evidence-generating medicine. In neuroinflammation, big data have the potential to aid diagnosis and prognosis, help predict drug efficacy and safety, design appropriate treatment regimens, and develop personalized disease management^{105,106} (**Fig. 2**). The types of big data discussed above are already driving advances in clinical management of neuroinflammation, and integration of more heterogeneous forms of big data in future, when analytics and integrative methodologies have been fully developed, are likely to facilitate further advances.

[H2] Diagnosis Use of high-throughput technologies to generate disease-relevant data has led to a greater appreciation of patient heterogeneity and the difficulty in reconciling these data with diagnoses made according to manifestations rather than mechanistic profiling of disease¹⁰⁷. Nevertheless, genomics are starting to be applied to diagnosis — particularly in the context of rare disorders for which a genetic aetiology is suspected — and concerted efforts are underway to determine the appropriate application of this approach. As an example, next-generation sequencing has revealed that previously unexplained cases of apparent spastic paraplegia with unusual characteristics (normal neuroimaging or non-specific demyelination) are

linked to mutations that cause aberrant interferon signalling³⁰. For this type of approach to reach its full clinical potential, we need more than the ability to analyse patient genomes and functionally interrogate identified variants to aid diagnosis and therapy. We also need databases of disease-associated and control genetic variation with which to compare data from patients^{52,108}, and electronic medical records and clinical databanks to determine which patients have a presentation that is consistent with a rare disease and might, therefore, benefit from genomic screening. For some conditions that are well characterized, genetic testing is already helping with diagnosis: in familial AD, for example, genetic testing of patients and relatives for high-penetrant variants in *APP*, *PSEN1* and *PSEN2* facilitates early diagnosis and healthcare provision, and allows early arrangement of future care¹⁰⁹.

For common, polygenic neurological diseases, the small effect size of most associated genetic variants means they cannot be used for predicting development of disease or for diagnosis. Efforts to assess such use of genetic data in MS include a study in which a weighted risk score was constructed on the basis of 16 known risk alleles and specific environmental parameters, including smoking and immune responsiveness to Epstein–Barr virus. The risk score had a modest ability to predict the risk of MS¹¹⁰, and further iterations of the score that incorporate more identified risk variants²⁵ might aid clinical implementation of the approach in routine diagnosis.

In addition to genetic data, molecular, cellular and other pathological signatures identified through omics analyses can enable stratification of patients on the basis of underlying pathways that drive their disease. In AD, for example, subgroups of genetic risk variants correlate more strongly with specific neuropathological features, such as neurofibrillary tangles, neuritic plaques and cerebral amyloid angiopathy¹¹¹, than with others, thereby providing a rationale for patient stratification on the basis of genetically determined pathological characteristics; such stratification might improve diagnosis and enable more specific therapeutic targeting. In future, improvements in big data analytics could allow integration of genomic, biomarker, imaging and health record big data so that methods such as machine learning could be used to identify previously unrecognized, distinct groups of patients¹¹².

Much remains to be achieved in big-data-driven diagnosis of neurological diseases, but innovative approaches to the prediction of disease are underway, although yet to yield insights. These initiatives include projects such as the 100K Wellness Project, the aim of which is to systematically phenotype healthy individuals over time,

determine which individuals develop disease, and retrospectively assess whether specific signatures could have effectively predicted the diagnosis¹¹³ (**Box 3**). A non-technological but fundamental aspect of such longitudinal investigations and large-scale biobanking efforts is volunteer participation, and the ability to make such ventures a reality reflects an increasing demand for more-proactive health monitoring and management^{114,115}.

[H2] Prognosis

The ability to predict disease progression and severity has important implications for the design and planning of disease management strategies, and big data have the potential to aid prognosis. To date, most research in this area has focused on interrogating the diversity of pathology in neurological disorders and trying to identify correlates with predictive value that can, by inference, provide insight into the parameters that dictate the development, severity and clinical manifestations in disease. Such investigations have relied on advances in structural and functional neuroimaging techniques, including PET and MRI¹¹⁶. In ALS, PET has been used to show that the presence of hypermetabolism — an indicator of microglial activation and astrogliosis — in the midbrain and medial temporal cortex distinguishes patients from healthy controls with an accuracy >90%, thereby demonstrating the diagnostic value of hypermetabolism and suggesting its potential for longitudinal interrogation¹¹⁷. In AD, a GWAS in which changes in brain amyloid burden were measured with ¹⁸F-florbetapir PET found an association between amyloid accumulation and a microglial activation gene that encodes IL-1 receptor accessory protein¹¹⁸. The same study showed that carriers of the associated variant were more likely to progress from mild cognitive impairment to AD, showing an accelerated cognitive decline and pronounced temporal cortex atrophy. In MS, a GWAS in which CNS lesion topology was measured suggested that variants in loci linked to immune cell function, myelination and neural growth may play a role in determining lesion location¹¹⁹. Identifying predictors of lesion distribution is particularly important, as the location of a lesion dictates the resulting clinical manifestations and, consequently, the healthcare required¹²⁰.

Mining of electronic medical records and merging them with other data also has potential in prognostic research, and initial analyses suggest that accurate information about disease severity can be extracted from such records. In one existing example, an algorithm that included codified and narrative information from routinely collected electronic medical records and adjusted for age of disease onset,

sex and disease duration accurately identified patients with MS, estimated their MS severity score, and successfully distinguished between patients with relapsing–remitting and progressive disease¹²¹. This algorithm provides an encouraging first step towards the use of big data to predict disease course and outcome, a growing area of research that requires large-scale and in-depth longitudinal investigations. For example, the UK Biobank initiative — which includes sociodemographic data, biophysical and biomarker measures, health and lifestyle factors and medical histories from self-reported surveys and hospitalization records from 500,000 adults¹²² (**Box 3**) — has been used to identify predictors of 5-year mortality in the general population¹²³. In addition, big data analytics are increasingly being used in predictive modelling to identify and manage patients with particularly high risks^{124–126}, and as data resources expand and analysis methods improve, increasingly meaningful predictions about the prognosis of lower-risk patients might become feasible.

