Findings Of This Thesis

This thesis aimed to identify genomic and metabolomic determinants of neurological and psychiatric disorders and their related endophenotypes by making use of various omics approaches. This chapter summarizes the main findings of this thesis, discusses the implication towards the understanding of molecular processes and pathways underlying these disorders and comments on future research.

OMICS OF NEURODEGENERATION

The five projects described in **Chapter 2** focus on endophenotypes of neurological and psychiatric disorders including brain volumetric measures obtained by magnetic resonance imaging (MRI) and cognitive ability.

Brain MRI provides an opportunity to study the complex architecture of brain structures and related disorders. In **Chapter 2.1**, I performed the genome-wide association study of lateral ventricular volume in 23,533 middle-aged to elderly individuals from 26 population-based cohorts participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. I identified, for the first time, association of lateral ventricular volume and genetic variants at 7 loci (3q28, 16q24.2, 7p22.3, 12q23.3, 22q13.1, 10p12.31 and 11q23.1) (**Figure 1**). Some of the identified loci have previously been linked to various pathologies including cerebrospinal tau/ptau levels, Alzheimer's disease risk, and cognitive decline (3q28),[1] tau pathology (12q23.3),[2] or small vessel disease and white matter lesions (16q24.2).[3] Additionally, several biological pathways emerged including regulation of cytoskeleton organization and S1P signaling.[4-6] The findings described in this chapter provide new insights into understanding complex genetic architecture underlying brain structures. However, identified associations even cumulatively do not account for a substantial fraction of heritability of lateral ventricular volume. In addition to increasing sample size, studying low-frequency and rare variants, incorporating interactions or investigating aggregate or multivariate effects holds promise to expand our knowledge about the genetic architecture of brain structures and related disorders.[7,8]

I next focused on cognitive function, an important predictor of health outcomes, including mortality and morbidity.[9-12] Even though some genes were discovered in specific domains of cognition,[13,14] capturing all cognitive domains into general cognitive function has been more successful in detecting genetic variants.[15] In **Chapter 2.2,** I performed the genome-wide association study of general cognitive function in 243,000 participants of European ancestry (EA) from CHARGE consortium and UK Biobank. I reported 32 novel genetic loci, bringing the total number of independent loci implicated in general cogni-

tive function up to 180.[15] I also showed that genetic risk score based on 180 loci was significantly associated with general cognitive function in 2,117 participants of African-American ancestry (AA) from CHARGE consortium, suggesting that findings in EA can be generalized to AA. Furthermore, I linked genes implicated in general cognitive function to circulating levels of metabolites and found association with tyrosine, an amino acid that plays an important role in synthesis of dopamine,[16,17] glycoprotein acetyl, a marker of acute phase reaction associated with future mortality and cognitive ability,[18,19] and 22:6 docosahexaenoic acid (DHA), a long-chain omega-3 polyunsaturated fatty acid, that has been associated with cognitive function and risk of Alzheimer's disease and dementia.[18] Using Mendelian randomization, I also showed that that genes determining circulating levels of tyrosine and glycoprotein acetyl also determine general cognitive function while DHA is rather a consequence of the physiological processes determining cognitive function. The findings described in this chapter provide new insights into general cognitive ability and demonstrate that further integration of genetic and molecular data with nongenetic data holds great potential to provide additional information about variation in cognitive ability. Discovery of casually associated metabolites provides insights into the pathways underlying general cognitive function and provides starting point for new preventive studies. However, the fact that studied metabolites represent only small proportion of circulating metabolites asks for future studies focusing on a wider spectrum of metabolites. Future efforts should also focus on improving the power of genome-wide association studies of metabolites as novel genetic instruments for running Mendelian randomization for these metabolites may be revealed. Last but not least the causal association between circulating metabolites and general cognitive function should be replicated in other ethnic groups. Future epidemiological research efforts should focus on longitudinal data to validate the cross –omics findings.

There is increasing interest in epigenetics studies that reflect the effect of the genome and exposome (e.g. diet, life style, medication). As cognitive function is also determined by environmental factors, and the complex balance between genes and environment is poorly understood, studying epigenetic signatures may provide insights into cognitive function. [20] In **Chapter 2.3**, we studied the association of blood-based DNA methylation and cognitive test scores in up to 6,809 healthy adults from 11 cohorts. We identified a significant association of two CpG sites and cognitive tests. Cg21450381, located in an intergenic region on chromosome 12 was associated with global cognitive function, whereas cg12507869, located in the *INPP5A* gene on chromosome 10, was associated with verbal fluency. The findings described in this chapter provided evidence for blood-based epigenetic signatures of cognitive function. However, methylation signatures for cognitive function are modest compared to other traits such as body mass index.[21] One of the reasons that may explain this is that the epigenetics in the blood is a poor

surrogate for the post-translational processes in the brain. Improving statistical power by adding additional samples and using newly developed methylation arrays with increased genome coverage may lead to novel discoveries in the future.[22] Furthermore, disentangling correlation from causation in epigenetics is also important from a genetic epidemiological perspective. Careful epidemiological follow-up studies could be used. Alternative approach is Mendelian randomization that makes use of the fact that the genetic drivers of methylation are being rapidly uncovered.[23] Mendelian randomization approach uses a genetic proxy for DNA methylation to evaluate causal relationship between the disease outcome or trait and epigenetic variation and has the potential to help distinguish between truly causal intervention targets and non-causal, which may be informative biomarkers.[24,25]

