

Review

Advances in Genomic Discovery and Implications for Personalized Prevention and Medicine: Estonia as Example

Bram Peter Prins ^{1,*†}, Liis Leitsalu ^{2,†}, Katri Pärna ^{2,3,4}, Krista Fischer ^{2,5}, Andres Metspalu ², Toomas Haller ^{2,†} and Harold Snieder ^{3,*‡}

¹ MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; bp406@medschl.cam.ac.uk (B.P.P.)

² Institute of Genomics, University of Tartu, 51010 Tartu, Estonia; k.parna@umcg.nl (K.P.); Liis.Leitsalu@ut.ee (L.L.); krista.fischer@ut.ee (K.F.); andres.metspalu@ut.ee (A.M.); toomas.haller@ut.ee (T.H.)

³ Department of Epidemiology, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands; k.parna@umcg.nl (K.P.); h.snieder@umcg.nl (H.S.)

⁴ Institute of Molecular and Cell Biology, University of Tartu, 51010 Tartu, Estonia; k.parna@umcg.nl (K.P.)

⁵ Institute of Mathematics and Statistics, University of Tartu, 50409 Tartu, Estonia; krista.fischer@ut.ee (K.F.)

* Correspondence: bp406@medschl.cam.ac.uk (B.P.P.); h.snieder@umcg.nl (H.S.)

† These authors contributed equally.

‡ These authors contributed equally.

Abstract: The current paradigm of personalized medicine envisages the use of genomic data to provide predictive information on the health course of an individual with the aim of prevention and individualized care. However, substantial efforts are required to realize the concept: enhanced genetic discoveries, translation into intervention strategies, and a systematic implementation in healthcare. Here we review how further genetic discoveries are improving personalized prediction and advance functional insights into the link between genetics and disease. In the second part we give our perspective on the way these advances in genomic research will transform the future of personalized prevention and medicine using Estonia as a primer.

Keywords: genetic discovery; genomic medicine; return of results; patient communication; biobank; Estonia

Citation: Prins, B.P.; Leitsalu, L.; Pärna, K.; Fischer, K.; Metspalu, A.; Haller, T.; Snieder, H. Advances in Genomic Discovery and Implications for Personalized Prevention and Medicine: Estonia as Example. *J. Pers. Med.* **2021**, *11*, 358. <https://doi.org/10.3390/jpm11050358>

Academic Editor: Kenneth S. Ramos

Received: 17 March 2021

Accepted: 25 April 2021

Published: 29 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Progress of Genomic Technologies Enabling Personalized Medicine

1.1. Overcoming the Limitations of Genetic Loci Discovery

1.1.1. Power Gains Through Sample Size Increase

Traditionally, one of the simplest ways to uncover substantially larger parts of the so-called ‘missing heritability’—next to focusing on structural variants rather than single nucleotide polymorphisms (SNPs)—is by simply increasing sample in genomic discovery studies, which in turn increases statistical power [1]. From very early on, this has been done through combining genome wide association study (GWAS) results of individual studies in meta-analyses, facilitated by the formation of large GWAS consortia, which are often phenotype-specific. The recognition by public bodies of earlier successes in genetic discoveries through consortia and the importance of data-driven healthcare, led to the creation of several biobanks, with the first initiatives having been established over 20 years ago [2,3]. Currently, one of the most well-known public biobanks is the UK Biobank effort, that aims to improve prevention, diagnosis, and treatment of illness, and the promotion of health through society [4]. Up to now, roughly 500,000 samples have been collected with both genetic data and a wide range of phenotypical measurements, which are available for a small fee, to the wider scientific community in its entirety. Similar large

public efforts are Lifelines in the Netherlands [5] and the Estonian Biobank [6]. An interesting source of biobank data comes from personal genomics companies such as 23andMe [7] where participants can opt to make their genomes available (anonymized) to the scientific community, combined with data retrieved from self-reported health surveys. As of 2020 23andMe had reached a 12 million customers milestone, what makes it one of the largest genetic research databases in the world.

1.1.2. Public Availability of GWAS (Results) Data

Significant increases in statistical power are furthermore facilitated by utilizing publicly available data [8]. Large groundbreaking projects funded by the public such as the Human Genome project typically made their findings publicly available. For institutional projects (i.e., within non-governmental organizations/universities), however, retaining the data privately in order to maximize the generation of scientific publications and authorships has historically been the rule rather than the exception. There is a trend toward more openly sharing data with researchers and the public [9–11]. The Wellcome Trust Case Control Consortium (WTCCC) study on seven common diseases [12] was a front-runner in this regard, making the best use of often publicly funded research. This is exactly why platforms such as the database of Genotypes and Phenotypes [13] or the European Genome Phenome (EGA) [14] archives have been set up: facilitating easy data access and sharing. The volume of data from GWAS from various consortia presents logistical challenges in sharing data, but summary statistics from many of these studies are now available publicly, for example, through the OpenGWAS project (<https://gwas.mrcieu.ac.uk/>). This allows easy combination of study statistics or replication, and can be done even in the case of unknown proportions of sample overlaps by methods such as METACARPA [15]. It further enables downstream analyses using GWAS summary statistics only, such as large-scale Mendelian Randomization-like analyses using platforms such as MR-Base [16], or LD score regressions using LD hub [17]. To ensure that the vast amounts of generated data are not only publicly available, but also are truly accessible, other models to improve the portability of data are being developed, such as methods for decentralized storage of scientific data [18], and cloud-based models of multi-center collaborations [19].

1.1.3. Increase in Resolution

Population Sequencing

Regardless of the sample size, genomic discoveries are dependent on variants that have been identified through sequencing. As we continue to sequence many more individuals, more variation is identified, and it has become clear that rarer variation may explain parts of the missing heritability for complex traits [20]. Therefore, improving genomic resolution is another feasible way to improve genomic discoveries, as has already been demonstrated early on by comparing HapMap to 1000 Genomes imputation for various traits and the observed improvement in terms of additional loci discovered and fine-mapping [21]. Sequence efforts continue to grow and projects such as the Haplotype Reference Consortium [22], which combine sequencing data from multiple cohorts, provide greatly enriched imputation panels, the creation of which will substantially improve genetic discoveries. A great number of population level whole genome sequence efforts are underway, such as in the UK ($n = 500,000$ [23]), Australia ($n = 200,000$ [24]), France ($n = 235,000$ [25]), and the USA ($n = 1,000,000$ [26]). Similar efforts to map and collect human genetic variation are ongoing for non-European populations, such as the African Genome Variation Project [27], H3Africa [28], Genome-Asia [29]. Multiple private sequence efforts are also underway, one of the most ambitious being the effort of the AstraZeneca to sequence 2 million human genomes [30], complemented by projects that will establish deeper and more accurate reference maps of the human genome [31].

