Fundamental Research 2 (2022) 918–928

Contents lists available at ScienceDirect

Fundamental Research

journal homepage: http://www.keaipublishing.com/en/journals/fundamental-research/

Review

Rare and undiagnosed diseases: From disease-causing gene identification to mechanism elucidation

Gang Wang[a,b,c,1], Yuyan Xu[a,b,1], Qintao Wang[a,b,1], Yi Chai[a,b,1], Xiangwei Sun[a,b], Fan Yang[c], Jian Zhang[a,b], Mengchen Wu[a,b], Xufeng Liao[b], Xiaomin Yu[a,b,*], Xin Sheng[a,b,*], Zhihong Liu[a,b,c,*], Jin Zhang[a,d,*]

[a] Zhejiang Laboratory for Systems & Precision Medicine, Zhejiang University Medical Center, Hangzhou 311121, China
[b] Zhejiang University School of Medicine, Hangzhou 310058, China
[c] National Clinical Research Center of Kidney Diseases, Jinling Hospital, Nanjing University School of Medicine, Nanjing, Jiangsu 210002, China
[d] Center for Stem Cell and Regenerative Medicine, Department of Basic Medical Sciences, The First Affiliated Hospital, Zhejiang University School of Medicine; Center of Gene/Cell Engineering and Genome Medicine of Zhejiang Province, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 16 July 2022
Received in revised form 4 September 2022
Accepted 5 September 2022
Available online 11 September 2022

Key words:
Rare and undiagnosed diseases
Next-generation sequencing
GWAS
WGS
Multiomics
iPSC
Organoid
Genome editing

ABSTRACT

Rare and undiagnosed diseases substantially decrease patient quality of life and have increasingly become a heavy burden on healthcare systems. Because of the challenges in disease-causing gene identification and mechanism elucidation, patients are often confronted with difficulty obtaining a precise diagnosis and treatment. Due to advances in sequencing and multiomics analysis approaches combined with patient-derived iPSC models and gene-editing platforms, substantial progress has been made in the diagnosis and treatment of rare and undiagnosed diseases. The aforementioned techniques also provide an operational basis for future precision medicine studies. In this review, we summarize recent progress in identifying disease-causing genes based on GWAS/WES/WGS-guided multiomics analysis approaches. In addition, we discuss recent advances in the elucidation of pathogenic mechanisms and treatment of diseases with state-of-the-art iPSC and organoid models, which are improved by cell maturation level and gene editing technology. The comprehensive strategies described above will generate a new paradigm of disease classification that will significantly promote the precision and efficiency of diagnosis and treatment for rare and undiagnosed diseases.

1. Introduction

With advances in next-generation sequencing and the development of various multiomics approaches, our understanding of the ontology of complicated diseases, particularly those involving genetic factors, has become increasingly dependent on analyses of patients at the molecular or gene/variant level. This trend challenges the current paradigm of disease diagnosis, classification and treatment. Patients with similar disease phenotypes diagnosed with a certain “syndrome” could possess completely different disease-causing mutations; therefore, “syndrome” can be considered a collection of many diseases if defined by underlying genetic causes. These diseases include rare diseases in which the causative mutations are confirmed and undiagnosed diseases in which the causal gene has not been found. Thus, the traditional way in which a complicated syndrome is defined needs to be reshaped, and whether a syndrome can be precisely classified depends on the ability to identify genetic causes.

Genomics studies provide insights into putative causal genes, and functional studies are used to confirm the role of the gene in the disease. Due to advances in iPSC, genome editing and organoid technologies, humanized models that can faithfully recapitulate human disease pathology have become more accessible. In this review, we summarize recent progress in identifying disease-causing genetic factors in rare and complicated undiagnosed diseases and the development of state-of-the-art human disease models to elucidate causality and disease mechanisms. This progress strengthens genotype-phenotype relationship knowledge, reshapes our understanding of disease ontology and guides our exploration of precise disease treatment.

* Corresponding authors.
E-mail addresses: yuxiaomin@zju.edu.cn (X. Yu), shengxin@zju.edu.cn (X. Sheng), liuzhihong@zju.edu.cn (Z. Liu), zhgene@zju.edu.cn (J. Zhang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.fmre.2022.09.002
2667-3258/© 2022 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

2. Identification of rare disease-causing genes

Rare diseases are those that affect fewer than one in two thousand people. There are between 5,000 and 8,000 rare diseases that impact 300 million individuals globally [1]. Among these diseases, 71.9% are genetic, of which 69.9% are exclusively pediatric [1]. Diagnosis is one of the most difficult aspects of rare diseases. Consequently, identifying genes that cause rare diseases is crucial for diagnosing and treating rare diseases.

Monogenic disorders, also known as Mendelian diseases, are caused by a single genetic mutation and can be inherited in different manners. Most rare diseases are monogenic, and many disorders previously assumed to polygenic were later revealed to be heterogeneous collections of monogenic diseases [2].

Historically, linkage mapping and candidate gene resequencing were employed to confirm the mutation based on the patient's phenotype. This took decades and required substantial expenditures. Recently, the diagnosis rate has improved due to the development of next-generation sequencing (NGS) technologies. Whole-exome sequencing (WES) and whole-genome sequencing (WGS) are becoming the most effective large-scale and unbiased diagnostic technologies [3–6].

2.1. Traditional genetic testing

Despite the increasing demand for tests for candidate disease genes, Sanger sequencing remains the gold standard for mutation detection in routine diagnostics. Some classic genetic illnesses, such as cystic fibrosis, are detected via Sanger sequencing [7,8]. In contrast to high-resolution Sanger sequencing, cytogenetic technologies have emerged as integral aspects of molecular diagnostics. Common trisomies, segmental aneuploidy, and copy number variants (CNVs) can be diagnosed by cytogenetic techniques. Some enhanced techniques can even detect individual nucleotides, blurring the line between molecular biology and cytogenetics [9].

Even though standard genetic testing can identify some rare diseases, it is inefficient and resource-intensive; hence, few families have access to this treatment. In addition, the variety of mutations in the population and the limited number of study samples for rare diseases significantly impact the generalizability of standard genetic testing methods.

2.2. Modern genetic testing

The advent of NGS has changed the landscape of rare disease research [10]. Since the completion of the Human Genome Project, the cost of WGS has decreased, allowing researchers to conduct WGS studies. [11]. Approximately 85% of known Mendelian diseases affect protein-coding regions. Therefore, although WES misses information from noncoding regions, it greatly reduces the cost of sequencing and allows the detection of most pathogenic mutations. WES has been shown to identify new disease-causing genes in 60% of previously reported studies [11].

2.3. Strategies for the identification of disease-causing genes

First, WES and WGS data must be cleaned to remove low-quality data, and then genome or exome mutations are obtained through standard alignment and variant calling. Mutations are annotated using classical databases to determine whether the mutation is likely pathogenic. Depending on the sequencing depth, WES data may contain hundreds of thousands or millions of variations, whereas WGS data may contain millions or even tens of millions.

After obtaining the original mutation data, it is necessary to filter the mutations, removing those with relatively high frequency in the population and those that are synonymous and removing or retaining mutations in relevant intronic regions.

The initial filtering step can yield thousands of candidate mutations that must be further filtered based on the inheritance pattern in distinct families.

2.4. Filter based on genetic pattern

Mendelian genetic analyses based on the disease prevalence in the patient's family are necessary to filter mutations further.

2.4.1. Autosomal dominant mutations

Autosomal dominant (AD) inheritance has distinct genetic characteristics: mutations occur in the autosomal chromosomes, and both the proband and the parent or offspring will carry the same mutation. Numerous variants have been identified through WES with dominant pathogenic mutations consistent with Mendelian inheritance.

Through WES of six autoinflammatory disease families, Zhou Q *et al.* [12] found that the gene encoding protein A20, *TNFAIP3*, had a truncation mutation, resulting in haploinsufficiency. This mutation prevented NF-κB inhibition and resulted in its sustained activation, causing autoinflammatory disease. Tao P *et al.* [12] performed WES on patients with recurrent fever. They found a causative autosomal dominant mutation in the *RIPK1* gene cleavage site preventing cleavage by caspase8; the resulting constitutive activation of RIPK1 caused the recurrent fever.

2.4.2. Autosomal recessive mutations

In autosomal recessive (AR) inheritance patterns, both parents of the proband carry a pathogenic mutation but are not affected, and the proband acquires two mutations from the two parents and becomes affected. In consanguineous families, the probability of offspring having autosomal recessive disorders is greatly increased.

Mutations inherited in AR patterns are mostly related to gene loss of function (LOF). Loss of one copy of the gene does not result in LOF; LOF occurs only when both copies are lost. Through WES analysis of 160 deaf families, Bademci G *et al.* [13] identified multiple genes known to contribute to the AR inheritance pattern of deafness.

Compound heterozygosity is also an AR inheritance pattern, with the exception that various mutations in the same gene are inherited from both parents. Inherited genetic disorders resulting from compound heterozygosity are generally also LOF disorders.

For example, Zhou Q *et al.* [14] found mutations in the *CECR1* gene encoding the ADA2 protein through WES in nine patients, and most of the nine patients' inheritance patterns were compound heterozygous. Obviously, compound heterozygosity is easier to recognize than AR and AD because the number is significantly lower.

2.4.3. X-linked mutations

The X-linked inheritance pattern is reasonably straightforward to recognize. In general, family analysis can reveal X-linked phenomena in patients. For example, Sankaran V G *et al.* [15] found a mutation in the hematopoietic transcription factor *GATA1* inherited in an X-linked pattern in Diamond-Blackfan anemia (DBA) patients by WES and in the proband's family, and only the male offspring were affected.

2.4.4. De novo mutations

An extremely common form of mutation is the de novo dominant mutation. De novo mutations are mutations that are not inherited. These mutations result in protein loss of function (LOF), gain of function (GOF), dominant negative effects or haplo-insufficiency. De novo dominant mutations have the same effect on protein function as AD mutations, the only difference being that AD mutations are inherited instead of de novo.