Even though genome-wide association studies, have been successful in identifying the genetic variants underlying cognitive ability, hypothesis-driven candidate gene design in which biologically relevant regions of the genome are studied in relation to cognitive ability could also shed light on pathways involved in cognitive ability.[26] With recent advances in high-throughput technologies, candidate gene approach is making its re-appearance in genetic epidemiology.[27] In **Chapter 2.4,** I used exome-sequencing data in order to study impact of rare genetic variants in the dystrophin gene (*DMD)* on cognitive ability in about 2,700 participants from two studied populations including family-based Erasmus Rucphen Family (ERF) study and population-based Rotterdam Study. I found a suggestive association of rs147546024:A>G and visuospatial ability in ERF study. However, I was not able to replicate this finding in the Rotterdam Study. I also found a missense variant rs1800273:G>A to be nominally associated with cognitive tests in ERF and Rotterdam Study. The variant, predicted to have a damaging effect on the protein, is present in the different isoforms which are expressed in the brain and which have a stabilizing effect on the GABA receptors recognized for regulation of cognition, emotions, and memory.[28-31] This chapter highlights the challenges of search and replication of rare variant associations. The replication of rare variants is even more challenging if the variants are identified in family-based studies and validation of findings in general population requires extremely large studies. Family-based studies have unique advantages such as enrichment of rare variants and control of population stratification.[32] This design holds great promise for success in searching for rare variants and the chances of success are even higher in genetic isolates since due to genetic drift and inbreeding over several generations, rare variants become more frequent over generations.[33-36]

As *DMD* gene has several different isoforms, **Chapter 2.5** focused on studying association between intelligence and structural mutation location and affected dystrophin isoforms including full-length dystrophin isoform (Dp427) and shorter dystrophin isoforms

(Dp260, Dp140, Dp116, and Dp71/Dp40).[37-39] The study population included patients with Duchenne muscular dystrophy (DMD), a fatal muscular dystrophy during childhood that leads to progressive muscular weakness and less well described nonprogressive central nervous system manifestations.[40] We found that mutations affecting expression of several isoforms including Dp427, Dp140 and Dp71/Dp40 were associated with higher frequency and severe cognitive impairment confirming the findings of previous studies that cumulative loss of dystrophin isoforms has an impact on intellectual ability.[41-45] Furthermore, we observed that expression of Dp140 isoform is not mainly affected by the mutations located in 5'UTR.[41] The findings of this chapter are relevant for personalized medicine initiatives as they allow recognition of the subgroup of patients with great risk for cognitive problems in whom early intervention and support in cognitive, emotional and behavioral development could be very useful.

OMICS OF NEUROVASCULAR PATHOLOGY

Four projects described in **Chapter 3** focus on omics of neurovascular pathology. In **Chapter 3.1** and **3.2**, I focused on endophenotypes of neurodegenerative disorders and stroke characterized by imaging, including carotid intima-media thickness (cIMT) and carotid artery calcification.

Carotid-intima media thickness is an established heritable marker for subclinical atherosclerosis that has been shown to predict future cardiovascular events.[46-48] As previous genome-wide and exome-wide studies identified only a few genetic regions that explain a small proportion of trait variance, and sequencing study of candidate regions yielded inconclusive results due to limited power, more powerful approaches for uncovering the role of rare variants are needed. In **Chapter 3.1**, I performed a genome-wide linkage analysis of individuals in the extremes of cIMT trait distribution (>90th percentile) followed by fine-mapping using exome-sequencing in a large family-based study from a genetically isolated population in the Netherlands. I observed significant evidence of linkage on chromosomes 2p16.3, 19q13.43, 20p13, and 21q22.12. Fine-mapping using exome-sequencing data identified a variant under the linkage peak at 2p16 mapped to *PNPT1* gene which has been characterized as a type I interferon-inducible early response gene.[49-51] Interestingly, several plausible candidate genes were noted under 19q13.43, 20p13, and 21q22.12 peaks, which are highly expressed in tissues relevant for atherosclerosis and linked to pathways implicated in the development of atherosclerosis or cardiovascular diseases. The results of this chapter provide novel insights into genetic architecture of cIMT by making use of extreme phenotype approach. This approach was reported to be better powered in rare variant studies as it reduces phenotypic

heterogeneity.[52] Future functional studies of identified candidate genes are needed to validate and explore the findings ultimately leading to biological pathways involved in the etiology of subclinical atherosclerosis.