[H1] Therapy: efficacy and response

The use of big data promises to allow identification of novel drug targets as a result of improved understanding of disease mechanisms from large volumes of biomedical information¹²⁷. The development of drugs against new targets can be fraught with complications owing to insufficient efficacy and/or problems with safety, both of which commonly cause suspension of clinical trials¹²⁸. Retrospective analysis indicates, however, that drug development programs progress further and have a greater likelihood of being successful if they are supported by human genetic evidence than if they are not¹²⁹. Moreover, naturally occurring genetic variants can be used to validate therapeutic targets before clinical trials: understanding the impact of variation in biological pathways – as driven by genetic variants – can help predict the effect of pharmacologically skewing such pathways. Therefore, genotype-to-phenotype relationships can be used as proxies to help estimate the outcome of drug targeting¹³⁰, particularly in the case of rare and low-frequency variants with a large impact on gene function and disease. However, even common variants with weaker effects might be informative in helping to predict drug efficacy and safety if combined with molecular and cellular functional follow-up studies. For example, this type of approach has revealed why drugs that block the activity of TNF lead to worsening of disease in MS when they have been successfully used to treat many other autoimmune diseases^{131,132}. This long-standing conundrum was solved by the discovery that an MS genetic risk variant drives expression of an endogenous TNF blocking protein¹³³. The impact of the variant is milder than that of the drug, but both

act on the same pathophysiological pathway. This discovery demonstrates that elucidating genetically determined neuroinflammatory mechanisms can be of predictive value in a therapeutic setting¹³³.

Pharmacogenomics is another obvious application of big data that could improve therapy. For example, the effects of genetic variants on IFN- β treatment response in patients with MS is an active area of investigation^{134,135}, and evidence suggests an association of several variants, including those in the *FHIT*, *GAPVD1* and *ZNF697* gene regions¹³⁵. Nevertheless, the findings of such studies need to have substantial predictive power to warrant modification of clinical practice so that genetic biomarkers can be used to inform drug administration. Ensuring this power necessitates pharmacogenomic analyses that include large numbers of patients for whom extensive clinical information is available; in MS, international collaborations and growing patient databases, such as MSBase¹³⁶, are expected to facilitate attainment of the required power¹³⁵.

Another example of efforts to generate and collate molecular and clinical data to guide treatment relates to the risk of developing progressive multifocal leukoencephalopathy (PML), a usually fatal opportunistic infection, for patients with MS who are treated with natalizumab¹³⁷. Although not on the typical scale of big data, the experience with PML sets a precedent for the systematic collation of samples and data from patients on a large scale to minimize the time needed to respond to emerging and changing medical needs. The success of natalizumab, which reduces the annualized relapse rate in MS by 68%¹³⁸, was hampered by reports of associated PML that resulted in voluntary withdrawal of the drug from the market in 2005. Subsequent work has shown that PML arises in immunosuppressed or immunocompromised individuals when JC virus infects the CNS, causing acute destruction of oligodendrocytes¹³⁹, although the exact mechanisms of viral reactivation are yet to be determined. At the time of natalizumab withdrawal, the incidence of PML was ~1 per 1,000 patients receiving the drug¹⁴⁰. Following the establishment of a comprehensive program for global risk management, natalizumab was reintroduced to the market 1 year later. This program has since led to the development of specific drug administration guidelines that include recommendations for assessment of anti-JC virus antibody levels, consideration of prior immunosuppressant use, and a limit to the total duration of natalizumab treatment so as to ameliorate MS but minimize the risk of PML¹³⁷.

In AD, investigations into the efficacy of non-steroidal anti-inflammatory drugs have included analysis of responses to drugs such as naproxen and celecoxib relative to the rate of cognitive decline¹⁴¹. This analysis suggests that patient stratification according to the degree of cognitive decline might be useful for assessing therapeutic efficacy, particularly as the rate of AD progression might reflect changes in the balance between proinflammatory and anti-inflammatory processes over time^{141,142}.

Much of the work described above has involved prospective studies, but a growing body of evidence shows the importance of evaluating, integrating and building on existing clinical data and knowledge¹⁰⁵. For example, large-scale longitudinal studies are critical for assessing drug efficacy and projecting cost-effectiveness; an example is the UK Multiple Sclerosis Risk Sharing Scheme that has evaluated the impact of IFN- β and glatiramer acetate on quality-adjusted life-years of >4,000 patients since 2002¹⁴³.

Existing information from failed trials should also be considered when designing and refining therapeutic strategies, a point demonstrated by the success of the anti-CD20 antibody ocrelizumab, the first drug to demonstrate some efficacy in patients with primary progressive MS¹⁴⁴. Trials of ocrelizumab stemmed from previous experience in unsuccessful trials of rituximab: despite a lack of overall efficacy, the results of these trials indicated that a subset of patients with inflammatory lesions might respond well to B-cell depletion¹⁴⁵.

Integration of heterogeneous data, such as coupling comparisons of pathophysiological mechanisms across diseases with information about drugs already in use to treat other conditions, can be invaluable for drug repositioning strategies. For example, acid-sensing ion channel 1 has a role in both macrophage and neuronal function, and can be blocked by amiloride, a drug that has an established safety profile and is used to treat hypertension^{146,147}. Repositioning strategies are already accelerating clinical trials of amiloride in neurological conditions¹⁴⁸. The prominent involvement of inflammation in a variety of neurological disorders also provides an opportunity to test drugs that are used to treat other immune-mediated disorders. For example, *in vivo* modelling suggests that inhibition of the proinflammatory IL-12 and IL-23 signalling pathways might be effective for treating AD¹⁴⁹, and the experience that has been amassed in targeting these cytokines in the context of autoimmune disease, with respect to the relative efficacy

and side effect profiles of drugs that inhibit this cytokine signaling¹⁵⁰, might help to determine dosing regimens and manage adverse responses in clinical trials of the same drugs in AD¹⁵¹.

Overall, the use of big data to inform diagnosis, prognosis and therapy should provide risk–benefit estimates and measures of performance for different healthcare strategies. As clinical interventions can have unintended outcomes, improvements in predictive modelling as a result of big data analytics is an attractive prospect. However, given the limitations of currently available big data sources, the ongoing methodological developments and the need for validation of big data analytics and real-time re-evaluation when applied to healthcare, big data is likely to support rather than dictate clinical decision-making in the short-term.