For example, Hood, R. L, *et al.* [16] found de novo heterozygous truncating mutations in SRCAP by WES of five Floating Harbor syndrome (FHS) patients. Eight other patients were found to carry the same mutation in a later study. Moghaddas F *et al.* [17] reported a de novo mutation

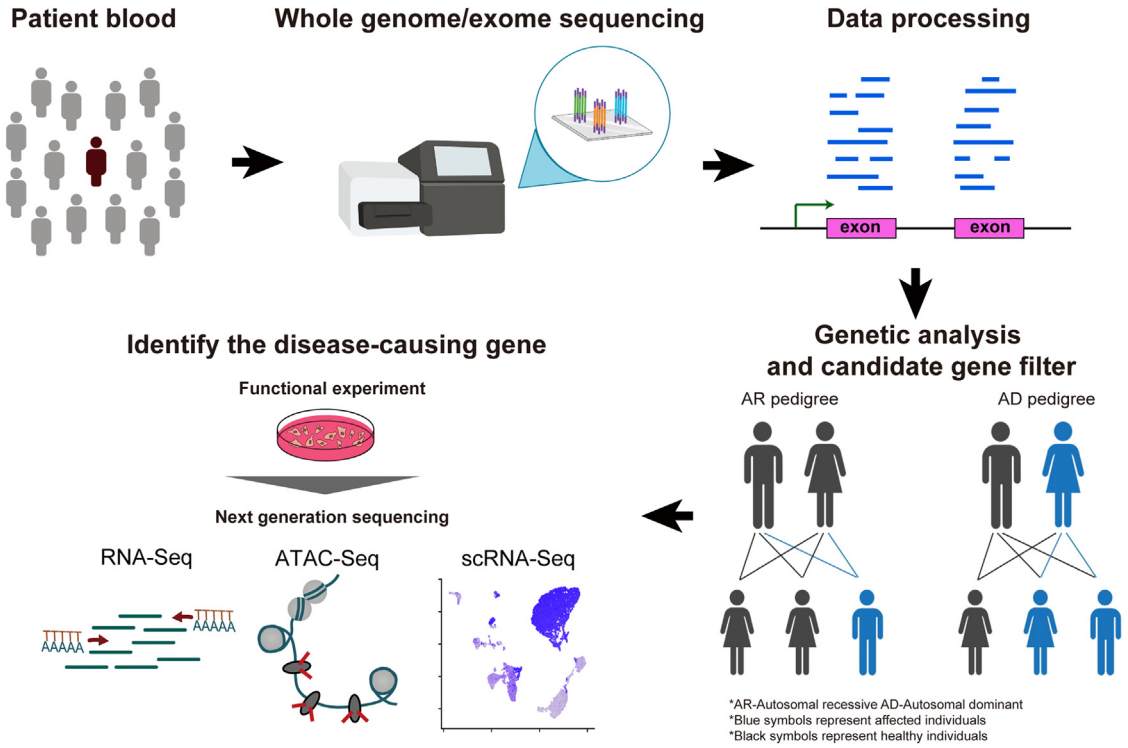

Fig. 1. Identification strategies for disease-causing genes or mutations associated with rare and undiagnosed complicated diseases.

in the MEFV gene, which affects the binding of the MEFV-encoded protein pyrin to its autoinhibitory protein 14-3-3, resulting in constitutive activation of the pyrin inflammasome.

2.4.5. Somatic mutations

Somatic mutation refers to the presence or absence of mutations in all tissues and cells of nonprobands, and NGS is advantageous over conventional techniques for identifying these mutations. Generally, AD mutation reads account for 50% of all reads, AR mutations for 100%, and somatic mutations for 1%-49%. However, due to the limitation of sequencing technology, the range can be appropriately increased ~15%.

Wang J *et al.* [18], for instance, found a somatic known pathogenic mutation in *NLRC4*. NLRC4 senses the invasion of external pathogens and is normally in a suppressed state, and mutations can lead to constitutive activation of NLRC4, causing autoinflammatory diseases.

Rare diseases can have a significant impact on patient survival and quality of life. Eighty percent of rare diseases are inherited, and the majority are caused by mutations in a single gene. Gene mutations encompass not only single base substitutions, deletions, and duplications but also structural abnormalities in vast parts of genes, deletions, duplications, and even chromosome number changes. Moreover, due to the influence of gene penetrance and epigenetics, diseases resulting from the same gene manifest differently in various patients. Therefore, genetic analysis-based clinical diagnosis is more accurate. With the rapid advancement of NGS in recent years, Sanger sequencing in conjunction with WES, WGS, RNA-Seq, scRNA-Seq, and other technologies enables the comprehensive and accurate genetic study of patients. For example, Frésard L *et al.* [19] used blood RNA–Seq to diagnose 7.5% of the 94 total cases, demonstrating the efficacy of their method.

However, acquiring precise and accurate disease-causing mutation information from NGS data is challenging. For instance, WES and WGS data contain millions of variants, and researchers must address how to find disease-causing phenotypes in these variants and how to develop more effective techniques to detect CNV and somatic mutations.

Nevertheless, with the continual improvement of diagnostic methods and accuracy of single-gene rare diseases, the constant improvement in the understanding and identification of an increasing number of rare disease mechanisms will assist a greater number of families affected by rare diseases (Fig. 1).

3. Identification of susceptible genes and mechanisms with GWAS-guided multiomics data integration approaches

3.1. GWAS improves the diagnosis and treatment of undiagnosed complicated diseases

Undiagnosed complicated diseases, such as primary immunodeficiency (PID) [20] and systemic lupus erythematosus (SLE) [21], are generally characterized by heterogeneous disease phenotypes, leading to diagnostic difficulty and underestimation of their true incidence. As most patients experience severe multisystem disorders and undergo a long and frustrating diagnostic processs [22], effective therapeutics are desperately needed. Drug targets, supported by human genetic evidence that explains the disease causation are twice likely to be approved [23,24], highlighting the critical importance of understanding disease genetics.

Multigenic diseases are influenced by accumulated contributions of hundreds of low-impact genetic variants. One of the ‘complex’ diseases, SLE, is a polygenic chronic autoimmune disease with the production of autoantibodies to nuclear antigens leading to multiple organ damage [25]. Lupus nephritis (LN) affects approximately 50% of SLE patients of Chinese descent and contributes to significant morbidity and mortality [26]. Both genetic and environmental factors heavily influence SLE. Large population-based genome-wide association studies (GWAS) or sequencing-based approaches have systematically identified thousands of disease- or phenotype-relevant genetic variants, suggesting potential risk genes for diseases. Approximately 180 susceptibility loci have been reported to be associated with SLE by GWAS, accounting for approximately 30% of genetic heritability [25]. Although only one previous study exclusively focused on identifying LN-associated genetic variants in European women [27], the inclusion of LN patients in other SLE GWASs provides important insights into potential pathogenic path-

ways contributing to both SLE and LN [26]. Susceptibility genes, such as *BLK, TNFSF4*, and *STAT4* in lymphocyte signaling; *IRF5, IKZF1*, and *TLR9* in innate immune signaling; *ACE and KLK* in intrarenal signaling; and *ITGAM and FCGR2A* in immune complex clearance, were identified as being associated with both SLE and LN [26] by GWAS.

The phenotypic heterogeneity of undiagnosed complicated diseases leads to the underestimation of their true incidence. Testing the association between genetic variants and phenotypes could provide additional insights into understanding phenotype variations contributing to disease incidence. For example, previous studies suggested an association between *IRF5* and eGFR (representing kidney function) [28], which provided indirect evidence supporting the reported association between *IRF5* and the incidence of lupus nephritis [29].

GWASs have remained largely underpowered to identify associations in rare diseases, as these diseases only occur in a smaller percentage of the general population. The association studies that focus on identifying genetic risk factors used for rare-disease diagnosis differ significantly from those focused on common complex diseases. The limited sample sizes available for each rare disease are the main challenge for GWAS. A compromise is to collectively recruit participants for GWAS according to a generalized disease definition that includes the rare disease of interest [20,30]. For instance, Thaventhiran *et al.* [20] constructed a heterogeneous cohort dominated by adult-onset, sporadic antibody-deficiency-associated PID (AD-PID), comprising CVID, combined immunodeficiency (CID), and isolated antibody deficiency to identify genetic variants associated with AD-PID. Furthermore, it is important to involve a combination of strategies to prioritize genes or variants associated with rare diseases. Methods such as genotype imputation, whole-exome sequencing (WES), and targeted sequencing are more cost-effective alternatives and could be considered [31].

3.2. Integrating multiomics data to understand disease pathogenesis

A better understanding of disease genetics translates to a more accurate prognosis and enables the provision of better support. Coding mutations could directly influence protein function by altering its amino acid composition, playing a pivotal role in disease pathophysiology. However, more than 90% of disease-associated variants identified by GWAS are located in noncoding genome regions. It remains unclear which cell types these variants are active in and how they impact specific pathways. The rapidly evolving next-generation sequencing technologies, such as RNA-sequencing and transposase-accessible chromatin sequencing, have already resulted in an increased discovery rate of susceptible genes and improved diagnosis of rare and undiagnosed complicated diseases. Integration of multiomics information would greatly assist the interpretation of GWAS data and capture the context-dependent or cell-type-dependent gene regulatory logic (Fig. 2), as regulatory SNPs might modulate gene expression through multiple epigenetic mechanisms, such as DNA methylation [32], miRNA recruitment [33], and transcription factor binding [34]. By systematic, integrated analysis of genetic and epigenetic information, Sheng *et al.* prioritized 40 loci, such as *IRF5, LACTB*, and *MUC1*, where methylation and gene-expression changes likely mediate the genotype effect on kidney function variations [32]. Hou *et al.* identified the rs2431697 variant as likely causal for SLE through disrupting a regulatory element, modulating miR-146a expression [33]. The integration of GWAS data with promoter capture Hi-C [35] (pcHi-C) data could potentially bridge common and rare variants to prioritize candidate monogenic genes for rare diseases, such as CVID [20]. Furthermore, integrating GWAS with single-cell RNA sequencing and single-nucleus assays for transposase-accessible chromatin with high-throughput sequencing data could shed light on cell-type-dependent genetic-driven epigenetic and transcriptomic changes causing disease phenotype variations. For example, Sheng *et al.* found a cell-type-specific regulatory effect of rs4292 on *ACE* in proximal tubule cells, which could cause blood pressure variations [34].

