Introduction

Many neurological disorders are said to be complex, meaning they are oligogenic and multifactorial. That is, they are the end result of the complex effects of several genes interacting with the environment. To identify their genetic background is just as important as finding the environmental factors, in view of their high population prevalence, severe course, serious impact on patients’ disability and progressive mental and physical de-adaptation.

One of the earliest ventures into understanding the genetic basis of these complex diseases was through investigating candidate genes with plausible biological function. The association between the disease and a specific allele of a single nucleotide polymorphism (SNP) within the functional candidate genes is analyzed between patients and controls. Most of these studies included relatively small numbers of patients and controls. Therefore, results have been conflicting or have not been replicated. Another main disadvantage of the hypothesis-based approach of candidate gene studies is that genes involved in the pathogenesis of a disease through unknown pathways are overlooked, while, the hypothesis-free approach by genome-wide association studies (GWAS) allow researchers to overcome this drawback because in these studies nearly all common variants in the entire genome can be tested for association with a disease (1,2). Additionally, in contrast to linkage studies, GWAS utilizes readily available case-control datasets rather than multiplex family based sets, permitting collection of much larger datasets (Figure 1).

Over the past few years, tremendous progress has been made in understanding the genetic basis of complex diseases and hundreds of genetic loci have been identified through the use of GWAS. In this article, we briefly review the role and findings of GWAS in some common neurological diseases. We will then discuss the present and future implications of these findings with regards to disease prediction, uncovering basic biology, and the development of potential therapy.
Stroke

Stroke, as a heterogeneous multifactorial disorder, is the world’s third leading cause of death and a major cause of adult chronic disability. Although epidemiological data from twin and family studies provide substantial evidence for a genetic basis for stroke, the contribution of genetic factors identified so far is small. More recently, the GWAS approach is just beginning to affect the study of stroke. As we all known, over 80% of strokes are ischaemic, as opposed to haemorrhagic. Ischaemic stroke (IS) itself includes several subtypes with differing pathophysiological mechanisms, the most common of which are large-vessel disease stroke, small-vessel disease stroke, and cardioembolic stroke. Most early stroke studies utilizing the GWAS examined whether associations detected by applying the GWAS to other cardiovascular diseases are also risk factors in stroke. Two loci (PITX2 and ZFHX3) associated with atrial fibrillation have both been shown to be risk factors for cardioembolic stroke (3,4). And a locus on chromosome 9p21 originally associated with coronary artery disease was shown to be a risk factor for large-vessel stroke (5). Indeed, it has demonstrated that alleles of 9p21 locus were associated with a broad variety of vascular diseases (6,7). What is worth more noting is an important finding from GWAS, that is different genetic risk factors appear to predispose to specific stroke subtypes. In Japanese populations, a variant in the protein kinase C family (PRKCH) was associated with small-vessel stroke (8). This association was further replicated in patients with subcortical silent brain infarction (9). Moreover, SNPs in CELSR1 gene have been identified as IS risk factors in Japanese individuals (10), while this has not yet been confirmed in the non-Japanese population. A meta-analysis of prospective population based cohort studies reported an association with the 12p13 region, thought to be with the NINJ2 gene, although this result was not replicated in a larger case-control sample (11,12). Recently, the Wellcome Trust Case Control Consortium 2 (WTCCC2) IS GWAS reported a novel association on chromosome 7p21 within the HDAC9 gene, which associated only with large-vessel ischaemic stroke (13). More importantly, through a largest meta-analysis of available IS GWAS datasets, four previously described associations were further confirmed and all these associations were specific to stroke subtypes; PITX2 and ZFHX3 were specific to cardioembolic stroke, whereas HDAC9 and the 9p21 locus were specific to large-vessel strokes (14).

Anyway, GWAS to date have identified robust associations with stroke. Most associations occur with specific stroke subtypes, emphasizing the importance of careful subtyping in genetic epidemiology studies. However, we must know that conventional environmental risk factors play an important role in stroke pathogenesis. Future studies should fully take into account the gene-environment interactions.