Addressing Neglected Parts of the Genome

Even though many genomes have been sequenced and analyzed, certain regions in the genome are largely neglected in genome-wide analyses, such as the Y-chromosome or mitochondrial chromosomes. This is often due to complications in genotype calling, imputation, and selection of test statistics, as well as a lower proportion of genes, and a lower coverage of current genotyping platforms compared with autosomal chromosomes [32–34]. There is ample evidence for genetic regions in these chromosomes to be involved in a wide range of traits and diseases, including auto-immune diseases [35–37]. Investing in the large-scale analyses of these chromosomes, through efforts such as YGEN, the first international consortium that will assess the influence of Y-chromosome variation on complex traits of public health or evolutionary interest, are likely to provide a multitude of additional insights into the contribution of variants on neglected chromosomes to complex traits [38]. Moreover, mitochondrial work at scale is currently undertaken [39]. In a similar way, certain genetic regions have been rarely analyzed due to the difficulty to type and impute these regions, amongst the best-known examples immunologically are important complex regions such as the HLA region and killer cell immunoglobulin-like receptors (KIRs). Recent methods have however enabled the proper imputation of these regions [40,41], allowing the detailed investigation of their role in human disease.

Fine-Mapping

With much higher resolution genomic data as compared to traditionally imputed GWAS chip data, it becomes more likely that a causal variant will be present in the set of analyzed variants. However, it remains difficult to identify these, because the vast majority of associated variants fall outside the coding regions as recognized early on [42], whereas the majority of annotation efforts have traditionally focused on coding variants [43]. Developments in fine-mapping approaches such as CADD [44] and Eigen [45] have integrated extensive functional genomic annotations into a per-variant score of functionality, including conservation scores, epigenomic annotations and protein-level consequence scores. This importantly enables the prioritization of variants that are intergenic or in the non-coding regions, where previously this was challenging. Other approaches such as CAVIARDB [46], are able to highlight the most likely causal variants in a locus, just using summary statistics and PICS [47], provides similar fine-mapping capabilities, but then also includes prior knowledge such as transcriptional and epigenomic data, promising much improved fine-mapping results. Further fine-mapping opportunities lie in genetic diversity between populations, where differences in LD structure may help to further narrow down genetic regions associated with a trait [48], thereby more accurately pinpointing the location of actual causal variants. Studying different populations can improve the discovery of rare risk variants in loci already highlighted by common variants found by GWAS, as has been demonstrated in, for example, Greek isolates [49], or the Icelandic population [50].

1.1.4. Increase in Throughput and Moving towards ‘Big Data’

Genetic discovery can thus be greatly aided by increasing sample sizes in combination with high-depth sequencing. This still comes at a substantial price, although the costs of whole genome sequencing have been rapidly declining, nearing the USD 1000 dollar genome barrier [51]. Newer developments offer further promise. One such technique is nanopore or ‘strand sequencing’ which allow sequencing of longer reads and faster phasing of genomes [52]. In concert with ever-increasing biobanking efforts are resulting in large amounts of data for which need to be matched by raw computational power and infrastructure. The quantity of data to organize and interpret has posed challenges, requiring the development of appropriate tools to address these issues. On the genomics analyses front, developments like Genotype Query Tools [53], provide novel data index-

ing strategies and much improved data compression, resulting in substantial query analysis performance improvements over current state-of-the-art tools of over 400-fold. In parallel, analytical algorithms are continuously optimized, whereby borrowing techniques from other fields of research such as signal processing [54] or artificial intelligence [55] offer further improvements in deciphering the human genomic architecture.

Developments on the analytical infrastructure front continue to contribute to ever more efficient genetic discovery. For example, sequencing pipelines, typically using algorithms running on high-end computer clusters on general CPUs, are being integrated in processors themselves (i.e., 'hard-coded'). The reconfigurable DRAGEN Bio-IT Processor, produced by Illumina [56], has hard-coded highly optimized algorithms for the full next generation sequencing (NGS) secondary analysis pipeline, which set world speed records for genomic data analysis [57].

Solving data processing problems in big data genomics requires supercomputing infrastructure and expertise, which are not always available in academic environments. In industry, genomic big data management is tackled in various ways. Advanced genomics companies make use of large-scale commercial cloud-computing infrastructures available such as Amazon Web Services (AWS) or Google Genomics, to process and analyze sequence data using clever open-source distributed tools such as Hadoop or Apache SPARK that bring the algorithms to the data versus the much slower opposite. With the ever increasing size of genomics data, big data approaches like the use of computing clouds and advanced analytics platforms are already being adopted by the scientific community [58,59], simply out of sheer necessity.

1.2. Increase in Understanding (from Genotype to Phenotype)

Discovery and interpretation of genetic trait variation requires insights from multiple layers of biological intermediates through which genetic loci exert their effects on the phenotype in order to be able to obtain causal and mechanistic insights. This means integrating genetic data with expression data (eQTL), regulatory elements and effectors, proteomes, metabolomes, and intermediate phenotypes, all of which are interacting biological levels that have their effects on the final outcome studied.

Large efforts have been undertaken to map epigenomic data, such as in the ENCODE [60] and Roadmap Epigenetics Project [61], Blueprint Epigenome [62], but also for the proteome (HPP) [63] data exists on a comparable large scale. Efforts such as the Human Cell Atlas [64] and HipSci [65], which enable the retrieval of these data simultaneously from pluripotent stem cells allow insights in developmental and differentiation mechanisms on a cellular level. However, another phenotypic level to be considered is the microbiome. Apart from more intuitive associations of the microbiome with disease, such as inflammatory bowel disease [66], there are also indications that it influences psychiatric [67] and cardiovascular outcomes [68], whereas the composition of the microbiome itself is also influenced by genetic variation of the host [69], which brings yet another dimension to integrative genotype–phenotype analyses. Various efforts have been undertaken to obtain, map and interpret the aforementioned data types.

Other efforts even aim to map entire organs, which have the potential to uncover genetic effects on a precise organic scale. Examples include the UK Digital Heart Project [70], having reconstituted entire human hearts from echocardiographs or ENIGMA [71] with the world's largest set of brain scans. Linking and integrating these biological data types will on the one hand undoubtedly improve our understanding of higher-order networks and mechanisms driving complex disease phenotypes across multiple tissues, but also bring along a great challenge to build statistical models, although promising methods are already available [72]. Many of the "omics" data can be readily integrated with genetic data using imputation methods. For example, epigenomic markers [73] or expression levels [74] that can be predicted based on cohort genotypes only. Imputation of epigenomes is useful as trait-associated variants affect regulatory regions in a cell-type dependent manner [75], and it is not always feasible to map every epigenetic mark in every tissue,

cell type and condition of interest. Imputation of epigenomes for specific cell-types enables cell-type/tissue specific follow-up analyses to understand phenotype specific regulatory consequences of variants identified. Similarly, eQTLs can be cell-type specific [76,77], and appropriate cellular context enables a better understanding of variant consequences on expression level in relevant tissue.