4. Elucidation of disease mechanisms with iPSC and organoid models

4.1. The advantages of establishing rare disease models based on iPSC technologies

Common models of rare diseases often rely on primary cells derived from patients or animal models. The main limitation of the former is that some primary cells are difficult to obtain, and the obtained primary cells cannot be expanded indefinitely. In addition, immortal cell lines often do not accurately reflect the state of primary cells with extended culture time in vitro, thus reducing the reliability of the primary cells used in the study. The main limitation of animal models is that different species have fundamental physiological differences, and animal models cannot fully simulate the phenotypes of some human diseases. For example, in Alzheimer's disease, the mouse model lacks many neuronal cells because of species differences between humans and mice [36]. The heart size and resting heart rate in most commonly used rodent animal models differ from those in humans [37]. Therefore, no animal model's physiological and pathological conditions are identical to those of humans. In addition, there are many rare diseases with multiple disease-causing genes. Polygenic rare disease often involves deletion, duplication or translocation of chromosomes, and it is more difficult to establish animal models of these diseases. Therefore, it is urgent to build a human disease model platform to complement animal models.

In 2006, Yamanaka reprogrammed somatic cells to a pluripotent state similar to embryonic stem cells by forcing the expression of four transcription factors (Oct4, Sox2, C-MYC, Klf4) [38] and obtained induced pluripotent stem cells (iPSCs). Subsequently, research on iPSCs has progressed rapidly, especially studies that use these cells as a model for various rare diseases. The advantages mainly lie in the fact that human iPSCs are highly similar to ESCs, not only in morphology, proliferation, surface antigen, gene expression, and epigenomic state but also in their ability to differentiate into the cell types of the three germ layers both in vitro and in teratoma [39]. iPSCs can be cultured in vitro for long periods and are essentially suitable for almost all cell biological and molecular analyses developed for "conventional" cell lines. As such, they provide a new window for studying basic gene functions and cellular processes, in addition to immortalized cell lines and in vivo animal studies. In addition, in previous studies of rare diseases, patient-specific cells were often compared with those of healthy donors. With the development of gene-editing technology, especially the development of CRISPR/Cas9 technologies, it is possible to avoid the effects of genetic background differences by engineering isogenic control cell lines.

Moreover, in 2009, Clevers and colleagues showed that stem cells resident in the adult intestine could proliferate and self-organize in vitro, leading to the concept of organoids [40]. Unlike two-dimensional models, an organoid is a three-dimensional structure formed by various cell types through self-organization and can better reflect the development and physiological process of a tissue or organ, yielding a cell composition and behavior closest to physiology. In addition, iPSCs are an unlimited source of organoids for disease modeling. Compared with animal models, organoids can reduce the complexity of experiments and are amenable to real-time imaging techniques. In summary, iPSCs can be differentiated into disease-related functional cells or organoids for constructing disease models. These models can not only reproduce the disease phenotype and allow the study of the pathology of diseases in vitro but can also be employed to conduct corresponding drug screening or gene editing to test the safety and efficacy of drugs or therapies for rare diseases (Fig. 3).

4.2. Application of organoid technology in rare diseases

Organoids are derived from iPSCs in vitro, comprising various cell types and an extracellular matrix similar to the structure and function of organs in vivo [41]. Compared with 2D cell culture, organoids can show

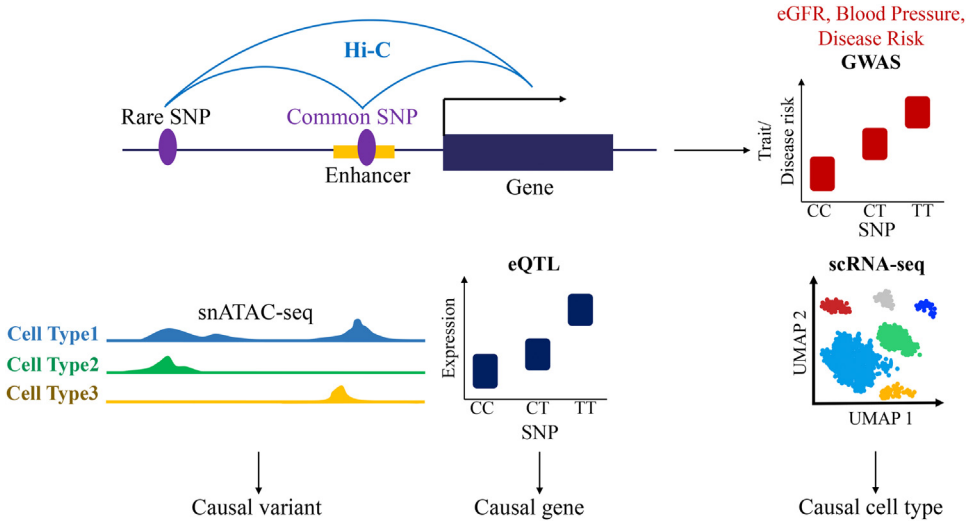

Fig. 2. Integration of multiomics information can define key disease-driving mechanisms. Genome-wide association studies (GWAS) can define traits/disease risk variants. Hi-C can bridge common and rare variants to prioritize candidate monogenic genes for rare diseases. Multiomics information, including single-nucleus open chromatin sequencing (snATAC-seq), expression quantitative trait loci (eQTL), and single-cell RNA sequencing (scRNA-seq), can prioritize causal variants, genes, and cell types responsible for rare and undiagnosed diseases.

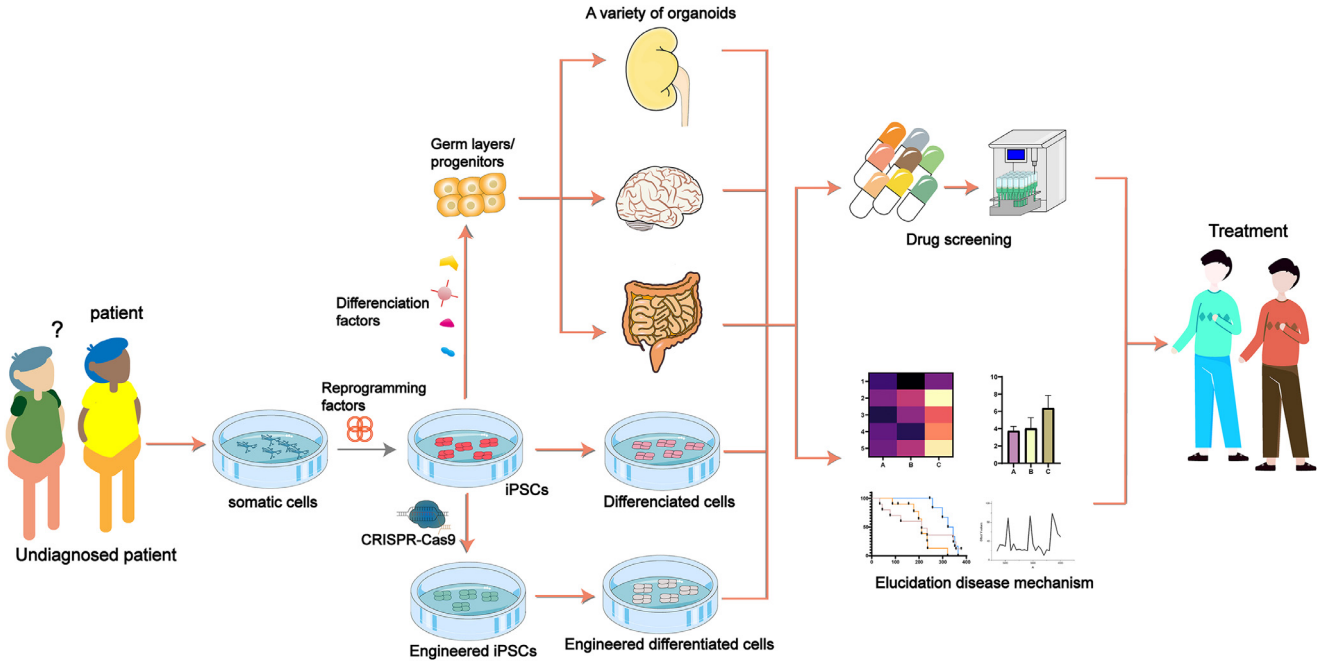

Fig. 3. iPSCs from patients with rare and undiagnosed complicated diseases can be directed to differentiate into different cell types or organoids in vitro. On this basis, combined with rapidly developed gene editing technology, we can use iPSCs and organoid models to study the mechanisms of rare and undiagnosed complicated diseases and conduct large-scale drug screening tests, leading to new treatments for patients.

a cell type composition closer to physiological conditions and reflect the interaction between different cell types and the extracellular matrix. In addition, compared with animal models, the advantage of organoids lies in their human genetic background, which can effectively reduce the deviation of experimental results caused by genetic differences among species. Because of the advantages of organoid models in vitro, the number of studies on their application for disease models is increasing. To date, organoids have been applied in studies on various organs, such as the brain [42], heart [43], and kidney [44].