Another proxy of carotid artery atherosclerosis studied in this thesis is carotid artery calcification. Carotid artery atherosclerosis is associated with stroke, dementia, and cognitive decline.[53,54] As the specific location of carotid atherosclerosis, i.e. extracranial versus intracranial, may develop under the influence of different metabolic risk factors, in **Chapter 3.2** I performed further in-depth investigation of metabolic determinants and extra- and intracranial carotid artery calcification (ICAC and ECAC) in 1,111 participants from the Rotterdam Study. The significant evidence for association was found between 3-hydroxybutyrate, a ketone body, and ICAC volume. Additionally, the metabolic association pattern of ICAC was found to be different compared to that of ECAC providing further evidence for location-specific differences in the etiology of atherosclerosis.[55,56] However, our study was not designed to resolve the question of reverse causation, emphasizing need to explore metabolomics in longitudinal studies. Also here, Mendelian randomization may contribute substantially to separate associations that are a cause or rather a consequence of disease. This was illustrated by Liu *et al.* who used genetic determinants robustly associated with plasma metabolite levels in order to investigate causal relationship between circulating metabolites and fasting glucose and type 2 diabetes.[57] This may help in translating findings from observational studies from association to causation.

I next focused on stroke, a neurological deficit of sudden onset. As risk determinants of stroke are various complex modifiable risk factors, detailed profiling of metabolic status facilitated by development of high-throughput technologies, could provide novel insights into metabolic changes and identify individuals with higher risk of stroke. In the most comprehensive study to date conducted within the China Kadoorie Biobank, several circulating compounds were associated with stroke, including lipids and lipoprotein particles of various sizes, glycoprotein acetyls, ketone bodies, glucose and docosahexaenoic acid.[58] As large metabolomics studies of stroke in individuals of European ancestry are lacking, in **Chapter 3.3**, I investigated association of circulating metabolites and risk of stroke in seven population-based cohorts including more than 1,790 incident stroke events among 38,797 participants. The significant associations were found between incident stroke and amino acid histidine, glycolysis-related metabolite pyruvate, acute phase reaction marker glycoprotein acetyls, cholesterol in high-density lipoprotein-2 and several other lipoprotein particles. Furthermore, amino-acid phenylalanine and total and free cholesterol in large high-density lipoprotein particles were associated with risk of ischemic stroke. Our results confirmed the association of glycoprotein acetyl and

ischemic stroke that was observed in individuals within the China Kadoorie Biobank, however, we also observed associations that are specific for Western societies.[58] Environmental and ethnic differences across populations or the confounders adjusted for could explain lack of the replication. The results of this chapter provide insights into understanding metabolic determinants of stroke and highlight potential of metabolomics approach in identifying potential targets for the prevention strategies. Future studies should focus on wider range of metabolites and collection of blood samples at multiple time points. As identified associations are starting point for relating metabolites to their biological role, new efforts should focus on integrating metabolomics and other –omics data in order to shed light on metabolic pathways underlying stroke.

Finally, the last project described in **Chapter 3** focused on relation between gut microbiota and human metabolome. As described in **Chapter 3.4,** we examined the association of gut microbiota on circulating metabolites in 2,309 individuals from two population-based cohort studies, Rotterdam Study and LifeLines DEEP. We found 32 microbial families and genera to be associated with a wide range of circulating metabolites including specific very-low density and high-density lipoprotein subfractions, serum lipid measures, glycolysis-related metabolites, amino acids, and acute phase reaction markers. These results provide insights into the role of microbiota in human metabolome, supporting the role of gut microbiota as a target for therapeutic and preventive interventions. However, the current challenges to study microbiota are large, involving reverse causality and bias due to confounding (e.g. the association of the gut microbiome with diabetes was found to be contributed largely to the effects of metformin).[59] Capturing the gut microbiota composition is easy for the distal part of the gut through feces, but microbiota of the proximal part of the gut is more difficult to characterize.[60,61] Perhaps the largest hurdle to overcome is that to date there are no studies that have stored feces samples, allowing prospective studies.[62] An alternative approach may be to use Mendelian randomization, as also the gut microbiome is determined by the human host genome.[63]

GENETIC STUDIES OF PSYCHIATRIC DISEASES

The two projects described in **Chapter 4** focus on genetic determinants of neurodevelopmental disorders including autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD).