[H1] From populations to individual patients

The possible uses of big data to improve healthcare provision suggest that precision medicine – which focuses on determining which therapeutic strategies will be effective for which individuals based on their molecular, clinical, environmental and lifestyle factors – is plausible for patients with neurological conditions. However, the summation of in-depth phenotypic, biometric and behavioral signatures of an individual over time provides the context in which their disease has developed and in which it must be treated. Combined with the increasing feasibility of large-scale, longitudinal studies, this access to data raises the question of whether personalized medicine – which focuses developing therapeutic strategies tailored to each individual patient – is a realistic expectation for the future. The logistics and economics of personalized, big-data-enabled medicine, especially for patients with common multifactorial diseases, are not trivial^{105,152}. Nevertheless, for common neurological disorders, personalized therapy could come to mean the integration of an individual's genetic and phenotypic biomarkers and other biometric measures through advanced analytics to support clinical decision-making and inform the design of combination therapy regimens. These regimens could include tailored selection and dosage of specific drugs to appropriately decrease neuroinflammation and simultaneously boost neuroreparative and neuroprotective mechanisms.

Concomitantly, the healthcare industry is likely to be transformed by use of mobile communications, cloud-based approaches, web-based applications and social media for data acquisition, storage and sharing. This technology will allow individuals to proactively monitor their health in real time and enable remote surveillance by

healthcare providers, thereby helping to maintain health and perhaps prevent disease. Research in this area is relatively limited, but examples of such technological application already exist. In MS, a tablet-based disability assessment tool — the Multiple Sclerosis Performance Test — has been developed to measure neurological and neuropsychological impairment and disability¹⁵³. For PD, studies are underway to assess the feasibility of home-based monitoring through wearable sensors and smartphone-based apps¹⁵⁴. For example, the Mobile Parkinson Observatory for Worldwide, Evidence-based Research (mPower) initiative (**Box 3**) is a clinical observational study conducted through a smartphone app that records high-resolution activity data from patients with PD and controls¹⁵⁵. Home-based monitoring is well tolerated and has the potential to provide direct feedback about each patient's condition¹⁵⁴, but large-scale use would create a substantial analytical load and a demand for real-time data interpretation by healthcare professionals.

Despite the potential benefits, the impetus to collate and analyse increasingly large volumes of detailed and heterogeneous information to provide medical benefit on a personalized level for an entire population raises multiple economic, ethical and legal concerns (**Box 4**). These concerns will need to be addressed before big data collection and analysis can proceed unhindered on this scale.

[H1] Conclusions

Neuroinflammation is emerging as a central process across neurological disorders, and these conditions collectively pose a huge public health problem. The generation and application of relevant big data provides a rational approach for facilitating fundamental progress in addressing this problem. In the context of basic biomedical research, omics are revealing the breadth and heterogeneity of neuroinflammatory responses and, consequently, how these responses might be targeted to ameliorate disease.

Nevertheless, the full potential of translating big data to the clinic is yet to be realized. As we develop a better understanding of how we can interrogate the growing body of information generated by big data approaches to answer specific medical questions, translational goals will be within our reach. To bridge the gap between the laboratory and the clinic in the application of big data to healthcare, cross-disciplinary training and sharing of knowledge will be imperative, as will standardization of methods and practices and the development of interfaces to accelerate data mining and

dissemination¹⁰⁶. At the same time, as big data herald a paradigm shift in medicine, the sociological implications must be considered, with care taken to safeguard information privacy and ensure worldwide benefits, particularly for neurological disorders, which have an enormous global impact^{156,157}.

Proposed display items:

Box 1 | Neurological disorders: big global problems need big data

Neurological disorders are increasing in prevalence and currently affect as many as one billion people worldwide, including all age groups and ethnicities, geographical locations and socioeconomic backgrounds¹⁵⁶. The long-term disability associated with many of these conditions means they have a serious impact not only on the patients themselves, but also on their families, friends and caregivers. Collectively, neurological disorders represent ~7% of the total global burden of disease, measured in disability-adjusted life-years, across all causes and ages⁵.

Neurological disorders can affect the CNS and/or PNS, and include neuropsychiatric disorders such as Alzheimer disease, other dementias and Parkinson disease; mental health conditions such as schizophrenia; autoimmune diseases such as multiple sclerosis and neuromyelitis optica spectrum disorder; stroke; migraines and other headache-associated conditions; sleep disorders including narcolepsy; epilepsy; motor neuron diseases such as amyotrophic lateral sclerosis; prion diseases and many others¹⁵⁷.

Neurological conditions can also arise from viral infections (such as HIV, herpes, influenza, JC virus, arboviruses and emerging viruses such as Zika), mycobacterial and bacterial infections (including meningeal, cerebral and vertebral tuberculosis and bacterial meningitis), and parasitic infections (neurocysticercosis, cerebral malaria and *Toxoplasma* encephalitis, for example). Neurological sequelae can also be associated with other factors, such as nutritional deficiencies, traumatic injury and as adverse effects of drugs¹⁵⁷.

The breadth of neurological disorders and their sequelae demonstrates the magnitude of the global health challenge that they pose. The involvement of neuroinflammation in practically all of these conditions presents a tangible opportunity to address this challenge, particularly as the integration of emerging CNS-relevant large-scale data sets with pre-existing information regarding inflammatory mechanisms in other contexts might accelerate the identification of optimal therapeutic approaches to neurological disorders.

Box 2 | Big data: methodological and practical challenges

Health-relevant data are currently underexploited in the improvement of healthcare provision owing to ineffective and inefficient collection, management and use of data

derived from research, the clinic and patients. These shortcomings lead to resource squandering, missed opportunities and even detriment to patients¹⁵⁸. An anticipated benefit of the big data drive is the rectification of these broad deficiencies in healthcare systems, but methodological and practical challenges need to be met before this goal can be achieved.

Limitations of big data can range from sampling biases, confounding factors and false positives to variability within and between data sets, diversity of data structure, and issues related to the existence, availability and accessibility of data required to address particular medical needs. Inconsistency of data can complicate the integration and analysis process, and can arise from a lack of standard data collection practices and changes in quality control, coding and nomenclature over time. Validation of analytical tools and methods is, therefore, critical but must occur systematically to ensure sustained performance²².

Advances in computational science are facilitating the development of tools and platforms that provide a suitable architectural framework for accommodating, organizing, integrating, transforming and analysing big data. For use in healthcare, such tools and platforms need to be scalable, reliable, user-friendly, privacy-enabled and security-enabled, and must ensure minimal lag time between data collection, processing and analytical output²².