Several research groups have published methods for producing brain organoids from iPSCs, including unguided and guided methods [45,46]. Brain organoids contain neural progenitor cells, neuronal cells and glial cells, which simulate the structure and function of fetal brain tissue [47]. Based on such advantages, many research groups use the brain organoids induced by patient-iPSCs to establish neuronal disease models and further study the pathogenesis. At present, brain organoid disease models are mainly used for studying early brain development diseases with obvious structural malformations, such as microcephaly and macrocephaly. The mechanism of microcephaly is usually caused by premature differentiation and reduced proliferation of neural progenitor cells, all of which can be reliably analyzed in brain organoids. The organoids induced by iPSCs of microcephaly patients model disorders of cell mitosis and a small volume of brain organoids [48]. CRISPR–Cas9 was used to establish PTEN-deficient iPSCs, which were

differentiated into brain organoids. The PTEN-deficient brain organoids reflected abnormal enlargement due to excessive proliferation of neural progenitor cells, which simulates a macrocephaly phenotype [49]. However, in addition to this obvious volume-change phenotype, the other phenotypes of brain diseases are limited to the cellular and molecular levels. Organoids derived from patients with autism spectrum disorder (ASD) showed that the ratio between excitatory and inhibitory neurons is dysregulated [50]. Organoids derived from Alzheimer's disease patients reflect amyloid deposition and abnormal endosomes [51]. Whether late-onset and aging-related diseases can be simulated by fetal-state organoids remains to be explored.

Several research groups have used the self-organization characteristics of cardiac lineage cells differentiated from iPSCs to construct cardiac organoids [52]. In addition, biocompatible scaffolds are also used to construct cardiac organoids, including cell hydrogel matrices, biomaterial-based microchambers, and 3D-printed biomaterials [53–55]. The cell type and structure of cardiac organoids are similar to those of the heart in vivo, which is very useful for studying genetic diseases. The cardiac organoid model induced by iPSCs from patients with familial cardiomyopathy can simulate insufficient myocardial contractile function [56]. Although brain and cardiac organoids have already been used to model many inherited disease phenotypes, most of these phenotypes are limited to molecular and cellular phenotypes. Cultured organoids in vitro cannot perform the physiological functions of this organ in vivo, which is the major disadvantage of cultured organoids for modeling disease phenotypes compared with animal models. How to make in vitro cultured organoids perform the physiological functions of organs is a research focus.

Induction methods for developing kidney organoids have also been established by many laboratories [57,58]. Kidney organoids include nephron progenitor cells, renal tubular cells, basement membrane, stromal cells and off-target cells. The cell types and structures are similar to those of the fetal kidney [59]. The phenotype simulation of hereditary kidney diseases using kidney organoids mainly includes polycystic structural changes and cellular and molecular phenotypes. Polycystic kidney disease (PKD) is an important cause of end-stage renal disease. PKD is classified into autosomal dominant polycystic kidney disease (ADPKD) and autosomal recessive polycystic kidney disease (ARPKD). Through gene editing technology, double allele truncation mutations of *PKD1* and *PKD2* were performed in iPSCs, which were subsequently induced into kidney organoids. Cyst formation was observed in the proximal tubules of kidney organoids [60]. NPHP is an autosomal recessive interstitial nephritis and the most common single-gene disease causing chronic kidney disease in adolescents. The primary cilia of kidney organoids carrying the *IFT140* mutation show an abnormal short pestle, while the gene-corrected organoid did not have this phenotype [61]. Congenital nephrotic syndrome is a severe hereditary nephropathy. Kidney organoids derived from congenital nephrotic syndrome patients showed that the nephrin encoded by *NPHS1* was not located on the cell surface, thus affecting the formation of the slit diaphragm [62]. The kidney maintains the balance of water, electrolytes and pH and excretes metabolic waste. The functional unit of the kidney is the nephron, which exerts glomerular filtration function and renal tubular reabsorption function. Mutations can result in various cell phenotypes and eventually lead to impairment of two functions. However, the above studies only modeled the molecular phenotype of the corresponding cell type and could not further model the changes in the phenotypes of the two renal functions in vivo. Therefore, making kidney organoids that perform the two functions in vitro needs further investigation. The development of organ-on-chip technology can provide some solutions to this problem.

Despite the great progress made in the application of organoids, major challenges remain. At present, brain, cardiac, kidney and other organoids are very simple and immature. Researchers have found that the maturity of brain and kidney organoids is close to that of fetal organs. In addition, it is accompanied by the absence of some cell types. Thus, developing mechanisms to further induce organoids to mature and make them more similar to in vivo adult organs is critical.

4.3. Significance of cell maturation for the application of iPSC organoid technology

Organoids cultured in vitro are similar to fetal organs, indicating that various organoid cell types are at an immature stage. These cell types cannot fully perform the corresponding physiological functions. These organoids are more suitable for modeling diseases in the early developmental stage. Organoid modeling is limited for diseases caused by damage to the physiological function of specific cell types in adults. Therefore, generating mature organoids that can perform the relevant physiological functions can improve the accuracy and applicability of organoids for disease modeling.

With the development of research, the concept of cell maturation has been updated accordingly. An accurate understanding of the concept of cell maturation guides the improvement of cell maturation protocols. Recently, the definition of maturity has been further explained [63]. Maturity may not be the end point of development, nor is it completely age-related, and it is not a one-way process. Maturity should be regarded as an adaptive process based on the maximum adaptation of genes and responses to the environment. Maturity characteristics can be classified as either anatomy (form, gene circuitry, and interconnectivity) or physiology (function, rhythms, and proliferation) [63].

Cell form is related to maturation. Mature red blood cells are in the shape of a biconcave disc, and the large surface area allows the cells to reform in a manner allowing them to pass through narrow capillaries. Red blood cells become spherical or crenate in hypotonic or hypertonic liquid, indicating the adaptability of their mature form to the environment [64]. Gene circuitry change is a marker of cell maturation. AMPK and mTOR signals mediate the transition from glycolysis to fatty acid oxidation, marking the maturation of cardiomyocytes [65]. Metabolic change is an adaptive state brought about by environmental factors. The highly ordered network organization (interconnectivity) among components in a complex system can determine maturity. With the maturation of podocytes in the glomerulus, the connection between the protruding podocytes forms a special filtration barrier, which becomes a slit diaphragm [66]. Exerting specific physiological functions is a sign of cell maturation. Mature pancreatic β cells selectively respond to hyperglycemia and increase insulin secretion. Selective reactions gradually appear after birth because β cells must constantly adapt to the body environment, which is rich in glucose and oxygen [67]. Rhythms are a sign of cell maturation. Mature hepatocytes adjust physiological functions according to the circadian feeding–fasting cycle [68]. Matching the 'biological clock' of mature hepatocytes associated with meal times optimizes the clearance and effect of circulating nutrients. Cell proliferation and maturation are usually negatively correlated during development. After birth, most cardiomyocytes exit the cell cycle during metabolic maturation. However, this adaptive state is reversible because quiescent cardiomyocytes can re-enter the cell cycle with injury and aging [69]. The above findings elucidate the effects of each maturity characteristic on the function of a specific cell type. However, the aforementioned six mature cell characteristics are not isolated, they are interrelated and provide support for the cells to perform physiological functions. Therefore, we can design a variety of combined environmental stimuli according to the six mature characteristics to produce cells that are more mature with characteristics more similar to those of the organ in vivo.

How can mature cells be developed? The general method is to enable cells in culture to receive stimulation factors related to development in vivo, such as chemical (nutrients, oxygen, and growth factors) and physical (mechanical, spatial, and electrical) stimuli. In addition, transplantation can also promote organoid morphological maturation.

Nutrients, oxygen and growth factors can change the state of cell maturation. Cardiomyocytes induced by iPSCs in vitro are closer to

those in the fetal state. If their metabolic environment is lipid-rich, cardiomyocytes will show more mature characteristics [70]. In addition, inhibiting hypoxia signaling by inhibiting HIF1α can promote adult-like anatomical and physiological changes in stem cell-derived cardiomyocytes [71]. Glucocorticoids can also promote the formation of specific morphological and physiological functions of stem cell-derived cardiomyocytes [72]. Mechanical, spatial, and electrical factors can change cell maturation state. When cultured on a gel resembling the rigidity and texture of adult heart tissue, stem cell-derived cardiomyocytes will demonstrate adult-like special contractility, genetic circuit and structural characteristics [73]. In addition, the differentiation and vascularization of kidney organoids can be accelerated by culturing them on hydrogels with the same stiffness as the ovo chorioallantoic membrane microenvironment [74]. Electricity is also an important factor. Mature morphology can be induced in neonatal cardiomyocytes cultured in vitro after stimulation with an electric field [75]. Inducing kidney organoids under flow conditions can result in substantial vascularization and maturation [76]. Intercellular contact and diffusible signals enable cells to adapt to the composition of the environment. Culture in a three-dimensional environment promotes tissue and cell maturation, but further maturation may require proximity or contact with supporting tissues. Therefore, transplanting organoids cultured in vitro into mice will induce maturation due to the corresponding support. Studies have shown that after kidney organoids are implanted subcutaneously in mice, their vascularization increases and the morphology of the glomerulus becomes more mature [77]. This ectopic transplantation can only promote morphological maturation. Studies have shown that after transplanting kidney organoids into the renal capsule of NSG mice, their maturity will further increase, resembling the state of the adult kidney [78]. The glomerulus forms a mature filtration barrier. The podocytes form a silt diaphragm. Also, transplanted kidney organoids can establish a blood circulation connection with mouse kidneys and exercise the function of renal filtration and reabsorption. Thus, transplanted kidney organoids more accurately represent kidney diseases. Based on the above studies, screening different physical and chemical stimuli and formulating joint stimulation strategies could make organoids more mature in vitro. Then, organoids could be transplanted into the corresponding organs of NSG mice, which induced vascularization in the organoids. We believe that the maturation level of organoids can be improved by combining in vivo and in vitro stimulation techniques.