1.3. Impact of Genetic Discoveries and Clinical Relevance

Translation of identified genetic variation or loci into pathogenic molecular mechanisms appears more feasible than ever before and allow for the realization of personalized medicine in practice. The need for such approaches is clear in clinical practice where the marked differences in response to therapies between different individuals are a recognized phenomenon, with drug response variability as a particularly well-studied case [78]. Medical conditions are personally unique when viewed at the molecular level; therefore, it only makes sense if treatment options follow suit. Generic treatment options can potentially have harmful consequences when “one size fits all” approaches are taken as evident from a wealth of drug response studies. As of today, GWAS findings have already proven to be clinically informative in a number of ways [79,80].

1.3.1. Risk Prediction and Causal Inference

A currently more directly applicable use of genomic discoveries is (genetic) risk prediction of disease, where individual-level risk estimates through genomic risk scores (GRS) may help in early intervention and improve diagnostic procedures. There have been numerous studies developing risk prediction models using genetic markers and a number of successful examples [81,82]. Notably, a recent study using nearly two million genetic variants allowed stratification of individuals with different trajectories of CAD solely based on genetic information, and performed better than conventional risk factors at predicting incidence CAD [83]. With up to 2 million variants as a model for the genetic architecture of CAD, it is virtually impossible to pinpoint specific causal mechanisms to be used for intervention. Nevertheless, it is possible to perform causal inference using genetic risk scores through Mendelian randomization (MR) analyses. In this specific use of genetic risk scores, care must be taken however to design a risk score that instruments a particular exposure that is hypothesized to impose a causal effect on the outcome. For example, as previously suspected, it has been confirmed that LDL levels are causal to CAD [84], stressing that LDL-cholesterol lowering interventions such as adjusted diets and LDL lowering medication will have effects, as demonstrated also in a recent meta-analysis of randomized, controlled trials (RCTs) [85]. Similar approaches will allow the discovery of previously unknown causal risk factors that have the potential to aid disease management.

1.3.2. Disease Stratification and Tailored Clinical Surveillance/Management

Another application comes in the form of disease classification, where for example Sirota and colleagues show that they were able to classify auto-immune diseases [86], identifying variants that make an individual susceptible to one class of autoimmune disease whilst simultaneously protecting from diseases in the other autoimmune class. This may enable clinicians to more optimally tailor treatment, as certain drugs for example are known to improve one type of autoimmune disorder, whilst having negative effects on another. A good example for this is infliximab, which is an antibody that binds to TNF. Typically prescribed and working well for rheumatoid arthritis and ankylosing spondylitis [87,88], it however has no efficacy and sometimes even worsens the condition in individuals with other autoimmune diseases such as multiple sclerosis [89].

1.3.3. Personalized Treatment

A major area where GWAS may play a role is in pharmacogenomics. As modern medication only recently appeared as an environmental factor, it will not have caused any negative evolutionary selection pressures on common variants associated with (severe) adverse drug reactions (ADRs). Successful discoveries related to inflammatory disorders include identification of loci for ADRs against Lumiracoxib [90], a drug that is prescribed for the treatment of osteoarthritis and RA, causing liver injury, and loci associated with ADRs against thiopurin [91,92], prescribed for autoimmune disorders such as Crohn's disease and rheumatoid arthritis, causing leukopenia and pancreatitis. In addition to the discovery of genetic risk loci for ADRs, tailoring treatment doses of drugs depending on an individual's genetic profile will also become valuable in clinical practice and has already been demonstrated to be feasible for asthma [93].

GWAS can also aid in identifying drug targets [94,95]. By making clever use of known well-established and up to date gene-drug target databases such as DrugBank [96], therapeutic target database (TTD) [97], and PharmGKB [98], one can overlap genes identified in loci in a meta-analysis, or their interacting genes [99] and filter out those that appear druggable for further study. As many of the compounds in these databases are already FDA approved, this creates a wealth of opportunities for drug repurposing and repositioning, bypassing the necessary lengthy and costly process of clinical safety trials.

2. Estonia as a Primer for Personalized Medicine

2.1. Favorable Circumstances

In all, genetic discoveries will continue to contribute in various ways to the realization of personalized medicine [100], whereby personal biomarker-specific profiles affected by genetic, clinical, and lifestyle factors are likely to play an important role [101] (Box 1). The recognition of both the importance and the feasibility of personalized medicine for common complex disorders for national public health has incentivized the Estonian government to start an ambitious biobanking initiative aiming to completely overhaul its health-care system based on biologically informed personalized care, of which the resulting precision medicine tools will be made available to everyone as part of the basic health insurance coverage.

Box 1. The advent of personalized medicine.

It has long been recognized that standardized medical treatments, so called "one-size-fits-all" approaches, may have serious limitations. The idea of personalized medicine is, therefore, not new, but can be seen as an ongoing development. Prior to the biotech revolution of the last decades, methods to determine detailed biological differences between humans and populations that could be utilized for optimizing healthcare were limited. Decoding human DNA through the completion of the Human Genome Project resulted in what can be regarded as the establishment of a blueprint for the construction, development, and maintenance of the human organism. The HapMap project that followed, made an inventory of DNA differences between individuals and provided the groundwork for starting to understand inter-individual differences in genomic risk for disease. This genomic data, in concert with other upcoming health data such as other omics measurements, electronic health records (EHRs), and smart wearable devices can ultimately be devised to develop personally optimized intervention strategies.

Genomic data in particular has been shown to be useful in guiding drug development. For example, lead drug developer Astra Zeneca showed that genetically informed decisions not only led to an increase in number of successful drug development projects, but also a reduced time to market [102]. Similarly, for personalized medicine, genetic profiling already helps optimizing drug dosage and avoiding adverse effects, such as done for S-methyltransferase (TPMT) genotype testing prior to administration of a class of drugs named thiopurines [103]. Additional genetic testing helps to select medication for individuals with specific genetic mutations, an example being melanoma patients with particular mutations in the BRAF gene developing resistance against the normally used single agent BRAF inhibitors, who display a better response with combined use of MEK inhibitors [104]. The advance of personalized medicine is demonstrated by the vast increase of approved medication entering European and American markets, which define it generally as a treatment that recommends or requires a biomarker pre-test to determine if a treatment

is suitable for a patient. In Germany, for example, according to this definition, there were no personalized drugs on the market before 1996. The decade that followed, saw 10 oncological drugs becoming available as personalized medicine, after which this number shot up to 63 in 2019, and diversified over various disease classes [105].