Box 3 | Large-scale data collation initiatives

Large-scale data collation initiatives are set to revolutionize our understanding of human health and disease. Notable ongoing ventures with relevance to neurological diseases include:

Human Connectome Project

An international consortium aiming to extensively map human brain circuitry in 1,200 healthy adults using noninvasive neuroimaging.

Mobile Parkinson Observatory for Worldwide, Evidence-based Research (mPower)

A US smartphone-based study to track health and symptoms during Parkinson disease progression.

NeurOmics

A European project aiming to link genotypic and phenotypic information from individuals with rare neurodegenerative diseases, and to create a phenotypic database at PhenoTips.

Neuropsychiatric Genome-scale and RDoC-Individualized Domains (N-GRID)

A US project collecting cells from neuropsychiatric disorder patients to enable investigation of how molecular profiles affect clinical features of disease and responses to small molecules.

OPTIMISE

A UK project to develop and use tools for comprehensive collection and integration of information from MS patients, including brain scans, genomics data, biomarkers, self-reported quality of life measures and data from movement tracking sensors.

UK Biobank

A biobank that has recruited and is monitoring 500,000 volunteers; genetic, clinical, para-clinical and epidemiological information from these individuals is being collated.

US Brain Research through Advancing Innovative Neurotechnologies® (BRAIN) Initiative

An initiative to accelerate development and implementation of technologies for a dynamic conceptualization of the brain, aiming to show how individual cells and complex neural circuits interact in time and space.

US Precision Medicine Initiative

An initiative including a study of the interplay between genetics and lifestyle factors in determining health outcomes in one million volunteers.

100K Wellness Project

A US-based project to recruit 100,000 healthy individuals, assess their molecular profiles and monitor them longitudinally.

100,000 Genomes Project

A UK-based project to integrate whole-genome sequencing and clinical data from 100,000 patients with cancer or rare diseases and their families to facilitate diagnosis and clinical care.

Box 4 | Big data in the clinic — economics, ethics, and legalities

The application of big data in the clinic is a nascent but growing field that will require substantial investments — both economically and logistically — to ensure collection of high-quality data, development of tools and approaches for analysis, provision of suitable infrastructure for effective implementation, and training of healthcare providers. Direct evidence for the concrete clinical utility of big data is limited, but the desired outcome of these investments is a substantial impact on population management and health system policy^{159,160} as a result of the ability to identify high-risk and high-cost patients¹²⁴, determine the most cost-effective therapeutic strategies and predict the most economically-demanding health burdens, as well as ways to counteract them.

To meet these goals requires increased availability of and access to varied data resources, and interconnection of these different types of data at the level of the individual patient. Patient privacy, information security and consent for data access and usage are, therefore, particularly pertinent ethical and legal issues^{161,162} that necessitate appropriate education of the general public, as well as healthcare professionals, about the risks and benefits of big data and the existing legislature that surrounds data protection and health insurance policy. An equally important consideration is that use of big data has the potential to identify novel, unanticipated

applications for existing information in real time; maximum exploitation of this aspect will likely require authorization for use of collected data unspecified future analyses without the need to obtain further consent. This issue might be partly addressed by increases in pro-active patient involvement in health monitoring, which might lead to interactive data management systems in which patients can alter their consent preferences throughout their lifetime¹⁶³.

Figures:

Figure 1 | Neuroinflammatory disease pathways and the impact of big data.

Neuroinflammation is implicated in many neurological disorders including autoimmune and primary neurodegenerative diseases, such as MS and AD, respectively. Research to date indicates that the neuroinflammatory processes involved in these conditions are highly diverse and big data analyses are fortifying this concept by helping to further reveal the heterogeneity: many different cell types and pathways can be involved and their roles vary through time. In MS and its murine model EAE, immune cell infiltration from the periphery into the CNS is characteristic, and along with a multitude of soluble mediators, drives much of the initial pathology by promoting demyelination. As disease progresses, infiltration wanes but microglia and astrocytes remain activated and neurodegeneration continues. In AD and in murine models of the disease, infiltration - particularly of adaptive immune cells - is not characteristic. Instead, microglia/macrophages and astrocytes are the main immune cells implicated. They may promote disease by producing pro-inflammatory and neurotoxic mediators, displaying an aberrant synaptic pruning function and disseminating tauopathy via exosomal shedding. Microglia may also have a neuroprotective role by phagocytosing A β and by producing neurotrophic factors. APP, amyloid precursor protein; ApoE, apolipoprotein E; BBB, blood-brain barrier; BDNF, brain-derived neurotrophic factor; CCL2, CC-chemokine ligand 2; CD8⁺ MAIT cell, CD8⁺ mucosa-associated invariant T cell; DAP12, DNAX-activation protein 12; GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN γ , interferon- γ ; IL-1 β , interleukin-1 β ; IL-6, interleukin-6; IL-10, interleukin-10; IL-17, interleukin-17; iNOS, inducible nitric oxide synthase; LRP1, LDLR-related protein 1; NADPH, nicotinamide adenine dinucleotide phosphate; NO, nitric oxide; ODC, oligodendrocyte; PGE2, prostaglandin E2; RAGE, receptor for advanced glycosylation end products; RNS, reactive nitrogen species; ROS, reactive oxygen species; TGF β , transforming growth factor β ; T_H1 cell, CD4⁺ T helper 1 cell; T_H17 cell, CD4⁺ T helper 17 cell; TLR, toll-like receptor; TNF, tumor necrosis factor; TREM2, triggering receptor expressed on myeloid cells 2.

Figure 2 | Big data: From the lab to the clinic and back.

Technological and analytical advances coupled with large-scale analyses of thousands of individuals are generating big data relating to intra- and inter-individual heterogeneity. Integrating this information along with epidemiological, clinical and -- clinical and other types of data facilitates the implementation of precision medicine by enabling patient stratification for more appropriate therapeutic intervention, aided by

the design of new treatment and disease management strategies as the understanding of disease mechanisms grows. The experience and knowledge amassed through precision medicine approaches feeds back into the pipeline of big data utilization to further modify and improve clinical practice. Monitoring of behavior and molecular profiling at the level of the individual through time generates personal big data to help guide clinical decision making for the application of personalized medicine. The challenges surrounding big data acquisition and utilization are not insubstantial, but if they are met then big data implementation in the healthcare system has the potential to facilitate the maintenance of health and prevention of disease development at the population level.

