Title

Functional genomics, genetic risk profiling and cell phenotypes in neurodegenerative disease

Permalink

https://escholarship.org/uc/item/1s41s58s

Author

Finkbeiner, Steven

Publication Date

2020-12-01

DOI

10.1016/j.nbd.2020.105088

Peer reviewed

Functional genomics, genetic risk profiling and cell phenotypes in neurodegenerative disease

Steven Finkbeiner[a,b,c,*]

[a]Center for Systems and Therapeutics, USA

[b]Taube/Koret Center for Neurodegenerative Disease Research, Gladstone Institutes, San Francisco, CA 94158, USA

[c]Departments of Neurology and Physiology, University of Califorina, San Francisco, CA 94158, USA

Abstract

Human genetics provides unbiased insights into the causes of human disease, which can be used to create a foundation for effective ways to more accurately diagnose patients, stratify patients for more successful clinical trials, discover and develop new therapies, and ultimately help patients choose the safest and most promising therapeutic option based on their risk profile. But the process for translating basic observations from human genetics studies into pathogenic disease mechanisms and treatments is laborious and complex, and this challenge has particularly slowed the development of interventions for neurodegenerative disease. In this review, we discuss the many steps in the process, the important considerations at each stage, and some of the latest tools and technologies that are available to help investigators translate insights from human genetics into diagnostic and therapeutic strategies that will lead to the sort of advances in clinical care that make a difference for patients.

1. Introduction

A major goal of biomedical research is to understand the causes of and find treatments for human disease. This has been tremendously challenging for neurodegenerative diseases, with failure rates of late-stage clinical trials reaching well over 90% (Arrowsmith and Miller, 2013). Combined with the fact that these trials are extremely expensive, the rational decision many pharmaceutical companies have reached has been to avoid investing in developing neurotherapeutics (Ringel et al., 2013).

To find clues about how to make the process more successful, several groups have performed retrospective analyses to discover the features of therapeutics programs that would have better predicted success or failure. One predictor of success is if the target has some genetic link to the disease for which a treatment is being developed (King et al., 2019; Plenge et al., 2013; Prinz et al., 2011; Wang et al., 2012). Estimates from the different studies vary, but

[*]Corresponding author: steve.finkbeiner@gladstone.ucsf.edu.

there may be up to a several-fold enrichment of approved drugs whose targets are genetically linked to disease compared to targets without such links. As new therapeutic approaches and drugs are being developed to treat neurodegenerative diseases that target genes rather than proteins, the importance of understanding the genetic risk factors and modifiers of disease is likely to grow.

More broadly, understanding the genetic underpinnings of neurodegenerative disease can offer unbiased insights into the underlying mechanisms of disease. These insights can be hugely important for improving diagnosis, prognosis and patient stratification in clinical trials, even if some of the specific mechanisms prove to be poor therapeutic targets themselves. For example, a major question for some of the major neurodegenerative diseases —Alzheimer disease (AD), Parkinson disease (PD), amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)—is the extent to which each of these clinical syndromes is a single entity that could reasonably be treated with a single modality, or whether they are a collection of diseases that demand approaches tailored to each underlying cause.

The answer to this question has dire implications for clinical trial design. Because clinical trials in neurodegenerative disease already are very expensive, few if any studies can afford to enroll enough patients to provide adequate power to uncover effects in subgroups of patients. So, it is conceivable that drugs that "fail" in efficacy trials actually do work in a subgroup of patients, but this effect is never detected. This may lead to abandonment of an effective drug and the conclusion that the target is wrong, ending further research. Depending on their design, genetic studies have the potential to help stratify patient populations, based on their symptom profile, progression rates, and molecular pathogenesis. In the future, when more than one treatment for a neuro degenerative disease becomes available, it may become possible to use genetic information from a specific patient to tailor their treatment program to minimize side effects and maximize safety and efficacy.

2. Genetics of neurodegenerative disease

For these reasons, enormous efforts are being made to elucidate the genetic underpinnings of neurodegenerative disease. The reader is referred to excellent reviews of these efforts and the specific findings that have emerged in this issue as well as in many recent publications (Uricchio, 2020; Reynolds et al., 2019; Nalls et al., 2019; Kunkle et al., 2019; Visscher et al., 2017; Morgan et al., 2017).

This review is primarily focused on using insights from genetics studies as a starting point to better understand causes of neurodegenerative diseases and eventually to find treatments (Fig. 1). There are a few general observations that can be made from genetic studies of neurodegenerative diseases that are relevant. First, the incidence and heritability of neurodegenerative disease varies substantially (Nalls et al., 2019; Kunkle et al., 2019; Manolio et al., 2009; Keller et al., 2012; Wingo et al., 2011; Huang et al., 2017; Simpson and Al-Chalabi, 2006; Al-Chalabi and Hardiman, 2013). AD and PD are common, whereas ALS, FTD and Huntington's disease (HD) are rare. All patients with HD have a mutation in a single gene, whereas the fraction of patients with AD, PD, ALS and FTD who have single gene mutations is much lower (~10% for AD, PD and ALS, and ~30% for FTD).

Indeed, the relatively small fraction of patients with familial neurodegenerative disease has played a critical role driving much of our current understanding of the genetics and mechanisms responsible for these diseases. Studies to discover autosomal dominant or recessive disease-causing mutations utilizing linkage with association and/or segregation analysis has provided unequivocal proof for the role of specific genes and gene mutations. For the most part, it has been the discovery of rare Mendelian causes of neurodegenerative disease that has enabled the development of genetic laboratory models of those diseases, which have led to most of the mechanistic insights into disease pathogenesis today (Pihlstrøm et al., 2017; Stamatakou et al., 2020).

Still, the precise relationship between rare Mendelian and common non-Mendelian forms of neurodegenerative disease is unknown. It is concerning that so many potential therapies that showed efficacy in laboratory models based on Mendelian forms of disease have failed in clinical trials in human populations with mostly sporadic forms of the disease. Fortunately, even though it is impossible to pinpoint a single responsible gene for sporadic forms of neurodegenerative disease, the heritable risk of these diseases is high (e.g., ~30% for PD, 40–60% for ALS). Some of this so-called "missing heritability" (Uricchio, 2020; Manolio et al., 2009; Keller et al., 2012) has been uncovered through genome-wide association studies (GWAS), which have found numerous genetic loci and common single nucleotide variants (SNV) associated with increased or decreased risk of disease. However, with a few exceptions, the magnitude of effect of any one SNV is very small, and the majority of the "missing heritability" for all the diseases remains unexplained.

There are some important limitations of existing genetics studies. The first is that they are not exhaustive. For the most part, existing studies have been performed on genomes from people of European or North American ancestry (Rosenberg et al., 2010; Genetics for all, 2019; Sirugo et al., 2019). Since the variants present in a human genome vary across different ethnic populations, studies of genomes from patients with other ancestries are likely to uncover new risk variants and genes relevant to neurodegenerative disease. This may lead to the implication of new genes and pathways, and help identify the causal SNVs within known GWAS loci (Edwards et al., 2013). Yet, even if all the genetic variants that affect a disease are eventually discovered, our understanding of the genetic regulation of disease mechanisms is still limited by the relevant variants that exist in the human genome.

The second is that most of the studies have been conventional GWAS focused on discovering associations between disease risk and genomic loci. By design, these studies are unlikely to discover rare variants that may have much larger effects and, for that reason, could be especially attractive therapeutic targets. Moreover, because of linkage disequilibrium (LD) and the sparseness of the genomic probes, it is not always straightforward to understand which SNVs mediate the effects of the GWAS loci. Increasingly, whole genome sequence (WGS) data are being collected, which substantially increases the resolution of the analysis. But the large increase in variants identified means that there are associated challenges with having enough power to identify those that are significant.

A third limitation of existing genetic studies is that the majority have focused on the genetic risk of developing disease, whereas other important features of pathogenesis, such

as progression, have been less studied. That is because segregating people with a diagnosis of a particular neurodegenerative disease from a healthy control population is relatively easy and inexpensive. By contrast, the clinical presentation and progression of neurodegenerative disease can be quite heterogeneous, even amongst patients who have been diagnosed with the same clinical syndrome. For the most part, it is less understood to what extent clinical presentation and progression rate can be explained by genetics, and to what extent they mirror the genetics that govern the risk of developing disease. Understanding the genetics of disease progression might be of particular importance for therapeutics development because targets linked to progression might be more relevant during the symptomatic phase of disease when clinical trials would normally be performed.

Even if the heritable portion of neurodegenerative diseases is eventually understood, the majority of cases will remain unexplained. Often, these have been attributed to "environmental" factors. And although "environmental" factors are frequently conceptualized as non-genetic influences, that may not be entirely true. One of the most replicated findings from behavioral genetics is that most measures of "environment" show significant genetic influence, and most associations between environmental measures and psychological traits are genetically mediated (Plomin et al., 2016). That is a critical consideration for genetic risk profiling for neurodegenerative disease because behavioral measures and psychological traits often feature prominently in the disease phenotypes against which GWAS are performed, and the risk of disease has been associated with "environmental" factors, such as educational attainment (Sharp and Gatz, 2011), which have substantial heritable and non-heritable genetic underpinnings (Plomin et al., 2016). That said, with a few exceptions such as pesticide exposure and head trauma, in epidemiology studies it has proven difficult to identify robust and reproducible non-genetic factors that account for significant portions of non-heritable disease risk. It is likely that some environmental factors only confer disease risk in the context of certain genomic variants, making them particularly challenging to discover. Still, a complete understanding of the genetic risk underlying neurodegenerative disease could be a foundational starting point for exploring environmental factors that synergize with genetics. Lastly, it must be acknowledged that some non-heritable disease risk may be essentially random, and will remain undiscovered even after exhaustive genetic and environmental studies are performed. Perhaps a better understanding of generic risk factors for neurodegenerative disease, such as aging, could lead to new strategies that improve human resilience and reduce the contribution of stochastic factors to neurodegenerative disease.

3. Moving from GWAS to causal variants

As noted above, GWAS results help establish the scope and scale of genetic contributions to neurodegenerative disease. And although so far there are few examples of taking specific GWAS loci from patients with neurodegenerative disease to validated mechanisms, the wealth of available genetic data has poised the field to do the difficult work to perform functional genomics and translate these discoveries into potential biomarkers and therapeutic targets (Ramanan and Saykin, 2013). With this in mind, it is important to note that GWAS methodology suffers two important limitations that typically prevent it from uncovering the causal variants, genes and pathways that are responsible for the GWA signals. First, the

majority of GWAS loci are in non-coding or intergenic regions, leaving unclear the precise gene(s) and mechanisms through which they act. Second, GWAS loci often encompass multiple variants in linkage disequilibrium, leaving unclear which variant or combination of variants is responsible.

Several strategies have been developed to address these limitations. The SNVs tested in GWAS typically act as a tag for a haplotype block that can span thousands of DNA bases (> 100 kB) and contain > 1,000 variants, hundreds of which might exhibit statistically significant associations with the phenotype of interest (Edwards et al., 2013). As such, a logical first step is to look for genes in the neighborhood of the SNV and contained within the haplotype block that might mediate its effect. A variety of computational approaches have been developed to help investigators look for co-localization of candidate genes with GWAS loci (Wu et al., 2019; Giambartolomei et al., 2014; Simovski et al., 2018; Hormozdiari et al., 2016).

However, because of the size of haplotype blocks and the number of variants they contain, it can be difficult based on GWAS data alone to link an SNV to the gene(s) that mediate its effect. In that case, a common next step can be fine mapping. One way to fine map is to use genotyping arrays with greater probe density for SNVs within a haplotype. With a sufficient sample size, often between 10,000 and 100,000, there can be enough examples in which individual SNVs within a haplotype are inherited independently that the risk conferred by the haplotype can be pinpointed to a specific SNV (Udler et al., 2010). However, it is common after fine mapping to still have 10–50 SNVs that are strong candidates, highly correlated with the best hit (Edwards et al., 2013). So, additional methods are often needed to elucidate causal SNVs and prioritize genes and pathways.