Many nations have already launched initiatives to promote personalized medicine [106], but Estonia in particular has shown to be on the forefront through its unique nation-wide integrated approach, connecting a national biobank with electronic health records with the aim to incorporate it into general healthcare.

This initiative is enabled by a number of aligning favorable circumstances. Firstly, the progressive political climate in Estonia and attitude of its population support the adaptation of new technological developments, including personalized medicine. The Estonian parliament was the first in the world to pass a special law to govern genome research in the general population, the Human Genes Research Act [107]. This law is favorable to the implementation of personalized medicine and provides the required legal framework for it. The various acting governments throughout the lifespan of the initiative have been overwhelmingly supportive of the initiative. Although, perhaps even more important in this context is the positive attitude of the general public. Regularly conducted surveys show that up to 75% of the respondents support the biobank, only 1% hold the opposing view, and around 77% are interested in genetic testing [108] (Figure 1).

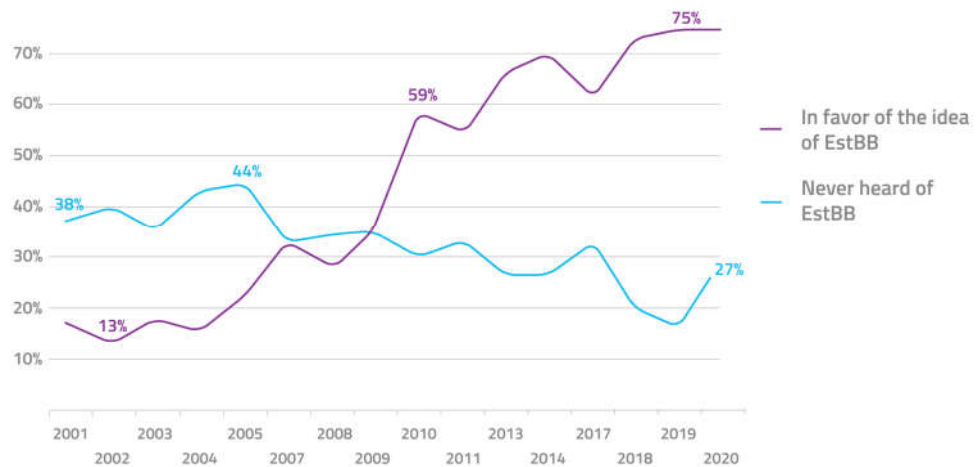


Figure 1. Public awareness and approval of the Estonian biobank over the past 2 decades.

Secondly, in the last decades Estonia has heavily invested in a first-class national IT infrastructure and know-how, which has digitized many aspects of daily life. At the center of this is the Estonian ID card—a computer-readable compulsory identification card implemented in 2002 that gives access to a nationwide governmental technical infrastructure called the X-road [109]. This is the central “digital highway” that gives access to all online services in the country. The X-road is highly adaptable in the sense that new functionalities can be added at any time. This includes the support for personalized medicine. The system can be used to link major Estonian health registries and hospitals and facilitate sharing of medical information for patients, such as health data history, hospital admissions, pharmacy prescriptions, to compile a full medical picture to which treatments can be tailored in a secure manner [110].

2.2. Biobank Cohort, Data Collected and Generated

The favorable circumstances culminated in the foundation of the National Biobank (“Estonian Biobank”, EstBB) in 2001, which was transferred to the University of Tartu as

the Estonian Genome Center (EGCUT) in 2007. From 2002, it has built up its databases and research portfolio ever since [6,111]. By 2011, the population-based biobank had 52,000 adult participants from the first wave of recruitment, and in 2018–2019 an additional 150,000 participants were recruited in the second wave. Altogether the biobank cohort constitutes 20% of the country's adult population. All projects described here are based on the first 52,000 subjects in the EstBB, not on the whole cohort of 202,000 as it is today.

Blood samples were collected from all participants, which were subsequently used for genotyping using the Illumina Infinium Global Screening Array, as well as for deep phenotyping. Whole genome sequence data is currently available for 3000 individuals, as well as whole exome sequencing for 2500 individuals and these numbers are continuously increasing. Furthermore, EstBB has measured transcriptomics (Illumina RNAseq), DNA methylation (Illumina 450 K methylation array), proteomics (Olink) as well as various metabolomics profiles (clinical biochemistry, NMR, mass spectrometry) for a large number of participants to generate precise information based on molecular profiles. In 2018, the EstBB launched a project for obtaining microbiome profiles from stool samples (over 2500 currently collected). Recently, a new contract was signed to expand the current Nightingale NMR metabolomics dataset (<https://nightingalehealth.com/> (accessed on 14 April 2021)) to all biobank participants in 2021–2022. In combination with regular supplementation of electronic health status updates from registries, large scale longitudinal studies are facilitated.

2.3. Translational Research in Genomics

The Estonian biobank legislation and consent allow return of scientific results to biobank participants. The first projects involved a genotype first approach where biobank participants with specific genetic findings were invited for further phenotyping and clinical assessment while offering disclosure of genetic findings to the participant [112,113].

In 2017, a different approach was introduced offering a selection of results by the biobank to participants who expressed interest in receiving them. As part of the process, participants will update some measurements, update their lifestyle information and sign the consent for return of results all through an online participant portal. Over a two-year period, 2957 participants received the report accompanied with face-to-face counseling. On top of disease risk prediction, EstBB is analyzing the potential impact of considering pharmacogenomics in patient care [114]. This supports not only more tailored medication dosages, but also allows avoiding ADRs and medication that might not even be effective for a particular individual. Currently, the team at EstBB is able to profile 10 genes involved in compound metabolism to give feedback about 31 different medications.

Furthermore, the biobank has put an extensive monitoring practice in place to ensure and improve the quality of the entire process, with a focus on assessing the effect and outcomes of receiving genomic risk information. Through the participant portal, each participant is asked to fill out three questionnaires regarding the process: one before the return of results visit, one immediately after it, and a third 6 months later. The results of this pilot will inform future projects involving return of genomic risk information (manuscript in preparation). For instance, when is face-to-face counseling necessary and when would other modes, such as using the participant/patient portal for risk communication, be sufficient?

The most important result of these early translational studies, however, was the subsequent decision of the Estonian government to invest substantially in order to increase the biobank fourfold, to 200,000 subjects, and have all participants genotyped using arrays, and sequence a subset of 3000 individuals to provide a population-based reference panel for imputation of the array data.