Tables: N/A

Movies: N/A

Links to web sites: N/A

Key points:

- Neurological disorders affect as many as a billion people worldwide, including people of all age groups and races, in different geographical locations and of different socioeconomic backgrounds.
- Neuroinflammation is emerging as a key process common to the majority of neurological conditions - be it as a causative factor or as a secondary response to nervous system insult.
- Big data generated through large-scale, high-throughput 'omics' technologies are changing our perception of the breadth and heterogeneity of neuroinflammatory responses and consequently how they may be exploited and targeted to ameliorate disease.
- For big data to be appropriately utilized to meet a diverse range of clinical needs, crosstalk between the scientific and medical disciplines is needed to allow the convergence of clinically targeted science and evidence-generating medicine.
- Big data pertaining to neuroinflammation may be of clinical utility in aiding patient diagnosis and prognosis and in helping to predict and monitor drug efficacy and safety through the implementation of precision medicine.
- Personalized medicine for patients with neurological diseases may involve combination therapy using drugs that inhibit neuroinflammation and promote neuroprotection, as well as more systematic health monitoring and management.

Competing interests:

The authors declare no competing interests.

References:

1. Streit, W. J. Microglia as neuroprotective, immunocompetent cells of the CNS. *Glia* **40**, 133–139 (2002).
2. Glass, C. K., Saijo, K., Winner, B., Marchetto, M. C. & Gage, F. H. Mechanisms underlying inflammation in neurodegeneration. *Cell* **140**, 918–934 (2010).
A seminal review of the molecular and cellular mechanisms that underpin inflammation across different neurodegenerative diseases.
3. Saijo, K., Crotti, A. & Glass, C. K. Regulation of microglia activation and deactivation by nuclear receptors. *Glia* **61**, 104–111 (2013).
4. Shechter, R., London, A. & Schwartz, M. Orchestrated leukocyte recruitment to immune-privileged sites: absolute barriers versus educational gates. *Nat. Rev. Immunol.* **13**, 206–218 (2013).
5. Chin, J. H. & Vora, N. The global burden of neurologic diseases. *Neurology* **83**, 349–351 (2014).
6. Marx, V. Biology: The big challenges of big data. *Nature* **498**, 255–260 (2013).
7. Bahcall, O. Precision medicine. *Nature* **526**, 335 (2015).
8. Dendrou, C. A., Fugger, L. & Friese, M. A. Immunopathology of multiple sclerosis. *Nat. Rev. Immunol.* **15**, 545–558 (2015).
9. Ransohoff, R. M. & Engelhardt, B. The anatomical and cellular basis of immune surveillance in the central nervous system. *Nat. Rev. Immunol.* **12**, 623–635 (2012).
10. Chard, D. T. *et al.* Brain atrophy in clinically early relapsing-remitting multiple sclerosis. *Brain* **125**, 327–337 (2002).
11. Popescu, B. F. & Lucchinetti, C. F. Pathology of demyelinating diseases. *Annu. Rev. Pathol.* **7**, 185–217 (2012).
12. Fischer, M. T. *et al.* NADPH oxidase expression in active multiple sclerosis lesions in relation to oxidative tissue damage and mitochondrial injury. *Brain* **135**, 886–899 (2012).
13. Crotti, A. & Glass, C. K. The choreography of neuroinflammation in Huntington's disease. *Trends Immunol.* **36**, 364–373 (2015).
14. Crotti, A. *et al.* Mutant Huntingtin promotes autonomous microglia activation via myeloid lineage-determining factors. *Nat. Neurosci.* **17**, 513–521 (2014).
15. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
16. Bradshaw, E. M. *et al.* CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nat. Neurosci.* **16**, 848–850 (2013).
17. Griciuc, A. *et al.* Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron* **78**, 631–643 (2013).
18. Chan, G. *et al.* CD33 modulates TREM2: convergence of Alzheimer loci. *Nat. Neurosci.* **18**, 1556–1558 (2015).
19. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
20. Hong, S. *et al.* Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* **352**, 712–716 (2016).

21. Song, S. *et al.* Major histocompatibility complex class I molecules protect motor neurons from astrocyte-induced toxicity in amyotrophic lateral sclerosis. *Nat. Med.* **22**, 397–403 (2016).
22. Raghupathi, W. & Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* **2**, 3 (2014).
23. Rumsfeld, J. S., Joynt, K. E. & Maddox, T. M. Big data analytics to improved cardiovascular care: promise and challenges. *Nat. Rev. Cardiol.* **13**, 350–359 (2016).
24. Green, E. D., Watson, J. D. & Collins, F. S. Human Genome Project: Twenty-five years of big biology. *Nature* **526**, 29–31 (2015).
25. Beecham, A. H. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1360 (2013).
26. Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
27. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
28. Cirulli, E. T. *et al.* Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436–1441 (2015).
29. Freischmidt, A. *et al.* Haploinsufficiency of *TBK1* causes familial ALS and fronto-temporal dementia. *Nat. Neurosci.* **18**, 631–636 (2015).
30. Crow, Y. J. *et al.* Mutations in ADAR1, IFIH1, and RNASEH2B presenting as spastic paraplegia. *Neuropediatrics* **45**, 386–393 (2014).
31. Marx, V. Human phenotyping on a population scale. *Nat. Methods* **12**, 711–714 (2015).
32. Moutsianas, L. *et al.* Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.* **47**, 1107–1113 (2015).
33. Menche, J. *et al.* Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
34. Tasan, M. *et al.* Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat. Methods* **12**, 154–159 (2015).
35. Parkes, M., Cortes, A., van Heel, D. A. & Brown, M. A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **14**, 661–673 (2013).
36. Cho, J. H. & Feldman, M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. *Nat. Med.* **21**, 730–738 (2015).
37. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
38. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* **368**, 117–127 (2013).
39. Jay, T. R. *et al.* TREM2 deficiency eliminates TREM2⁺ inflammatory macrophages and ameliorates pathology in Alzheimer's disease mouse models. *J. Exp. Med.* **212**, 287–295 (2015).
40. Kleinberger, G. *et al.* TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Sci. Transl. Med.* **6**, 243ra86 (2014).

