Next-generation sequencing in drug development: target identification and genetically stratified clinical trials

Abolfazl Doostparast Torshizi[1,‡] **and Kai Wang**[1,2,3,‡]

[1] Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[2] Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[3] Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Next-generation sequencing (NGS) enabled high-throughput analysis of genotype–phenotype relationships on human populations, ushering in a new era of genetics-informed drug development. The year 2017 was remarkable, with the first FDA-approved gene therapy for cancer (Kymriah™) and for inherited diseases (LUXTURNA™), the first multiplex NGS panel for companion diagnostics (MSK-IMPACT™) and the first drug targeting a genetic signature rather than a disease (Keytruda®). We envision that population-scale NGS with paired electronic health records (EHRs) will become a routine measure in the drug development process for the identification of novel drug targets, and that genetically stratified clinical trials could be widely adopted to improve power in precision-medicine-guided drug development.

Introduction

A large number of human diseases are known to be caused or influenced by genetic factors, which can provide potential insights into disease pathogenesis, leading to the development of novel treatment strategies [1]. For over four decades, genetic-linkage-based studies have been successfully applied to the identification of causal genetic factors in Mendelian disorders [2]. Despite all of these breakthroughs, many common diseases do not exhibit Mendelian inheritance but represent complex multifactorial inheritance patterns, making genetic-linkage-based studies less successful at capturing allelic determinants of such disorders [3]. To circumvent such limitations, genome-wide association studies (GWAS) were then widely employed to characterize susceptibility loci associated with complex phenotypes. Ever since 2005 when the first GWAS was published on patients with age-related macular degeneration [4], GWAS have been employed in a variety of human diseases and traits [5,6]. Thus far, almost 10,000 strong associations [e.g., statistically significant associations with *P*-value threshold of 5×10^{-8} excluding single-nucleotide polymorphisms (SNPs) in linkage disequilibrium] have been reported between genetic variants and one or more complex traits [6]. Among these findings, there are several examples of disease-associated genes that have been identified as being effective drug targets, such as *HMGCR* being the target of statins which is associated with serum cholesterol levels [7]. GWAS can also enable the discovery of biological pathways that confer susceptibility to diseases [8,9]. Some early examples include the confirmation of interleukin (IL)-12/IL-23 pathways in inflammatory bowel diseases [10], and the discovery of the autophagy pathway in inflammatory bowel diseases through the *ATG16L1*/*IRGM* associations [11,12]; indeed, it has now become clear that the modulation of autophagy has strong therapeutic implications for drug development [13]. Pleiotropic SNPs for multiple related diseases have also been reported in GWAS [5,6], sometimes with opposing effects [14]. For example, SNPs discovered at the *IL23R* locus were associated with several autoimmune conditions such as ankylosing spondylitis, inflammatory bowel disease and psoriasis [15–17]. Such pleiotropic genes discovered by GWAS can be useful for drug repurposing or enable basket trials that recruit patients with diverse diseases. Monoclonal

Corresponding author: Wang, K. (wangk@email.chop.edu)
‡ Prior address: Department of Biomedical Informatics and Institute for Genomic Medicine, Columbia University Medical Center, New York, NY 10032, USA.

https://doi.org/10.1016/j.drudis.2018.05.015

antibodies targeting IL-23 and/or IL-12 are already in clinical trials for the treatment of various autoimmune diseases [18–20]. A recent study demonstrated that, among well-studied indications, the proportion of drug mechanisms being supported by direct genetic evidence across the drug development pipeline can increase from 2% at the preclinical stage to 8.2% for approved drugs [21]. Therefore, human genetic studies on well-phenotyped populations can guide the selection of the best targets and indications, with a measurable impact on the successful development of new drugs.

However, GWAS have several intrinsic limitations, such as the inability to account for a large proportion of genetic variance in complex traits and the difficulty to identify casual genes or mutations from proxy markers [6,22]. These limitations can be partially attributed to the fact that most GWAS only target common SNPs, yet common and complex disorders such as schizophrenia are genetically heterogeneous with only a proportion of genetic risks caused by SNPs of small-to-moderate effect sizes [23]. Therefore, even with carefully designed exome arrays, GWAS are also unable to or inefficient at detecting loci harboring very rare variations with larger effect sizes. In fact, despite considerable and unquestionable successes of GWAS, it has become clear that many disease-associated genetic loci with therapeutic implications have not been discovered by GWAS [24]. An illustrative example is the association of hypertension drug target genes with blood pressure [25]. Despite the discovery of over 30 genes before the GWAS era, only a small fraction has been re-identified by GWAS [26]. The need for extremely large cohort sizes to capture disease-associated variants using GWAS is also a major drawback, leading to a diminishing return after a study of the first several thousand subjects has identified the strongest determinants among common variants [27]. For instance, for a long time, no large-scale GWAS were able to identify *DRD2*, a traditional target of most antipsychotic drugs in the market, to be associated with schizophrenia until the largest GWAS was kicked off involving 37,000 schizophrenia individuals and 113,000 controls [28]. The ultimate challenge of effective use of GWAS in drug development is well probed by an omnigenic model [29], where regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes. Finding the set of 'core' genes with biologically interpretable roles in disease thus becomes especially important to guide drug development efforts.

The advent of next-generation sequencing (NGS) technologies and their applications on human populations have led to the identification of a large catalog of rare and common variants [30–32]. The use of NGS on large isolated population cohorts [33] and on specific disease cohorts [34–37] has led to the discovery of a number of new disease genes. Although NGS cannot completely address the limitations of GWAS, its avoidance of proxy markers allows direct biological interpretation of rare genetic mutations in the context of disease phenotypes. In particular, NGS can be leveraged to identify genetic variations that can inactivate drug target genes. Such mutations can mimic the action of therapeutic antagonism of these targets, thus providing a means to infer possible clinical effects of antagonist drugs on such gene products.