Based on the above discussion, great progress has been made in using cultured organoids induced from patient-derived iPSCs to model disease phenotypes, but this application is limited by immaturity. Therefore, we should improve these models further to make them more mature and closer to the physiological function state in vivo. A more mature model could reflect a more accurate disease phenotype. Thus, using a mature model to explore the mechanisms of disease and screen drugs will produce more reliable data.

4.4. Establishing rare disease models based on gene editing technology

The differences among human cell lines due to genetic heterogeneity between patients and healthy controls has always been a limitation of the further application of iPSC technology, and the emergence of CRISPR–Cas9 (clustered regularly interspaced short palindromic repeats (CRISPR) associated nuclease 9) technology can largely solve this problem [79,80]. Compared to previous gene editing tools, including zinc finger nucleases and transcription activator-like effector nucleases, the Cas9 nuclease is guided to the genomic sequence of interest to generate a DSB by a guide RNA sequence (gRNA), with higher levels of target specificity and activity [81–83]. In the field of rare diseases, the main benefit of CRISPR–Cas9 technology is the ability to introduce pathogenic mutations into wild-type iPSCs through gene editing and eliminate such mutations in patient iPSCs, or vice versa, thereby creating genetic control for iPSC-based disease modeling. These two complementary strategies can test whether a genetic variant is necessary or sufficient for disease-relevant cellular phenotypes. Numerous examples of genetic engineering of iPSCs using this system exist, including the generation of isogenic cell lines with specific mutations. Gene editing using CRISPR–Cas9 could provide potential genetic and pathological insights into rare diseases. By correcting mutations associated with various kinds of rare diseases in patient-derived iPSCs, CRISPR/Cas9 has been used to verify the effects of mutations and observe phenotypic differences between patient-derived and gene-corrected iPSCs. For example, iPSC models based on gene editing technology have been used in several diseases, including amyotrophic lateral sclerosis (ALS), Hermansky-Pudlak syndrome, and Duchenne muscular dystrophy [84–86]. In 2011, Guang-Hui Liu *et al.* corrected the mutation of C>T in Hutchinson-Gilford progeria syndrome (the formation of HGPS is caused by a dominant-negative C-G to T-A mutation (c.1824 C>T; p. G608G) in LMNA, the gene that encodes nuclear lamin A. This mutation causes a new form of RNA splicing that produces progerin (a toxic protein that induces premature aging [87,88]) in patient-derived iPSCs using a helper adenovirus vector (HDAdV), thus solving the problem of exogenous proteins and reducing the possibility of off-target effects [89]. Helper-dependent adenoviral (HD-Ad) vectors are based on adenovirus serotype 5 with all viral coding sequences deleted [90]. In 2016, David R. Liu and his team developed a single clip editor (ABE) that can convert A–T in target gene loci into G—C [91]. Subsequently, they further upgraded ABE to integrate dCas9 (a catalytically inactive version of Cas9, termed dead Cas9 (dCas9) with TadA (adenine deaminase), which can convert A-T base pairs into G-C base pairs without causing DNA double-strand breaks, achieving site-directed repair of genomic point mutations [92]. Using this technique, their team corrected mutations in HGPS mouse models and patient-derived fibroblasts to reduce progerin protein expression and prolong the lifespan of the model mice [93]. The combination of CRISPR–Cas9 and organoid technology further broadens the application of organoid technology in many aspects. Human intestinal organoids derived from CF patients containing F508del caused the misfolding of CFTR channel protein, which led to its rapid degradation. This mutation was accurately corrected to the normal sequence by CRISPR–Cas9. Engineered organoids with corrected CFTR amino acid sequences demonstrated CFTR channel activity in vitro, demonstrating a causal relationship between mutations and disease phenotypes and the possibility of using similar strategies to generate autologous organoids for transplantation [94].

In general, there are many ways to realize gene perturbation based on CRISPR tools, including gene editing-based homologous directed repair, gene disruption based on small deletions introduced by nonhomologous end-joining, base editors and prime editing. The first two rely on the Cas9 protein to produce DNA double-strand breaks, while the latter involve dCas9 or Cas9 nickase, in which one of two catalytic sites is inactivated, resulting in cleavage of only one of the DNA strands [91]. In addition, the CRISPR-Cas system can be used to temporarily interfere with gene function without changing the genome sequence, mainly referring to CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) [95–97]. dCas9 recruits transcription suppressor structures to gene transcription start sites to inhibit gene transcription, referred to as CRISPR interference. The dCas9–sgRNA complex can also serve as a recruitment platform for transcriptional activator domains of specific transcription start sites in the genome, effectively creating a highly specific custom transcription factor. This approach is referred to as CRISPRa. CRISPR-based approaches can be used for gene editing by targeting specific genes with sgRNA-enabled genome-wide screening from the scaling of CRISPR-based gene perturbation, as mentioned above. Such screening methods combined with iPSC and a screening method based on single-cell sequencing or phenotyping by survival analysis or FACS can be used to query the causal relationship between a large number of genes (such as all ~ 20,000 human genes encoded by proteins) and rare diseases and disease phenotypes. Thus, these approaches have the potential to provide insights into the mechanisms through which a disease variant

affects relevant cellular function and reveal potential therapeutic targets. Several rare diseases, including Parkinson's disease, Huntington's disease and amyotrophic lateral sclerosis, are all neurodegenerative. A major obstacle to developing effective therapies for neurodegenerative diseases is our incomplete understanding of the underlying cellular mechanisms. Genetic screening of neurons derived from human induced pluripotent stem cells could shed light on these mechanisms. Genome-wide screening using CRISPRi and CRISPRa provides complementary biological insights and may reveal potential therapeutic targets [98]. Using CRISPRi- and CRISPRa-based screening platforms, Martin Kampmann and his team conducted genome-wide screening of human neurons from iPSCs, revealing pathways that regulate neuronal responses to chronic oxidative stress, which are associated with several of the neurodegenerative disorders mentioned above [99]. For example, they found that deletion of the PSAP gene, which encodes prosaponins in lysosomes, triggers lipofuscin formation, increases ROS levels, leads to increased lipid peroxidation and triggers ferroptosis in neurons [99]. In addition, using this screening platform can verify the causal relationship between genes and disease phenotypes. For example, mutations in FBXO7, a gene associated with Parkinson's disease, lead to complex respiratory damage and ROS production, consistent with the results of a previous study [99]. Many rare diseases, particularly neurodegenerative diseases, are characterized by the selective vulnerability of particular cells. For example, the mutated genes that cause ALS are not exclusively expressed in motor neurons, but only motor neurons show vulnerability, suggesting that specific factors in the motor neuron cell environment make them more susceptible to ALS disease-causing genes. Thus, the selective vulnerability of specific neuron subtypes is disease-specific. Elucidating the mechanisms of selective vulnerability is of great significance for studying the disease mechanisms and revealing potential therapeutic targets [100].

In addition to gene editing technology playing a role in building rare disease models, some researchers have even proposed the direct use of CRISPR technology for treating rare diseases. A recent CRISPR-based in vivo therapy has been shown to effectively reduce the serum concentration of misfolded transthyretin protein, thereby alleviating the symptoms of ATTR amyloidosis [101]. This finding greatly increases our confidence that CRISPR technology can be used as a potential therapy for rare diseases with a one-time, systematic delivery that produces long-lasting effects. However, the ethical issues associated with this process and the possibility of off-target effects are still worth considering.

4.5. Drug screening for rare diseases based on patient-derived iPSCs

The greatest challenge in drug discovery is building the right model. At present, the most commonly used model is the animal model, but it not only involves ethical problems but also has genetic background and physiological and metabolic differences from the human model. It takes a long time to build the animal model and test the efficiency of the drugs due to the growth cycles of the animal itself. In addition, many rare diseases are multigenic, leading to substantial individual differences even among patients with the same type of rare diseases. However, patient-derived iPSCs can be used as a model to screen out targeted drugs and realize truly individualized treatment. Moreover, these iPSC-based drug screening platforms provide a new tool for the toxicological effects and side effects of small molecule drugs on humans.

For example, Parkinson's disease is a rare nervous system disorder. The most important pathologic change in Parkinson's disease is the progressive loss of A9 dopaminergic neurons in the substantia nigra, which leads to the loss of dopamine and the dysregulation of fine motor control in the basal ganglia. The clinical manifestations of Parkinson's disease, including bradykinesia, muscle rigidity and rest tremor, pathologically involve the presence of Lewy body aggregates comprised of α-synuclein [102,103]. Most patients with Parkinson's disease have sporadic disease; less than 10% have a family history of the disease, resulting in great heterogeneity. Genetically, PD has been linked to multiple genes, including *SCNA, LRRK2, PINK1, PARK2* (*parkin*), *GBA1, DJ-1, PARK9* (*ATP13A2*), *SJ-1* and *VPS35*. Therefore, further development of neurons and neural organs from patient-derived iPSCs represents a new modeling approach for studying PD pathology and testing drugs. Common iPSC models of PD include α-synuclein (SNCA) models, LRRK2 models, PINK1 and Parkin models, GBA models and other models of sporadic cases. Mutations in *SCNA* and *LRRK2* lead to autosomal dominant forms of familial Parkinson's disease, and mutations in *PINK1, Parkin*, and *GBA1* are associated with autosomal recessive inheritance and early-onset Parkinson's disease [104–107]. Many studies have shown that the iPSC model of PD can be used for drug development. IPSC-derived neurons (iPSC-DA neurons) from PD patients with GBA mutations demonstrated reduced protein levels and GCase activity. Researchers using the high-content screening (HCS) method found that NCGC758 and NCGC607, two small molecular chaperones that bind to mutant GBA in various PD models of iPSC-DA neurons, could restore glucoside esterase activity and reduce substrate accumulation in lysosomes by promoting correct folding of GBA [108,109]. In addition, quinazoline compounds are known to be GCase inhibitors. Many other drugs have been tested on the iPSC-DA model. For example, isoxasole could rescue α-synuclein A53T iPSC-DA neurons from nitrosative stress by increasing mitochondrial respiration and biosynthesis through the MEF2C-PGC1α signaling pathway [110]. Coenzyme Q, rapamycin, and the LRRK2 kinase inhibitor GW5074 have also been shown to improve the mitochondrial function of iPSC-DA neurons [111]. Based on the iPSC model, small molecule drugs have been screened for many other rare diseases. Alessandra *et al.* established an iPSC cell line from patients with Marfan syndrome (MFS) and further differentiated it into MFS-IPSC-derived smooth muscle cells. In addition to finding that inhibiting the TGF-β signaling pathway can rescue the expression of fibrillin-1 accumulation and matrix metalloproteinase, they also identified new therapeutic targets (P38 and KLF4) for MFS and provided an innovative human platform for the testing of new drugs [112]. Based on Hutchinson-Gilford progeria syndrome (HGPS) iPSC-based vascular smooth muscle cells (VSMCs) and mesenchymal stem cells (MSCs), S Blondel *et al.* identified several compounds referred to as monoaminopyrimidines by testing 21,608 small molecules using a new screening method. These compounds target two key enzymes of the farnesylation process, farnesyl pyrophosphate synthase and farnesyl transferase, and rescue in vitro phenotypes associated with HGPS [113]. Similarly, through HCS, Alessandra Lo Cicero and colleagues identified seven compounds that promote osteogenic differentiation of HGPS-IPSC-derived MSCs, and among these, all-trans retinoic acid and 13-cis-retinoic acid can reduce the expression of progerin, which has the potential to treat HGPS [114].