41. Wang, Y. *et al.* TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* **160**, 1061–1071 (2015).
42. Paloneva, J. *et al.* Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. *Am. J. Hum. Genet.* **71**, 656–662 (2002).
43. Guerreiro, R. J. *et al.* Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. *JAMA Neurol.* **70**, 78–84 (2013).
44. Rayaprolu, S. *et al.* TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. *Mol. Neurodegener.* **8**, 19 (2013).
45. Cady, J. *et al.* TREM2 variant p.R47H as a risk factor for sporadic amyotrophic lateral sclerosis. *JAMA Neurol.* **71**, 449–453 (2014)..
46. Ferrari, R. *et al.* Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* **13**, 686–699 (2014).
47. Maruyama, H. *et al.* Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **465**, 223–226 (2010).
48. Girard, S. L. *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860–863 (2011).
49. Xu, B. *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat. Genet.* **43**, 864–868 (2011).
50. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
51. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
52. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
53. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat. Neurosci.* **19**, 571–577 (2016).
54. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
55. Shaltouki, A., Peng, J., Liu, Q., Rao, M.S. & Zeng, X. Efficient generation of astrocytes from human pluripotent stem cells in defined conditions. *Stem Cells* **31**, 941–952 (2013).
56. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature* **533**, 95–99 (2016).
57. Ward, L. D. & Kellis, M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* **30**, 1095–1106 (2012).
58. Farh, K. K.-H. *et al.* Genetic and epigenetic fine-mapping of causal variants in autoimmune disease. *Nature* **518**, 337–343 (2015).
59. Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.* **8**, 509–524 (2013).
60. Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212 (2014).

61. Huynh, J. L. *et al.* Epigenome-wide differences in pathology-free regions of multiple-sclerosis-affected brains. *Nat. Neurosci.* **17**, 121–130 (2014).
62. De Jager, P. L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at *ANK1*, *BIN1*, *RHBDF2* and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
63. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014).
64. Hickman, S. E. *et al.* The microglial sensome revealed by direct RNA sequencing. *Nat. Neurosci.* **16**, 1896–1905 (2013).
65. Butovsky, O. *et al.* Identification of a unique TGF- β -dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
66. Yamasaki, R. *et al.* Differential roles of microglia and monocytes in the inflamed central nervous system. *J. Exp. Med.* **211**, 1533–1549 (2014).
67. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
68. Estes, M. L. & McAllister, A. K. Immune mediators in the brain and peripheral tissues in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 469–486 (2015).
69. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
70. Molofsky, A. V. *et al.* Astrocyte-encoded positional cues maintain sensorimotor circuit integrity. *Nature*, **509**, 189–193 (2014).
71. Palanichamy, A. *et al.* Immunoglobulin class-switched B cells form an active immune axis between CNS and periphery in multiple sclerosis. *Sci. Transl. Med.* **6**, 248ra106 (2014).
72. Stern, J. N. H. *et al.* B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci. Transl. Med.* **6**, 248ra107 (2014).
73. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
74. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
75. Frei, A. P. *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* **13**, 269–275 (2016).
76. Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
77. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
78. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).
79. Delude, C. M. Deep phenotyping: The details of disease. *Nature* **527**, S14–S15 (2015).
80. Battle, A. *et al.* Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667 (2015).
81. Ho, P. P. *et al.* (2012) Identification of naturally occurring fatty acids of the myelin sheath that resolve neuroinflammation. *Sci. Transl. Med.* **4**, 137ra73.
82. Mayo, L. *et al.* Regulation of astrocyte activation by glycolipids drives chronic CNS inflammation. *Nat. Med.* **20**, 1147–1156 (2014).

83. Asai, H. *et al.* Depletion of microglia and inhibition of exosome synthesis halt tau propagation. *Nat. Neurosci.* **18**, 1584–1593 (2015).
 84. Virgin, H. W. The virome in mammalian physiology and disease. *Cell* **157**, 142–150 (2014).
 85. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
 86. Erny, D. *et al.* Host microbiota constantly control maturation and function of microglia in the CNS. *Nat. Neurosci.* **18**, 965–977 (2015).
 87. Berer, K. *et al.* Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* **479**, 538–541 (2011).
 88. Jangi, S. *et al.* Alterations of the human gut microbiome in multiple sclerosis. *Nat. Commun.* **7**, 12015 (2016).
 89. Belbasis, L., Bellou, V., Evangelou, E., Ioannidis, J. P. A. & Tzoulaki, I. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. *Lancet Neurol.* **14**, 263–273 (2015).
 90. Volk, H. E. *et al.* Autism spectrum disorder: interaction of air pollution with the MET receptor tyrosine kinase gene. *Epidemiology* **25**, 44–47 (2014).
 91. Heuer, L., Braunschweig, D., Ashwood, P., Van de Water, J. & Campbell, D. B. Association of a *MET* genetic variant with autism-associated maternal autoantibodies to fetal brain proteins and cytokine expression. *Translat. Psychiatry* **1**, e48 (2011).
 92. Sejnowski, T. J., Churchland, P. S. & Movshon, J. A. Putting big data to good use in neuroscience. *Nat. Neurosci.* **17**, 1440–1441 (2014).
 93. Poldrack, R. A. & Gorgolewski, K. J. Making big data open: data sharing in neuroimaging. *Nat. Neurosci.* **17**, 1510–1517 (2014).
 94. Lichtman, J. W., Pfister, H. & Shavit, N. The big data challenges of connectomics. *Nat. Neurosci.* **17**, 1448–1454 (2014).
 95. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
- An innovative study describing the generation of a 'universal' map of the human cerebral cortex by integrating multi-modal imaging data and the application of a machine-learning classifier to automatically identify brain regions.***
96. Lio, D. *et al.* Interleukin-10 promoter polymorphism in sporadic Alzheimer's disease. *Genes Immun.* **4**, 234–238 (2003).
 97. Chakrabarty, P. *et al.* IL-10 alters immunoproteostasis in *APP* mice, increasing plaque burden and worsening cognitive behavior. *Neuron* **85**, 519–533 (2015).
 98. Guillot-Sestier, M-V. *et al.* *Il10* deficiency rebalances innate immunity to mitigate Alzheimer-like pathology. *Neuron* **85**, 534–548 (2015).
 99. Colonna, M. & Wang, Y. TREM2 variants: new keys to decipher Alzheimer disease pathogenesis. *Nat. Rev. Neurosci.* **17**, 201–207 (2016).
 100. Takahashi, K., Rochford, C. D. & Neumann, H. Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *J. Exp. Med.* **201**, 647–657 (2005).
 101. Turnbull, I. R. *et al.* Cutting edge: TREM-2 attenuates macrophage activation. *J. Immunol.* **177**, 3520–3524 (2006).