Alzheimer’s disease (AD)

AD is a complex multifactorial neurodegenerative disease that is the leading cause of dementia in the elderly. Genetic susceptibility at multiple loci and interactions among these genes influence the risk of developing AD; recent estimates of heritability range from 58-79%. For many years amyloid precursor protein (APP) and the presenilin genes 1 and 2 (PSEN1, PSEN2) have been the only unequivocally established susceptibility genes for early-onset familial AD, and apolipoprotein E (APOE) the only confirmed susceptibility gene for common late-onset AD (LOAD). Since 2009, large GWAS changed the face of the complex genetics of AD. At least nine novel risk loci (CRI, BIN1, CLU, PICALM, CD2AP, CD33, EPHA1, ABCA7 and MS4A cluster) were uncovered, which along with APOE4 contribute
a high proportion of genetic risk for LOAD (15-19). Interestingly, these new identified susceptibility genes are not random, but show patterns of putative functional relation, which increase understanding of the molecular underpinnings of AD by showing prominent roles of the lipid-processing pathway (APOE, CLU, and ABCA7), the immune system (CLU, CR1, ABCA7, CD33, and EPHA1), and the synaptic-cell-functioning pathways (PICALM, BIN1, CD33, and CD2AP) (20-22). Moreover, a recent study has demonstrated that multiple AD-associated genes with evidence for positive selection encode proteins that physically interact within an independently defined protein-protein interaction network, suggesting such loci variants may have functional consequences and be involved in executing a shared molecular function (23). Besides, the combination of information about SNPs at different AD risk genes led to significantly associated genotype patterns that affected episodic memory performance (24). In short, a single locus direct effect is after all small. Further research investigating potential functional gene variants and their mechanistic role in AD is very needed.

**Parkinson’s disease (PD)**

PD is a chronic neurodegenerative disorder that leads to motor and cognitive disability. Although the causes of PD are thought to be primarily environmental, recent studies suggest that a number of genes influence susceptibility. To date, several GWAS for PD have been reported (25-32) and four meta-analyses have been performed (33-36). All GWAS indicate a strong association to the SNCA gene and most studies also confirm an association with the MAPT gene (25-31). Early in 2011, the International Parkinson’s Disease Genomics Consortium (IPDGC) performed a meta-analysis of datasets from five PD GWAS across the United States and Europe in order to identify novel risk loci for PD (33). Besides six previously identified (MAPT, SNCA, HLA-DRB5, BST1, GAK and LRRK2), five novel loci (ACMSD, STK39, MCCCI1/LAMP3, SYT11 and CCDC62/HIP1R) were identified. A second meta-analysis by the IPDGC in collaboration with the Wellcome Trust Case Control Consortium 2 revealed another five PD risk loci (PARK16, STXB1, FGF20, STBD1 and GPNMB) (34). The third and most comprehensive meta-analysis included data from seven million polymorphisms originating either from GWAS datasets or from smaller scale PD association studies (35). Many previously reported risk loci were confirmed and evidence for a new risk variant in the ITGA8 gene was found (35). Along with the deepening in PD meta-analysis, another PD susceptibility locus RIT2 was also identified (36).

To sum up, GWAS identified a number of susceptibility loci in sporadic PD, meta-analyses greatly expanded the list of proposed associations, and further studies replicated and confirmed most of these well-established loci in independent populations (37). By applying the genome-wide pathway analysis, three candidate SNPs, two genes (MAPT and HSD3B7), and 21 pathways involving protein domain specific binding and neurogenesis have also been identified (38). Further efforts to pinpoint functional variants and understand the biological implications of each risk locus are warranted.