Another strategy has been to use bioinformatic tools to integrate GWAS and additional publicly-available data, based on the concept that causal variants from GWAS presumably confer risk by affecting gene expression, which perturbs the function of biological pathways. One web-based platform to do this is called FUMA, which broadly breaks the task into two parts: 1) mapping the SNV to gene(s), and then 2) relating the implicated gene to its normal function and pathogenic mechanism (Watanabe et al., 2017). With FUMA, genomic loci are first characterized to identify independent SNVs and candidate SNVs in LD to define lead SNVs and genomic risk loci. FUMA then predicts the functional impact of SNVs and annotates them by integrating information related to each SNV position and any known expression quantitative trait loci (eQTL) or chromatin interaction. Functionally annotated SNVs are then mapped to genes with positional, eQTL and chromatin interaction mapping data (Hormozdiari et al., 2016). Whereas eQTL data tend to have distance boundaries, with 90% of SNPs affecting eQTLs within 15 kB of the 5′ and 3′ gene boundaries (Pickrell et al., 2010), chromatin interaction mapping data can better account for long-range interactions between the disease-associated regions and nearby or distant genes. The goal of these efforts is to predict effects of an SNV, ranging from its local effects on gene function up to cell- and tissue-specific effects. In general, these various approaches each have scoring mechanisms, and the user can adjust the stringency and weighting of each score to reflect different models. Once a list of genes has been prioritized based on functional SNV annotation, FUMA annotates the genes by drawing on information about the gene, its implication

in human disease, its tissue-specific expression, and its role in biological pathways and functions.

In genetic risk profiling, one must consider "missing heritability," and the concept of polygenic disease, and the implications it could have for finding causal variants. The disease-associated variants discovered by GWAS are common in the general population, the total number of risk variants can be large, and most can be inherited independently of each other. As such, there will be some distribution of disease-associated variants within human populations, leading to some individuals having a greater and some a lesser number of risk variants for a disease. The prevailing view is that the higher the burden of risk variants an individual has, the higher their risk of developing disease.

One way to quantify the overall risk for an individual is to simply add the risk of each SNV, as determined by GWAS, in their genome. Regression models can be tested and SNV weighting can be adjusted to maximize the predictive value of these so-called polygenic risk scores. For some disorders, this approach has led to models with high predictive value (Escott-Price et al., 2015), and in some cases, the predictive power can be improved if additional variants from GWAS data are included in the model even if they did not achieve genome-wide significance on their own. The argument for including these variants in the analysis is that the statistical threshold for declaring certain variants as having achieved genome-wide significance is set high because of multiple testing and type I errors. However, this approach is controversial. Critics argue that at the lower limit of effect size, the entire genome may make contributions to complex phenotypes by generic mechanisms that may not be biologically meaningful and are not amenable to therapeutic intervention (Boyle et al., 2017).

The development of polygenic risk scores with regression methods is primarily aimed at maximizing the predictive accuracy of phenotype from genotype by deriving the optimal weights to assign various SNVs. However, the method *per se* does not address whether biological cause-and-effect relationships might exist amongst the variants. Conventional experimental approaches to address such questions in model systems are described below, but because of important limitations in the predictive value of laboratory models of disease, investigators have sought to develop bioinformatic approaches to investigate the presence of causal relationships amongst GWAS hits relying solely on human genetic data. A major approach for doing this is causal modeling (Ewald et al., 2020; Mattis and Gloyn, 2020), in which variation in gene co-expression is analyzed across patients and samples to infer potentially causal relationships between genomic variants and expression of specific genes, and amongst genes within networks (Uricchio, 2020; Barbeira et al., 2018).

One approach to more firmly root the investigation of polygenic effects in biology is pathway analysis (White et al., 2019; Wang et al., 2010). This is distinct from methods to map an SNV to a gene, such as FUMA, which rely on pathway analysis to place individual SNVs into a biological context. Rather, the focus here is on analyzing collections of SNVs across the genome to find pathways that are relatively over represented by the hits. In fact, pathways represented or largely driven by a single SNV should probably be disregarded for this type of analysis. Likewise, pathway analysis is distinct from polygenic

risk score calculation in that an effort is made to incorporate *a priori* information about the genes through which the SNVs are presumably acting to discover meaningful biological connections amongst the SNVs. For example, an enrichment in disease-associated SNVs in genes already known to encode proteins in a defined biological pathway could suggest that disease risk is being conferred through perturbation of that pathway, offering a potential pathogenic mechanism and providing greater confidence that the identified SNVs play a causal role. By the same token, common variants in the same individual that happen to act through the same biological pathway might be more impactful on disease risk than the same number of variants acting in different pathways because of the potential for non-linear synergistic interaction between SNVs on biological function. It is conceivable that some biological functions are so highly conserved that only SNVs with very small effects are tolerated. If so, they may not be detected on their own as risk variants in GWAS, but might still confer substantial risk in synergistic combination with other SNVs in the same pathway. In this case, the additive model underlying most polygenic risk scores may fail to accurately predict the overall risk of certain patterns of variants.

Looking for interactions between SNVs has been challenging because the number of possible combinations is enormous and increases exponentially as the number of simultaneously acting SNVs increases. More recently, machine learning (ML) approaches have been applied to GWAS data to investigate polygenic models of disease, and have shown promise (Behravan et al., 2018). One advantage of ML approaches is that they are well suited to looking for patterns in data, including non-linear relationships. In addition, it is possible to estimate the robustness and reproducibility of an algorithm by repeatedly training on different subsets of the data and validating the algorithm on subsets that have been withheld. This avoids limitations of conventional statistical tests, which would be underpowered, even for investigating most pairwise SNV interactions.

4. Experimental evaluation of prioritized genes and causal variants

If the aforementioned approaches are doing their job, they should enable an investigator to take GWAS hits and begin to develop a list of presumed causal variants and the genes, combinations of genes, and biological pathways through which those variants act to confer risk of whatever phenotype was used to do the GWAS. Still, all these approaches are imperfect. Knowledge about normal gene functions, 3D interactions within the genome, cell type-specific expression, and functional pathways is still very much being developed. Moreover, some of these relationships presumably change in the context of disease, and it is unclear if accurate predictions in a disease context can be made based on data from a healthy context.

For all these reasons, it is critical to develop a prioritized list of causal variants that are further evaluated experimentally. Causal variants presumably confer disease risk by impacting the transcriptome and proteome, either by changing the function of the gene in which the SNV is found or, more commonly, by affecting the expression of one or more genes (eQTLs) (Nica and Dermitzakis, 2013). To elucidate this, several core techniques such as bulk and single cell RNA sequencing (scRNAseq), assays for transposase-accessible chromatin using sequencing (ATAC seq), and proteomics are frequently applied (Edwards

et al., 2013). Bulk transcriptomics is very sensitive, making it possible to develop a comprehensive list of differentially expressed genes in the presence or absence of a causal variant or a perturbation of the gene through which it presumably acts. However, if the causal variant only affects a gene that is expressed in a cell type-specific fashion, it may be difficult to detect the signal by bulk analysis of a heterogenous population. Indeed, it is estimated that 50–90% of eQTLs are tissue dependent, implying the presence and importance of cell type-specific expression (Nica et al., 2011; Dimas et al., 2009). Also, the target genes of eQTLs can be coding or non-coding RNAs, and genotype can also influence splicing and alternate isoform production, so detection methods need to be chosen with this in mind (Kumar et al., 2013; Lalonde et al., 2011). Weighted gene expression coregulatory network analysis (WGCNA) is one method that has been used successfully to tease out transcriptional networks from bulk transcriptomic data, which often correspond to cell type-specific gene expression profiles (Pei et al., 2017; Langfelder and Horvath, 2008). ScRNAseq (or single nuclei RNAseq) can be extremely useful, especially if the relevant cell type is a minor subset of the cells in the general population. However, the read depth of scRNAseq is significantly less than for bulk RNA seq, which limits the analysis to genes that are relatively highly expressed.

Elucidating the relationship between eQTLs and the genes they regulate is an important goal for understanding how SNVs affect disease risk. If an SNV affects promoter, enhancer or silencer elements and transcription factor accessibility, ATAC seq may reveal corresponding changes in chromatin accessibility (Lowe et al., 2019; Oulhen et al., 2019; Shashikant and Ettensohn, 2019). On the other hand, enhancers and silencers can reside more than 1 kb from their target genes and exert their regulatory influence through long-range interactions mediated by the formation of chromatin loops (Sexton et al., 2009). That is one reason that the gene closest to an SNV may not in fact be the principle target. Fortunately, a number of approaches including established proximity-ligation based methods, such as Hi-C, and newer ligation-free techniques, can elucidate 3D genome topology (Kempfer and Pombo, 2019; Kaul et al., 2020). These can lead to a better understanding of the physical relationship between a domain containing an SNV and other domains in the genome it may be influencing.

Although the data from RNA seq, ATAC seq, proteomics and other analytical measures from the same sample should clearly be related, it is important to keep in mind that these are independent snap shots of different levels of biology. Using statistical methods to analyze data from a single method can be prone to uncovering spurious correlations that are not causal. For example, > 90% of randomly selected gene signatures are correlated with cancer outcomes (Venet et al., 2011). These random signatures occur because expression of a large fraction of the genome is highly correlated with a marker of proliferation. However, simply looking for overlap in the subset of differentially expressed genes identified by different methods is not an effective strategy either, since multi-Omic studies often find that the overlap between each pair of assays can be less than expected by chance (Yeger-Lotem et al., 2009). That likely happens because some macromolecules may be principally regulated by changes in their levels (e.g., transcripts), whereas others may be regulated primarily by other mechanisms (e.g., post-translational modifications) and because of the presence of complex feedback loops. One strategy to overcome some of these limitations and tease out

true biological signals and signatures is to integrate data from multiple approaches, such as through the OMICs integrator, which uses Prize-collecting Steiner forest algorithms to identify networks that span the available transcriptomic, epigenomic, proteomic and other data (Tuncbag et al., 2016). In addition to avoiding the pitfalls of looking simply at statistical correlations within a data type or overlap between data types, it has the added advantage of being able to uncover novel networks and pathways, which may be especially important in the context of disease, where pathways may be deranged and existing knowledge about them based on studies of normal cells may not be fully applicable.

5.   Systems cell biology, functional genomics, and genetic risk profiling

The approaches described above can, in principle, be applied to human tissue. The tremendous advantage of doing so is the obvious relevance of using human tissue to study neurodegenerative diseases, which primarily if not exclusively affect humans, as well as the opportunities to integrate a variety of measurements from genomics to clinical phenotype in the same individual. But there are also some practical challenges with this strategy in relation to neurodegenerative disease. The main obstacle is that gaining access to brain tissue is not trivial; it usually occurs post-mortem, and the availability of high quality post-mortem tissue is limited. The interval between death and tissue harvest (so-called post-mortem interval, PMI) is often used to estimate tissue quality. Yet, the circumstances immediately surrounding the death of the individual may be more important than PMI. Other measures, such as the RNA integrity number (RIN), may be better markers of tissue quality (Stan et al., 2006).

The other nagging concern about using post-mortem human brain to study neurodegenerative disease is that most patients who come to autopsy are rather end stage. This can mean that most of the cells with the greatest susceptibility for disease have already been lost, taking with them the signals one might hope to detect with multi-Omics approaches. Often in their place, a reactive gliosis can develop whose relevance to disease-specific pathogenic mechanisms can be unclear. Lastly, studies on human tissue, for the most part, are limited to observational measurements. Observational studies are critical for establishing foundational knowledge and generating hypotheses, but it is important to validate or invalidate the models that emerge with perturbation experiments to further establish which apparent correlations are due to causal relationships.

For all these reasons, there is avid interest in developing model systems in which to interrogate the effect of SNVs on genes and pathways and their relevance to disease mechanisms. Historically, these efforts have focused on non-human models, especially mouse models of disease (Pankevich et al., 2014). In general with non-human model systems, there can be a tradeoff between ease-of-use, cost and perceived relevance to human biology, and the "best" model may depend on what question the model is being used to address (Xi et al., 2011; Roberson, 2012; Shulman et al., 2014; Lee et al., 2012; Fernagut and Tison, 2012). That said, critics of the use of non-human model systems point out that their predictive value for human clinical trials has a poor track record (Pankevich et al., 2014; Mullane et al., 2014; Mullane and Williams, 2013; McGonigle and Ruggeri, 2014; Crook and Housman, 2011; Mullane and Williams, 2019; Ransohoff, 2018). In turn, some

have rebutted this criticism by blaming the failures on poor scientific practices, highlighting the fact that many studies done in academic labs use small cohorts of mice and are underpowered, fail to implement critical quality control measures to avoid genetic drift, and fail to hew to the highest standards of animal husbandry (Perrin, 2014).