102. Hamerman, J. A. *et al.* Cutting edge: inhibition of TLR and FcR responses in macrophages by triggering receptor expressed on myeloid cells (TREM)-2 and DAP12. *J. Immunol.* **177**, 2051–2055 (2006).
 103. Kawabori, M. *et al.* Triggering receptor expressed on myeloid cells 2 (TREM2) deficiency attenuates phagocytic activities of microglia and exacerbates ischemic damage in experimental stroke. *J. Neurosci.* **35**, 3384–3396 (2015).
 104. Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E. & Martone, M. E. Big data from small data: data-sharing the ‘long tail’ of neuroscience. *Nat. Neurosci.* **17**, 1442–1447 (2014).
 105. Schneeweiss, S. Learning from Big Health Care Data. *N. Engl. J. Med.* **370**, 2161–2163 (2014).
 106. Chaussabel, D. & Pulendran, B. A vision and a prescription for big data-enabled medicine. *Nat. Immunol.* **16**, 435–439 (2015).
 107. Dendrou, C. A., Bell, J. I. & Fugger, L. Weighing in on autoimmune disease: Big data tip the scale. *Nat. Med.* **19**, 138–139 (2013).
 108. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
 109. American College of Medical Genetics and Genomics Board of Directors. Clinical utility of genetic and genomic services: a position statement of the American College of Medical Genetics and Genomics. *Genet. Med.* **17**, 505–507 (2015).
 110. De Jager, P. L. *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol.* **8**, 1111–1119 (2009).
 111. Beecham, G. W. *et al.* Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet.* **10**, e1004606 (2014).
 112. Shivade, C. *et al.* A review of approaches to identifying patient phenotype cohorts using electronic health records. *J. Am. Med. Inform. Assoc.* **21**, 221–230 (2014).
 113. Hood, L. & Price, N. Demystifying disease, democratizing healthcare. *Sci. Transl. Med.* **6**, 225ed5 (2014).
 114. Hood, L. & Auffray, C. Participatory medicine is revolutionizing healthcare. *Genome Med.* **5**, 110 (2013).
 115. Aronson, S. J. & Rehm, H. L. Building the foundation for genomics in precision medicine. *Nature* **526**, 336–342 (2015).
- A thorough review of how genomics are being used to help direct clinical decision-making and how they will accelerate progress towards precision medicine.***
116. Stoessl, A. J. *et al.* Imaging insights into basal ganglia function, Parkinson's disease and dystonia. *Lancet* **384**, 532–544 (2014).
 117. Chio, A. *et al.* Neuroimaging in amyotrophic lateral sclerosis: insights into structural and functional changes. *Lancet Neurol.* **13**, 1228–1240 (2014).
 118. Ramanan, V. K. *et al.* GWAS of longitudinal amyloid accumulation on ¹⁸F-florbetapir PET in Alzheimer's disease implicates microglial activation gene *IL1RAP*. *Brain* **138**, 3076–3088 (2015).

119. Gourraud, P.-A. *et al.* A genome-wide association study of brain lesion distribution in multiple sclerosis. *Brain* **136**, 1012–1024 (2013).
120. Kearney, H. *et al.* Cervical cord lesion load is associated with disability independently from atrophy in MS. *Neurology* **84**, 367–373 (2015).
121. Xia, Z. *et al.* Modeling disease severity in multiple sclerosis using electronic health records. *PLoS One*. **8**, e78927 (2013).
122. Thompson, S. G. & Willeit, P. UK Biobank comes of age. *Lancet* **386**, 509–510 (2015).
123. Ganna, A. & Ingelsson, E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* **386**, 533–540 (2015).
124. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A. & Escobar, G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff. (Millwood)* **33**, 1123–1131 (2014).
125. Lee, J. & Maslove, D. M. Customization of a severity of illness score using local electronic medical record data. *J. Intensive Care Med.* pii: 0885066615585951 (2015).
126. Hu, Z. *et al.* Online prediction of health care utilization in the next six months based on electronic health record information: a cohort and validation study. *J. Med. Internet Res.* **17**, e219 (2015).
127. Manolio, T. A. Bringing genome-wide association findings into clinical use. *Nat. Rev. Genet.* **14**, 549–558 (2013).
128. Hay, M. *et al.* Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014).
129. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
130. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug. Discov.* **12**, 581–594 (2013).
131. van Oosten, B. W. *et al.* Increased MRI activity and immune activation in two multiple sclerosis patients treated with the monoclonal anti-tumor necrosis factor antibody cA2. *Neurology* **47**, 1531–1534 (1996).
132. The Lenercept Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group. TNF neutralization in MS: results of a randomized, placebo-controlled multicenter study. *Neurology* **53**, 457–465 (1999).
133. Gregory, A. P. *et al.* TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* **488**, 508–511 (2012).
134. Bustamente, M. F. *et al.* Pharmacogenomic study in patients with multiple sclerosis - Responders and nonresponders to IFN- β . *Neurol. Neuroimmunol. Neuroinflamm.* **2**, e154 (2015).
135. Mahurkar, S. *et al.* Response to interferon-beta treatment in multiple sclerosis patients: a genome-wide association study. *Pharmacogenomics J.* doi: 10.1038/tpj.2016.20 (2016).
136. Butzkueven, H. *et al.* MSBase: an international, online registry and platform for collaborative outcomes research in multiple sclerosis. *Mult. Scler.* **12**, 769–774 (2006).