Even though ASD is a heritable disorder and most genetic variance is attributed to common genetic variants, not many loci have been identified.[64,65] The largest effort to

date including more than 16,000 individuals failed to identify new common genetic variants.[66] This suggests that common variants individually have low impact in the ASD, as seen in other psychiatric disorders. However, their joint effect may be substantial.[67] Therefore, in **Chapter 4.1**, I performed a single-variant and gene-based genome-wide association studies in a sample of 160 families with at least one child affected with non-syndromic ASD using both binary phenotype and a quantitative autistic trait. The majority of patients in the study had a normal intelligence, unlike the most ASD cohorts in which rates of intellectual disability are ranging from 30 to 50%.[68] I identified a novel gene *TTC25* associated with quantitative autistic trait in gene-based analysis, replicated this association in an independent sample of general population, and confirmed association of ASD with the known 5p14.1 locus.[69,70] This chapter demonstrates power of endophenotypes to identify a genetic signal, the strength of phenotypic homogeneity of the sample (normal intelligence of majority of the sample) and advantage of gene-based test compared to single-variant analysis.

ADHD is also heritable disorder and a substantial proportion of genetic variance is attributed to common genetic variants.[71-73] However, the first risk loci have been described recently.[74] As ADHD is the extreme end of a continuous ADHD symptoms scores, novel variants could be discovered by focusing on quantitative ADHD symptoms. In **Chapter 4.2**, we sought to leverage the power of population studies of ADHD symptoms in adults in order to discover disease-relevant genes in nine cohorts including about 15,000 individuals. The most strongly associated variant in a genome-wide meta-analysis was mapped to *STXBP5-AS1*. This association was confirmed in the replication analysis of childhood ADHD symptom scores (n~15,000).[75] Even though the function of *STXBP5-AS1* is currently unknown, this lncRNA overlaps in anti-sense with *STXBP5* encoding a protein involved in synaptic function by regulating neurotransmitter release.[76,77] The results of this chapter provide novel insights into the genetic underpinnings of ADHD symptoms implicating synaptic function regulation through *STXBP5-AS1* and potentially *STXBP5* in ADHD symptom etiology.

New insights into the genetics of neurological and psychiatric disorders and related endophenotypes described in this thesis are summarized in **Figure 1.** Previously reported genomic regions are shaded.

FROM OMICS TO TRANSLATION

With the development of high-throughput technologies and omics approaches, our understanding of disease pathophysiology is improving. The major expectation is that

Figure 1. Association of neurological and psychiatric disorders with SNPs accross the genome. Previously reported genomic regions (GWAS catalog as of April 2018) are shaded, whereas the new insights are depicted by arrows. Colors represent studied traits.

these approaches will provide valuable information for prevention programs, earlier disease diagnosis, and personalized treatments taking into account individual variability in context of precision or personalized medicine.[78,79] The research presented in this thesis provides novel insight into the genetic and metabolic determinants or neurological and psychiatric disorders. Even though genetic variants have small effect sizes, going beyond genetic loci in combination with other –omics markers may help classify individuals with higher risk in precise manner.[80]

Genome-wide sequencing studies showed that most individuals carry at least some potentially deleterious variants in their genome.[81] However, the effects of these mutations on individuals are not well understood and to identify their function will be a tall order as there are millions of variants to be mapped in the medium-sized population. As the time and cost would be huge, a major question is the value which refers to doing things at high quality, safely, and at reasonable cost. However, there are certain medical condi-

tions for which genetic testing can provide new opportunities for patients' management but precision or personalized medicine has not seen widespread adoption because of difficulties in achieving a balance between providing personalized care at a population level and delivering standardized care. In **Chapter 5.2** we present how outcome-based healthcare system design developed by the Value-based Healthcare Programme at the University of Oxford could deliver better patient and population-level outcomes and personalized care for cardiovascular disease in a standard way. This chapter does not address a genetic disorder involved in neurodegeneration or neurovascular pathology but focuses on inherited heart rhythm disorder. As a proof of principle, we applied this approach to long QT syndrome (LQTS). We designed two outcome-based systems to focus on patient outcomes, which means that they are service agnostic, context-independent and applicable in a variety of healthcare organizations irrespective of resource constraints.

FUTURE RESEARCH

The extensive research efforts in the past decades have made a progress in understanding the complex architecture of neurological and psychiatric disorders. Identification of numerous common and rare variants underlying these disorders was facilitated by larger sample sizes and development of relatively inexpensive SNP arrays.[82] Resources such as large biobanks that collect biological material and phenotype data made it possible to dramatically increase sample size for some of the traits.[83,84] Population-wide biobanks have been developed in several countries including UK (UK Biobank, n = 0.5 million),[85] Estonia (Estonian Biobank, n = 52,000),[86] USA (Million Veteran Program, n = 1 million),[87] China (Kadoorie Biobank, n = 0.5 million).[88,89] These biobanks will be a large resource for studying neurological and psychiatric disorders and related endophenotypes in the future. Additionally, the majority of studies to date have been conducted in participants of European ancestry. Therefore, future studies should also focus on generalization of the findings in other ancestries and multi-ethnic studies.[90] Increasing the sample size by adding additional samples may lead to novel genetic discoveries and expansion of our knowledge on novel pathways underlying these disorders. Furthermore, novel genetic variants together may improve classification of neurological and psychiatric disorders and facilitate identification of individuals at high risk. As demonstrated by van der Lee *et al.* cumulative effect of common genetic variants modified the risk of Alzheimer's disease and all-cause dementia beyond the *APOE* genotype and contributed to better risk prediction.[91]