**Epilepsy**

Epilepsy is one of the most common neurological disorders characterized by recurrent unprovoked seizures due to neuronal hyperexcitability and abnormal synchronization. Despite advances in epilepsy research, the heterogeneous and complex molecular mechanisms involved in epileptogenesis remain elusive. Genetic factors play a predominant role in 40% of epilepsies, however, early candidate gene-based studies in epilepsy proved ineffective in identifying genetic risk factors. And GWAS of epilepsy patients have been largely negative (39), with the exception of several putative susceptibility loci discovered in Han Chinese focal epilepsy and European Caucasian genetic generalized epilepsy (GGE) patients (40,41). In the combined analysis of the 2-stage Han Chinese GWAS of symptomatic epilepsy (40), the strongest signals were observed with two highly correlated SNPs, rs2292096 and rs6660197, on 1q32.1 in the CAMSAP1L1 gene, which encodes a cytoskeletal protein. More recently, the EPICURE study published a first large GWAS in GGE and reported the allelic associations at 2p16.1 (rs13026414) and 17q21.3 (rs72823592) (41). Further replication studies in independent populations are necessary to validate these novel association findings. Meanwhile, we should be aware that these susceptibility loci appear likely to account for only a small fraction of the heritability of epilepsy, fuelling the effort to search for alternative genetic contributors is quite essential.

**Multiple sclerosis (MS)**

MS is a chronic neuroinflammatory autoimmune disease believed to arise from complex interactions of both
environmental and genetic factors. The importance of genetic in MS is best illustrated by observations of strong familial clustering and a significantly increased risk in first-degree relatives. Meanwhile, some studies in monozygotic and dizygotic twins also indicate a strong genetic component. To date, alleles of the human leukocyte antigen (HLA) are known to make the single strongest contribution to MS susceptibility (42). Additionally, many loci of more modest effect have been identified in recent GWAS. In 2007, the first MS GWAS identified the first non-HLA regions with genome-wide significance. IL7Ra and IL2Ra were found to be significantly associated with MS (43). Simultaneously, the IL7Ra gene was confirmed to be associated in other MS cohorts (44,45). In the years that followed, a new series of GWAS and meta-analyses were performed in different MS cohorts, steadily adding more regions to the list of confirmed MS-associated loci (46-55). What is worth noting is a larger-scale collaborative GWAS reported in 2011, which replicated almost all of the previously suggested associations and identified at least 29 novel susceptibility loci (56). More importantly, a recent analysis shown that approximately 30% of the genetic variation in liability to MS may be directly explained by variants represented by current GWAS arrays (57). And further genome-wide pathway analysis in identification of candidate SNPs, pathways, and hypothetical biological mechanisms (58) will continue to be conducted.

It is also remarkable that the vast majority of the confirmed MS-associated loci are located close or inside genes encoding immune system-related molecules, strongly supporting the hypothesis that MS is primarily an immune-mediated disease. Moreover, one-third of the identified loci were reportedly associated with at least one other autoimmune disease, strengthening the notion that common disease mechanism(s) may underlie most if not all autoimmune diseases (59,60). It is very interesting that a core of genes is shared among multiple autoimmune diseases, and some are disease specific. Future investigation of the shared susceptibility genes might be of great value.

**Migraine**

Migraine is an episodic and disabling neurological disorder and has a strong genetic basis, with heritability estimates of 40-57%. The two most prevalent forms are migraine without aura (MO) and migraine with aura (MA). Many linkage studies and candidate gene studies have suggested causative genes in MO and MA, but few have been replicated. Recent attempts using GWAS have yielded four SNPs, located on chromosome 8q22.1, 2q37.1, 12q13.3 and 1p36.32, that are significantly associated with MA and/ or MO (61,62). Meanwhile, some meta-analysis replicated and confirmed these findings in independent populations (63,64). In another recent GWAS, three additional SNPs located at 1q22 and 3p24 have shown convincing association with MO as well (65). However, all the aforementioned associations conferred only a small to moderate change in risk for migraine. Ongoing large international collaborations to identify additional gene variants for migraine might be very needed in the future. Incidentally, Cox et al. recently did a pedigree-based GWAS in an isolated population of Norfolk Island with a high prevalence of migraine and several novel variants in migraine susceptibility were identified (66). Although the associations might be specific only for the isolated population, we still believe this finding will bring some insights and inspiration to future research.