137. Bloomgren, G. *et al.* Risk of natalizumab-associated progressive multifocal leukoencephalopathy. *N. Engl. J. Med.* **366**, 1870–1880 (2012).
138. Polman, C. H. *et al.* A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N. Engl. J. Med.* **354**, 899–910 (2006).
139. Tan, C. S. & Koralnik, I. J. Progressive multifocal leukoencephalopathy and other disorders caused by JC virus: clinical features and pathogenesis. *Lancet Neurol.* **9**, 425–437 (2010).
140. Yousry, T. A. *et al.* Evaluation of patients treated with natalizumab for progressive multifocal leukoencephalopathy. *N. Engl. J. Med.* **354**, 924–933 (2006).
141. Leoutsakos, J-M. S. *et al.* Effects of non-steroidal anti-inflammatory drug treatments on cognitive decline vary by phase of pre-clinical Alzheimer disease: findings from the randomized controlled Alzheimer's Disease Anti-inflammatory Prevention Trial. *Int. J. Geriatr. Psychiatry* **27**, 364–374 (2012).
142. Wichmann, M. A. *et al.* NSAID use and incident cognitive impairment in a population-based cohort. *Alzheimer Dis. Assoc. Disord.* **30**, 105–112 (2016).
143. Palace, J. *et al.* Effectiveness and cost-effectiveness of interferon beta and glatiramer acetate in the UK Multiple Sclerosis Risk Sharing Scheme at 6 years: a clinical cohort study with natural history comparator. *Lancet Neurol.* **14**, 497–505 (2015).
- A unique study demonstrating how large-scale, longitudinal data collection can be critical for assessing drug efficacy and projecting cost-effectiveness.***
144. Sheridan, C. Anti-CD20 antibody wows in multiple sclerosis. *Nat. Biotechnol.* **33**, 1215–1216 (2015).
145. Hawker, K. *et al.* Rituximab in patients with primary progressive multiple sclerosis: results of a randomized double-blind placebo-controlled multicenter trial. *Ann. Neurol.* **66**, 460–471 (2009).
146. Friese, M. A. *et al.* Acid-sensing ion channel-1 contributes to axonal degeneration in autoimmune inflammation of the central nervous system. *Nat. Med.* **13**, 1483–1489 (2007).
147. Fu, E. *et al.* Immune interventions in stroke. *Nat. Rev. Neurol.* **11**, 524–535 (2015).
148. McKee, J. B. *et al.* Amiloride Clinical Trial in Optic Neuritis (ACTION) protocol: a randomised, double blind, placebo controlled trial. *B.M.J. Open* **5**, e009200 (2015).
149. Vom Berg, J. *et al.* Inhibition of IL12/IL23 signaling reduces Alzheimer's disease-like pathology and cognitive decline. *Nat. Med.* **18**, 1812–1819 (2012).
150. Sandborn, W. J. *et al.* Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *N. Engl. J. Med.* **367**, 1519–1528 (2012).
151. Papatriantafyllou, M. Immunotherapy: immunological bullets against Alzheimer's disease. *Nat. Rev. Drug Disc.* **12**, 24 (2013).
152. Joyner, M. J. & Paneth, N. Seven questions for personalized medicine. *J.A.M.A.* **314**, 999–1000 (2015).
153. Rudick, R. A. *et al.* The Multiple Sclerosis Performance Test (MSPT): an iPad-based disability assessment tool. *J. Vis. Exp.* **88**, e51318 (2014).

154. Ferreira, J. J. *et al.* Quantitative home-based assessment of Parkinson's symptoms: the SENSE-PARK feasibility and usability study. *B.M.C. Neurol.* **15**, 89 (2015).
155. Bot, B. M. *et al.* The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci. Data* **3**, 160011 (2016).
A clinical observational study about Parkinson disease conducted solely through an iPhone app interface, demonstrating the applicability of device-dependent patient monitoring.
156. Cottler, L. B., Zunt J., Weiss, B., Kamal, A. K. & Vaddiparti, K. Building global capacity for brain and nervous system disorders research. *Nature* **527**, S207–S213 (2015).
157. World Health Organization. *Neurologic Disorders: Public Health Challenges* (WHO, 2006).
158. Committee on the Learning Health Care System in America. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America* (National Academies Press, 2013).
159. Roski, J., Bo-Linn, G. W. & Andrews, T. A. Creating value in health care through big data: opportunities and policy implications. *Health Aff. (Millwood)* **33**, 1115–1122 (2014).
160. Burwell, S. M. Setting value-based payment goals — HHS efforts to improve U.S. health care. *N. Engl. J. Med.* **372**, 897–899 (2015).
161. Murdoch, T. B. & Detsky, A. S. The inevitable application of big data to health care. *JAMA* **309**, 1351–1352 (2013).
162. Gray, E. A. & Thorpe, J. H. Comparative effectiveness research and big data: balancing potential with legal and ethical considerations. *J. Comp. Eff. Res.* **4**, 61–74 (2015).
163. Peddicord, D., Waldo, A. B., Boutin, M., Grande, T. & Gutierrez, L. Jr. A proposal to protect privacy of health information while accelerating comparative effectiveness research. *Health Aff. (Millwood)* **29**, 2082–2090 (2010).

Author biographies:

Calliope A. Dendrou is a postdoctoral research fellow in the Oxford Centre for Neuroinflammation, Nuffield Department of Clinical Neurosciences, University of Oxford (UK). She received her Ph.D. in 2010 from the University of Cambridge (UK) and previously trained at Imperial College London (UK). Her work focuses on leveraging genetic and functional data across diseases with an immunological or

inflammatory component to identify shared and distinct pathophysiological mechanisms for therapeutic targeting.

Gil McVean is Professor of Statistical Genetics at the University of Oxford and Acting Director of Oxford's Big Data Institute within the Li Ka Shing Centre for Health Information and Discovery (UK). After undergraduate training in Zoology, he worked in Cambridge (UK) and Edinburgh (UK) before joining Oxford in 2000. His research focuses on understanding the molecular and evolutionary processes that shape genetic variation in populations and the genotype-to-phenotype relationship. He has played a leading role in the HapMap and 1000 Genomes Projects and currently works on organisms from HIV to malaria.

Lars Fugger is an honorary consultant clinical immunologist at the John Radcliffe Hospital in Oxford (UK), Professor of Neuroimmunology at the University of Oxford, and Director of the Oxford Centre for Neuroinflammation. He received his M.D. in 1987 at the University of Copenhagen (Denmark) and subsequently trained at Stanford University (California, USA). His laboratory is working on understanding of how genes confer risk to neuroinflammatory diseases, with a strong focus on translating such findings into the clinical setting.