The pharmaceutical industry is no exception in opting for NGS technologies in diverse stages of drug development, because these technologies have brought about unique opportunities to ameliorate the discovery of novel therapeutic targets [38,39], facilitate the precise design of clinical trials on target populations more likely to benefit from the treatment [40,41], and characterize novel indications of existing drugs that are approved or under development [42–44]. While acknowledging the combined use of various types of NGS-based genomic techniques (including transcriptome and methylation profiling) in drug development, in this review we will only discuss the use of DNA sequencing in the pharmaceutical industry for target discovery and for genetically stratified clinical trials. Additionally, given the existence of several reviews on the oncology space [45,46], this article will place more emphasis on inherited diseases.

Target identification

It has been clear that the wealth of human genetic information can be leveraged to identify drug targets, validate therapeutic hypotheses and predict the potential safety of inhibitory compounds aimed at molecular targets [47,48]. NGS has the potential to uncover many mutations associated with genetic diseases and identify target genes for future drug development endeavors. Compared with GWAS which rely on proxy markers for unknown causal variants or genes, NGS is gaining momentum as a means of choice for drug target identification efforts. For such purposes, a particular trend in the field is to sequence well-phenotyped populations coupled with longitudinal electronic health records (EHRs) as a test bed to identify genes associated with a variety of phenotypic traits (Fig. 1a). A pioneering example is the DiscovEHR study as a collaboration between Regeneron and Geisinger Health System [49], in which whole-exome sequencing was performed on 50,726 subjects with paired EHRs. By leveraging rich phenotype information such as lipid levels extracted from EHRs, such studies can examine associations between loss-of-function (LoF) variants in candidate drug targets and selected phenotypes of interest. Some of the associations between predicted LoFs (pLoFs) in drug targets of low-density lipoprotein cholesterol (LDL-C) levels were confirmed. For example, pLoF mutations in *NPC1L1* (the drug target of ezetimibe [50]) and *PCSK9* (the drug target of alirocumab and evolocumab) were confirmed to be significantly associated with LDL-C levels [49]. Additionally, such LoF variants linked to EHR data can uncover novel associations, such as the association of LoF variants in *CSF2RB* with basophil and eosinophil counts [49]. In another study to investigate whether and how whole-genome sequencing (WGS) data can be used to implement genomic medicine, we examined WGS data and lifetime EHRs from 300 deceased patients [51]. Among them, five carried pathogenic or likely pathogenic variants in cancer-predisposing genes (*APC*, *BRCA1*, *BRCA2*, *NF1* and *TP53*). Based on EHRs, each of the five patients had one or more different types of cancers, fully consistent with their genetic profiles. Therefore, this is a clear demonstration that the discovery of pathogenic or likely pathogenic germline mutations from population-wide WGS correlates with clinical outcome on EHRs, thus the use of WGS could have clinical impacts to improve healthcare delivery.

While recognizing the importance of population-scale genome sequencing to find associations with common diseases or traits, NGS analysis on extreme phenotypes of therapeutic implications is also a growing opportunity to discover underlying drivers of