On the other hand, there has been no shortage of studies showing substantial differences between the physiology of humans and non-human model organisms, including in the sequence of homologous genes that affect drug responses (Mattson et al., 1991; Mertens et al., 2013), transcriptional regulation and profiles (Lin et al., 2014; Church et al., 2009), cell types (Hodge et al., 2019), and tissue composition and responses (e.g., inflammation (Seok et al., 2013)). Each species is a highly evolved, complex system, and it may be unreasonable to expect that simply placing a disease-associated human gene or mutation into a non-human organism will faithfully and sufficiently mimic the patho-physiology of human disease. Rather than judging the quality of non-human disease models based on their ability to recapitulate human disease pathology, a better approach may be to acknowledge that evolution may make that unrealistic, and instead search for and utilize species-appropriate orthologous phenotypes (McGary et al., 2010; Perlman, 2016). Within those constraints, non-human systems could be powerful models with great predictive power for functional genomics. Indeed, numerous examples exist in which yeast, fly, and mouse models have been used successfully to discover cross-species convergent pathogenic mechanisms of disease including roles for specific genetic modifiers, RNA metabolism, and stress granule formation (Martinez et al., 2016; Campioni and Finkbeiner, 2015; Barmada et al., 2015; Kim et al., 2014; Armakola et al., 2012; Aron et al., 2018).

The discovery of a series of technologies to generate human brain cells from peripheral cells from patients has opened up a number of exciting avenues for creating fully human preclinical models to investigate genetic risk profiles and functional genomics. One of the most important has been the development of induced pluripotent stem cell (iPSC) technology (Okita et al., 2011). iPSC lines can be differentiated into a variety of cell types found in the brain or periphery which maintain the genetics of the person from whom they were derived (Abud et al., 2017). Moreover, the advent of genome editing provides a way to introduce or revert specific genetic variants in iPSC lines with defined genomes. This approach can be used to isolate the effects of specific genetic variants as well as investigate the interaction between defined variants.

As such, iPSCs offer an opportunity to study the effects of one or more variant in the context of a defined genome on the transcriptome, epigenome, proteome, and phenome in stem cells or different differentiated cell types (Cobb et al., 2018; Kaus and Sareen, 2015; Haston and Finkbeiner, 2016; Kaye and Finkbeiner, 2013). For example, this has been an effective approach to demonstrate that certain genomic variants that confer risk of PD act as eQTLs for α-synuclein (Soldner et al., 2016). iPSC lines from PD patients with mutations in leucine-rich repeat kinase 2 (*LRRK2*) exhibit survival deficits (Skibinski et al., 2014), increased sensitivity to oxidative stress (Nguyen et al., 2011), and mitochondrial DNA damage that can be reversed with gene correction of the LRRK2 mutation (Sanders et al., 2014). Others have used patient-derived neurons to demonstrate that Miro1, a protein on the surface of mitochondria important for mitochondrial mobility, is abnormal

in a broad spectrum of PD patients (Hsieh et al., 2019). The fact that iPSC models have uncovered convergent mechanisms of disease, including roles for aging, epigenetics, autophagy/mitophagy, excitability, and mitochondrial dysfunction, which also appear in mice and humans, is encouraging (Aron et al., 2018; Haston and Finkbeiner, 2016; Barmada et al., 2014; Ocampo et al., 2016; Fang et al., 2019a; Osellame et al., 2013; Ekstrand and Galter, 2009; Martín-Maestro et al., 2019; Penney et al., 2020; Schaeffer et al., 2012; Wainger and Cudkowicz, 2015; Devlin et al., 2015; Selvaraj et al., 2018; Zanette et al., 2002; Kim et al., 2017).

Furthermore, in neurons differentiated from HD patients who carry a disease-associated triplet repeat expansion, transcriptomic analysis revealed deficits in gene regulation by the transcription factor NeuroD1 that were associated with impaired neurite outgrowth and survival (HD iPSC Consortium, 2017; HD iPSC Consortium, 2012). These abnormalities were reversed in iPSC-derived neurons and a mouse model of HD with a pharmacological tool, ISX-9, known to boost NeuroD1 activation. Neurons differentiated from HD iPSCs also exhibited bioenergetic deficits that could be reversed with glycolytic metabolites (Kedaigle et al., 2019), and changes in proteostasis that could be mitigated by UBR5, a ubiquitin ligase (Koyuncu et al., 2018).

This approach can also be used to illuminate the differing molecular mechanisms that drive heterogenous diseases. For example, iPSC lines from an ALS patient with a disease-causing hexanucleotide mutation in *C9ORF72* were used to investigate the role of methylation at the *C9ORF72* locus (Esanov et al., 2016). Motor neurons differentiated from an ALS patient with a mutation in Tar DNA Binding Protein 43 (TDP-43) showed abnormal aggregation of TDP43 and survival deficits (Bilican et al., 2012). Motor neurons from ALS patient iPSC lines with a disease-causing mutation in an RNA binding protein RNABP1 showed abnormalities in stress granule (SG) formation and survival that could be reversed pharmacologically (Fang et al., 2019b). Similarly, iPSC lines with an ALS-associated mutation in Fused in Sarcoma (FUS) also showed abnormal SG formation, which could be reversed by inducing the protein clearance pathway autophagy (Marrone et al., 2018).

iPSC models have also been used to study different dementias. The genetic variant encoding the apolipoprotein E4 (APOE4) isoform significantly elevates an individual's risk for developing AD. Interestingly, human γ-amino butyric acid (GABA)-ergic neurons harboring APOE4 exhibit elevated tau phosphorylation and increased degeneration (Wang et al., 2018). These phenotypes could be reversed with a compound that changes the conformation of APOE4 to mimic the structure of APOE3, which does not increase the risk of AD. Other AD iPSC models manifest intracellular stress associated with intracellular Aβ peptides (Kondo et al., 2013). Several groups have engineered iPSC lines to model tauopathies (Iovino et al., 2015), neurofibrillary tangle formation and transcellular propagation (Reilly et al., 2017). And forebrain neurons made from iPSC lines derived from patients with tau (*MAPT*) mutations and frontotemporal lobar degeneration (FTLD / FTD) show abnormalities in subcellular vesicle trafficking, stress pathways (Wren et al., 2015) and Ca^{2+} signaling that could be reversed with genome engineering and genetic correction of the *MAPT* mutation (Imamura et al., 2016).

In the preceding examples, the iPSC lines were derived from patients with defined Mendelian genetic mutations, or engineered based on known, highly penetrant mutations. However, as noted above, they leave open the question of whether the more common forms of disease are caused by dysfunction of the same pathways, and highlight the potential value of establishing cellular models of idiopathic disease. There have been some reports of the use of iPSC lines from patients without a known mutation or family history of neurodegenerative disorders, who presumably have idiopathic forms of disease. This would be a particularly exciting development because these patients make up the majority of cases of AD, PD, FTD and ALS, and without faithful models of idiopathic disease, it is difficult to elucidate the underlying causes experimentally. In an early study of forebrain-like neurons differentiated from ALS patient iPSCs, a subset showed abnormalities in immunolabeling for TDP-43 (Burkhardt et al., 2013). That is potentially significant because accumulation of TDP-43 is a pathological hallmark found in brain and spinal cord tissue of ~98% of ALS patients. In another study, motor neurons differentiated from iPSC lines derived from patients with mostly sporadic / idiopathic ALS were monitored with time lapse imaging and then immunolabeled, and were found to have abnormalities in survival, neurite arborization, and TDP-43 (Fujimori et al., 2018). iPSC models have also been developed from AD patients with both sporadic and familial forms of disease to investigate pathogenic mechanisms (Israel et al., 2012).

6. Challenges of cell models for functional genomics

As exciting as iPSC technology is, there are a host of important considerations and some inherent limitations that need to be approached thoughtfully. Some of the initial methods of reprogramming involved the use of integrating viruses to express the reprogramming factors. Unfortunately, expression of integrated reprogramming factors sometimes persisted despite the application of directed differentiation protocols, and this could cause substantial variability in the extent of differentiation and heterogeneity in the composition of the differentiated cultures. Indeed, the extent of epigenetic variation between iPSC lines may account for a significant proportion of the variability in the different lineages that a given iPSC line can produce (Nishizawa et al., 2016). Fortunately, more recently developed footprint-free reprogramming methods have largely eliminated this issue (Churko et al., 2013). Still, investigators should be aware of the reprogramming methods used to generate their lines, and may want to independently assess the pluripotency and genome quality of each line before investing significant effort to use it as an experimental platform.

Another issue concerns iPSC differentiation. Patient-derived iPSC lines may not exhibit significant phenotypes until they are differentiated (HD iPSC Consortium, 2012; Fernandes et al., 2016). This may reflect important changes in disease-relevant tissue-specific eQTLs (Reynolds et al., 2019; Barbeira et al., 2018; Nica et al., 2011; Dimas et al., 2009) as cells progress down a specific lineage, or more general changes in biological pathways, such as proteostasis, that may be a feature of the stem cell state (Vilchez et al., 2012). Some protocols, such as those utilizing forced expression or activation of neurogenin 2, may have the benefit of generating a more homogeneous cell culture comprising uniform cell types. This can be helpful if variability associated with culture heterogeneity or batch is interfering with the detection of relatively small signals generated by genetic variants. On the other

hand, it also may be the case that certain phenotypes are only manifest as cell type or maturity is sufficiently specified or achieved. In our experience, determining the optimal differentiation protocol for phenotypic effect size can require some empirical head-to-head comparisons. Regardless, attention to batch variability is key, and any steps that can be taken to scale up differentiations and bank cells, ideally as late as possible in the differentiation, and to simplify protocols without sacrificing the quality of the differentiation, will likely reduce variability and increase the sensitivity of the functional genomic work.

One specific limitation of iPSCs for investigating genetic risk profiling and functional genomics of neurodegenerative disease is related to the role of aging. Aging is the number one risk factor for a variety of neurodegenerative diseases (Kennedy et al., 2014) including AD, PD, ALS and FTD. The mechanistic link between aging and neurodegenerative disease is a focus of much research (McCormick et al., 2015). The prevailing view is that aging may be associated with DNA and protein damage (Lodato et al., 2017; Xie et al., 2015; Li et al., 2013) which impairs cell function, directly affecting viability (Tsakiri et al., 2013), as well as epigenetic changes (Zhang et al., 2015; Hannum et al., 2013) that affect the transcriptome and proteome and constrict the homeostatic capacity of cells to manage stresses. In fact, by measuring a handful of epigenetic marks from a tissue, including tissue used as the starting material to make iPSCs, it is possible to generate a surprisingly accurate estimate of the chronological age of the individual from whom the tissue came (Horvath, 2013). On the other hand, if those epigenetic marks are measured again after cells from the tissue sample have been reprogrammed to pluripotency, the epigenetic "age" of the cells appears to be essentially zero. Indeed, the reprogramming process involves removal of a wide variety of epigenetic marks that are critical for governing cell-specific gene expression that defines cell types. The extent to which these marks cause aging, and the extent to which the removal of them truly reverses the physiological changes that occur with aging, is unclear. Nevertheless, it has been shown that some aging-associated declines in biological pathways (Streit et al., 2004; Bergamini et al., 2007) directly related to neurodegenerative disease, such as proteostasis (Keller et al., 2004; Cuervo et al., 2005; Massey et al., 2006; Rubinsztein et al., 2011), are improved in stem cells (Vilchez et al., 2012). If some of the genetic loci that confer risk of neurodegenerative disease do so by interacting with pathways that are important for aging, and if "aging" is required to reveal their effects, it may be difficult to use differentiated iPSCs or non-human models that do not age similarly to humans to investigate them.

Several groups have tried to develop methods to address this potential limitation. One approach has simply been to allow differentiated cells to remain in culture for extended periods of time. This may help cells achieve a level of maturity that more closely resembles analogous cells found in adult humans, but it is unclear whether meaningful aging occurs, since aging appears to progress in real time (Lam et al., 2017). Another approach has been to try to accelerate or inhibit aging by manipulating molecules such as telomerase or lamin A, which have been implicated in normal aging or in diseases of apparent accelerated aging, such as Hutchinson-Gilford progeria syndrome (HGPS) (Miller et al., 2013; Brunauer and Kennedy, 2015; Shah et al., 2013; Theodoris et al., 2017). However, some have expressed concern about the relevance of these approaches for studying neurodegenerative diseases, since lamin A is not expressed in the central nervous system, patients with HGPS exhibit

aging of some but not all tissues, and children with HGPS do not exhibit cognitive deterioration (Ullrich and Gordon, 2015). As an alternative, new methods have been developed to directly reprogram adult cells to neurons without first making them pluripotent. This strategy retains many of the epigenetic marks of aging in the reprogrammed neurons that are seen in the tissue from the donor, raising hopes that they may be able to provide a better system in which to study disease-associated risk variants in the context of aging (Mertens et al., 2015; Huh et al., 2016).