As mentioned above, with the development of 3D organoid technology, researchers have started to use this platform for drug screening research, even though there are still some limitations. Taking Parkinson's disease as an example, first, 3D neuronal organoids are highly heterogeneous, even when derived from the same cell line [115,116]; Second, it takes much longer and is more expensive to culture 3D neuronal organoids than 2D organoids. Third, the current progress in organoid technology has been in cortical and forebrain organoids, while research on the differentiation of midbrain organoids remains scarce. However, recent studies have shown that cultured hiPSCs can differentiate into a large multicellular organoid-like structure that contains distinct layers of neuronal cells expressing characteristic markers of the human midbrain [47]. This 3D-like midbrain organoid can form electrically active and functionally mature midbrain dopaminergic neurons and melanin, unlike mouse models [47]. Additionally, since the entire process of organoid maturation takes place in vitro, it may provide insights into developmental characteristics related to disease progression. As neuronal organoids are aggregates of different brain regions, they provide an externalizing model for the association between brain regions. Therefore, compared with other models, neuronal organoids are more suitable for screening therapeutic drugs. It has been shown that the midbrain-like organs derived from LRRK2-G2019S iPSCs exhibit upregulation of

the thiol-oxidoreductase TXNIP and an increase in the floor during the development plate marker FOXA2 [117,118]. Silvia *et al.* optimized organoid culture conditions by 3D microfluidics and found that LRRK2 inhibitor 2 (Inh2) restored the dopaminergic phenotype with organoids [119].

Many other rare diseases have been studied in vitro with three-dimensional organoid models. IPSCs derived from patients with cystic fibrosis (an autosomal linear genetic disorder caused by inactivation of the transmembrane regulator *CTFR* gene mutation) can differentiate into epithelial organoids, cholangiocyte-like cells (CLCs), and pancreatic organs [94,120,121]. Using these organoid models, pathological phenotypes of cystic fibers can be reproduced in vitro, including defective expression of mature CFTR protein, impaired chloride channel activity, and rapid swelling upon opening of the CFTR channel in a cyclic adenosine monophosphate (cAMP)-dependent manner through the addition of forskolin (FSK). In addition, these organoids can be used as drug screening models in vitro. Using CLCs, Fotios *et al.* found that the experimental CF drug VX809 could rescue the phenotype of CF-CLCs [120]. Beekman and colleagues established an organoid biobank derived from the rectum of 71 CF patients with 28 different CFTR genotypes. Based on this organoid sample bank, they developed a personalized medicine approach. FSK-induced swelling tests have been used to screen clinical patients who respond to CF modulators. Two patients with a rare and unidentified F508del/G1249R phenotype were found to respond in vitro to ivacaftor, a specific regulator of CF [122]. In summary, patient-derived iPSCs and iPSC-derived organoids have become the new platform for drug screening in the human system.

5. Conclusion and future perspective

In summary, we provided a review of (1) the identification of various types of disease-causing mutations for rare diseases, (2) the implications of GWAS-guided multiomic data integration in causal gene identification of undiagnosed complicated diseases, and (3) the key functions of disease models used to demonstrate the causality of associated mutations, study mechanisms and test new drugs or therapeutics. In the future, more advanced technologies will further promote our understanding of rare and undiagnosed complicated diseases and their treatment. For instance, the third generation of sequencing will reveal more genomic structure variations, the artificial intelligence approach will identify many potential candidate causal genes, and the state-of-the-art organoid models will be employed in more contexts for subtle disease phenotype modeling or multiorgan disease modeling. Beyond redefining disease classification, this new paradigm will also transition the way of drug development. “From-human-to-human”, namely, finding targets from human genomics data, validating target functions with humanized models, and testing the efficacy and safety of new targeted drug candidates within humanized models, is an upcoming trend in drug development. Ultimately, these evolving perspectives on how we understand and treat diseases will benefit patients and improve their care.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgment

We thank Core Facilities of Zhejiang University Medical Center for the technical support. This work is supported by the National Natural Science Foundation of China (31871453, 91857116), the Natural Science Foundation of Jiangsu Province (BK20210150), the National Key Research and Development Program of China (2018YFC1005002, 2018YFA0107100, 2018YFA0107103), the Zhejiang Natural Science Foundation of China (LR19C120001) and the Zhejiang Innovation Team Grant (2019R01004).

References

[1] S. Nguengang Wakap, D.M. Lambert, A. Olry, et al., Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database, Eur. J. Hum. Genet. 28 (2020) 165–173.
[2] J. McClellan, M.C. King, Genetic heterogeneity in human disease, Cell 141 (2010) 210–217.
[3] S.B. Ng, E.H. Turner, P.D. Robertson, et al., Targeted capture and massively parallel sequencing of 12 human exomes, Nature 461 (2009) 272–276.
[4] K. Huang, De novo paradigm: the ultimate answer to the paradox in mental retardation? Clin. Genet. 79 (2011) 427–428.
[5] Y. Yang, D.M. Muzny, J.G. Reid, et al., Clinical whole-exome sequencing for the diagnosis of mendelian disorders, N. Engl. J. Med. 369 (2013) 1502–1511.
[6] K.M. Boycott, M.R. Vanstone, D.E. Bulman, et al., Rare-disease genetics in the era of next-generation sequencing: discovery to translation, Nat. Rev. Genet. 14 (2013) 681–691.
[7] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors. 1977, Biotechnology 24 (1992) 104–108.
[8] J. Brewington, J.P. Clancy, Diagnostic testing in cystic fibrosis, Clin. Chest. Med. 37 (2016) 31–46.
[9] M.R. Speicher, N.P. Carter, The new cytogenetics: blurring the boundaries with molecular biology, Nat. Rev. Genet. 6 (2005) 782–792.
[10] E.R. Mardis, A decade’s perspective on DNA sequencing technology, Nature 470 (2011) 198–203.
[11] D.C. Koboldt, K.M. Steinberg, D.E. Larson, et al., The next-generation sequencing revolution and its impact on genomics, Cell 155 (2013) 27–38.
[12] P. Tao, J. Sun, Z. Wu, et al., A dominant autoinflammatory disease caused by non–cleavable variants of RIPK1, Nature 577 (2020) 109–114.
[13] G. Bademci, J. Foster 2nd, N. Mahdieh, et al., Comprehensive analysis via exome sequencing uncovers genetic etiology in autosomal recessive nonsyndromic deafness in a large multiethnic cohort, Genet. Med. 18 (2016) 364–371.
[14] Q. Zhou, D. Yang, A.K. Ombrello, et al., Early-onset stroke and vasculopathy associated with mutations in ADA2, N. Engl. J. Med. 370 (2014) 911–920.
[15] V.G. Sankaran, R. Ghazvinian, R. Do, et al., Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia, J. Clin. Invest. 122 (2012) 2439–2443.
[16] R.L. Hood, M.A. Lines, S.M. Nikkel, et al., Mutations in SRCAP, encoding SNF2-related CREBBP activator protein, cause Floating-Harbor syndrome, Am. J. Hum. Genet. 90 (2012) 308–313.
[17] F. Moghaddas, R. Llamas, D. De Nardo, et al., A novel Pyrin-Associated Autoinflammation with neutrophilic dermatosis mutation further defines 14-3-3 binding of pyrin and distinction to familial mediterranean fever, Ann. Rheum. Dis. 76 (2017) 2085–2094.
[18] J. Wang, Q. Ye, W. Zheng, et al., Low-ratio somatic NLRC4 mutation causes late-onset autoinflammatory disease, Ann. Rheum. Dis. (2022).
[19] L. Fresard, C. Smail, N.M. Ferraro, et al., Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts, Nat. Med. 25 (2019) 911 919.
[20] J.E. Thaventhiran, H. Lango Allen, O.S. Burren, et al., Whole-genome sequencing of a sporadic primary immunodeficiency cohort, Nature 583 (2020) 90–95.
[21] D. Nehar-Belaid, S. Hong, R. Marches, et al., Mapping systemic lupus erythematosus heterogeneity at the single-cell level, Nature Immunol. 21 (2020) 1094–1106.
[22] U.-K.I. Umlai, D.K. Bangarusamy, X. Estivill, et al., Genome sequencing data analysis for rare disease gene discovery, Briefing. Bioinform. 23 (2022) bbab363.
[23] E.A. King, J.W. Davis, J.F. Degner, 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 Genetic. 15 (2019) e1008489.
[24] M.R. Nelson, H. Tipney, J.L. Painter, et al., The support of human genetic evidence for approved drug indications, Nature Genetic. 47 (2015) 856–860.
[25] E. Ha, S.-C. Bae, K. Kim, in *Seminars in Immunopathology*.1-18 (Springer).
[26] C. Mohan, C. Putterman, Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis, Nature Rev. Nephrol. 11 (2015) 329–341.
[27] S.A. Chung, E.E. Brown, A.H. Williams, et al., Lupus nephritis susceptibility loci in women with systemic lupus erythematosus, J. Am. Soci. Nephrol. 25 (2014) 2859–2870.
[28] M. Wuttke, Y. Li, M. Li, et al., A catalog of genetic loci associated with kidney function from analyses of a million individuals, Nature Genetic. 51 (2019) 957.
[29] J. Guiteras, É. Ripoll, N. Bolaños, et al., The gene silencing of IRF5 and BLYSS effectively modulates the outcome of experimental lupus nephritis, Mol. Therapy-Nucl. Acid. 24 (2021) 807–821.
[30] B.B. Sun, M.I. Kurki, C.N. Foley, et al., Genetic associations of protein-coding variants in human disease, Nature 603 (2022) 95–102.
[31] A. Sazonovs, J. Barrett, Rare-variant studies to complement genome-wide association studies, Annu. Rev. Genomic. Hum. Genetic. 19 (2018) 97–112.
[32] X. Sheng, C. Qiu, H. Liu, et al., Systematic integrated analysis of genetic and epigenetic variation in diabetic kidney disease, Proc. Natl. Acad. Sci. 117 (2020) 29013–29024.
[33] G. Hou, I.T. Harley, X. Lu, et al., SLE non-coding genetic risk variant determines the epigenetic dysfunction of an immune cell specific enhancer that controls disease-critical microRNA expression, Nature Commun. 12 (2021) 1–19.
[34] X. Sheng, Y. Guan, Z. Ma, et al., Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments, Nature Genetic. 53 (2021) 1322–1333.