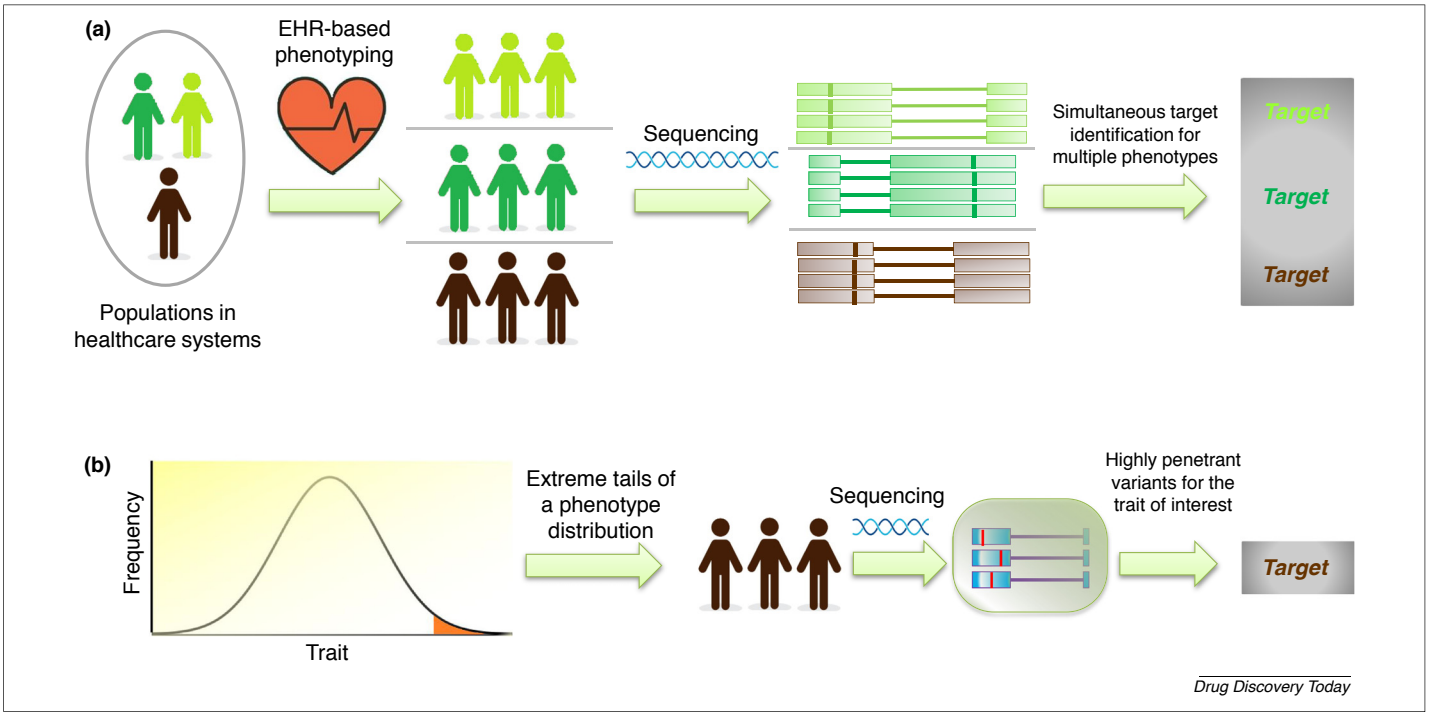

FIGURE 1

Simplified illustrations on how next-generation sequencing (NGS) helps target identification for drug development. (a) NGS on well-phenotyped populations from healthcare systems can potentially reveal phenotype-specific drug targets for multiple diseases or traits simultaneously; (b) NGS can be leveraged on extreme tails of a phenotype distribution to identify genetic targets of a specific phenotype of interest.

such phenotypes (Fig. 1b). Such extreme phenotypes can be extreme tails of quantitative traits, or sometimes on isolated individuals or families with supernatural traits. For the latter, a widely known example is that genetic mutations in the gene *LRP5* cause bones to bear extreme density (Box 1), which results in the development of therapies for osteoporosis by modulating the Wnt signaling pathway [52]. For quantitative traits, one strategy to overcome the challenge of large sample sizes to provide adequate power is to sequence individuals at both ends of a phenotype distribution. For example, a study design based on exome sequencing, where percentiles were estimated from a distribution of 1,322 early pseudomonas infection controls [53], was conducted to discover rare variants in *DCTN4* that are associated with the time to first *Pseudomonas aeruginosa* airway infection, chronic *P. aeruginosa* infection, and mucoid *P. aeruginosa* in individuals with cystic fibrosis [54]. By focusing on extreme phenotypes at both ends of a phenotype distribution, the study successfully inferred rare variants in a candidate gene with a moderate sample size [54].

The year 2017 marks an important year for gene therapy in the history of drug development, heralding a new era where genetic modulation of disease-relevant genes results in novel therapeutic strategies. In August 2017, the FDA approved the first ever chimeric antigen receptor (CAR)-T cell therapy: Kymriah™ (CTL019), for children and young adults with B cell acute lymphoblastic leukemia (ALL) that is refractory or has relapsed at least twice, ushering in a new approach to the treatment of cancer and other serious immune-related diseases. This first-in-class therapy, a collaboration between Novartis, University of Pennsylvania (Penn) and the Children's Hospital of Philadelphia (CHOP), is an innovative

BOX 1

Genetic mutations on extreme phenotypes translated to novel biological insights

Genetic causes of many rare disorders or phenotypes are uncovered by investigation of small groups of samples. For example, in 2002, scientists identified a link between *LRP5* mutations in individuals suffering from osteoporosis, a disease characterized by low bone mass. *LRP5* is a protein acting in the Wnt signaling pathway, which also acts as the target of Dkk (another developmental protein acting in the Wnt signaling pathway) [91]. Later on, in a family from Connecticut whose bone densities were supernatural, genetic mapping of 20 family members identified a mutation in *LRP5* that leads to supernatural bone density. In contrast to osteoporosis where a disruptive mutation led to the loss of bone density, seven members of this family carried an activating genetic mutation making their bones so hard that they had never experienced any kind of broken bones in their lives.

This is an example of how identification of genetic mutations causing extreme phenotypes can yield novel insights into the underlying architecture of rare diseases, opening new horizons to develop novel therapeutic agents. For drug development, it can be hypothesized that medications that mimic effects of the genetic mutation in this family have potential to increase bone density and can be further studied to develop new drugs. Although this finding was not achieved by NGS technologies, today NGS can be used to conduct similar experiments on well-phenotyped populations at an affordable cost and at outstanding speed and accuracy, providing great opportunities to identify novel drug targets from extreme phenotypes.