**Amyotrophic lateral sclerosis (ALS)**

ALS is a fatal disease characterized by motor neuron degeneration, which leads to muscle atrophy and paralysis. Typically, this progressive muscle wasting results in death from respiratory failure within 2-4 years after onset of symptoms. The majority of ALS cases are sporadic; approximately 5% of cases show a family history of disease (67). Although several causative genes have been identified in familial ALS, the genetic contributors to sporadic ALS have been more difficult to identify. GWAS in sporadic ALS have implicated a number of genes or regions (DPP6, ITPR2, UNC13A, FGGY, ELP3, KIF4AP3, 9p21.2) (68-76), but replication of these findings in independent populations has proven difficult (77-81). A possible exception is the 9p21.2 region, which was recently replicated by two large independent studies and seems to be important in both familial and sporadic ALS (82,83). Additionally, the genetic risk factor rs12608932 in UNC13A has been identified as a modifier of survival among ALS patients, which might provide promising therapeutic targets to slow disease progression (84,85). Moreover, environmental exposures also appear to play a role in ALS incidence. Following reports of an increased incidence of ALS in U.S. veterans, interestingly, a GWAS was designed to identify genetic factors that may contribute to ALS in a sample of U.S. veterans (86), while no SNPs reached genome-wide significance for either ALS.
outcome or survival. Further investigation of the nature of the relationship between military service and ALS merits are still needed.

**Frontotemporal lobar degeneration (FTLD)**

FTLD is one of the common causes of presenile dementia worldwide, and familial forms occur in 30%-50% of all frontotemporal dementia (FTD) patients. It manifests, clinically, with behavioural changes and language impairment and is pathologically associated with tau- or ubiquitin-positive inclusions detected in neurons and glial cells of the frontal and temporal lobes in the brain. Up to 50% of familial cases show association with chromosome 17: mutations in the microtubule-associated protein tau (MAPT) and progranulin (GRN) are considered as the main genetic cause of FTD. Rare defects in VCP, CHMP2B, TDP-43 and FUS genes have been found in a small number of families (87).

Recently, a first GWAS in FTLD identified common genetic variability at the TMEM106B gene on chromosome 7p21.3 as a potential important risk modifying factor for FTLD with pathologic inclusions of TAR DNA-binding protein (FTLD-TDP), the most common pathological subtype in FTLD (88). To gather additional evidence for the implication of TMEM106B in FTLD risk, multiple replication studies in geographically distinct populations were set up (89,90). It has demonstrated that the most significant association of TMEM106B polymorphisms with risk of FTLD-TDP was observed in patients with GRN mutations. Subsequent studies suggested an inverse correlation between TMEM106B expression and GRN levels in patient serum (90). Accordingly, a biochemical and cell biological basis for the understanding of the pathological role of TMEM106B in FTLD was also presented (91). Additional studies to further unravel the molecular role of TMEM106B in disease onset and progression are still needed.

It is particularly worth mentioning that FTLD may be found alone or accompanied by corticobasal degeneration, parkinsonism, or ALS. As many as 50% of ALS patients have cognitive and behavioral deficits and about half of FTLD cases have motor neuron impairment. Along this line, Vass et al. investigated whether this FTLD-TDP associated TMEM106B risk genotypes could also contribute to the risk and disease presentation of ALS. Although no association between the TMEM106B genotypes and ALS was demonstrated, the protective allele at TMEM106B did appear to be associated with preserved cognition in ALS patients (92). This finding, although preliminary, are of interest in light of the TDP-43 proteinopathy spectrum that binds FTLD and ALS. In addition, gene linkage to chromosome 9p has already been established in familial FTLD-ALS. And recently, a genetic locus associated with FTLD and ALS was identified as a hexanucleotide repeat located on chromosome 9, open reading frame 72 (C9ORF72) (74,83,93), which causes both familial and sporadic types of these two disease and can be believed as the most common cause of familial FTLD-ALS. As the pathophysiological mechanisms by which this mutation leads to FTD and ALS begin to be understood, we have reason to believe new targets for disease-modifying therapies will likely be unveiled (94).