Another key limitation is that most iPSC differentiation protocols are optimized to produce a two-dimensional (2D) culture dominated by a single cell type, whereas the brain is made of many cell types that exist in particular three-dimensional (3D) spatial and stoichiometric relationships. That means that, in some cases, it is impossible to form key anatomical structures (e.g., neuromuscular junction), and critical disease-relevant processes (e.g., neuroinflammation) may be absent (Marttinen et al., 2018). Even though the diseases under consideration are often dominated by neuron-specific degeneration, new data are increasingly pointing to the importance of cell non-autonomous processes that are more complex to model (Brandscheid et al., 2017; Lim et al., 2018; Rudnick et al., 2017; Ilieva et al., 2009; McCauley and Baloh, 2018). Especially pertinent to the focus of this review, major GWAS for AD and PD revealed variants and pathways that have strongly implicated both innate and adaptive immune biology (Nalls et al., 2019; Kunkle et al., 2019; Huang et al., 2017; Lambert et al., 2013). Protocols have been developed to make and study a variety of neuronal and non-neuronal cell types from iPSCs that are relevant to neurodegenerative disease (Abud et al., 2017; Thonhoff et al., 2018; Muffat et al., 2016; Pa ca et al., 2015; Marchetto et al., 2008; Emdad et al., 2012; Park et al., 2017). However, protocols to make neurons and glia are typically quite different from each other, so it can be difficult to generate co-cultures from a single protocol with the desired stoichiometry. One solution has been to generate monocultures of two or more cell types and then combine them to create co-cultures (Serio et al., 2013). Others have leveraged the self-organizing properties of differentiating iPSCs to generate 3D structures, frequently referred to as spheroids or organoids, that resemble brain tissue in some respects (Di Lullo and Kriegstein, 2017). Another option has been to use microphysiological or organ-on-a-chip systems, which create interfaces or microfluidic connections between cultures of different cell types (Park et al., 2020; Kratz et al., 2019; Raimondi et al., 2019). These systems enable certain factors and metabolites produced by one cell type to contact other cell types in the system, which may be uniquely useful for studying drug metabolism, and may facilitate levels of iPSC differentiation and maturation that are not possible with simpler systems. Yet, with all the potential benefits of using more complex systems, there are drawbacks. Frequently, the time and resources required to use these approaches can be substantially greater than simpler systems. In addition, as system complexity increases, the batch-to-batch variability often increases and throughput decreases, so the investigator needs to weigh carefully the tradeoff between cost, sensitivity, throughput and physiological relevance of each model system. A tiered approach that initially uses simpler systems for larger scale screens and more complex ones for secondary evaluation and validation may be an appealing option.

7.   Applications of imaging to functional genomics and genetic risk profiling

Many of the genes implicated by GWAS are differentially expressed amongst subpopulations of cells within the nervous system. Some techniques, such as scRNA seq, can define critical disease-related variation in the transcriptome at single cell resolution. But most single cell OMICs techniques require that tissue be dissociated into single cell suspensions prior to analysis, largely destroying important information about the spatial relationships among different cell types in the tissue from which they came. That could be important if some variation in cell-specific gene expression results from cell non-autonomous effects of other cell types *in situ*. All OMICs assays are terminal, and it can be difficult to accurately infer from snap shots the underlying dynamic biology. Finally, and despite increasingly sophisticated data integration techniques, it remains difficult to "reconstruct" the biology of cells and tissues using only molecular biology measurements.

Fortunately, an array of powerful imaging techniques are available to visualize biology in cells and tissues, including dynamic measurements in live cells over indefinite time periods. An approach our group has pioneered is high throughput automated longitudinal single cell imaging (Arrasate and Finkbeiner, 2005; Linsley et al., 2019). A fully automated robotic microscope is enclosed in an environmental chamber coupled to a robotic incubator. Multi-well plates of cells are stored in the incubator and automatically transferred to and from the microscope stage through a computer-controlled robotic arm. Importantly, the system is designed to map the location of each cell in the culture so that the instrument can return and re-image the same cell as often and for as long as the investigator wants. Cells can be transduced with biosensors such as fluorescent dyes or proteins engineered to identify cell types or measure specific cell structures or functions to discover the impact of genomic variants on cell biology (Finkbeiner et al., 2015). At the end of the live cell imaging portion of an experiment, cells can be fixed, labeled (e.g., with dyes, fluorescently conjugated antibodies, etc.), and reimaged, enabling investigators to expand the number of single cell measurements that can be made.

To facilitate dynamic deep phenotyping, we have assembled an array of nearly 300 biosensors that we named the "Physical Exam of the Cell," specially tailored to visualize diverse biologies relevant to neurodegenerative disease (Finkbeiner et al., 2015). These biosensors can be used in two main ways. One way is to do conventional hypothesis-driven experiments. For example, an investigator may hypothesize that a particular genomic variant or gene implicated in neurodegenerative disease confers risk via certain biological pathways or structures. To test these hypotheses, biosensors can be chosen to visualize those biologies in live cells containing or lacking the SNV, or with or without perturbation of the implicated genes.

A second, potentially more novel and exciting application, is to use arrays of biosensors to explore genetic risk profiling and the role of polygenic mechanisms in conferring risk of disease. If "missing" heritability is conferred by combinations of variants that conspire to act on critical pathways relevant to specific neurodegenerative disease, it may be the case that the effect size of any one variant alone is insufficient to detect using OMICs and imaging

studies. In addition, given limitations in knowledge of biological pathways and how they are altered in the context of disease, it may be difficult to accurately predict which biologies should be investigated based only on knowledge of the substituent SNVs. If "disease" is an emergent property of an abnormal system of disease-associated SNVs, it may be necessary to study them as a system, in the context of single cells, in order to discover the pathways they alter and understand how they confer risk of disease (Fig. 2). Our strategy is to select tiered sets of biosensors to first broadly uncover key cytopathologies, and then follow those up with more specialized sets of biosensors designed to explore specific pathways and systems. This approach is patterned after the way the physical exam and sets of laboratory tests are used in medicine to arrive at a specific diagnosis from more general symptoms and complaints.

Another exciting potential application of this approach is investigating patient stratification. For the most part, the diagnosis of neurodegenerative diseases such as AD, PD and ALS is made based on clinical manifestations, and it remains an open question whether patients with sporadic or idiopathic forms of the disease all share the same underlying mechanism for their clinical syndrome or whether distinct substrata exist. As described above, the answer to this question is critical because most clinical trials in neurodegenerative disease are not powered to detect effects in patient subpopulations. That could mean that even if a trial agent works for a subset of patients, its effect may not be demonstrable statistically across the whole population because of the large group of non-responders and the inherent variability in many clinical endpoints. However, deep cellular phenotyping with OMICs or cell-based imaging may reveal molecular or cellular signatures that suggest biologically meaningful and mechanism-related patient clusters. Knowledge of such clusters can be used to design clinical trials that enroll patients who could be reasonably expected to benefit from a specific trial agent, increasing the overall chance of detecting an efficacy signal.

8. Image analysis

Images can be a particularly rich data source. Imaging can be used to measure some of the same readouts as OMICs technologies, including RNA, epigenetic changes, proteins and post-translational modifications. But imaging has the enormous advantage of being able to measure different types of macromolecules simultaneously, across space and time, and at different levels of resolution—from atomic resolution with cryoelectron tomography (Frangakis and Förster, 2004; Jiang and Ludtke, 2005) to measurements within defined cell types in complex tissues *in vivo* (Linsley et al., 2019; Keller and Ahrens, 2015). Methods are being developed to retain some spatial information in OMICs studies without using imaging, but the detection and spatial resolution of these techniques is currently limited (Weinstein et al., 2019).

A number of open-source and commercially available software tools have been developed to help investigators quantify cellular phenotypes from imaging data (Carpenter et al., 2006; Chiang et al., 2015; Schindelin et al., 2012). In general, these approaches process the images in a stepwise fashion, first segmenting pixels within the image that correspond to objects of interest in the foreground from those that make up the background. Contiguous pixels in the foreground are grouped as objects, and these are evaluated further by convolutional software

filters to identify the subset of objects that meet a set of criteria to define them as objects of particular interest and worthy of further analysis. For example, the dimensions of cell bodies typically fall within a certain range, so software filters can be used to identify the subset of segmented objects that fit those criteria and are likely cells, and exclude objects that are much bigger or smaller and that may correspond to artifacts or debris. Objects of particular interest can then be further analyzed to quantify features of them that form phenotypes, which may be differentially affected by genomic variants and genes associated with disease. These methods can be applied to images of the same cell over time to understand the cascade of phenotypic changes that flow from an SNV or genetic perturbation. In turn, statistical methods including survival analysis or Cox proportional hazards analysis and Bayesian hierarchical approaches can be applied to longitudinal single cell phenotypic data to unravel cause-and-effect relationships amongst different time-dependent phenotypes and to understand which, if any, early phenotypic changes predict future ones with particular relevance to neurodegenerative disease (Skibinski et al., 2014; Barmada et al., 2014; Finkbeiner et al., 2015; Miller et al., 2010; Miller et al., 2011; Tsvetkov et al., 2013; Skibinski et al., 2016; Shaby et al., 2016).

The fact that imaging is a relatively fast and inexpensive type of data to create makes it possible to generate very large amounts of data, which enables additional levels of analysis. In particular, machine learning and especially deep learning approaches are powerful methods for developing algorithms to measure features in images, often with better accuracy than human observers can achieve, and, in some cases, identifying features that humans can't reliably see unaided (Webb, 2018; Grys et al., 2017; Waller and Tian, 2015; Armañanzas and Ascoli, 2015; Sommer and Gerlich, 2013; Christiansen et al., 2018; Yang et al., 2018). In a recent study, deep learning convolutional neural networks were trained to accurately perform pixel-wise predictions of a range of fluorescent labels of cell structure, state and type from images of unlabeled cells (Christiansen et al., 2018)—a feat that would be difficult or impossible for humans. Remarkably, in that study, the deep learning networks were able to extract useful information for making predictions from images that were not in the focal plane and likely would have been disregarded as useless out-of-focus images by a human observer. This example underscores the potential of machine learning-based computer vision methods to find valuable information in images that is currently ignored, and to discover important patterns in imaging datasets that are simply too large and complex for humans to comprehend. With this in mind, there is a lot of excitement about using biosensors in live cells or labeling schemes in fixed cells (Bray et al., 2016) in a somewhat disease-agnostic way that could be harnessed by machine learning to discover completely new phenotypes or phenotypic classification schemes. Perhaps someday, approaches like these will enable more multidimensional and integrated analyses of cellular phenotypes in response to genetic and small molecular perturbations that bear a clear connection to phenotypes in patients, and that will predict better which interventions are likely to work in clinical trials.

9. Investigating functional genomics and genetic risk profiling with perturbation analysis

The preceding sections describe a logical information flow from studying genotype-phenotype associations in humans, to the discovery of specific genetic loci that demonstrate a significant phenotypic association, to a list of genomic variants and genes through which those loci act to lead to phenotypes of interest. Still, these approaches all rely on observations and associations, and the evidence that a specific genomic variant or gene plays a causal role in disease pathogenesis and warrants consideration as a therapeutic target can be substantially strengthened with perturbation analysis.

Typically, the simplest, fastest and cheapest way to evaluate potential therapeutic targets is to perform gene level perturbations and measure whether the phenotype of interest is modulated in the expected way. In most cases, a gene level perturbation will have a larger effect than the SNV that implicated the gene, so it also can be a sensitive way to screen for mechanisms to investigate further. The methods to do this have evolved rapidly. Antisense oligonucleotides (ASOs), which are designed to bind a specific target mRNA and trigger host responses to lower its levels, have been available for years. They remain relevant because ASOs have been designed that modulate the levels of genes that cause specific neurodegenerative diseases, and they have been shown to have beneficial effects in animal models and in clinical trials (Evers et al., 2015; Mercuri et al., 2018; Finkel et al., 2017; Finkel et al., 2016). In the laboratory, libraries of small-interfering RNAs (siRNA) have been developed to do forward genetic studies in model systems to discover genotype-phenotype relationships and to perturb genes implicated by human studies (Elia et al., 2019). However, because siRNAs are prone to bind to genes other than their intended target, a variety of strategies have been developed to avoid confusing "off-target" effects (Elia and Finkbeiner, 2012). Often, multiple siRNAs are designed against different regions of the same gene target, based on the idea that it would be statistically unlikely for siRNAs with different sequences to have the same phenotypic effect unless that effect was through the common target gene. To further reduce the likelihood of off-target effects, it is not uncommon to test newly-designed siRNAs that were not included in an original screen, or to use siRNA-resistant alleles of the target gene to demonstrate that the phenotypic effects of the siRNA can be rescued by the target gene. A major appeal of siRNAs is that they are relatively inexpensive and, in many cases, can be transduced into cells efficiently without special transfection agents. In addition, pools of siRNA against specific combinations of genes can be designed and delivered alone or in combination with plasmids overexpressing other genes to investigate polygenic hypotheses.