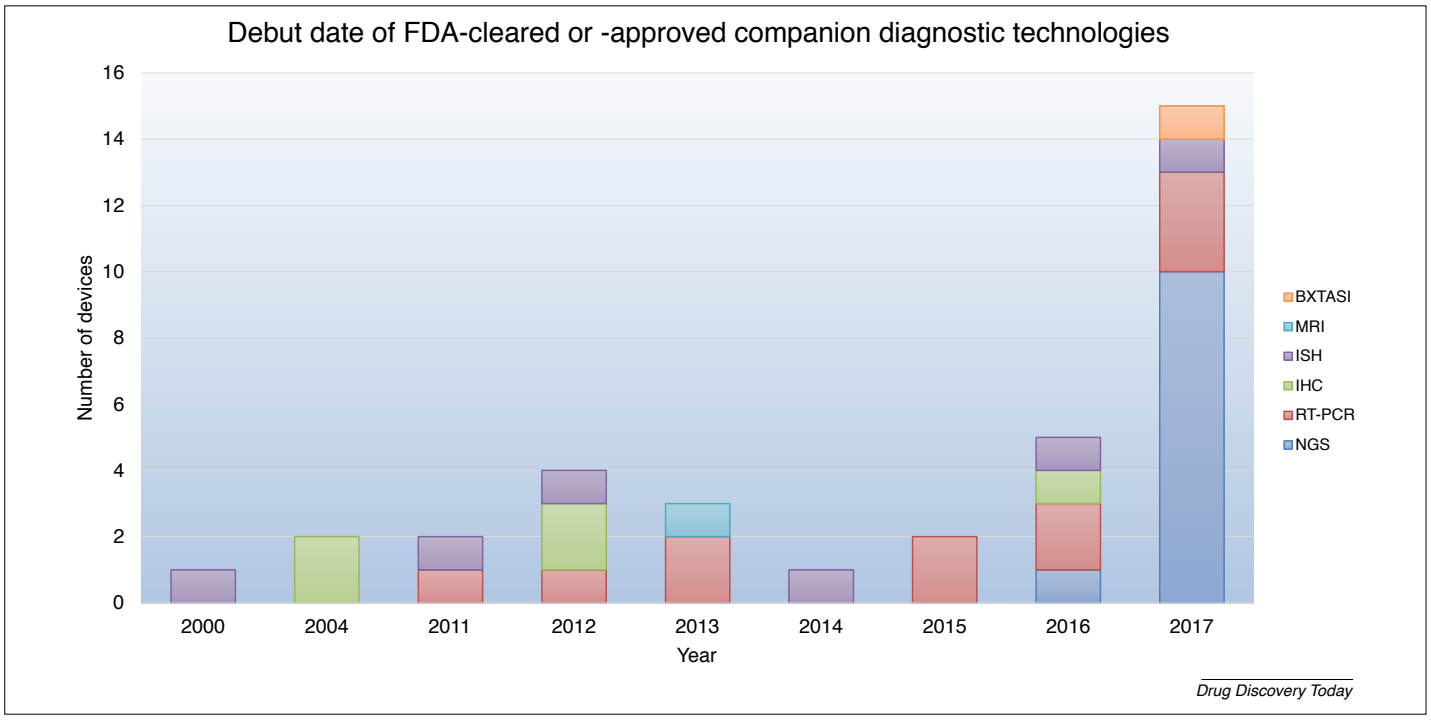

FIGURE 2

Debut date of FDA-cleared or -approved companion diagnostic devices from 2000 to 2017. Next-generation sequencing (NGS)-based companion diagnostic devices have begun to be increasingly adopted over the past two years. Abbreviations: IHC, immunohistochemistry; RT-PCR, real-time PCR; ISH, nucleotide-based *in situ* hybridization (such as fluorescence, chromogenic or dual *in situ* hybridization); BXTASI, BenchMark XT automated staining instrument; MRI, magnetic resonance imaging.

immunocellular therapy that injects an individual's T cells with DNA that encodes a CAR, which directs T cells to seek and destroy tumor cells. This new line of therapy, enabled by genetic modification, demonstrated the effectiveness, showing an 81% overall remission rate in the ELIANA trial [55], in a patient population with limited treatment options and historically poor outcomes [56]. Similarly, >50% of people treated with the second FDA-approved CAR-T therapy in October 2017, Kite Pharma's Yescarta®, are in complete remission [57]. Subsequently, in December 2017, the FDA approved the first gene therapy for inherited diseases. This therapy, known as LUXTURNA™, was developed at CHOP and Penn in collaboration with Spark Therapeutics, for the treatment of patients with confirmed biallelic *RPE65* mutation-associated retinal dystrophy, through the injection of a viral vector with a corrective gene. *RPE65* and its contribution to retinal diseases was first described in 1993 long before the emergence of GWAS and NGS [58]; however, NGS can significantly expedite the identification of individuals who carry mutations in the gene for the clinical trial, and it will greatly help care-providers identify patients who are suitable to be treated with LUXTURNA™. Altogether, they demonstrated the importance of developing genetic-modulation-based therapeutic strategies using knowledge gleaned from human genetics research, yet future application of these therapies will require the use of NGS to select appropriate patient populations who are most likely to benefit from treatment.

In summary, human genetics information has enabled the discovery of several drug targets, even before the GWAS and NGS era. However, with the recent adoption of NGS on well-phenotyped populations paired with large-scale EHRs, as well as the development of therapeutic strategies targeting specific genes (such as gene therapy and oligonucleotide therapy), we believe that NGS will shape the future landscape of drug development and significantly speed up the identification of novel drug targets.

NGS in clinical trials: companion diagnostics, basket trials and genetically stratified trials

Most currently available drugs are 'one size fits all', but it is well known that individuals differ in their responses to drugs and their susceptibility to adverse drug reactions, which could in part be caused by genetic factors. Pharmacogenomics refers to the process of optimizing treatment efficacy and minimizing adverse drug reactions using genomic information, and this topic has been extensively reviewed previously [59–61], so we will not discuss it here. A special application area of pharmacogenomics is companion diagnostics, which represents a specific diagnostic test that is tied to the expected efficacy of a drug, in the sense that they are a 'companion' to a particular drug before the drug is prescribed, and that the test is predictive in assessing the probable effectiveness of the drug treatment. The full list of FDA-cleared or -approved companion diagnostic devices is given online (https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm301431.htm). Based on this list, although SNP genotyping, real-time PCR and immunohistochemistry represent the most common modalities, NGS assays are increasingly approved as companion diagnostics devices (Fig. 2). For example, several NGS assays based on single gene sequencing were developed as companion diagnostic devices, such as the BRACAnalysis CDx test (submission: P140020) by Myriad Genetic Laboratories

and the FoundationFocus CDxBRCA test (submission: P160018) by Foundation Medicine for ovarian cancer. These assays enabled detection of *BRCA1/2* alterations in ovarian tumor tissues, to help identify cancer patients for whom treatment with Rubraca™ (rucaparib) can be more effective.

In November 2017, the MSK-IMPACT assay (submission: DEN170058), which stands for integrated mutation profiling of actionable cancer targets, became the first FDA-approved tumor-profiling multiplex panel based on NGS technology. MSK-IMPACT was developed by the Memorial Sloan-Kettering Cancer Center as a hybridization capture-based NGS panel that is capable of detecting all protein-coding mutations, copy number alterations (CNAs) and selected promoter mutations and structural rearrangements in 341 (more recently expanded to 410) cancer-associated genes [62]. The test was evaluated in a study that compiles tumor and matched normal sequence data from a large cohort of >10,000 patients with advanced cancer coupled with available pathological and clinical information [63]. Remarkably, 11% of the patients with MSK-IMPACT tests were enrolled in genomically matched clinical trials.