[35] B.M. Javierre, O.S. Burren, S.P. Wilder, et al., Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters, Cell 167 (2016) 1369–1384 e1319.
[36] K.D. Onos, S.J. Sukoff Rizzo, G.R. Howell, et al., Toward more predictive genetic mouse models of Alzheimer's disease, Brain Res. Bull. 122 (2016) 1–11.
[37] R.L. Hamlin, R.A. Altschuld, Extrapolation from mouse to man, Circulation: Cardiovascular Imaging 4 (2011) 2–4.
[38] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, Cell 126 (2006) 663–676.
[39] K. Takahashi, K. Tanabe, M. Ohnuki, et al., Induction of pluripotent stem cells from adult human fibroblasts by defined factors, Cell 131 (2007) 861–872.
[40] T. Sato, R.G. Vries, H.J. Snippert, et al., Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche, Nature 459 (2009) 262–265.
[41] M. Li, J.C. Izpisua Belmonte, Organoids - preclinical models of human disease 380 (2019) 569–579.
[42] I. Chiaradia, M.J.N.n. Lancaster, Brain organoids for the study of human neurobiology at the interface of in vitro and in vivo 23 (2020) 1496–1508.
[43] H. Kim, R. Kamm, G. Vunjak-Novakovic, et al., Progress in multicellular human cardiac organoids for clinical applications 29 (2022) 503–514.
[44] R.J.N.r.N. Nishinakamura, Human kidney organoids: progress and remaining challenges 15 (2019) 613–624.
[45] J.G. Camp, F. Badsha, M. Florio, et al., Human cerebral organoids recapitulate gene expression programs of fetal neocortex development, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 15672–15677.
[46] H. Sakaguchi, T. Kadoshima, M. Soen, et al., Generation of functional hippocampal neurons from self-organizing human embryonic stem cell-derived dorsomedial telencephalic tissue, Nat. Commun. 6 (2015) 8896.
[47] J. Jo, Y. Xiao, A.X. Sun, et al., Midbrain-like organoids from human pluripotent stem cells contain functional dopaminergic and neuromelanin-producing neurons, Cell Stem Cell 19 (2016) 248–257.
[48] R. Li, L. Sun, A. Fang, et al., Recapitulating cortical development with organoid culture in vitro and modeling abnormal spindle-like (ASPM related primary) microcephaly disease, Protein Cell 8 (2017) 823–833.
[49] Y. Li, J. Muffat, A. Omer, et al., Induction of expansion and folding in human cerebral organoids, Cell Stem. Cell 20 (2017) 385–396 e383.
[50] H. Choi, J. Song, G. Park, et al., Modeling of autism using organoid technology, Mol. Neurobiol. 54 (2017) 7789–7795.
[51] M. Jorfi, C. D'Avanzo, R.E. Tanzi, et al., Human Neurospheroid arrays for in vitro studies of Alzheimer's disease, Sci. Rep. 8 (2018) 2450.
[52] B. Nugraha, M.F. Buono, L. von Boehmer, et al., Human cardiac organoids for disease modeling, Clin. Pharmacol. Ther. 105 (2019) 79–85.
[53] J.U. Lind, T.A. Busbee, A.D. Valentine, et al., Instrumented cardiac microphysiological devices via multimaterial three-dimensional printing, Nat. Mater. 16 (2017) 303–308.
[54] M. Tiburcy, J.E. Hudson, P. Balfanz, et al., Defined engineered human myocardium with advanced maturation for applications in heart failure modeling and repair, Circulation 135 (2017) 1832–1847.
[55] P. Hoang, J. Wang, B.R. Conklin, et al., Generation of spatial-patterned early-developing cardiac organoids using human pluripotent stem cells, Nat. Protoc. 13 (2018) 723–737.
[56] K.C. Yang, A. Breitbart, W.J. De Lange, et al., Novel adult-onset systolic cardiomyopathy due to MYH7 E848G mutation in patient-derived induced pluripotent stem cells, JACC Basic Transl. Sci. 3 (2018) 728–740.
[57] M. Takasato, P.X. Er, H.S. Chiu, et al., Generation of kidney organoids from human pluripotent stem cells, Nat. Protoc. 11 (2016) 1681–1692.
[58] R. Morizane, A.Q. Lam, B.S. Freedman, et al., Nephron organoids derived from human pluripotent stem cells model kidney development and injury, Nat. Biotechnol. 33 (2015) 1193–1200.
[59] A.N. Combes, L. Zappia, P.X. Er, et al., Single-cell analysis reveals congruence between kidney organoids and human fetal kidney, Genome Med. 11 (2019) 3.
[60] B.S. Freedman, C.R. Brooks, A.Q. Lam, et al., Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids, Nat. Commun. 6 (2015) 8715.
[61] T.A. Forbes, S.E. Howden, K. Lawlor, et al., Patient-iPSC-derived kidney organoids show functional validation of a ciliopathic renal phenotype and reveal underlying pathogenetic mechanisms, Am. J. Hum. Genet. 102 (2018) 816–831.
[62] S. Tanigawa, M. Islam, S. Sharmin, et al., Organoids from nephrotic disease-derived iPSCs identify impaired NEPHRIN localization and slit diaphragm formation in kidney podocytes, Stem. Cell Rep. 11 (2018) 727–740.
[63] J.R. Alvarez-Dominguez, D.A. Melton, Cell maturation: hallmarks, triggers, and manipulation, Cell 185 (2022) 235–249.
[64] D.E. Discher, New insights into erythrocyte membrane organization and microelasticity, Curr. Opin. Hematol. 7 (2000) 117–122.
[65] W. Hang, B. He, J. Chen, et al., Berberine ameliorates high glucose-induced cardiomyocyte injury via AMPK signaling activation to stimulate mitochondrial biogenesis and restore autophagic flux, Front. Pharmacol. 9 (2018) 1121.
[66] K. Ichimura, S. Kakuta, Y. Kawasaki, et al., Morphological process of podocyte development revealed by block-face scanning electron microscopy, J. Cell Sci. 130 (2017) 132–142.
[67] C. Huang, E.M. Walker, P.K. Dadi, et al., Synaptotagmin 4 regulates pancreatic beta cell maturation by modulating the Ca(2+) sensitivity of insulin secretion vesicles, Dev. Cell 45 (2018) 347–361 e345.
[68] L. Bideyan, R. Nagari, P. Tontonoz, Hepatic transcriptional responses to fasting and feeding, Genes Dev. 35 (2021) 635–657.
[69] O. Bergmann, R.D. Bhardwaj, S. Bernard, et al., Evidence for cardiomyocyte renewal in humans, Science 324 (2009) 98–102.
[70] C. Correia, A. Koshkin, P. Duarte, et al., Distinct carbon sources affect structural and functional maturation of cardiomyocytes derived from human pluripotent stem cells, Sci. Rep. 7 (2017) 8590.
[71] C. Gentillon, D. Li, M. Duan, et al., Targeting HIF-1alpha in combination with PPA-Ralpha activation and postnatal factors promotes the metabolic maturation of human induced pluripotent stem cell-derived cardiomyocytes, J. Mol. Cell Cardiol. 132 (2019) 120–135.
[72] S.S. Parikh, D.J. Blackwell, N. Gomez-Hurtado, et al., Thyroid and glucocorticoid hormones promote functional T-tubule development in human-induced pluripotent stem cell-derived cardiomyocytes, Circ. Res. 121 (2017) 1323–1330.
[73] T.K. Feaster, A.G. Cadar, L. Wang, et al., Matrigel mattress: a method for the generation of single contracting human-induced pluripotent stem cell-derived cardiomyocytes, Circ. Res. 117 (2015) 995–1000.
[74] E. Garreta, P. Prado, C. Tarantino, et al., Fine tuning the extracellular environment accelerates the derivation of kidney organoids from human pluripotent stem cells, Nat. Mater. 18 (2019) 397–405.
[75] M. Radisic, H. Park, H. Shing, et al., Functional assembly of engineered myocardium by electrical stimulation of cardiac myocytes cultured on scaffolds, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 18129–18134.
[76] K.A. Homan, N. Gupta, K.T. Kroll, et al., Flow-enhanced vascularization and maturation of kidney organoids in vitro, Nat. Methods 16 (2019) 255–262.
[77] A.S. Shankar, Z. Du, H.T. Mora, et al., Human kidney organoids produce functional renin, Kidney Int. 99 (2021) 134–147.
[78] C.W. van den Berg, L. Ritsma, M.C. Avramut, et al., Renal subcapsular transplantation of PSC-derived kidney organoids induces Neo-vasculogenesis and significant glomerular and tubular maturation In Vivo, Stem. Cell Rep. 10 (2018) 751–765.
[79] C. DeBoever, H. Li, D. Jakubosky, et al., Large-scale profiling reveals the influence of genetic variation on gene expression in human induced pluripotent stem cells, Cell Stem. Cell 20 (2017) 533–546 e537.
[80] H. Kilpinen, A. Goncalves, A. Leha, et al., Common genetic variation drives molecular heterogeneity in human iPSCs, Nature 546 (2017) 370–375.
[81] M. Bibikova, K. Beumer, J.K. Trautman, et al., Enhancing gene targeting with designed zinc finger nucleases, Science 300 (2003) 764.
[82] J.C. Miller, S. Tan, G. Qiao, et al., A TALE nuclease architecture for efficient genome editing, Nature Biotechnol. 29 (2011) 143–148.
[83] A. Pickar-Oliver, C.A. Gersbach, The next generation of CRISPR–Cas technologies and applications, Nature Rev. Mol. Cell Biol. 20 (2019) 490–507.
[84] C.S. Young, M.R. Hicks, N.V. Ermolova, et al., A single CRISPR-Cas9 deletion strategy that targets the majority of DMD patients restores dystrophin function in hiPSC-Derived muscle cells, Cell Stem. Cell 18 (2016) 533–540.
[85] Y. Korogi, S. Gotoh, S. Ikeo, et al., In Vitro disease modeling of Hermansky-pudlak syndrome Type 2 using human induced pluripotent stem cell-derived alveolar organoids, Stem Cell Reports 12 (2019) 431–440.
[86] A. Bhinge, S.C. Namboori, X. Zhang, et al., Genetic correction of SOD1 mutant iPSCs reveals ERK and JNK activated AP1 as a driver of neurodegeneration in amyotrophic lateral sclerosis, Stem. Cell Rep. 8 (2017) 856–869.
[87] A. De Sandre-Giovannoli, R. Bernard, P. Cau, et al., Lamin a truncation in Hutchinson-Gilford progeria, Science 300 (2003) 2055.
[88] M. Eriksson, W.T. Brown, L.B. Gordon, et al., Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome, Nature 423 (2003) 293–298.
[89] G.H. Liu, K. Suzuki, J. Qu, et al., Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs, Cell Stem. Cell 8 (2011) 688–694.
[90] T.G. Park, J.H. Jeong, S.W. Kim, Current status of polymeric gene delivery systems, Adv. Drug. Deliv. Rev. 58 (2006) 467–486.
[91] A.C. Komor, Y.B. Kim, M.S. Packer, et al., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, Nature 533 (2016) 420–424.
[92] N.M. Gaudelli, A.C. Komor, H.A. Rees, et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, Nature 551 (2017) 464–471.
[93] L.W. Koblan, M.R. Erdos, C. Wilson, et al., In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice, Nature 589 (2021) 608–614.
[94] G. Schwank, B.K. Koo, V. Sasselli, et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients, Cell Stem. Cell 13 (2013) 653–658.
[95] L.A. Gilbert, M.H. Larson, L. Morsut, et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes, Cell 154 (2013) 442–451.
[96] M.L. Maeder, S.J. Linder, V.M. Cascio, et al., CRISPR RNA–guided activation of endogenous human genes, Nature Methods 10 (2013) 977–979.
[97] P. Perez-Pinera, D.D. Kocak, C.M. Vockley, et al., RNA-guided gene activation by CRISPR-Cas9–based transcription factors, Nature Methods 10 (2013) 973–976.
[98] M. Kampmann, A CRISPR approach to neurodegenerative diseases, Trends Mol. Med. 23 (2017) 483–485.
[99] R. Tian, A. Abarientos, J. Hong, et al., Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to Ferroptosis, Nat. Neurosci. 24 (2021) 1020–1034.
[100] M. Kampmann, CRISPR-based functional genomics for neurological disease, Nat. Rev. Neurol. 16 (2020) 465–480.
[101] J.D. Gillmore, E. Gane, J. Taubel, et al., CRISPR-Cas9 In Vivo gene editing for transthyretin amyloidosis, N. Engl. J. Med. 385 (2021) 493–502.
[102] M.G. Spillantini, M.L. Schmidt, V.M.Y. Lee, et al., α-Synuclein in Lewy bodies, Nature 388 (1997) 839–840.
[103] L.V. Kalia, A.E. Lang, Parkinson's disease, The Lancet 386 (2015) 896–912.
[104] T. Kitada, S. Asakawa, N. Hattori, et al., Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism, Nature 392 (1998) 605–608.
[105] H. Shimura, N. Hattori, S.-I. Kubo, et al., Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase, Nature Genetics 25 (2000) 302–305.
[106] J. Neumann, J. Bras, E. Deas, et al., Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease, Brain 132 (2009) 1783–1794.