More recently, clustered regular interspaced short palindromic repeat (CRISPR)-based methods of gene modulation have gained wide use because they offer improvements in the specificity of gene targeting and the ability to perform a range of elegant genomic engineering feats (Wang et al., 2015; Zhang et al., 2014; Wang et al., 2014; Shalem et al., 2014). So-called guide RNAs (gRNAs) are designed to recruit Cas 9 proteins to specific genomic loci, and depending on the Cas 9 and peptide domains to which it is fused, recruitment may lead to gene mutation, correction, activation or suppression. Improving

gRNA design is an ongoing effort, and specificity continues to grow (Doench et al., 2016). The ability to use these systems to introduce specific genomic variants into the genome or to revert a potential causal SNV to a non-pathogenic variant is particularly powerful because it ostensibly provides a way to generate and characterize cells that are largely identical except for the SNV under investigation. In the same vein, multiple SNVs can be introduced in various combinations to explore polygenic mechanisms of disease pathogenesis. In 2020, it still remains relatively expensive and laborious to perform genome editing, but improvements in the quality, speed and cost of the technology are rapidly being made.

CRISPR-based approaches have also been combined with single cell transcriptomics in a method called CROPseq to do rapid functional genomics (Datlinger et al., 2017). In one version, pools of gRNAs are designed against an array of gene targets with built-in sequence "bar codes". Cells harboring Cas 9 and transduced with these libraries are given time to achieve modulation of the target genes and then are subjected to scRNAseq. Along with information on each cell's transcriptome, the RNAseq data contains the bar code information of the gRNA that triggered the response. These data can then be used to rapidly define the gene networks that the perturbation of specific genes affect, and begin to develop a substantial understanding of the likely functional consequences of genomic variants.

10. Conclusions

The importance of unbiased human genetics for establishing the role of genetics in disease, and laying the foundation for identifying pathogenic mechanisms and opportunities for effective therapeutic intervention has only grown over time. The methods to collect and analyze human genetic data are very well established, and the depth of insights that can be gleaned are remarkable. But much work must usually be done to take these insights and develop them into an understanding of the pathogenic mechanisms of disease. Investigators first need to understand the genetic mechanisms through which disease associated loci work. Fine mapping, co-localization, causal modeling, pathway analysis and machine learning are all important bioinformatic approaches to analyze disease-associated loci and nominate a prioritized list of variants and genes that may mediate the observed effects.

With a more manageable list to investigate, multi-Omics approaches and imaging can be effective lines of study to generate deep and comprehensive understanding of human and non-human models with putative disease-associated variants and gene mutations. Statistical and deep learning methods can be used to integrate the data from these studies and discover molecular and cellular signatures of specific genomic variants. Perturbation analysis, including modulating the levels of candidate genes or introducing or reverting putative causal SNVs with genome engineering, can provide additional evidence for or against a pathogenic role for a candidate gene or variant. These techniques can be used to understand the functional genomics of a specific variant as well as investigate the collective effect of multiple variants, thereby providing an approach to test hypotheses of polygenic mechanisms of disease and elucidate the molecular underpinnings of specific genetic risk profiles.

In conclusion, we suggest that application of the types of methods and pipelines outlined in this review to human genetic data will lead to improved prioritization of therapeutic targets in the coming decade. The overarching goal is to base the prioritization on firm genetic foundations and convincing computational and experimental evidence demonstrating causal relationships to disease. Still, it is important to keep in mind that a therapeutic target based on a genomic variant is not truly validated until modulation in humans is shown to effectively mitigate disease. Hopefully, rooting discovery in genetics and implementing a rigorous approach to prioritize genetic variants to find druggable targets will eventually lead to greater rates of clinical validation and successful therapies.

Acknowledgments

References

Abud EM, et al. , 2017. iPSC-derived human microglia-like cells to study neurological diseases. Neuron 94 278–293.e279. [PubMed: 28426964]

Al-Chalabi A, Hardiman O, 2013. The epidemiology of ALS: A conspiracy of genes, environment and time. Nat. Rev. Neurol 9, 617–628. [PubMed: 24126629]

Armakola M, et al. , 2012. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. Nat. Genet 44, 1302–1309. [PubMed: 23104007]

Armañanzas R, Ascoli GA, 2015. Towards the automatic classification of neurons. Trends Neurosci. 38, 307–318. [PubMed: 25765323]

Aron R, et al. , 2018. Deubiquitinase Usp12 functions noncatalytically to induce autophagy and confer neuroprotection in models of Huntington's disease. Nat. Commun 9, 3191. [PubMed: 30266909]

Arrasate M, Finkbeiner S, 2005. Automated microscope system for determining factors that predict neuronal fate. Proc. Natl. Acad. Sci. U.S.A 102, 3840–3845. [PubMed: 15738408]

Arrowsmith J, Miller P, 2013. Trial watch: phase II and phase III attrition rates 2011–2012. Nat. Rev. Drug Discov 12, 569. [PubMed: 23903212]

Barbeira AN, et al. , 2018. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun 9, 1825. [PubMed: 29739930]

Barmada SJ, et al. , 2014. Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. Nat. Chem. Biol 10, 677–685. [PubMed: 24974230]

Barmada SJ, et al. , 2015. Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. Proc. Natl. Acad. Sci. U.S.A 112, 7821–7826. [PubMed: 26056265]

Behravan H, et al. , 2018. Machine learning identifies interacting genetic variants contributing to breast cancer risk: A case study in Finnish cases and controls. Sci. Rep 8, 13149. [PubMed: 30177847]

Bergamini E, Cavallini G, Donati A, Gori Z, 2007. The role of autophagy in aging: Its essential part in the anti-aging mechanism of caloric restriction. Ann. N. Y. Acad. Sci 1114, 69–78. [PubMed: 17934054]

Bilican B, et al. , 2012. Mutant induced pluripotent stem cell lines recapitulate aspects of TDP-43 proteinopathies and reveal cell-specific vulnerability. Proc. Natl. Acad. Sci. U.S.A 109, 5803–5808. [PubMed: 22451909]

Boyle EA, Li YI, Pritchard JK, 2017. An expanded view of complex traits: From polygenic to omnigenic. Cell 169, 1177–1186. [PubMed: 28622505]

Brandscheid C, et al. , 2017. Altered gut microbiome composition and tryptic activity of the 5xFAD Alzheimer's mouse model. J. Alzheimers Dis 56, 775–788. [PubMed: 28035935]

Bray MA, et al. , 2016. Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat. Protocols 11, 1757–1774. [PubMed: 27560178]

Brunauer R, Kennedy BK, 2015. Progeria accelerates adult stem cell aging. Science (New York, N.Y.) 348, 1093–1094.

Burkhardt MF, et al. , 2013. A cellular model for sporadic ALS using patient-derived induced pluripotent stem cells. Mol. Cell. Neurosci 56, 355–364. [PubMed: 23891805]

Campioni MR, Finkbeiner S, 2015. Going retro: Ancient viral origins of cognition. Neuron 86, 346–348. [PubMed: 25905805]

Carpenter AE, et al. , 2006. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 7, R100. [PubMed: 17076895]

Chiang M, et al. , 2015. Analysis of *in vivo* single cell behavior by high throughput, human-in-the-loop segmentation of three-dimensional images. BMC Bioinformatics 16, 397. [PubMed: 26607933]

Christiansen E, et al. , 2018. *In silico* labeling: Predicting fluorescent labels in unlabeled images. Cell 173, 1–12. [PubMed: 29570990]

Church DM, et al. , 2009. Lineage-specific biology revealed by a finished genome as- sembly of the mouse. PLoS Biol. 7, e1000112. [PubMed: 19468303]

Churko JM, Burridge PW, Wu JC, 2013. Generation of human iPSCs from human peripheral blood mononuclear cells using non-integrative Sendai virus in chemically defined conditions. Methods Mol. Biol (Clifton, N.J.) 1036, 81–88. [PubMed: 23807788]

Cobb MM, Ravisankar A, Skibinski G, Finkbeiner S, 2018. iPS cells in the study of PD molecular pathogenesis. Cell Tissue Res. 373, 61–77. [PubMed: 29234887]

Crook ZR, Housman D, 2011. Huntington′s disease: Can mice lead the way to treatment? Neuron 69, 423–435. [PubMed: 21315254]

Cuervo AM, et al. , 2005. Autophagy and aging: The importance of maintaining "clean" cells. Autophagy 1, 131–140. [PubMed: 16874025]

Datlinger P, et al. , 2017. Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods 14, 297–301. [PubMed: 28099430]

Devlin AC, et al. , 2015. Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. Nat. Commun 6, 5999. [PubMed: 25580746]

Di Lullo E, Kriegstein AR, 2017. The use of brain organoids to investigate neural development and disease. Nat. Rev. Neurosci 18, 573–584. [PubMed: 28878372]

Dimas AS, et al. , 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science (New York, N.Y.) 325, 1246–1250.

Doench JG, et al. , 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol 34, 184–191. [PubMed: 26780180]

Edwards SL, Beesley J, French JD, Dunning AM, 2013. Beyond GWASs: Illuminating the dark road from association to function. Am. J. Hum. Genet 93, 779–797. [PubMed: 24210251]

Ekstrand MI, Galter D, 2009. The MitoPark mouse - an animal model of Parkinson's disease with impaired respiratory chain function in dopamine neurons. Parkinsonism Related Disorders 15 (Suppl. 3), S185–S188. [PubMed: 20082987]

Elia LP, Finkbeiner S, 2012. In OMICs in Neurology and Neuroscience Submitted (Oxford University Press).

Elia LP, Mason AR, Alijagic A, Finkbeiner S, 2019. Genetic regulation of neuronal progranulin reveals a critical role for the autophagy-lysosome pathway. J. Neurosci 39, 3332–3344. [PubMed: 30696728]

Emdad L, D'Souza SL, Kothari HP, Qadeer ZA, Germano IM, 2012. Efficient differentiation of human embryonic and induced pluripotent stem cells into functional astrocytes. Stem Cells Dev. 21, 404–410. [PubMed: 21631388]

Esanov R, et al. , 2016. C9orf72 promoter hypermethylation is reduced while hydroxymethylation is acquired during reprogramming of ALS patient cells. Exp. Neurol 277, 171–177. [PubMed: 26746986]

Escott-Price V, et al. , 2015. Polygenic risk of Parkinson disease is correlated with disease age at onset. Ann. Neurol 77, 582–591. [PubMed: 25773351]

Evers MM, Toonen LJ, van Roon-Mom WM, 2015. Antisense oligonucleotides in therapy for neurodegenerative disorders. Adv. Drug Deliv. Rev 87, 90–103. [PubMed: 25797014]

Ewald H, et al. , 2020. Nonrandomized studies using causal-modeling may give different answers than RCTs: A meta-epidemiological study. J. Clin. Epidemiol 118, 29–41. [PubMed: 31704350]

Fang EF, et al. , 2019a. Mitophagy inhibits amyloid-β and tau pathology and reverses cognitive deficits in models of Alzheimer's disease. Nat. Neurosci 22, 401–412. [PubMed: 30742114]

Fang MY, et al. , 2019b. Small-molecule modulation of TDP-43 recruitment to stress granules prevents persistent TDP-43 accumulation in ALS/FTD. Neuron 103, 802–819. [PubMed: 31272829]

Fernagut PO, Tison F, 2012. Animal models of multiple system atrophy. Neuroscience 211, 77–82. [PubMed: 21963351]

Fernandes HJ, et al. , 2016. ER stress and autophagic perturbations lead to elevated extracellular α-Synuclein in GBA-N370S Parkinson's iPSC-Derived Dopamine Neurons. Stem Cell Rep. 6, 342–356.

Finkbeiner S, Frumkin M, Kassner PD, 2015. Cell-based screening: Extracting meaning from complex data. Neuron 86, 160–174. [PubMed: 25856492]

Finkel RS, et al. , 2016. Treatment of infantile-onset spinal muscular atrophy with nu-sinersen: A phase 2, open-label, dose-escalation study. Lancet (London, England) 388, 3017–3026.