The NGS-based diagnostic tests such as MSK-IMPACT support the development of basket trials [64,65], in which therapies are offered to patients whose tumors test positive for certain mutations regardless of cancer type or location where the cancer originated. In May 2017, for the first time in history, the FDA granted approval to a drug as a treatment for adult and pediatric patients whose cancers have a specific genetic signature (biomarker), rather than the location in the body where the tumor originated. This drug, Keytruda® (pembrolizumab) from Merck, is indicated for the treatment of unresectable or metastatic solid tumors that have been identified as having a biomarker as microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR). This indication covers patients with solid tumors that have progressed following prior treatment and who have no satisfactory alternative treatment options and patients with colorectal cancer that has progressed following treatment with certain chemotherapy drugs. The approval of Keytruda® represents the start of a new phase in precision oncology. Multiple similar trials in the oncology space are ongoing; for example, the STARTRK-2 trial sponsored by Ignyta targets patients with over a dozen different cancers who harbor *NTRK1*/2/3, *ROS1* or *ALK* gene rearrangements. Furthermore, several informatics tools, such as iCAGES [66], DawnRank [67] and CARE [68], have already been developed or are in active development to handle upcoming challenges in basket trials to help optimize treatment regimens for patients with specific genomic signatures.

Although successful in cancer, whether the basket trial approach can be applied to treatments of rare inherited diseases with shared molecular etiologies could be much less straightforward, and probably requires deep understanding of the genetic basis of disease manifestation for individual patients and appropriate definition of clinical endpoints. Several particularly compelling cases might be the use of stop-codon read-through drugs to treat genetic diseases caused by nonsense/frameshift mutations, and the use of protein misfolding drugs to treat genetic diseases caused by missense mutations that alter protein folding [69]. For example, in one study, a small molecule called Sephin1 safely prevented the motor, morphological and molecular defects of two otherwise unrelated protein-misfolding diseases, including Charcot-Marie-Tooth 1B and amyotrophic lateral sclerosis [70]. Additionally, many genes or genetic mutations for monogenic diseases can be pleiotropic in nature [71–73]. For example, a hexanucleotide repeat expansion in *C9orf72* accounts for ~22% of familial cases of amyotrophic lateral sclerosis (ALS) and ~12% of familial cases of frontotemporal dementia (FTD) [74,75], whereas deletions in *NRXN1* can cause a wide spectrum of psychiatric and neurological disorders such as autism, schizophrenia and epilepsy [76–78]. It is foreseeable that basket trial approaches could be potentially applied to these areas as well, by grouping patients with similar causal genetic mutations but distinct clinical phenotypes together for testing novel therapeutic agents.

Another major opportunity for the applications of NGS in drug development lies in genetically stratified design of clinical trials, because an ideal system for validating a therapeutic hypothesis is the one being designed exactly in the context that the drug has been targeted for. Unlike a *post hoc* genome sequencing of all participants in clinical trials to identify biomarkers for drug efficacy or adverse events, a genetically stratified clinical trial only recruits patients *a priori* based on the presence of specific mutations who are more likely to respond to the drug to be tested. The mutations could be in the same gene or different genes in the same biological pathway, such as the trial recruiting carriers of copy number variations in the glutamatergic pathway [79] based on their known association with ADHD [80]. A clear advantage is that, by testing on a subset of participants with specific genotypes who are most likely to respond to treatment, the power of trials with fixed sample size can be significantly improved [81].

A notable example of leveraging NGS in stratifying and selecting patients into clinical trials is the recruitment of patients with a rare form of Alzheimer's disease (AD) from the Dominantly Inherited Alzheimer Network (DIAN) [82]. The DIAN trial was designed as a blinded and placebo-controlled trial targeting asymptotic to mildly symptotic mutation-carrying individuals who have not yet developed AD [83]. In fact, one possible explanation of the failure of many potential AD drugs is that they are administered too late in the course of the disease progression [83], because AD pathology is present in the brain long before the onset of the disease itself [84]. Autosomal-dominant Alzheimer's disease (ADAD) individuals carrying a dominant mutation in one of three genes, amyloid precursor protein (*APP*), presenilin 2 (*PSEN2*) or presenilin 1 (*PSEN1*), are predicted to develop AD later in life (Fig. 3a). Thus, only participants carrying the highly penetrant mutations in these three genes are eligible to participate in the DIAN clinical trial. By focusing on a subset of individuals who are more likely to develop AD and benefit from early intervention, these studies are more likely to achieve positive outcomes with limited sample sizes. In principle, such a subsetting mechanism increased the participant's likelihood of receiving active drug (75%) compared with the traditional designs (50%) [83]. Given the failure of a Phase III trial of solanezumab in patients with mild AD [85,86] and that the AD pathology is present in the brain long before the disease onset, a genetically stratified design to focus on the patients who are most likely to benefit from early intervention could achieve higher efficacy compared with traditional 'one size fits all' designs.

Similar to DIAN, another genetically stratified clinical trial is the Generation Program for Alzheimer's disease [87] (Fig. 3b), which represents a collaboration between Novartis, Amgen and the Banner