**Restless legs syndrome (RLS)**

RLS is a sensorimotor disorder in which affected individuals suffer from uncomfortable sensations and an urge to move their lower limbs; it occurs mainly in resting situations during the evening or at night. The high familial aggregation of RLS suggests a genetic component. Through GWAS analysis, RLS was identified an association with variants in gene MEIS1, BTBD9, and MAP2K5/LBXCOR1 on chromosomes 2p, 6p and 15q (95,96). These findings were further confirmed in several studies (97,98). Recently, another GWAS identified two additional RLS susceptibility loci within an intergenic region on chromosome 2p14 and 16q12.1 (99). Opportunities now exist to study the functionality of the RLS risk susceptibility loci and the potential molecular pathways involved. Some illustrative examples include the study reporting that the MEIS1 risk variant influences iron metabolism (100) and BTBD9 regulates brain dopamine levels and controls iron homeostasis through the iron regulatory protein-2 (101), adding to growing body of evidence that dopaminergic neurotransmission and iron dysregulation might contribute to the pathogenesis of RLS.

**Intracranial aneurysm (IA)**

The rupture of IA is the most common cause of subarachnoid hemorrhage associated with high morbidity and mortality, while its pathogenesis is still unknown. Because catastrophic hemorrhage is commonly the first sign of disease, early identification is very essential. Two GWAS in European and Japanese cohorts have successfully identified common variants located on chromosomes 8q11,
9p21, 10q24, 13q13 and 18q11 that are associated with IA (102,103). Some independent GWAS in the Japanese population was also conducted (104,105). A recent largescale GWAS replicated and validated rs10757272 on CDKN2BAS at chromosome 9p21.3 to be significantly associated with IA, meanwhile, a novel functional SNP, rs6842241, near EDNRA at chromosome 4q31.22 was found (105), which might affect the expression of EDNRA and subsequently result in the IA susceptibility. In addition, a meta-analysis also yielded genome-wide significance for SNP in CDKN2BAS on chromosome 9p (rs6475606) and supported previously reported SNP in SOX17 on chromosome 8q (rs1072737) contributing to IA susceptibility (106).

Among the loci above, 9p21 locus seems most intriguing because the SNPs identified by the GWAS of IA are in linkage disequilibrium with SNPs associated with several other arterial diseases (coronary artery disease, abdominal aortic aneurysm, and peripheral arterial disease) (7), and this risk locus overlaps an annotated noncoding RNA, called ANRIL (or CDKN2BAS) (107). More importantly, 9p21 variation is found associated with the site distribution of IA, especially IA of the posterior communicating artery and the posterior circulation than in the anterior circulation, which are clinically important (108).

Human prion diseases

Prion diseases are fatal neurodegenerative diseases of humans and animals caused by the misfolding and aggregation of prion protein (PrP). Mammalian prion diseases are under strong genetic control but few risk factors are known apart from the PrP gene locus (PRNP). In order to providing other (non-PRNP) genetic modifiers for human prion disease, many studies were conducted based on candidates derived from close functional links to PrP, screening human genes orthologous to mouse candidates, and also human GWAS (109). From a GWAS in variant Creutzfeldt-Jakob disease (vCJD), it has been demonstrated that SNPs upstream of the RARB and STMN2 loci were candidates for vCJD association (110). Interestingly, markers on human chromosome 8 in this human vCJD study share conservation of synteny with the region of cattle chromosome 14, where an association was reported with classical bovine spongiform encephalopathy (111). Moreover, another large GWAS in human prion diseases provide evidence for several non-PRNP genetic risk factors as well, although none of these loci achieved genome-wide significance in meta-analyses between regions or aetiologies. It is unsurprising that PRNP genotypes were found strongly associated with risk in all categories and regions of human prion diseases (112). Further genetic analyses are required to provide convincing evidence for association at these non-PRNP loci. Meanwhile, the mechanism by which the existing loci exert their influence on nearby genes and potential molecular pathways involved need to be investigated.