Finkel RS, et al. , 2017. Nusinersen versus sham control in infantile-onset spinal muscular atrophy. N. Engl. J. Med 377, 1723–1732. [PubMed: 29091570]

Frangakis AS, Förster F, 2004. Computational exploration of structural information from cryo-electron tomograms. Curr. Opin. Struct. Biol 14, 325–331. [PubMed: 15193312]

Fujimori K, et al. , 2018. Modeling sporadic ALS in iPSC-derived motor neurons identifies a potential therapeutic agent. Nat. Med 24, 1579–1589. [PubMed: 30127392]

Genetics for all, 2019. Nat. Genet 51. pp. 579. [PubMed: 30926967]

Giambartolomei C, et al. , 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383. [PubMed: 24830394]

Grys BT, et al. , 2017. Machine learning and computer vision approaches for phenotypic profiling. J. Cell Biol 216, 65–71. [PubMed: 27940887]

Hannum G, et al. , 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell 49, 359–367. [PubMed: 23177740]

Haston KM, Finkbeiner S, 2016. Clinical trials in a dish: The potential of pluripotent stem cells to develop therapies for neurodegenerative diseases. Annu. Rev. Pharmacol. Toxicol 56, 489–510. [PubMed: 26514199]

HD iPSC Consortium, 2012. Induced pluripotent stem cells from patients with Huntington's disease show CAG-repeat-expansion-associated phenotypes. Cell Stem Cell 11, 264–278. [PubMed: 22748968]

HD iPSC Consortium, 2017. Developmental alterations in Huntington′s disease neural cells and pharmacological rescue in cells and mice. Nat. Neurosci 20, 648–660. [PubMed: 28319609]

Hodge RD, et al. , 2019. Conserved cell types with divergent features in human versus mouse cortex. Nature 573, 61–68. [PubMed: 31435019]

Hormozdiari F, et al. , 2016. Colocalization of GWAS and eQTL signals detects target genes. Am. J. Hum. Genet 99 ga 2.7.20.

Horvath S, 2013. DNA methylation age of human tissues and cell types. Genome Biol. 14, R115. [PubMed: 24138928]

Hsieh CH, et al. , 2019. Miro1 marks Parkinson′s disease subset and Miro1 Reducer rescues neuron loss in Parkinson's models. Cell Metab. 30, 1131–1140. [PubMed: 31564441]

Huang KL, et al. , 2017. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. Nat. Neurosci 20, 1052–1061. [PubMed: 28628103]

Huh CJ, et al. , 2016. Maintenance of age in human neurons generated by microRNA- based neuronal conversion of fibroblasts. eLife 5.

Ilieva H, Polymenidou M, Cleveland DW, 2009. Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. J. Cell Biol 187, 761–772. [PubMed: 19951898]

Imamura K, et al. , 2016. Calcium dysregulation contributes to neurodegeneration in FTLD patient iPSC-derived neurons. Sci. Rep 6, 34904. [PubMed: 27721502]

Iovino M, et al. , 2015. Early maturation and distinct tau pathology in induced pluripotent stem cell-derived neurons from patients with MAPT mutations. Brain 138, 3345–3359. [PubMed: 26220942]

Israel MA, et al. , 2012. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature 482, 216–220. [PubMed: 22278060]

Jiang W, Ludtke SJ, 2005. Electron cryomicroscopy of single particles at sub-nanometer resolution. Curr. Opin. Struct. Biol 15, 571–577. [PubMed: 16140524]

Kaul A, Bhattacharyya S, Ay F, 2020. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. Nature Protocols 15, 991–1012. [PubMed: 31980751]

Kaus A, Sareen D, 2015. ALS patient stem cells for unveiling disease signatures of motoneuron susceptibility: Perspectives on the deadly mitochondria, ER stress and calcium triad. Front. Cell. Neurosci 9, 448. [PubMed: 26635528]

Kaye JA, Finkbeiner S, 2013. Modeling Huntington's disease with induced pluripotent stem cells. Mol. Cell. Neurosci 56C, 50–64.

Kedaigle AJ, et al. , 2019. Bioenergetic deficits in Huntington's disease iPSC-derived neural cells and rescue with glycolytic metabolites. Hum. Mol. Genet 29, 1757–1771.

Keller PJ, Ahrens MB, 2015. Visualizing whole-brain activity and development at the single-cell level using light-sheet microscopy. Neuron 85, 462–483. [PubMed: 25654253]

Keller JN, et al. , 2004. Autophagy, proteasomes, lipofuscin, and oxidative stress in the aging brain. Int. J. Biochem. Cell Biol 36, 2376–2391. [PubMed: 15325579]

Keller MF, et al. , 2012. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. Hum. Mol. Genet 21, 4996–5009. [PubMed: 22892372]

Kempfer R, Pombo A, 2019. Methods for mapping 3D chromosome architecture. Nat. Rev. Genet 21, 207–226. [PubMed: 31848476]

Kennedy BK, et al. , 2014. Geroscience: Linking aging to chronic disease. Cell 159, 709–713. [PubMed: 25417146]

Kim HJ, et al. , 2014. Therapeutic modulation of eIF2α phosphorylation rescues TDP-43 toxicity in amyotrophic lateral sclerosis disease models. Nat. Genet 46, 152–160. [PubMed: 24336168]

Kim J, et al. , 2017. Changes in the excitability of neocortical neurons in a mouse model of amyotrophic lateral sclerosis are not specific to corticospinal neurons and are modulated by advancing disease. J. Neurosci 37, 9037–9053. [PubMed: 28821643]

King EA, Davis JW, Degner JF, 2019. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet. 15, e1008489. [PubMed: 31830040]

Kondo T, et al. , 2013. Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Abeta and differential drug responsiveness. Cell Stem Cell 12, 487–496. [PubMed: 23434393]

Koyuncu S, et al. , 2018. The ubiquitin ligase UBR5 suppresses proteostasis collapse in pluripotent stem cells from Huntington's disease patients. Nat. Commun 9, 2886. [PubMed: 30038412]

Kratz SRA, Holl G, Schuller P, Ertl P, Rothbauer M, 2019. Latest trends in biosensing for microphysiological organs-on-a-chip and body-on-a-chip systems. Biosensors 9.

Kumar V, et al. , 2013. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genetics 9, e1003201. [PubMed: 23341781]

Kunkle BW, et al. , 2019. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat. Genet 51, 414–430. [PubMed: 30820047]

Lalonde E, et al. , 2011. RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome Res. 21, 545–554. [PubMed: 21173033]

Lam RS, Topfer FM, Wood PG, Busskamp V, Bamberg E, 2017. Functional maturation of human stem cell-derived neurons in long-term cultures. PloS One 12, e0169506. [PubMed: 28052116]

Lambert JC, et al. , 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet 45, 1452–1458. [PubMed: 24162737]

Langfelder P, Horvath S, 2008. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. [PubMed: 19114008]

Lee Y, Dawson VL, Dawson TM, 2012. Animal models of Parkinson's disease: Vertebrate genetics. Cold Spring Harb. Perspect. Med 2.

Li W, et al. , 2013. Activation of transposable elements during aging and neuronal de- cline in Drosophila. Nat. Neurosci 16, 529–531. [PubMed: 23563579]

Lim S, Kim HJ, Kim DK, Lee SJ, 2018. Non-cell-autonomous actions of alpha-sy-nuclein: Implications in glial synucleinopathies. Prog. Neurobiol 169, 158–171. [PubMed: 30173732]

Lin S, et al. , 2014. Comparison of the transcriptional landscapes between human and mouse tissues. Proc. Natl. Acad. Sci. USA 111, 17224–17229. [PubMed: 25413365]

Linsley JW, et al. , 2019. Automated four-dimensional long term imaging enables single cell tracking within organotypic brain slices to study neurodevelopment and degeneration. Commun. Biol 2, 155. [PubMed: 31069265]

Lodato MA, et al. , 2017. Aging and neurodegeneration are associated with increased mutations in single human neurons. Science (New York, N.Y.) 359, 555–559.

Lowe EK, Cuomo C, Voronov D, Arnone MI, 2019. Using ATAC-seq and RNA-seq to increase resolution in GRN connectivity. Methods Cell Biol. 151, 115–126. [PubMed: 30948003]

Manolio TA, et al. , 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753. [PubMed: 19812666]

Marchetto MCN, et al. , 2008. Non-cell-autonomous effect of human SOD1^{G37R} astrocytes on motor neurons derived from human embryonic stem cells. Cell Stem Cell 3, 649–657. [PubMed: 19041781]

Marrone L, et al. , 2018. Isogenic FUS-eGFP iPSC reporter lines enable quantification of FUS stress granule pathology that is rescued by drugs inducing autophagy. Stem Cell Rep. 10, 375–389.

Martinez FJ, et al. , 2016. Protein-RNA networks regulated by normal and ALS-asso-ciated mutant HNRNPA2B1 in the nervous system. Neuron 92, 780–795. [PubMed: 27773581]

Martín-Maestro P, et al. , 2019. Autophagy induction by bexarotene promotes mitophagy in presenilin 1 familial Alzheimer's disease iPSC-derived neural stem cells. Mol. Neurobiol 56, 8220–8236. [PubMed: 31203573]

Marttinen M, et al. , 2018. Molecular mechanisms of synaptotoxicity and neuroinflammation in Alzheimer's disease. Front. Neurosci 12, 963. [PubMed: 30618585]

Massey AC, Zhang C, Cuervo AM, 2006. Chaperone-mediated autophagy in aging and disease. Curr. Top. Dev. Biol 73, 205–235. [PubMed: 16782460]

Mattis KK, Gloyn AL, 2020. From genetic association to molecular mechanisms for islet-cell dysfunction in type 2 diabetes. J. Mol. Biol 432, 1551–1578. [PubMed: 31945378]

Mattson MP, Rychlik B, You JS, Sisken JE, 1991. Sensitivity of cultured human embryonic cerebral cortical neurons to excitatory amino acid-induced calcium influx and neurotoxicity. Brain Res 542, 97–106. [PubMed: 2054664]

McCauley ME, Baloh RH, 2018. Inflammation in ALS/FTD pathogenesis. Acta Neuropathol. 137, 715–730. [PubMed: 30465257]

McCormick MA, et al. , 2015. A comprehensive analysis of replicative lifespan in 4,698 single-gene deletion strains uncovers conserved mechanisms of aging. Cell Metab. 22, 895–906. [PubMed: 26456335]

McGary KL, et al. , 2010. Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc. Natl. Acad. Sci. USA 107, 6544–6549. [PubMed: 20308572]

McGonigle P, Ruggeri B, 2014. Animal models of human disease: Challenges in enabling translation. Biochem. Pharmacol 87, 162–171. [PubMed: 23954708]

Mercuri E, et al. , 2018. Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. New England J. Med 378, 625–635. [PubMed: 29443664]

Mertens J, et al. , 2013. APP processing in human pluripotent stem cell-derived neurons is resistant to NSAID-based gamma-secretase modulation. Stem Cell Rep. 1, 491–498.