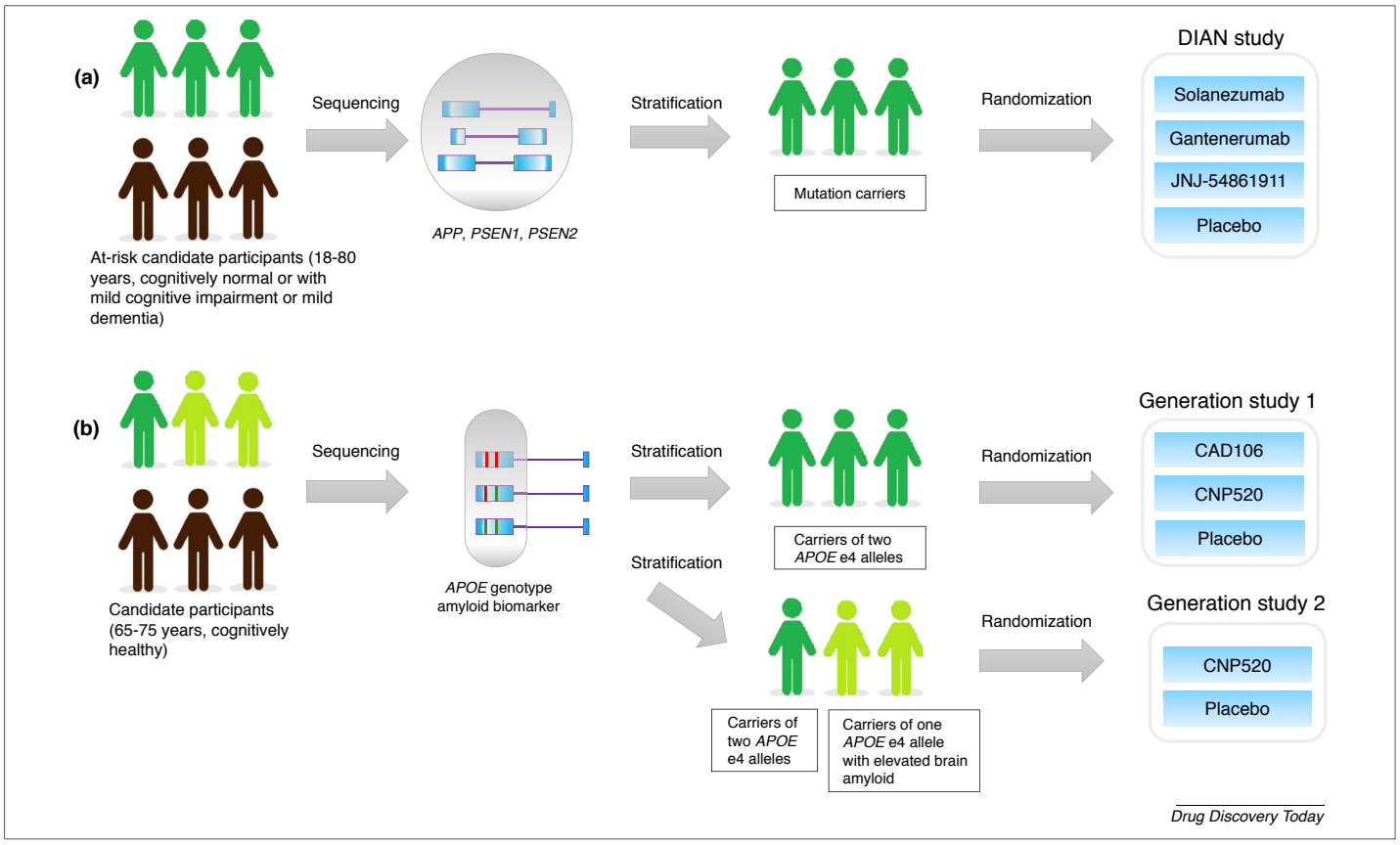

FIGURE 3

An illustration of geneticially stratified clinical trials applied to Alzheimer's disease (AD). (a) The Dominantly Inherited Alzheimer Network (DIAN) trial led by Washington University recruits study participants who are known to have an AD-causing mutation or have a family history of autosomal-dominant AD caused by the presence of causative mutation in the *APP*, *PSEN1* or *PSEN2* genes. (b) The Generation Program (a collaboration between Novartis, Amgen and the Banner Alzheimer's Institute) includes two clinical trials: Generation Study 1 (genetic biomarker) and Generation Study 2 (genetic and clinical biomarkers). The trials enroll people between age 60 and 75 who do not have symptoms of AD but are known to carry the risk-conferring e4 alleles in the *APOE* gene.

Alzheimer's Institute (BAI). The Generation Program comprises two pivotal Phase II/III studies with similar designs to assess the efficacy and safety of CNP520 (a beta-site-amyloid precursor protein-cleaving enzyme inhibitor) in a cognitively unimpaired population at increased risk for developing AD based on status of the *APOE* e4 risk allele. After age, the e4 allele is the greatest risk factor for developing sporadic AD, conferring an increased risk of 3–4 and 8–12 times for one or two copies of the allele, respectively [88]. Generation Study 1 (NCT02565511) recruits participants who carry two *APOE* e4 risk alleles; given that the e4 allele frequency is roughly 10–15% in various ethnicity groups [88–90], only a small fraction of the population is eligible to participate in the study. Generation Study 2 (NCT03131453) recruit participants who either carry two copies of the e4 allele in the *APOE* gene or carry one *APOE* e4 allele but with elevated brain amyloid. By focusing on people at particularly high risk for developing the disease owing to their genetic status and biomarker profiles, these studies potentially achieve greater power to evaluate whether anti-amyloid treatments might prevent or delay the emergence of symptoms of AD.

Concluding remarks

In this review, we have briefly discussed a few aspects of using NGS in drug discovery and development, with a specific focus on target identification and genetically stratified clinical trials. Adoption of these technologies will expand our understanding of the genetic basis of human diseases, help identify novel compounds or re-use existing compounds, increase drug efficacy and reduce adverse reactions, and ultimately facilitate personalized therapies. Currently, the pharmaceutical industry is performing population-scale NGS analysis to discover new knowledge about human diseases, as well as focused NGS analysis on clinical trials to identify biomarkers for drug efficacy or safety. In the years to come, we expect to see newly developed and marketed compounds facilitated by the use of NGS techniques, and that more genetically stratified clinical trials could be conducted to recruit specific subsets of participants based on their genetic profiles and other clinical biomarkers.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

This study is supported in part by NIH grant MH108728, Columbia University and CHOP Research Institute to K.W.; the authors would like to thank Dr. Rui Chen (Baylor College of Medicine) for providing valuable feedback, and the two anonymous reviewers for their insightful comments that helped improve the manuscript.