[107] M.E. Gegg, D. Burke, S.J.R. Heales, et al., Glucocerebrosidase deficiency in substantia nigra of parkinson disease brains, Annal. Neurol. 72 (2012) 455–463.
[108] E. Aflaki, D.K. Borger, N. Moaven, et al., A new glucocerebrosidase chaperone reduces -synuclein and glycolipid levels in iPSC-derived dopaminergic neurons from patients with gaucher disease and parkinsonism, J. Neurosci. 36 (2016) 7441–7452.
[109] J.R. Mazzulli, F. Zunke, T. Tsunemi, et al., Activation of -glucocerebrosidase reduces pathological -synuclein and restores lysosomal function in Parkinson's patient midbrain neurons, J. Neurosci. 36 (2016) 7693–7706.
[110] S.D. Ryan, N. Dolatabadi, S.F. Chan, et al., Isogenic human iPSC Parkinson's model shows nitrosative stress-induced dysfunction in MEF2-PGC1alpha transcription, Cell 155 (2013) 1351–1364.
[111] O. Cooper, H. Seo, S. Andrabi, et al., Pharmacological rescue of mitochondrial deficits in iPSC-derived neural cells from patients with familial Parkinson's disease, Sci. Transl. Med. 4 (2012) 141ra190.
[112] A. Granata, F. Serrano, W.G. Bernard, et al., An iPSC-derived vascular model of Marfan syndrome identifies key mediators of smooth muscle cell death, Nat. Genet. 49 (2017) 97–109.
[113] S. Blondel, A.L. Egesipe, P. Picardi, et al., Drug screening on Hutchinson Gilford progeria pluripotent stem cells reveals aminopyrimidines as new modulators of farnesylation, Cell Death. Dis. 7 (2016) e2105.
[114] A. Lo Cicero, A.-L. Jaskowiak, A.-L. Egesipe, et al., A high throughput phenotypic Screening reveals compounds that counteract premature osteogenic differentiation of HGPS iPS-derived mesenchymal stem cells, Sci. Rep. 6 (2016).
[115] S. Velasco, A.J. Kedaigle, S.K. Simmons, et al., Individual brain organoids reproducibly form cell diversity of the human cerebral cortex, Nature 570 (2019) 523–527.
[116] S.J. Yoon, L.S. Elahi, A.M. Pasca, et al., Reliability of human cortical organoid generation, Nat. Methods 16 (2019) 75–78.
[117] H. Kim, H.J. Park, H. Choi, et al., Modeling G2019S-LRRK2 Sporadic Parkinson's disease in 3D midbrain organoids, Stem. Cell Rep. 12 (2019) 518–531.
[118] L.M. Smits, L. Reinhardt, P. Reinhardt, et al., Modeling Parkinson's disease in midbrain-like organoids, NPJ Parkinsons Dis 5 (2019) 5.
[119] S. Bolognin, M. Fossépré, X. Qing, et al., 3D cultures of Parkinson's disease-specific dopaminergic neurons for high content phenotyping and drug testing, Adv. Sci. 6 (2019).
[120] F. Sampaziotis, M. Cardoso De Brito, P. Madrigal, et al., Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation, Nature Biotechnol. 33 (2015) 845–852.
[121] S. Simsek, T. Zhou, C.L. Robinson, et al., Modeling cystic fibrosis using pluripotent stem cell-derived human pancreatic ductal epithelial cells, Stem. Cells Transl. Med. 5 (2016) 572–579.
[122] J.F. Dekkers, G. Berkers, E. Kruisselbrink, et al., Characterizing responses to CFTR–modulating drugs using rectal organoids derived from subjects with cystic fibrosis, Sci. Transl. Med. 8 (2016) 344ra384.

Gang Wang is an attending doctor at the National Clinical Research Center of Kidney Diseases of Jinling Hospital. He got his Ph.D. from Naval Medical University in 2018. His current research interests are 1) the Establishment of an iPS cell bank for rare kidney diseases; 2) kidney organoids disease modeling.

Jin Zhang is a Principle Investigator at Zhejiang University School of Medicine. He got his Ph.D. from UCLA in 2011, and did his postdoc at Harvard Medical School from 2012 to 2017. He was also a scientist at Amgen Inc. and led the team to establish an iPSC platform for novel drug target discovery. His current research interests are 1) the molecular mechanisms of stem cell pluripotency; 2) metabolic regulation of stem cell fate; 3) iPSC disease modeling. His publications in *Cell Stem Cell, Nature Metabolism, Cell Metabolism,* etc have been cited more than 2000 times.