Mertens J, et al. , 2015. Directly reprogrammed human neurons retain aging-associated transcriptomic signatures and reveal age-related nucleocytoplasmic defects. Cell stem cell 17, 1–14. [PubMed: 26140600]

Miller J, et al. , 2010. Quantitative relationships between huntingtin levels, polyglutamine length, inclusion body formation, and neuronal death provide novel insight into Huntington's disease molecular pathogenesis. J. Neurosci 30, 10541–10550. [PubMed: 20685997]

Miller J, et al. , 2011. Identifying polyglutamine protein species *in situ* that best predict neurodegeneration. Nat. Chem. Biol 7, 925–934. [PubMed: 22037470]

Miller JD, et al. , 2013. Human iPSC-based modeling of late-onset disease via progerininduced aging. Cell Stem Cell 13, 691–705. [PubMed: 24315443]

Morgan S, et al. , 2017. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. Brain 140, 1611–1618. [PubMed: 28430856]

Muffat J, et al. , 2016. Efficient derivation of microglia-like cells from human pluripotent stem cells. Nat. Med 22, 1358–1367. [PubMed: 27668937]

Mullane K, Williams M, 2013. Alzheimer′s therapeutics: Continued clinical failures question the validity of the amyloid hypothesis-but what lies beyond? Biochem. Pharmacol 85, 289–305. [PubMed: 23178653]

Mullane K, Williams M, 2019. Preclinical models of Alzheimer′s disease: Relevance and translational validity. Curr. Protoc. Pharmacol 84, e57. [PubMed: 30802363]

Mullane K, Winquist RJ, Williams M, 2014. Translational paradigms in pharmacology and drug discovery. Biochem. Pharmacol 87, 189–210. [PubMed: 24184503]

Nalls MA, et al. , 2019. Identification of novel risk loci, causal insights, and heritable risk for Parkinson′s disease: A meta-analysis of genome-wide association studies. Lancet. Neurol 18, 1091–1102. [PubMed: 31701892]

Nguyen HN, et al. , 2011. LRRK2 mutant iPSC-derived DA neurons demonstrate increased susceptibility to oxidative stress. Cell Stem Cell 8, 267–280. [PubMed: 21362567]

Nica AC, Dermitzakis ET, 2013. Expression quantitative trait loci: Present and future. Philos. Trans. R. Soc. Lond. B. Biol. Sci 368, 20120362. [PubMed: 23650636]

Nica AC, et al. , 2011. The architecture of gene regulatory variation across multiple human tissues: The MuTHER study. PLoS Genetics 7, e1002003. [PubMed: 21304890]

Nishizawa M, et al. , 2016. Epigenetic variation between human induced pluripotent stem cell lines is an indicator of differentiation capacity. Cell Stem Cell 19, 341–354. [PubMed: 27476965]

Ocampo A, et al. , 2016. *In vivo* amelioration of age-associated hallmarks by partial reprogramming. Cell 167, 1719–1733. [PubMed: 27984723]

Okita K, et al. , 2011. A more efficient method to generate integration-free human iPS cells. Nat. Methods 8, 409–412. [PubMed: 21460823]

Osellame LD, et al. , 2013. Mitochondria and quality control defects in a mouse model of Gaucher disease–links to Parkinson′s disease. Cell Metab. 17, 941–953. [PubMed: 23707074]

Oulhen N, Foster S, Wray G, Wessel G, 2019. Identifying gene expression from single cells to single genes. Methods Cell Biol. 151, 127–158. [PubMed: 30948004]

Pankevich DE, Altevogt BM, Dunlop J, Gage FH, Hyman SE, 2014. Improving and accelerating drug development for nervous system disorders. Neuron 84, 546–553. [PubMed: 25442933]

Park GS, et al. , 2017. Emulating host-microbiome ecosystem of human gastrointestinal tract *in vitro*. Stem Cell Rev. Rep 13, 321–334. [PubMed: 28488235]

Park D, Lee J, Chung JJ, Jung Y, Kim SH, 2020. Integrating organs-on-chips: Multiplexing, scaling, vascularization, and innervation. Trends Biotechnol. 38, 99–112. [PubMed: 31345572]

Pa ca AM, et al. , 2015. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. Nat. Methods 12, 671–678. [PubMed: 26005811]

Pei G, Chen L, Zhang W, 2017. WGCNA application to proteomic and metabolomic data analysis. Methods Enzymol. 585, 135–158. [PubMed: 28109426]

Penney J, Ralvenius WT, Tsai LH, 2020. Modeling Alzheimer's disease with iPSC-derived brain cells. Mol. Psychiatry 25, 148–167. [PubMed: 31391546]

Perlman RL, 2016. Mouse models of human disease: An evolutionary perspective. Evol. Med. Public Health 2016, 170–176. [PubMed: 27121451]

Perrin S, 2014. Preclinical research: Make mouse studies work. Nature 507, 423–425. [PubMed: 24678540]

Pickrell JK, et al. , 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768–772. [PubMed: 20220758]

Pihlstrøm L, Wiethoff S, Houlden H, 2017. Genetics of neurodegenerative diseases: An overview. Handb. Clin. Neurol 145, 309–323. [PubMed: 28987179]

Plenge RM, Scolnick EM, Altshuler D, 2013. Validating therapeutic targets through human genetics. Nat. Rev. Drug Discov 12, 581–594. [PubMed: 23868113]

Plomin R, DeFries JC, Knopik VS, Neiderhiser JM, 2016. Top 10 replicated findings from behavioral genetics. Perspect. Psychol. Sci 11, 3–23. [PubMed: 26817721]

Prinz F, Schlange T, Asadullah K, 2011. Believe it or not: How much can we rely on published data on potential drug targets? Nat. Rev. Drug Discov 10, 712. [PubMed: 21892149]

Raimondi I, et al. , 2019. Organ-on-a-chip *in vitro* models of the brain and the blood-brain barrier and their value to study the microbiota-gut-brain axis in neurodegeneration. Front. Bioeng. Biotechnol 7, 435. [PubMed: 31998702]

Ramanan VK, Saykin AJ, 2013. Pathways to neurodegeneration: Mechanistic insights from GWAS in Alzheimer′s disease, Parkinson′s disease, and related disorders. Am. J. Neurodegener. Dis 2, 145–175. [PubMed: 24093081]

Ransohoff RM, 2018. All (animal) models (of neurodegeneration) are wrong. Are they also useful? J. Exp. Med 215, 2955–2958. [PubMed: 30459159]

Reilly P, et al. , 2017. Novel human neuronal tau model exhibiting neurofibrillary tangles and transcellular propagation. Neurobiol. Dis 106, 222–234. [PubMed: 28610892]

Reynolds RH, et al. , 2019. Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability. NPJ Parkinsons Dis. 5, 6. [PubMed: 31016231]

Ringel M, Tollman P, Hersch G, Schulze U, 2013. Does size matter in R&D productivity? If not, what does? Nat. Rev. Drug Discov 12, 901–902. [PubMed: 24136396]

Roberson ED, 2012. Mouse models of frontotemporal dementia. Ann. Neurol 72, 837–849. [PubMed: 23280835]

Rosenberg NA, et al. , 2010. Genome-wide association studies in diverse populations. Nat. Rev. Genet 11, 356–366. [PubMed: 20395969]

Rubinsztein DC, Mariño G, Kroemer G, 2011. Autophagy and aging. Cell 146, 682–695. [PubMed: 21884931]

Rudnick ND, et al. , 2017. Distinct roles for motor neuron autophagy early and late in the SOD1(G93A) mouse model of ALS. Proc. Natl. Acad. Sci. USA 114, E8294–E8303. [PubMed: 28904095]

Sanders LH, et al. , 2014. LRRK2 mutations cause mitochondrial DNA damage in iPSC- derived neural cells from Parkinson's disease patients: reversal by gene correction. Neurobiol. Dis 62, 381–386. [PubMed: 24148854]

Schaeffer V, et al. , 2012. Stimulation of autophagy reduces neurodegeneration in a mouse model of human tauopathy. Brain 135, 2169–2177. [PubMed: 22689910]

Schindelin J, et al. , 2012. Fiji: An open-source platform for biological-image analysis. Nat. Methods 9, 676–682. [PubMed: 22743772]

Selvaraj BT, et al. , 2018. C9ORF72 repeat expansion causes vulnerability of motor neurons to Ca(2+)-permeable AMPA receptor-mediated excitotoxicity. Nat. Commun 9, 347. [PubMed: 29367641]

Seok J, et al. , 2013. Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc. Natl. Acad. Sci. U.S.A 110, 3507–3512. [PubMed: 23401516]

Serio A, et al. , 2013. Astrocyte pathology and the absence of non-cell autonomy in an induced pluripotent stem cell model of TDP-43 proteinopathy. Proc. Natl. Acad. Sci. USA 110, 4697–4702. [PubMed: 23401527]

Sexton T, Bantignies F, Cavalli G, 2009. Genomic interactions: Chromatin loops and gene meeting points in transcriptional regulation. Seminars Cell Develop. Biol 20, 849–855.

Shaby BA, Skibinsk IG, Ando M, LaDow ES, Finkbeiner S, 2016. A three-groups model for high-throughput survival screens. Biometrics 72, 936–944. [PubMed: 26821783]

Shah PP, et al. , 2013. Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. Genes Dev. 27, 1787–1799. [PubMed: 23934658]

Shalem O, et al. , 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science (New York, N.Y.) 343, 84–87.

Sharp ES, Gatz M, 2011. Relationship between education and dementia: An updated systematic review. Alzheimer Dis. Assoc. Disord 25, 289–304. [PubMed: 21750453]

Shashikant T, Ettensohn CA, 2019. Genome-wide analysis of chromatin accessibility using ATAC-seq. Methods Cell Biol. 151, 219–235. [PubMed: 30948010]

Shulman JM, et al. , 2014. Functional screening in Drosophila identifies Alzheimer′s disease susceptibility genes and implicates tau-mediated mechanisms. Hum. Mol. Genetics 23, 870–877.

Simovski B, et al. , 2018. Coloc-stats: A unified web interface to perform colocalization analysis of genomic features. Nucleic. Acids Res 46, W186–W193. [PubMed: 29873782]

Simpson CL, Al-Chalabi A, 2006. Amyotrophic lateral sclerosis as a complex genetic disease. Biochim. Biophys. Acta 1762, 973–985. [PubMed: 16973338]

Sirugo G, Williams SM, Tishkoff SA, 2019. The missing diversity in human genetic studies. Cell 177, 26–31. [PubMed: 30901543]

Skibinski G, Nakamura K, Cookson MR, Finkbeiner S, 2014. Mutant LRRK2 toxicity in neurons depends on LRRK2 levels and synuclein but not kinase activity or inclusion bodies. J. Neurosci 34, 418–433. [PubMed: 24403142]

Skibinski G, et al. , 2016. Nrf2 mitigates LRRK2- and α-synuclein-induced neurodegeneration by modulating proteostasis. Proc. Natl. Acad. Sci. USA 114, 1165–11270. [PubMed: 28028237]

Soldner F, et al. , 2016. Parkinson-associated risk variant in distal enhancer of alpha- synuclein modulates target gene expression. Nature 533, 95–99. [PubMed: 27096366]

Sommer C, Gerlich DW, 2013. Machine learning in cell biology - teaching computers to recognize phenotypes. J. Cell Sci 126, 5529–5539. [PubMed: 24259662]

Stamatakou E, et al. , 2020. Mendelian neurodegenerative disease genes involved in autophagy. Cell Discov. 6, 24.

Stan AD, et al. , 2006. Human postmortem tissue: What quality markers matter? Brain Res. 1123, 1–11. [PubMed: 17045977]

Streit WJ, Sammons NW, Kuhns AJ, Sparks DL, 2004. Dystrophic microglia in the aging human brain. Glia 45, 208–212. [PubMed: 14730714]

Theodoris CV, et al. , 2017. Long telomeres protect against age-dependent cardiac disease caused by NOTCH1 haploinsufficiency. J. Clin. Invest 127, 1683–1688. [PubMed: 28346225]

Thonhoff JR, Simpson EP, Appel SH, 2018. Neuroinflammatory mechanisms in amyotrophic lateral sclerosis pathogenesis. Curr. Opin. Neurol 31, 635–639. [PubMed: 30048339]

Tsakiri EN, et al. , 2013. Proteasome dysfunction in Drosophila signals to an Nrf2-dependent regulatory circuit aiming to restore proteostasis and prevent premature aging. Aging Cell 12, 802–813. [PubMed: 23738891]

Tsvetkov AS, et al. , 2013. Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration. Nat. Chem. Biol 9, 586–592. [PubMed: 23873212]

Tuncbag N, et al. , 2016. Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. PLoS Comput. Biol 12, e1004879. [PubMed: 27096930]

Udler MS, Tyrer J, Easton DF, 2010. Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. Genet. Epidemiol 34, 463–468. [PubMed: 20583289]

Ullrich NJ, Gordon LB, 2015. Hutchinson-Gilford progeria syndrome. Handb. Clin. Neurol 132, 249–264. [PubMed: 26564085]

Uricchio LH, 2020. Evolutionary perspectives on polygenic selection, missing heritability, and GWAS. Hum. Genet 139, 5–21. [PubMed: 31201529]

Venet D, Dumont JE, Detours V, 2011. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput. Biol 7, e1002240. [PubMed: 22028643]

Vilchez D, et al. , 2012. Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. Nature 489, 304–308. [PubMed: 22972301]

Visscher PM, et al. , 2017. 10 years of GWAS siscovery: Biology, function, and translation. Am. J. Hum. Genet 101, 5–22. [PubMed: 28686856]

Wainger BJ, Cudkowicz ME, 2015. Cortical hyperexcitability in Amyotrophic Lateral Sclerosis: C9orf72 repeats. JAMA Neurol. 72, 1235–1236. [PubMed: 26348624]

Waller L, Tian L, 2015. Computational imaging: Machine learning for 3D microscopy. Nature 523, 416–417. [PubMed: 26201593]

Wang K, Li M, Hakonarson H, 2010. Analysing biological pathways in genome-wide association studies. Nat. Rev. Genet 11, 843–854. [PubMed: 21085203]

Wang ZY, Fu LY, Zhang HY, 2012. Can medical genetics and evolutionary biology inspire drug target identification? Trends Mol. Med 18, 69–71. [PubMed: 22172275]

Wang T, Wei JJ, Sabatini DM, Lander ES, 2014. Genetic screens in human cells using the CRISPR-Cas9 system. Science (New York, N.Y.) 343, 80–84.