References

1 Kingsmore, S.F. *et al.* (2008) Genome-wide association studies: progress and potential for drug discovery and development. *Nat. Rev. Drug Discov.* 7, 221–230
2 Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33 (Suppl), 228–237
3 Freimer, N. and Sabatti, C. (2004) The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. *Nat. Genet.* 36, 1045–1051
4 Klein, R.J. *et al.* (2005) Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385–389
5 MacArthur, J. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45, D896–D901
6 Visscher, P.M. *et al.* (2017) 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22
7 Kathiresan, S. *et al.* (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* 41, 56–65
8 Wang, K. *et al.* (2007) Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* 81, 1278–1283
9 Wang, K. *et al.* (2010) Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* 11, 843–854
10 Wang, K. *et al.* (2009) Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn disease. *Am. J. Hum. Genet.* 84, 399–405
11 Parkes, M. *et al.* (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* 39, 830–832
12 Hampe, J. *et al.* (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.* 39, 207–211
13 Hooper, K.M. *et al.* (2017) Inflammatory bowel disease drugs: a focus on autophagy. *J. Crohns Colitis* 11, 118–127
14 Wang, K. *et al.* (2010) Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. *Hum. Mol. Genet.* 19, 2059–2067
15 Gaffen, S.L. *et al.* (2014) The IL-23-IL-17 immune axis: from mechanisms to therapeutic testing. *Nat. Rev. Immunol.* 14, 585–600
16 Duerr, R.H. *et al.* (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 314, 1461–1463
17 Liu, Y. *et al.* (2008) A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet.* 4, e1000041
18 Feagan, B.G. *et al.* (2017) Induction therapy with the selective interleukin-23 inhibitor risankizumab in patients with moderate-to-severe Crohn's disease: a randomised, double-blind, placebo-controlled phase 2 study. *Lancet* 389, 1699–1709
19 Feagan, B.G. *et al.* (2016) Ustekinumab as induction and maintenance therapy for Crohn's disease. *N. Engl. J. Med.* 375, 1946–1960
20 Kopp, T. *et al.* (2015) Clinical improvement in psoriasis with specific targeting of interleukin-23. *Nature* 521, 222–226
21 Nelson, M.R. *et al.* (2015) The support of human genetic evidence for approved drug indications. *Nat. Genet.* 47, 856–860
22 Du, Y. *et al.* (2012) Genome-wide association studies: inherent limitations and future challenges. *Front. Med.* 6, 444–450
23 Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427
24 Cao, C. and Moult, J. (2014) GWAS and drug targets. *BMC Genomics* 15 (Suppl. 4), 5
25 Johnson, A.D. *et al.* (2011) Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. *Hypertension* 57, 903–910
26 Ehret, G.B. (2010) Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. *Curr. Hypertens. Rep.* 12, 17–25
27 Goldstein, D.B. (2009) Common genetic variation and human traits. *N. Engl. J. Med.* 360, 1696–1698
28 Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427
29 Boyle, E.A. *et al.* (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169, 1177–1186
30 Genomes Project (2010) A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073
31 Tennessen, J.A. *et al.* (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64–69
32 Auton, A. *et al.* (2015) A global reference for human genetic variation. *Nature* 526, 68–74
33 Gudbjartsson, D.F. *et al.* (2015) Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* 47, 435–444
34 Iossifov, I. *et al.* (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221
35 Yuen, R.K. *et al.* (2015) Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* 21 (2), 185–191
36 De Rubeis, S. *et al.* (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215
37 RK, C.Y. *et al.* (2017) Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* 20, 602–611
38 Woollard, P.M. *et al.* (2011) The application of next-generation sequencing technologies to drug discovery and development. *Drug Discov. Today* 16, 512–519
39 Vicini, P. *et al.* (2016) Precision medicine in the age of big data: the present and future role of large-scale unbiased sequencing in drug discovery and development. *Clin. Pharmacol. Ther.* 99, 198–207
40 Siu, L.L. *et al.* (2015) Next-generation sequencing to guide clinical trials. *Clin. Cancer Res.* 21, 4536–4544
41 Horak, P. *et al.* (2016) Integrating next-generation sequencing into clinical oncology: strategies, promises and pitfalls. *ESMO Open* 1, e000094
42 Rastegar-Mojarad, M. *et al.* (2015) Opportunities for drug repositioning from phenome-wide association studies. *Nat. Biotechnol.* 33, 342–345
43 Wang, Z.Y. and Zhang, H.Y. (2013) Rational drug repositioning by medical genetics. *Nat. Biotechnol.* 31, 1080–1082
44 Delavan, B. *et al.* (2018) Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discov. Today* 23, 382–394
45 Cummings, C.A. *et al.* (2016) The role of next-generation sequencing in enabling personalized oncology therapy. *Clin. Transl. Sci.* 9, 283–292
46 Blumenthal, G.M. *et al.* (2016) Next-generation sequencing in oncology in the era of precision medicine. *JAMA Oncol.* 2, 13–14
47 Stitziel, N.O. and Kathiresan, S. (2017) Leveraging human genetics to guide drug target discovery. *Trends Cardiovasc. Med.* 27, 352–359
48 Kamb, A. *et al.* (2013) Human genetics as a foundation for innovative drug development. *Nat. Biotechnol.* 31, 975–978
49 Dewey, F.E. *et al.* (2016) Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 354, 6814
50 Myocardial Infarction Genetics Consortium (2014) Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N. Engl. J. Med.* 371, 2072–2082
51 He, K.Y. *et al.* (2016) Pathogenic mutations in cancer-predisposing genes: a survey of 300 patients with whole-genome sequencing and lifetime electronic health records. *PLoS One* 11, e0167847
52 Zhang, W. and Drake, M.T. (2012) Potential role for therapies targeting DKK1, LRP5, and serotonin in the treatment of osteoporosis. *Curr. Osteoporos. Rep.* 10, 93–100
53 Treggiari, M.M. *et al.* (2009) Early anti-pseudomonal acquisition in young patients with cystic fibrosis: rationale and design of the EPIC clinical trial and observational study. *Contemp. Clin Trials* 30, 256–268
54 Emond, M.J. *et al.* (2012) Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat. Genet.* 44, 886–889
55 Maude, S.L. *et al.* (2018) Tisagenlecleucel in children and young adults with B-cell lymphoblastic leukemia. *N. Engl. J. Med.* 378, 439–448
56 Ronson, A. *et al.* (2016) Treatment of relapsed/refractory acute lymphoblastic leukemia in adults. *Curr. Oncol. Rep.* 18, 39
57 Neelapu, S.S. *et al.* (2017) Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. *N. Engl. J. Med.* 377, 2531–2544
58 Cai, X. *et al.* (2009) RPE65: role in the visual cycle, human retinal disease, and gene therapy. *Ophthalmic Genet.* 30, 57–62
59 Relling, M.V. and Evans, W.E. (2015) Pharmacogenomics in the clinic. *Nature* 526, 343–350
60 Weinshilboum, R. and Wang, L. (2004) Pharmacogenomics: bench to bedside. *Nat. Rev. Drug Discov.* 3, 739–748
61 Lee, J.W. *et al.* (2014) The emerging era of pharmacogenomics: current successes, future potential, and challenges. *Clin. Genet.* 86, 21–28
62 Cheng, D.T. *et al.* (2015) Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* 17, 251–264
63 Zehir, A. *et al.* (2017) Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* 23, 703–713