With development of SNP arrays and statistical imputation of unobserved variants, both common and less frequency variants could be assessed in the population increasing power of association studies and facilitating discovery of new loci.[92-94] On the other hand, whole-exome and whole-genome sequencing are expected to identify rare-variants. Even though studies of rare variants also require large datasets, these datasets are still small compared to datasets used for discovery of common variants. Focusing on family-based designs or studying extreme cases would be more efficient approach. For example, I showed in **Chapter 3.1** that studying extreme cases in a family-based study using linkage analysis could identify novel loci in a smaller sample. However, this approach should be complemented with deep sequencing. The power could also be boosted by combining the alleles of similar impact in a gene or a region.[82] This approach may also be highly relevant for personalized medicine - as discussed above - evaluating the health threat of damaging mutation is a tall order for rare variants in the population. However, within a family such a variant is not rare and segregates with a probability of 50%. Thus 50% of first-degree relatives and 25% of second-degree relative are carriers and those in other generations can provide key clues.

Apart from identifying new loci associated with neurological and psychiatric diseases focus in coming years should also be on understanding how these loci contribute to the disease. As illustrated in **Chapter 2.1**, most of the variants are located in non-coding regions of the genome. Understanding regulatory components of genome became recently available by projects such as ENCODE,[95] Epigenome RoadMap,[96] and GTEx.[82,97] For neurological and psychiatric disorders appropriate tissue-specific resources are also important and essential. These methods are useful for prioritizing genes from GWAS loci for functional follow-up with the ultimate objective to enable more effective prevention and treatment strategies of disease.[82]

Furthermore, other single omics approaches including proteomics, metabolomics, and microbiomics could also provide information about the biological processes. These fields could also greatly benefit from large population-based samples as increasing the sample size may lead to novel discoveries. In this thesis, we showed that large studies identified metabolites associated with stroke (**Chapter 3.3**) and gut microbiota (**Chapter 3.4**). The identified metabolites could be further studied, for example in relation to genetic determinants. Future studies should integrate multiple levels of data in multi-omics studies. This would overcome limited information derived from single omics approaches and could provide additional biological insight useful in understanding complex disorders (**Figure 2**). Exploration of genome, transcriptome, metabolome and microbiome levels could yield important conclusions that will be basis for precision medicine.[20]

Integrative multi-omics approaches should also focus on appropriate target tissues. One of the alternative approaches to examine tissue specificity for cells and tissues that

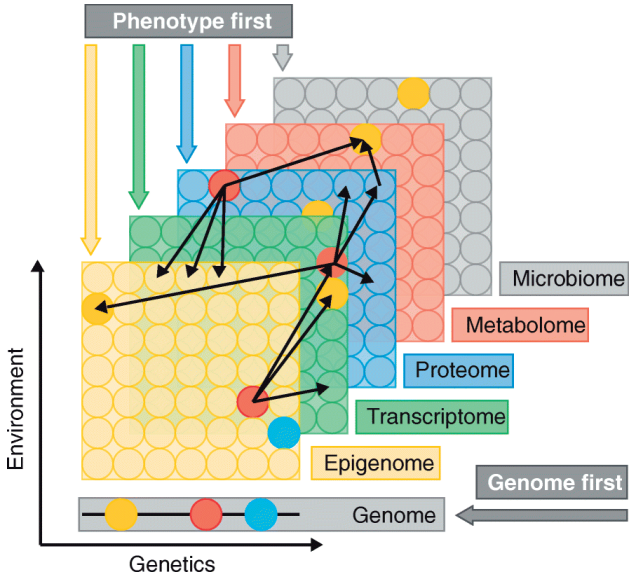

Figure 2. Multiple omics data. Multiple omics data types are depicted by layers. Pool of molecules collected within each of the layers is depicted by circles. All the layers except genome reflect genetic and environmental regulation. Arrows represent potential interactions between the molecules in different layers, whereas interaction within the layers are not shown. Source: Hasin *et al*.[20]

are difficult to obtain includes use of induced pluripotent stem (iPS) cell technology.[98] This technology may help to differentiate easily accessible cells into different brain cells. These cells with relevant genetic background or subjected to gene-editing through CRISP-Cas9 technology could be further used to reconstruct diseased brain models using organ-on-chip technology.[98-100]

Lastly, future translational research efforts may benefit from longitudinal measures. Collection of quantitative traits at different time points can reduce type I error and increase statistical power compared to a single measurement and could also identify determinants for age of onset.[101,102] Similarly, longitudinal profiling in a single individual could also be beneficial as it could provide consistent monitoring of dynamic changes in multi-omics components in relation to disease status and preventive interventions.[103]

CONCLUDING REMARKS

In this thesis, I have used various omics approaches in order to provide novel insights into the pathophysiology of complex neurological and psychiatric disorders and related endophenotypes. Discovery of novel genetic determinants underlying brain structures, cognitive ability, and neurodevelopmental disorders, as well as link between circulating metabolites and neurovascular pathology or gut microbiota are some of the highlights of this work. With the development of high-throughput technologies and integration of various approaches in the future, novel insights into molecular mechanisms underlying these disorders will be provided as well as valuable information for prevention programs, earlier disease diagnosis, and personalized treatments.