2.4. National Personalized Medicine Pilot Projects

While previous projects were conducted in a research context, these national personalized medicine pilots involve biobank participants as patients in clinical practice. To prepare for a nation-wide personalized medicine system (Figure 2), Estonia has just finished the two pilot projects involving participants of the EstBB, but conducted in the hospital settings and by hospital-based physicians and primary care providers (GP's) (manuscript is in preparation). The main goals of these pilots were to test the feasibility of different approaches in clinical practice and demonstrate the benefits of incorporating molecular profiles in disease management in current clinical practice and in hospital IT structures.

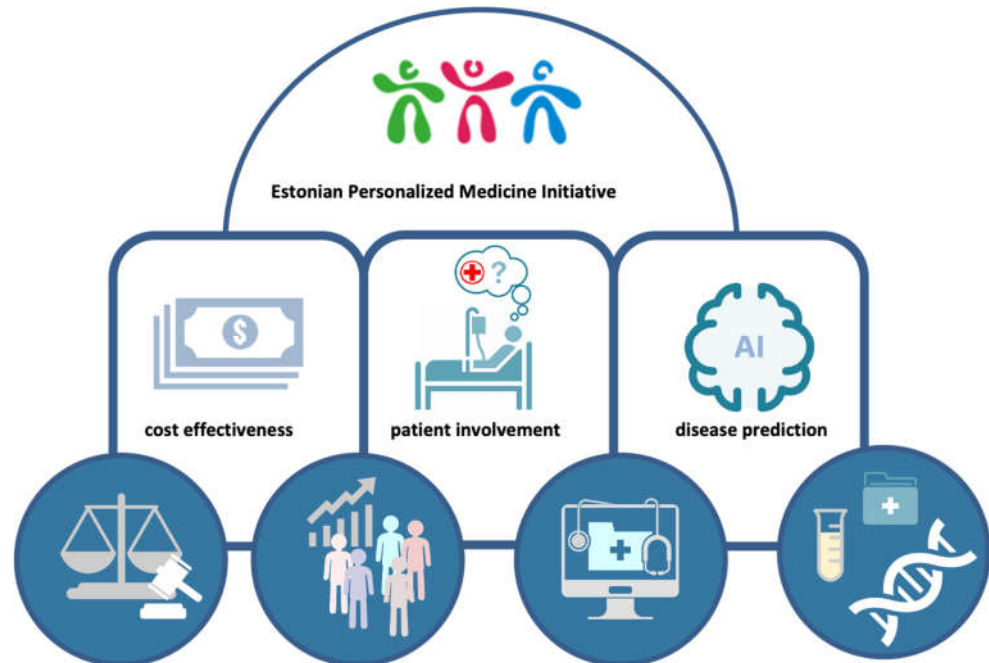


Figure 2. Implementation of personalized medicine. Four core elements (circles) that represent strategic advantages enabling the Estonian Personalized Medicine Initiative to be effectively realized.

Both of these national pilot projects involved 1000 biobank participants. The first project focused on personalized risk prediction and management of breast cancer and involves geneticists in combination with either primary care physicians or oncologists. The second project involved family physicians and focuses on CAD.

The Estonian Biobank research projects as well as the national personalized medicine clinical pilots are expected to lead to the gradual incorporation of personalized genome-based medicine into general health care in Estonia. The initial steps in this direction have been taken by the Ministry of Social Affairs of the Estonian Government when it announced the launch of the Personal Medicine Initiative in 2016 [115].

EstBB currently already actively participates in providing the genomic input both through genomic research and by providing the genotype information. Meanwhile other partners produced the necessary IT infrastructure changes to allow genomic data to be considered not separately but along with other health information, this includes pharmacogenomics integrated with the electronic prescription system, for instance.

As changes in medical care and health information of biobank participants is regularly updated the risk predictions can be improved as new data comes in (Figure 3).

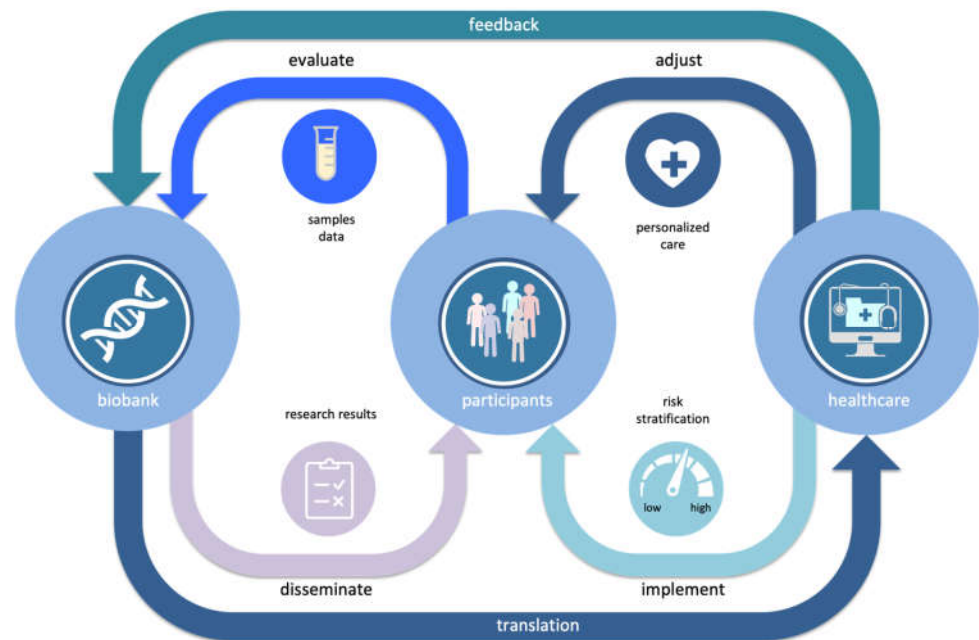


Figure 3. Learning healthcare cycle, where through continuous bi-directional interaction between a population biobank and health care systems new personalized genome-based information from biobank is gradually incorporated into the general health care.

3. Challenges and Pitfalls of Personalized Medicine Implementation in Estonia and Beyond

3.1. Scientific Challenges

The current pilot projects focus on tackling common complex diseases that have the highest burden in the nation and are known to have a substantial genetic basis, being type 2 diabetes, breast cancer, and CAD. One of the first major goals is to develop efficient predictors for the risk of each disease by integrating genetic, health, and environmental factors. The performance of the existing risk-prediction algorithms of the disease can be enhanced by the addition of newly identified genetic components, making the resulting risk estimates more personalized and accurate [116,117]. This in turn allows stratification of individuals into categories of different risk levels, allowing to differentiate between risk-reducing interventions targeted at those categories. Ultimately this will lead to an increase of cost-efficiency of interventions and an improvement of their efficacy.