64 Cunanan, K.M. *et al.* (2017) Basket trials in oncology: a trade-off between complexity and efficiency. *J. Clin. Oncol.* 35, 271–273
65 Woodcock, J. and LaVange, L.M. (2017) Master protocols to study multiple therapies, multiple diseases, or both. *N. Engl. J. Med.* 377, 62–70
66 Dong, C. *et al.* (2016) iCAGES: integrated CAncer GEnome Score for comprehensively prioritizing driver genes in personal cancer genomes. *Genome Med.* 8, 135
67 Hou, J.P. and Ma, J. (2014) DawnRank: discovering personalized driver genes in cancer. *Genome Med.* 6, 56
68 Jiang, P. *et al.* (2018) Genome-scale signatures of gene interaction from compound screens predict clinical efficacy of targeted cancer therapies. *Cell Syst.* 6, 343–354
69 Brooks, P.J. *et al.* (2014) Expanding rare disease drug trials based on shared molecular etiology. *Nat. Biotechnol.* 32, 515–518
70 Das, I. *et al.* (2015) Preventing proteostasis diseases by selective inhibition of a phosphatase regulatory subunit. *Science* 348, 239–242
71 Ittisoponpisan, S. *et al.* (2017) Landscape of pleiotropic proteins causing human disease: structural and system biology insights. *Hum. Mutat.* 38, 289–296
72 Sivakumaran, S. *et al.* (2011) Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* 89, 607–618
73 Gratten, J. and Visscher, P.M. (2016) Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med.* 8, 78
74 DeJesus-Hernandez, M. *et al.* (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256
75 Renton, A.E. *et al.* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268
76 Moller, R.S. *et al.* (2013) Exon-disrupting deletions of NRXN1 in idiopathic generalized epilepsy. *Epilepsia* 54, 256–264
77 Rujescu, D. *et al.* (2009) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* 18, 988–996
78 Ching, M.S. *et al.* (2010) Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am. J. Med. Genet. B: Neuropsychiatr. Genet.* 153B, 937–947
79 Connolly, J.J. *et al.* (2015) ADHD & pharmacotherapy: past, present and future: a review of the changing landscape of drug therapy for attention deficit hyperactivity disorder. *Ther. Innov. Regul. Sci.* 49, 632–642
80 Elia, J. *et al.* (2011) Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nat. Genet.* 44, 78–84
81 Mead, S. *et al.* (2016) Clinical trial simulations based on genetic stratification and the natural history of a functional outcome measure in Creutzfeldt–Jakob disease. *JAMA Neurol.* 73, 447–455
82 Morris, J.C. *et al.* (2012) Developing an international network for Alzheimer research: the dominantly inherited Alzheimer network. *Clin. Investig.* 2, 975–984
83 Moulder, K.L. *et al.* (2013) Dominantly inherited Alzheimer network: facilitating research and clinical trials. *Alzheimers Res. Ther.* 5, 48
84 Monsell, S.E. *et al.* (2017) Genetic comparison of symptomatic and asymptomatic persons with Alzheimer disease neuropathology. *Alzheimer Dis. Assoc. Disord.* 31, 232–238
85 Doody, R.S. *et al.* (2014) Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N. Engl. J. Med.* 370, 311–321
86 Honig, L.S. *et al.* (2018) Trial of solanezumab for mild dementia due to Alzheimer's disease. *N. Engl. J. Med.* 378, 321–330
87 Lopez Lopez, C. *et al.* (2017) The Alzheimer's Prevention Initiative Generation Program: evaluating CNP520 efficacy in the prevention of Alzheimer's disease. *J. Prev. Alzheimers Dis.* 4, 242–246
88 Heffernan, A.L. *et al.* (2016) The neurobiology and age-related prevalence of the epsilon4 allele of apolipoprotein E in Alzheimer's disease cohorts. *J. Mol. Neurosci.* 60, 316–324
89 Bang, O.Y. *et al.* (2003) Important link between dementia subtype and apolipoprotein E: a meta-analysis. *Yonsei Med. J.* 44, 401–413
90 Farrer, L.A. *et al.* (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* 278, 1349–1356
91 Boyden, L.M. *et al.* (2002) High bone density due to a mutation in LDL-receptor-related protein 5. *N. Engl. J. Med.* 346, 1513–1521