REFERENCES

1. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* **78**, 256-68 (2013).
2. Lasagna-Reeves, C.A. *et al.* Reduction of Nuak1 Decreases Tau and Reverses Phenotypes in a Tauopathy Mouse Model. *Neuron* **92**, 407-418 (2016).
3. Traylor, M. *et al.* Genetic Variation at 16q24.2 Is Associated With Small Vessel Stroke. *Annals of Neurology* **81**, 383-394 (2017).
4. Mishra, A. & MacGregor, S. A Novel Approach for Pathway Analysis of GWAS Data Highlights Role of BMP Signaling and Muscle Cell Differentiation in Colorectal Cancer Susceptibility. *Twin Research and Human Genetics* **20**, 1-9 (2017).
5. Shen, H.Y. *et al.* Coupling between endocytosis and sphingosine kinase 1 recruitment. *Nature Cell Biology* **16**, 652 (2014).
6. Ma, S., Santhosh, D., Kumar, T.P. & Huang, Z. A Brain-Region-Specific Neural Pathway Regulating Germinal Matrix Angiogenesis. *Developmental Cell* **41**, 366 (2017).
7. Mufford, M.S. *et al.* Neuroimaging genomics in psychiatry-a translational approach. *Genome Med* **9**, 102 (2017).
8. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
9. Deary, I.J., Weiss, A. & Batty, G.D. Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychological Science in the Public Interest***.11**, pp (2010).
10. Batty, G.D., Deary, I.J. & Gottfredson, L.S. Premorbid (early life) IQ and later mortality risk: systematic review. *Ann Epidemiol* **17**, 278-88 (2007).
11. Wraw, C., Deary, I.J., Gale, C.R. & Der, G. Intelligence in youth and health at age 50. *Intelligence. Vol.53 2015, pp. 23-32. Nov-Dec Intelligence* (2015).
12. Hagenaars, S.P., Gale, C.R., Deary, I.J. & Harris, S.E. Cognitive ability and physical health: a Mendelian randomization study. *Scientific Reports* **7**(2017).
13. Ibrahim-Verbaas, C.A. *et al.* GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Mol Psychiatry* **21**, 189-197 (2016).
14. Debette, S. *et al.* Genome-wide studies of verbal declarative memory in nondemented older people: the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium. *Biol Psychiatry* **77**, 749-63 (2015).
15. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* **9**, 2098 (2018).
16. Jongkees, B.J., Hommel, B. & Colzato, L.S. People are different: tyrosine's modulating effect on cognitive control in healthy humans may depend on individual differences related to dopamine function. *Frontiers in Psychology* **5**(2014).
17. Jongkees, B.J., Hommel, B., Kuhn, S. & Colzato, L.S. Effect of tyrosine supplementation on clinical and healthy populations under stress or cognitive demands-A review. *Journal of Psychiatric Research* **70**, 50-57 (2015).
18. van der Lee, S.J. *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimers Dement* (2018).
19. Lawler, P.R. *et al.* Circulating N-Linked Glycoprotein Acetyls and Longitudinal Mortality Risk. *Circ Res* **118**, 1106-15 (2016).

20. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biology* **18** (2017).
21. Dick, K.J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990-8 (2014).
22. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, 208 (2016).
23. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nature Genetics* **49**, 131-138 (2017).
24. Mill, J. & Heijmans, B.T. From promises to practical strategies in epigenetic epidemiology. *Nature Reviews Genetics* **14**, 585-594 (2013).
25. Relton, C.L. & Smith, G.D. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *International Journal of Epidemiology* **41**, 161-176 (2012).
26. Zondervan, K.T. & Cardon, L.R. Designing candidate gene and genome-wide case-control association studies. *Nat Protoc* **2**, 2492-501 (2007).
27. Patnala, R., Clements, J. & Batra, J. Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC Genet* **14**, 39 (2013).
28. Mohler, H. Role of GABAA receptors in cognition. *Biochem Soc Trans* **37**, 1328-33 (2009).
29. Millan, M.J. *et al.* Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat Rev Drug Discov* **11**, 141-68 (2012).
30. Fritschy, J.M., Schweizer, C., Brunig, I. & Luscher, B. Pre- and post-synaptic mechanisms regulating the clustering of type A gamma-aminobutyric acid receptors (GABAA receptors). *Biochem Soc Trans* **31**, 889-92 (2003).
31. Muntoni, F., Torelli, S. & Ferlini, A. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurology* **2**, 731-740 (2003).
32. Yan, Q. *et al.* The impact of genotype calling errors on family-based studies. *Sci Rep* **6**, 28323 (2016).
33. Gudmundsson, J. *et al.* A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nature Genetics* **44**, 1326-1329 (2012).
34. Stacey, S.N. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nature Genetics* **43**, 1098-U85 (2011).
35. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-99 (2012).
36. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M. & Aulchenko, Y.S. The effect of genetic drift in a young genetically isolated population. *Annals of Human Genetics* **69**, 288-295 (2005).
37. Lidov, H.G., Byers, T.J., Watkins, S.C. & Kunkel, L.M. Localization of dystrophin to postsynaptic regions of central nervous system cortical neurons. *Nature* **348**, 725-8 (1990).
38. Record Owner, N.L.M. Dystrophins, utrophins, and associated scaffolding complexes: role in mammalian brain and implications for therapeutic strategies.
39. Anderson, J.L., Head, S.I., Rae, C. & Morley, J.W. Brain function in Duchenne muscular dystrophy. *Brain* **125**, 4-13 (2002).
40. Emery, A. & Muntoni, F. *Duchenne muscular dystrophy*, (Oxford University Press, New York, 2003).
41. Taylor, P.J. *et al.* Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy. *PLoS One* **5**, e8803 (2010).
42. Daoud, F. *et al.* Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression. *Human Molecular Genetics* **18**, 3779-3794 (2009).

43. Moizard, M.P. *et al.* Severe cognitive impairment in DMD: obvious clinical indication for Dp71 isoform point mutation screening. *Eur J Hum Genet* **8**, 552-6 (2000).
44. Moizard, M.P. *et al.* Are Dp71 and Dp140 brain dystrophin isoforms related to cognitive impairment in Duchenne muscular dystrophy? *Am J Med Genet* **80**, 32-41 (1998).
45. Chelly, J., Khelfaoui, M., Francis, F., Cherif, B. & Bienvenu, T. Genetics and pathophysiology of mental retardation. *Eur J Hum Genet* **14**, 701-13 (2006).
46. Polak, J.F. *et al.* Carotid-Wall Intima-Media Thickness and Cardiovascular Events. *New England Journal of Medicine* **365**, 213-221 (2011).
47. Lorenz, M.W., Markus, H.S., Bots, M.L., Rosvall, M. & Sitzer, M. Prediction of clinical cardiovascular events with carotid intima-media thickness - A systematic review and meta-analysis. *Circulation* **115**, 459-467 (2007).
48. Den Ruijter, H.M. *et al.* Common Carotid Intima-Media Thickness Measurements in Cardiovascular Risk Prediction A Meta-analysis. *Jama-Journal of the American Medical Association* **308**, 796-803 (2012).
49. Leszczyniecka, M. *et al.* Identification and cloning of human polynucleotide phosphorylase, hPNPase old-35, in the context of terminal differentiation and cellular senescence. *Proc Natl Acad Sci U S A* **99**, 16636-41 (2002).
50. Leszczyniecka, M., Su, Z.Z., Kang, D.C., Sarkar, D. & Fisher, P.B. Expression regulation and genomic organization of human polynucleotide phosphorylase, hPNPase(old-35), a Type I interferon inducible early response gene. *Gene* **316**, 143-56 (2003).
51. Goossens, P. *et al.* Myeloid Type I Interferon Signaling Promotes Atherosclerosis by Stimulating Macrophage Recruitment to Lesions. *Cell Metabolism* **12**, 142-153 (2010).
52. Peloso, G.M. *et al.* Phenotypic extremes in rare variant study designs. *Eur J Hum Genet* **24**, 924-30 (2016).
53. Bos, D. *et al.* Intracranial Carotid Artery Atherosclerosis and the Risk of Stroke in Whites The Rotterdam Study. *Jama Neurology* **71**, 405-411 (2014).
54. Bos, D. *et al.* Atherosclerotic calcification is related to a higher risk of dementia and cognitive decline. *Alzheimers & Dementia* **11**, 639-647 (2015).
55. Bos, D. *et al.* Genetic Loci for Coronary Calcification and Serum Lipids Relate to Aortic and Carotid Calcification. *Circulation-Cardiovascular Genetics* **6**, 47-U72 (2013).
56. Bos, D. *et al.* Comparison of Atherosclerotic Calcification in Major Vessel Beds on the Risk of All-Cause and Cause-Specific Mortality The Rotterdam Study. *Circulation-Cardiovascular Imaging* **8** (2015).
57. Liu, J. *et al.* A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. *Diabetes* **66**, 2915-2926 (2017).
58. Holmes, M.V. *et al.* Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol* **71**, 620-632 (2018).
59. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262-266 (2015).
60. Donaldson, G.P., Lee, S.M. & Mazmanian, S.K. Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* **14**, 20-32 (2016).
61. Sartor, R.B. Gut microbiota: Optimal sampling of the intestinal microbiota for research. *Nat Rev Gastroenterol Hepatol* **12**, 253-4 (2015).
62. Mai, V. & Morris, J.G., Jr. Need for prospective cohort studies to establish human gut microbiome contributions to disease risk. *J Natl Cancer Inst* **105**, 1850-1 (2013).