In this context, however, one of the challenges for researchers is to develop an efficient genetic predictor capturing the polygenic nature of the heritable component of the disease. The traditional approach for developing genetic risk prediction algorithms encompasses the use of only a handful of genome-wide significant SNPs associated with a disease. The biostatistics team at the EstBB that focuses on GRS methodology has demonstrated that a polygenic risk score or ‘predictor’ that captures the effects of a large number of genetic variants in an additive manner, has good predictive ability [117], an approach now widely adopted [118]. Moreover, while conventionally the genetic variants are weighted by their estimated regression coefficients from GWAS, the EstBB biostatistics team has additionally demonstrated that further additional weighting can reduce biases in such weights caused by what is known in statistics as the “winner’s curse” phenomenon [119] and further improves the predictive accuracy. Such a doubly-weighted GRS has already been developed for type 2 diabetes, where it has been shown that the actual risk level for individuals with similar levels of conventional predictors (age, body mass index, physical activity, and diet) may differ by more than three times depending on the level of the GRS [117].

3.2. Data Security and Privacy Challenges

A major concern when conducting genetic research using the EstBB data and samples is honoring the need for privacy of the participants. Under no circumstances should the participant confidentiality be compromised unless he/she reveals it personally.

Most countries have no specific law that regulates usage of genomic data, thereby leaving a legal void that can lead to risky circumstances. For example, genetic information can be valuable for insurance or employment purposes, as genetic tests can be used to detect inherited conditions that might be costly to treat for the healthcare system, and put a burden on employers. This generates concerns regarding for example employment decisions, possibly the invocation of charging higher insurance premiums or even denial of coverage [120], fears that are not unreasonable [121]. These fears can have various consequences. Individuals denying genetic testing can miss out on health benefits these tests can facilitate. Lack of trust can also result in skepticism toward research in genomics and thereby result in decreased study participation, or even a reduction in funding.

Estonia, however, implemented a law in 2000 explicitly concerning the population biobank and genomics data [107]. Specifically, it regulates how genome data are allowed to be gathered and stored, to whom it belongs, and how it is allowed to be used in scientific and commercial research. The key principle is that the owner of the genomic data is always the individual. This is the basis of a robust system based on trust and transparency. All participant data is pseudonymized, and only a handful of individuals have access to the key that enables to link different types of data (phenotypes with genotype) [110]. Digital security is further enhanced by data storage on systems that are completely separate from the internet, making the system virtually unhackable from the outside.

3.3. Challenges in Communicating Research Results

Identifying high risk individuals early, focusing more on disease prevention, while also increasing patients' participation and responsibility in taking care of their own health is expected to help decrease the burden of common complex disease in the long run. One important issue is the automated decision support algorithms and AI tools analyzing the vast and complex data and providing the report for the physician who has to communicate the content to the individual (who is or is not yet a patient). The complex nature of risk determination and stratification through genomic and phenotype data, as well as the probabilistic nature of the results are very likely to be beyond the grasp of the general population.

Therefore, a challenge lies in the communication of these results, which are the aggregate of multiple layers of health-related information (genetic, clinical, psychological, etc.). To minimize the time spent on interpreting complex information, participants should have this information presented in an understandable format and language, with guiding visualizations and explanations (Figure 4 report example). Care must be taken that results are returned together with proper medical support to provide interpretation and translation, which includes genetic counseling, easy access to medical specialists and follow-up.



Type 2 diabetes

Type 2 diabetes is a complex disorder characterised by a high level of blood sugar and at least partial insulin resistance. The contributing factors to type 2 diabetes include both lifestyle and genetic predisposition.

Genetic predisposition

Based on your genetic profile, you are at **average** risk of developing diabetes.



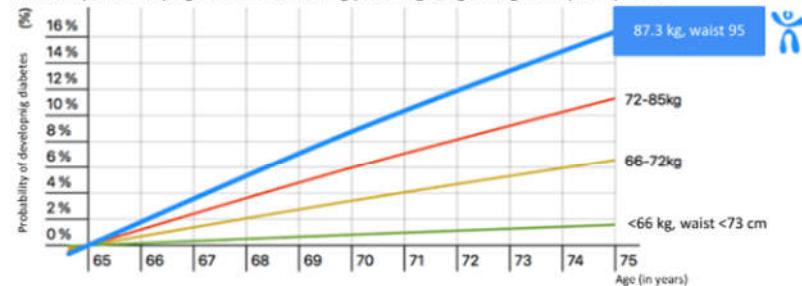
Seven out of 10 people are at lower hereditary risk than you. Two out of 10 people are at higher hereditary risk than you.

Effect of lifestyle and diseases on general risk of diabetes

Your lifestyle and existing conditions place you at **very high** risk of developing diabetes.

Effect of body weight on general risk of diabetes

Probability of developing diabetes considering your weight, age and genetic predisposition.



Total risk of type 2 diabetes

Your current weight places you at **very high** risk of developing type 2 diabetes. The probability of developing the disease in the next 10 years is **24%** and before the age of 70 **13,1%**.

In order to reduce the risk, it is recommended that you:

- maintain a healthy diet and monitor your food portions;
- exercise daily (at least 30 minutes) and engage in other regular physical activities at a moderate level of intensity;
- pay regular attention to your weight (avoid or reduce overweight) and consider consulting a nutritionist, if necessary;
- avoid smoking and excessive alcohol consumption; and
- consider having your diabetes risk assessed by your general practitioner (GP) or another health care professional.

1. Läll K, Magi R, Morris A, Metspalu A, Fischer K. Personalized Risk Prediction for Type 2 Diabetes: the Potential of Genetic Risk Scores. *Genetics in Medicine*. 2017;19(3):322-9.

Response issued:
Report received:

Prepared: 17 May 2018 12:22
3 / 13



Figure 4. Example page in the concept report for an Estonian biobank participant. In a personalized report, the participant will receive information on how her/his risk of the disease will increase as a function of age, how much the genetic component has contributed to risk and to what extent it could be altered by lifestyle modifications or medical intervention.

4. Future Outlook for Personalized Medicine in Estonia and Beyond

The EstBB project as well as the national personalized medicine clinical pilots are expected to lead to the gradual incorporation of personalized genome-based medicine into general health care in Estonia (Box 2). The initial steps in this direction have been taken by the Ministry of Social Affairs of the Estonian Government when it announced the launch of the national Personal Medicine Initiative in 2016 together with a long list of

other stakeholders, including three other ministries, the main Estonian universities and hospitals, as well as private IT solution providers [115]. This project is led by the Institute for Health and Development and will end in 2023. It is currently developing the first three genetically informed decision tools to be used in clinical practice, which consist of polygenic risk scores for breast/ovarian cancer, CAD, and pharmacogenomic traits.