Moyamoya disease (MMD)

Moyamoya, a poorly understood arteriopathy derives its name from the characteristic angiographic pattern. The name means ‘hazy puffs of smoke’ describing the small vessel collateral system that develops in response to hyperplastic stenosis and occlusion of the distal internal carotid and proximal vessels of the circle of Willis. MMD occurs worldwide, but its prevalence is highest in East Asian countries and 15% of MMD cases have a family history (113). Genetic analyses in familial MMD and GWAS represent promising strategies for underling its pathogenesis. A recent GWAS was performed in Japanese MMD patients, resulting in a strong association of chromosome 17q25-ter with MMD risk (114). And candidate locus on 17q25 was revealed associated with MMD before. What’s more, a single haplotype consisting of seven SNPs at the RNF213 locus was tightly associated with MMD and mutational analysis of RNF213 revealed a founder mutation greatly increases MMD risk (114). Therefore, RNF213 can be identified as a susceptibility gene for MMD and its possible role in vascular development (115) should be discussed further.

Conclusions and future directions

Significant progress has been made in recent years in unraveling the genetic roots of complex diseases. GWAS have played a major role in this effort as they have allowed for the discovery of novel genetic associations that are not based on prior knowledge. The studies reviewed above provide a synopsis of the major advances in several topics related to neurological diseases.

There are a number of key concepts that can be gleaned from these studies. Firstly, larger sample sizes are needed to detect variants with small effect sizes. Therefore, a metaanalysis of available GWAS datasets and further GWAS in different populations are of great value. Despite large collaborative efforts, so far, GWAS findings explain only
a small proportion of the heritability of complex diseases, which makes genetic risk prediction tests currently unfeasible for these diseases. It is evident that a proportion of the genetic predisposition will be explained by highly penetrant rare variants, while these rare variants will not be captured with GWAS. Further research, including new approaches to detect rare variants using next generation sequencing and structural variants, including copy number variants such as insertions and deletions, might improve genetic risk prediction for these diseases.

What’s more, reviewing some compelling loci based on current data from GWAS, genes that have been linked to more than one disease entity are emphasized. It is very interesting that a core of genes is shared among multiple diseases in a general category. Such overlapping findings could point to several common genetic and mechanistic denominators, especially in terms of neurodegeneration, vascular disease or autoimmune diseases. Unveiling the identity of these shared genetic factors will not only improve our understanding of the underlying pathophysiology, but also may lead to new avenues for preventing and treating these devastating diseases.

It is particularly worth noting that GWAS is not a method to conclude something. “Significantly associated” does not mean that the association identified by GWAS is significant in biology, medicine, or actual life. It’s just a method to map the gene and presents candidates of the causal variant. To identify and prove the causality, we need to find the best associated variations by fine mapping around the smoking gun GWAS signal, followed by functional biological studies to understand the underlying biology for them. Further efforts are necessary to make full use of its potential power to medicine. Therefore, identification of the significantly associated SNP is just ‘end of the beginning’ to get to the final goal: treatment of patients with common diseases. However, at present, after identification of ‘marker’ SNPs by GWASs, we have no golden road to find ‘causal’ variants. We have no customized method to identify functional variants. By studying the meaning of allelic differences of causal variants, we can clarify the molecular pathogenesis of the disease. With the knowledge of the pathogenesis, we can walk directly to develop innovative treatments and discover new drugs. So, there is still a long and winding road before us.

Although current examples of immediate clinical impact from GWAS studies are limited, we strongly believe that GWAS have begun to revolutionize our understanding of the genetic underpinnings of human disease. It’s just the beginning and it gives us wonderful start points. Future work should prioritize the studies designed to ascertain rarer SNPs with possibly larger functional effects, analysis of potential pathway and further study of the modification of clinical phenotype by genetic polymorphism. We believe that starting with GWAS results will increase the likelihood of successful biological and therapeutic discoveries in neurology in the years to come.

**Acknowledgements**

This work was supported by grants from the National Natural Science Foundation of China (81000544, 81171209), the Shandong Provincial Natural Science Foundation, China (ZR2010HQ004, ZR2011HZ001), the Medicine and Health Science Technology Development Project of Shandong Province (2011WSA02018, 2011WSA02020), and the Shandong Provincial Outstanding Medical Academic Professional Program.

**Disclosure:** The authors declare no conflict of interest.

**References**

8. Kubo M, Hata J, Ninomiya T, et al. A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk...
34. International Parkinson’s Disease Genomics Consortium