63. Bonder, M.J. *et al.* The effect of host genetics on the gut microbiome. *Nat Genet* **48**, 1407-1412 (2016).
64. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nature Genetics* **46**, 881-885 (2014).
65. Smoller, J.W. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-1379 (2013).
66. Anney, R.J.L. *et al.* Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism* **8** (2017).
67. Anney, R. *et al.* Individual common variants exert weak effects on the risk for autism spectrum disorderspi. *Human Molecular Genetics* **21**, 4781-4792 (2012).
68. Geschwind, D.H. Advances in Autism. *Annual Review of Medicine* **60**, 367-380 (2009).
69. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528-533 (2009).
70. Ma, D.Q. *et al.* A Genome-wide Association Study of Autism Reveals a Common Novel Risk Locus at 5p14.1. *Annals of Human Genetics* **73**, 263-273 (2009).
71. Brikell, I., Kuja-Halkola, R. & Larsson, H. Heritability of attention-deficit hyperactivity disorder in adults. *Am J Med Genet B Neuropsychiatr Genet* **168**, 406-413 (2015).
72. Lee, S.H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* **45**, 984 (2013).
73. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *bioRxiv* (2017).
74. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv* (2017).
75. Middeldorp, C.M. *et al.* A Genome-Wide Association Meta-Analysis of Attention-Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. *J Am Acad Child Adolesc Psychiatry* **55**, 896-905 e6 (2016).
76. Yizhar, O. *et al.* Tomosyn inhibits priming of large dense-core vesicles in a calcium-dependent manner. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2578-2583 (2004).
77. Sakisaka, T. *et al.* Dual inhibition of SNARE complex formation by tomosyn ensures controlled neurotransmitter release. *Journal of Cell Biology* **183**, 323-337 (2008).
78. Chen, R. & Snyder, M. Promise of personalized omics to precision medicine. *Wiley Interdisciplinary Reviews-Systems Biology and Medicine* **5**, 73-82 (2013).
79. Collins, F.S. & Varmus, H. A new initiative on precision medicine. *N Engl J Med* **372**, 793-5 (2015).
80. Fall, T., Mendelson, M. & Speliotes, E.K. Recent Advances in Human Genetics and Epigenetics of Adiposity: Pathway to Precision Medicine? *Gastroenterology* **152**, 1695-1706 (2017).
81. Sohail, M. *et al.* Negative selection in humans and fruit flies involves synergistic epistasis. *Science* **356**, 539-542 (2017).
82. Visscher, P.M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics* **101**, 5-22 (2017).
83. Nielsen, J.B. *et al.* Genome-wide association study of 1 million people identifies 111 loci for atrial fibrillation. *bioRxiv* (2018).
84. Jansen, P.R. *et al.* Genome-wide Analysis of Insomnia (N=1,331,010) Identifies Novel Loci and Functional Pathways. *bioRxiv* (2018).

85. Elliott, P., Peakman, T.C. & Biobank, U.K. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* **37**, 234-44 (2008).
86. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology* **44**, 1137-1147 (2015).
87. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).
88. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* **40**, 1652-66 (2011).
89. De Souza, Y.G. & Greenspan, J.S. Biobanking past, present and future: responsibilities and benefits. *AIDS* **27**, 303-12 (2013).
90. Carlson, C.S. *et al.* Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol* **11**, e1001661 (2013).
91. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *Lancet Neurol* **17**, 434-444 (2018).
92. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-97 (2007).
93. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: Using Sequence and Genotype Data to Estimate Haplotypes and Unobserved Genotypes. *Genetic Epidemiology* **34**, 816-834 (2010).
94. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* **39**, 906-913 (2007).
95. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
96. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
97. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
98. Wijmenga, C. & Zhernakova, A. The importance of cohort studies in the post-GWAS era. *Nat Genet* **50**, 322-328 (2018).
99. Yi, Y., Park, J., Lim, J., Lee, C.J. & Lee, S.H. Central Nervous System and its Disease Models on a Chip. *Trends Biotechnol* **33**, 762-776 (2015).
100. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281-2308 (2013).
101. Wu, Z., Hu, Y. & Melton, P.E. Longitudinal data analysis for genetic studies in the whole-genome sequencing era. *Genet Epidemiol* **38 Suppl 1**, S74-80 (2014).
102. Rasmussen-Torvik, L.J. *et al.* Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose. *Genet Epidemiol* **34**, 665-73 (2010).
103. Lau, E. & Wu, J.C. Omics, Big Data, and Precision Medicine in Cardiovascular Sciences. *Circ Res* **122**, 1165-1168 (2018).