Box 2. A future doctor visit.

When the Estonian National Personalized Medicine Initiative is in full effect, the healthcare system in Estonia will be unlike any other system in the world. An individual born in 21st century Estonia, will have their genome sequenced in high resolution early in adult life, for perhaps no more than GBP 100. It then will be scanned for potential genomics risks for an array of common complex diseases, as well as more high-risk Mendelian disorders and polymorphisms causing adverse drug reactions. The results will be encrypted and stored in a secure database as part of the national biobank. These data are accessible for the individual, who retains ownership and absolute control over accessibility, but also for primary care and specialist physicians and pharmacists, on the condition that the owner provided access permission. Doctor visits and regular health checks populate the electronic health record database, and intelligent learning algorithms will then utilize this new information to improve disease prediction models. During visits, the GP will provide updated genomic risk profiles if needed, identify high risk lifestyle elements that need to be adjusted, and then will devise lifestyle advice and monitoring priorities based on this information. When medication is needed, potential suitable drugs for the treatment will first be cross-checked against a pharmacogenetic profile for potential adverse reactions and potentially needed dose-adjustments. The prescription is being prepared even before the participant leaves the consultation room, and upon arrival at the pharmacy, a national ID card is scanned and medication picked-up straightaway. Meanwhile, genomic and other health-related data is continuously growing and being used (anonymized) by international scientific consortia to identify previously unknown disease mutations and pharmacogenetically important genes. Once confirmed, these are fed back into the database, and individuals with high-risk mutations are informed and invited for a consult to provide suitable preventative care.

Some aspects of the national Personal Medicine Initiative already have tangible impact: at the moment 99% of health data has been digitized, 99% of prescriptions are digital, and 100% of healthcare billing is digital [122]. Moreover, specifically in terms of genomics-informed healthcare, selected findings in the framework of current pilot projects, such as pharmacogenetic profiles, high-risk mutations, interpretation of polygenic risk scores, are already being offered to involved participants. The national personalized medicine services are continuously further improved by, for example, adding and developing follow-up smart IT solutions such as clinical decision support software, but also increasing investments in health management and surveillance and education are being made.

As certain genetic discoveries continue to be made, along with an ever-increasing body of biological insights that are accepted and deposited in electronic health databases, the best practices in personalized medicine will change with time. As such it is key to periodically re-evaluate the medical risk information in light of newly accumulated knowledge and adjust intervention and communication strategies accordingly. Continued education and training of health professionals is also highly important so they are qualified to effectively apply genomic tests and communicate the garnered results to their patients. While family physicians are optimistic about envisioning genomics in their everyday practice they too recognize that additional training is necessary [123]. To prepare the physicians, genomic healthcare workshops are prepared, covering topics on genetic risk scores, pharmacogenomics, and genetic counselling. Over the next two years 900 medical professionals, mainly family physicians and nurses, will be able to participate. Additionally, the National Institute for Health and Development is sending out a monthly newsletter to keep health-care workers informed of developments in personalized medicine arena.

The potential of the Estonian personalized medicine initiative is boosted by the Estonian biobank with over 200,000 participants, the latter representing nearly 20% of the

adult population. In the longer term the plan is to genotype most of the population, with a customized genotyping chip that includes Estonia-specific genetic variation, allowing the development of more precise genetic instruments for medical decision making. The European Union recognizes the need for knowledge transfer and supports these initiatives through ample funding in Horizon Europe programme from 2021 to 2027. Estonia is currently at the forefront of implementing eHealth services and genomic data in health care in the EU, and is sharing its knowledge on national digital healthcare and biobanking by participating actively in the EU Member States' driven initiative "Towards access to at least 1 Million Genomes in the EU by 2022" [124].

Though the future of personalized medicine is looking bright, the universal key to the success of national personalized medicine projects will lie in education and demonstrating the public benefits of genomic screening, and an unrelenting effort to execute these efforts in a human rights framework that guarantee the ethical and secure use of data, and the privacy and ultimate ownership of participants.

Author Contributions: B.P.P. and H.S.: drafted Section 1, L.L. and T.H. drafted Section 2. B.P.P. and T.H. drafted Figure 1. K.P. and B.P.P. drafted Figure 2. B.P.P. and L.L. drafted Figures 3 and 4. All authors edited and approved the final version. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by core funding from: the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG/13/13/30194; RG/18/13/33946) and the National Institute for Health Research (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust) (*)—B.P.P., the Estonian Research Council grant PUT (PRG555)—L.L. and grant PRG687(A.M.) by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012)—T.H., A.M. and Project No. 2014-2020.4.01.16-0024)—K.P. by the European Union through Horizon 2020 grant No. 633589—T.H. and grant No. 777107—PRESICE4Q K.F., T.H.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The views expressed in this review are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

- Spencer, C.C.A.; Su, Z.; Donnelly, P.; Marchini, J. Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *PLoS Genet.* **2009**, *5*, e1000477, doi:10.1371/journal.pgen.1000477.
- Gulcher, J.; Stefansson, K. An Icelandic Saga on a Centralized Healthcare Database and Democratic Decision Making. *Nat. Biotechnol.* **1999**, *17*, 620–620, doi:10.1038/10796.
- Riboli, E.; Kaaks, R. The EPIC Project: Rationale and Study Design. European Prospective Investigation into Cancer and Nutrition. *Int. J. Epidemiol.* **1997**, *26* (Suppl. S1), S6–S14, doi:10.1093/ije/26.suppl_1.s6.
- Sudlow, C.; Gallacher, J.; Allen, N.; Beral, V.; Burton, P.; Danesh, J.; Downey, P.; Elliott, P.; Green, J.; Landray, M.; et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **2015**, *12*, e1001779, doi:10.1371/journal.pmed.1001779.
- Scholtens, S.; Smidt, N.; Swertz, M.A.; Bakker, S.J.L.; Dotinga, A.; Vonk, J.M.; van Dijk, F.; van Zon, S.K.R.; Wijmenga, C.; Wolfenbutter, B.H.R.; et al. Cohort Profile: LifeLines, a Three-Generation Cohort Study and Biobank. *Int. J. Epidemiol.* **2015**, *44*, 1172–1180, doi:10.1093/ije/dyu229.
- Leitsalu, L.; Haller, T.; Esko, T.; Tammesoo, M.-L.; Alavere, H.; Snieder, H.; Perola, M.; Ng, P.C.; Mägi, R.; Milani, L.; et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **2015**, *44*, 1137–1147, doi:10.1093/ije/dyt268.
- Check Hayden, E. The Rise and Fall and Rise Again of 23andMe. *Nat. News* **2017**, *550*, 174, doi:10.1038/550174a.
- GSK. GSK and 23andMe Sign Agreement to Leverage Genetic Insights for the Development of Novel Medicines. Available online: <https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/> (accessed on 11 April 2021).
- Nature Genetics. No Impact without Data Access. *Nat. Genet.* **2015**, *47*, 691, doi:10.1038/ng.3351.
- Barsh, G.S.; Cooper, G.M.; Copenhaver, G.P.; Gibson, G.; McCarthy, M.I.; Tang, H.; Williams, S.M. PLoS Genetics Data Sharing Policy: In Pursuit of Functional Utility. *PLoS Genet.* **2015**, *11*, e1005716, doi:10.1371/journal.pgen.1005716.