Wang T, et al. , 2015. Identification and characterization of essential genes in the human genome. Science (New York, N.Y.) 350, 1096–1101.

Wang C, et al. , 2018. Gain of toxic apolipoprotein E4 effects in human iPSC-derived neurons is ameliorated by a small-molecule structure corrector. Nat. Med 24, 647–657. [PubMed: 29632371]

Watanabe K, Taskesen E, van Bochoven A, Posthuma D, 2017. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun 8, 1826. [PubMed: 29184056]

Webb S, 2018. Deep learning for biology. Nature 554, 555–557.

Weinstein JA, Regev A, Zhang F, 2019. DNA microscopy: Optics-free spatio-genetic imaging by a stand-alone chemical reaction. Cell 178, 229–241. [PubMed: 31230717]

White MJ, et al. , 2019. Strategies for pathway analysis using GWAS and WGS data. Curr. Protoc. Hum. Genet 100, e79. [PubMed: 30387919]

Wingo TS, Cutler DJ, Yarab N, Kelly CM, Glass JD, 2011. The heritability of amyotrophic lateral sclerosis in a clinically ascertained United States research registry. PloS one 6, e27985. [PubMed: 22132186]

Wren MC, et al. , 2015. Frontotremporal dementia-associated N279K tau mutant disrupts subcellular vesicle trafficking and induces cellular stress in iPSC-derived neural stem cells. Mol. Neurodegen 10, 46.

Wu Y, et al. , 2019. Colocalization of GWAS and eQTL signals at loci with multiple signals identifies additional candidate genes for body fat distribution. Hum. Mol. Genet 28, 4161–4172. [PubMed: 31691812]

Xi Y, Noble S, Ekker M, 2011. Modeling neurodegeneration in zebrafish. Curr. Neurol. Neurosci. Rep 11, 274–282. [PubMed: 21271309]

Xie Z, et al. , 2015. Early telomerase inactivation accelerates aging independently of telomere length. Cell 160, 928–939. [PubMed: 25723167]

Yang SJ, et al. , 2018. Assessing microscope image focus quality with deep learning. BMC Bioinform. 15, 77.

Yeger-Lotem E, et al. , 2009. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. Nat. Genet 41, 316–323. [PubMed: 19234470]

Zanette G, et al. , 2002. Different mechanisms contribute to motor cortex hyperexcitability in amyotrophic lateral sclerosis. Clin. Neurophysiol 113, 1688–1697. [PubMed: 12417221]

Zhang F, Wen Y, Guo X, 2014. CRISPR/Cas9 for genome editing: progress, implications and challenges. Hum. Mol. Genet 23, R40–R46. [PubMed: 24651067]

Zhang W, et al. , 2015. Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. Science (New York, N.Y.) 348, 1160–1163.

From Disease-Related Genetic Association to Causal Genes, Pathways, Mechanisms and Therapeutic Targets
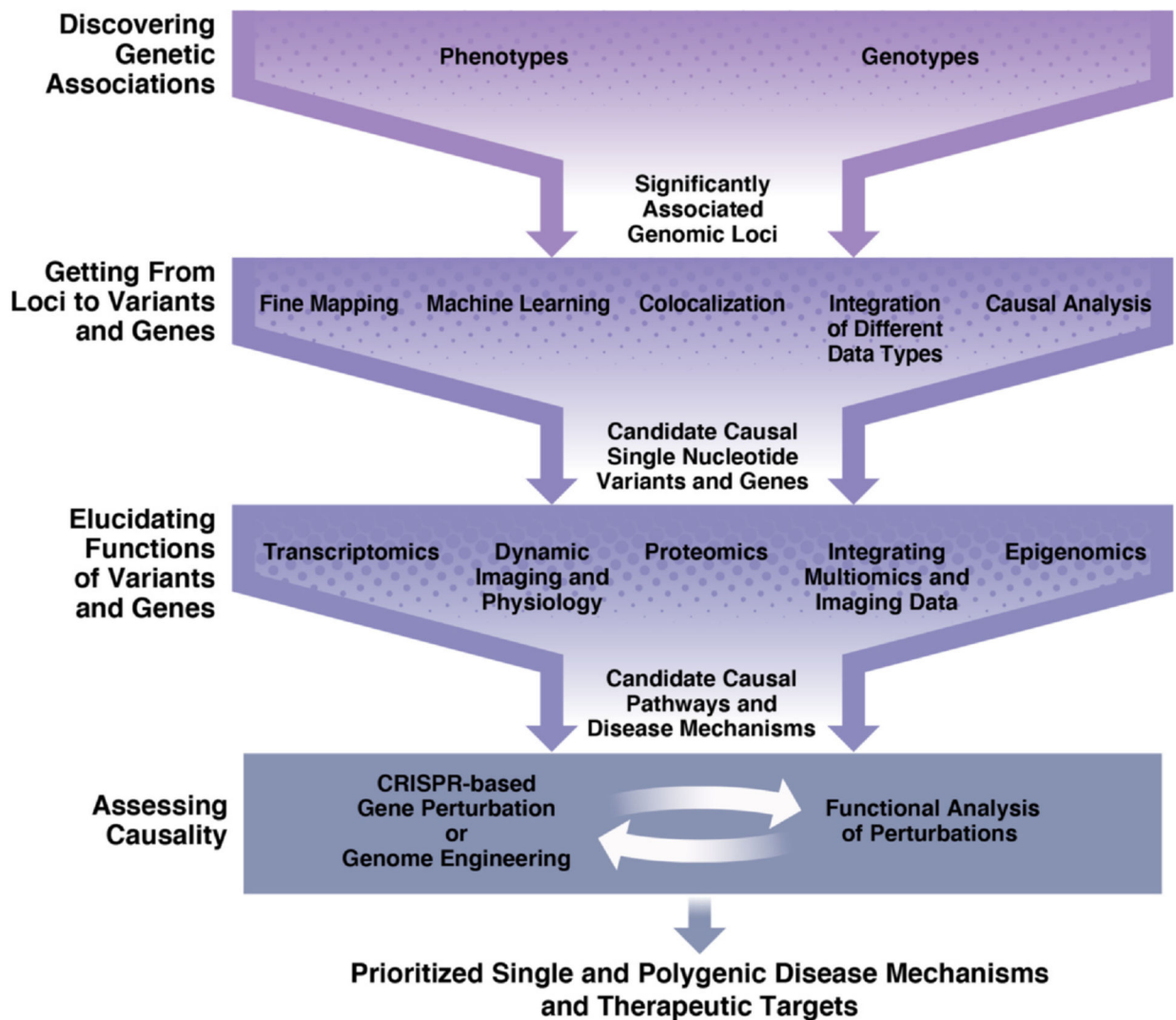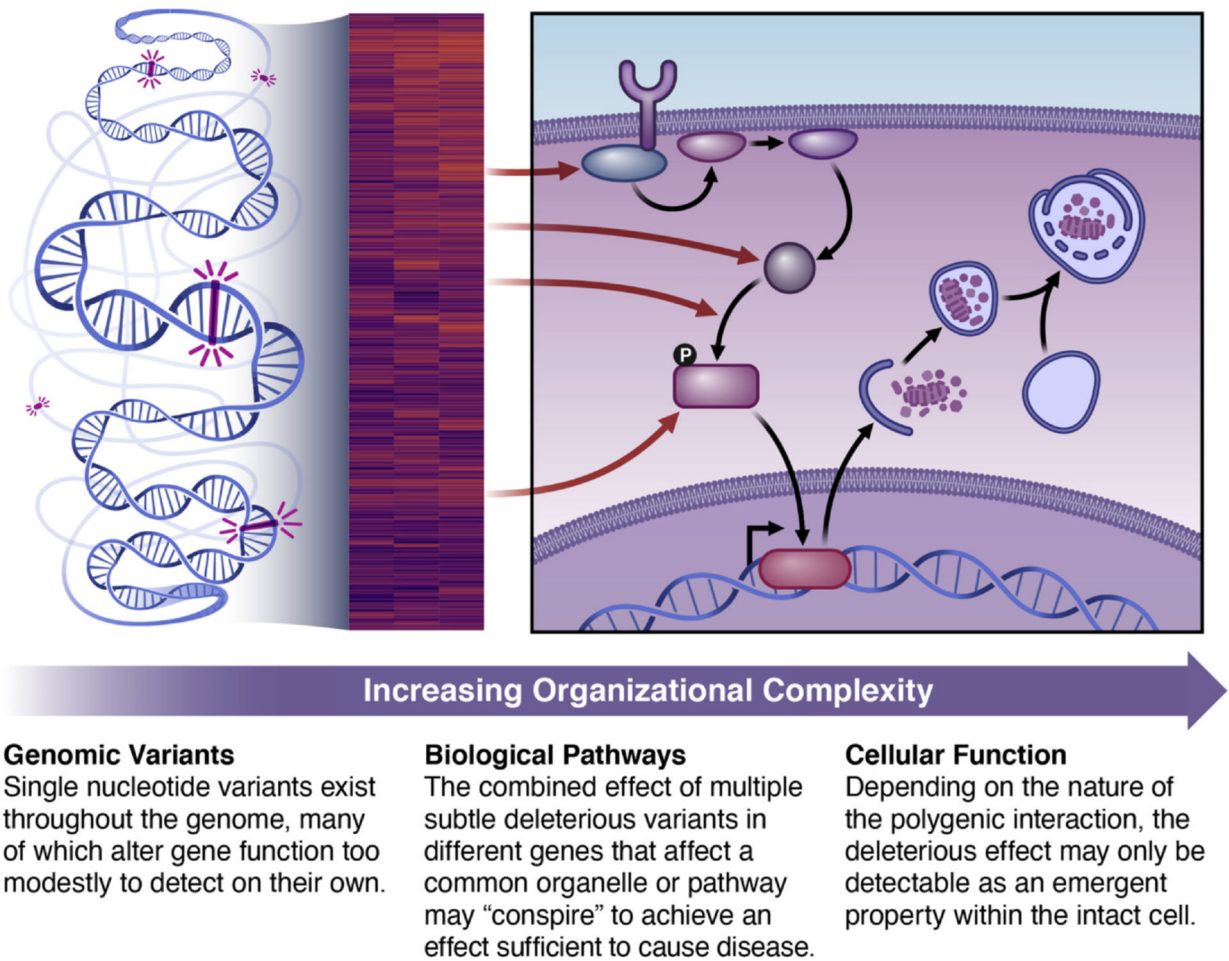

Fig. 1.

A Work Flow for Functional Genomics for Neurodegenerative Disease. The schematic delineates a series of steps and approaches available to investigators to use findings from human genetics studies, such as genome-wide association studies, and to elucidate the underlying mechanisms that explain the observed genetic associations. Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeats.

Increasing Organizational Complexity

Genomic Variants
Single nucleotide variants exist throughout the genome, many of which alter gene function too modestly to detect on their own.

Biological Pathways
The combined effect of multiple subtle deleterious variants in different genes that affect a common organelle or pathway may "conspire" to achieve an effect sufficient to cause disease.

Cellular Function
Depending on the nature of the polygenic interaction, the deleterious effect may only be detectable as an emergent property within the intact cell.

Fig. 2.
Neurodegenerative Disease as an Emergent Property of a Complex System. The heritable risk of developing amyotrophic lateral sclerosis, Alzheimer and Parkinson disease is much larger than what can be explained by single disease-causing mutations. This suggests that multiple genetic variants may act in combination to confer significant risk of disease. Conventional approaches that rely on measuring effects of individual variants one-at-a-time, such as genome-wide association studies, may be too insensitive to detect their effects. Standard OMICs methodologies can provide a more comprehensive view of the state of a biological system, but are still typically limited in the range of macromolecules they can detect and to the measurement of one type of macromolecule at one point in time. If different variants act at different levels (e.g., transcription, chromatin conformation, translation, post-translational, metabolism, etc.), and are subject to complex dynamic feedforward and feedback relationships, it may be difficult to detect combinatorial effects with a single or even a series of OMICs techniques. In that scenario, pathogenesis may be better understood as an emergent property of a complex system that might be best detected with methods, such as imaging, that are suited to the dynamic study of intact live cells.