11. Paltoo, D.N.; Rodriguez, L.L.; Feolo, M.; Gillanders, E.; Ramos, E.M.; Rutter, J.L.; Sherry, S.; Wang, V.O.; Bailey, A.; Baker, R.; et al. Data Use under the NIH GWAS Data Sharing Policy and Future Directions. *Nat. Genet.* **2014**, *46*, 934–938, doi:10.1038/ng.3062.
12. Wellcome Trust Case Control Consortium Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3000 Shared Controls. *Nature* **2007**, *447*, 661–678, doi:10.1038/nature05911.
13. Mailman, M.D.; Feolo, M.; Jin, Y.; Kimura, M.; Tryka, K.; Bagoutdinov, R.; Hao, L.; Kiang, A.; Paschall, J.; Phan, L.; et al. The NCBI DbGaP Database of Genotypes and Phenotypes. *Nat. Genet.* **2007**, *39*, 1181–1186, doi:10.1038/ng1007-1181.
14. Lappalainen, I.; Almeida-King, J.; Kumanduri, V.; Senf, A.; Spalding, J.D.; Ur-Rehman, S.; Saunders, G.; Kandasamy, J.; Caccamo, M.; Leinonen, R.; et al. The European Genome-Phenome Archive of Human Data Consented for Biomedical Research. *Nat. Genet.* **2015**, *47*, 692–695, doi:10.1038/ng.3312.
15. Gilly, A. Metacarpa: META-Analysis in C++ Accounting for Relatedness, Using Arbitrary Precision Arithmetic. Available online: <https://bitbucket.org/agilly/metacarpa/src/master/> (accessed on 14 April 2021).
16. Hemani, G.; Zheng, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-Base Platform Supports Systematic Causal Inference across the Human Phenome. *eLife* **2018**, *7*, e34408, doi:10.7554/eLife.34408.
17. Zheng, J.; Erzurumluoglu, A.M.; Elsworth, B.L.; Kemp, J.P.; Howe, L.; Haycock, P.C.; Hemani, G.; Tansey, K.; Laurin, C.; Pourcain, B.S.; et al. LD Hub: A Centralized Database and Web Interface to Perform LD Score Regression That Maximizes the Potential of Summary Level GWAS Data for SNP Heritability and Genetic Correlation Analysis. *Bioinformatics* **2017**, *33*, 272–279, doi:10.1093/bioinformatics/btw613.
18. Langille, M.G.I.; Eisen, J.A. BioTorrents: A File Sharing Service for Scientific Data. *PLoS ONE* **2010**, *5*, e10071, doi:10.1371/journal.pone.0010071.
19. Brody, J.A.; Morrison, A.C.; Bis, J.C.; O’Connell, J.R.; Brown, M.R.; Huffman, J.E.; Ames, D.C.; Carroll, A.; Conomos, M.P.; Gabriel, S.; et al. Analysis Commons, a Team Approach to Discovery in a Big-Data Environment for Genetic Epidemiology. *Nat. Genet.* **2017**, *49*, 1560–1563, doi:10.1038/ng.3968.
20. Walter, K.; Min, J.L.; Huang, J.; Crooks, L.; Memari, Y.; McCarthy, S.; Perry, J.R.B.; Xu, C.; Futema, M.; UK10K Consortium; et al. The UK10K Project Identifies Rare Variants in Health and Disease. *Nature* **2015**, *526*, 82–90, doi:10.1038/nature14962.
21. Wood, A.R.; Perry, J.R.B.; Tanaka, T.; Hernandez, D.G.; Zheng, H.-F.; Melzer, D.; Gibbs, J.R.; Nalls, M.A.; Weedon, M.N.; Spector, T.D.; et al. Imputation of Variants from the 1000 Genomes Project Modestly Improves Known Associations and Can Identify Low-Frequency Variant-Phenotype Associations Undetected by HapMap Based Imputation. *PLoS ONE* **2013**, *8*, e64343, doi:10.1371/journal.pone.0064343.
22. McCarthy, S.; Das, S.; Kretzschmar, W.; Delaneau, O.; Wood, A.R.; Teumer, A.; Kang, H.M.; Fuchsberger, C.; Danecek, P.; Sharp, K.; et al. A Reference Panel of 64,976 Haplotypes for Genotype Imputation. *Nat. Genet.* **2016**, *48*, 1279–1283, doi:10.1038/ng.3643.
23. Wellcome Trust Sanger Institute Communications Team 500,000 Whole Human Genomes Will Be a Game-Changer for Research into Human Diseases-Wellcome Sanger Institute. Available online: https://www.sanger.ac.uk/news_item/500000-whole-human-genomes-will-be-game-changer-research-human-diseases/ (accessed on 15 April 2021).
24. Health, A.G.D. of Genomics Health Futures Mission. Available online: <https://www.health.gov.au/initiatives-and-programs/genomics-health-futures-mission> (accessed on 14 April 2021).
25. Lévy, Y. Genomic Medicine 2025: France in the Race for Precision Medicine. *Lancet Lond. Engl.* **2016**, *388*, 2872, doi:10.1016/S0140-6736(16)32467-9.
26. Denny, J.C.; Rutter, J.L.; Goldstein, D.B.; Philippakis, A.; Smoller, J.W.; Jenkins, G.; Dishman, E.; All of Us Research Program Investigators. The “All of Us” Research Program. *N. Engl. J. Med.* **2019**, *381*, 668–676, doi:10.1056/NEJMs1809937.
27. Gurdasani, D.; Carstensen, T.; Tekola-Ayele, F.; Pagani, L.; Tachmazidou, I.; Hatzikotoulas, K.; Karthikeyan, S.; Iles, L.; Pollard, M.O.; Choudhury, A.; et al. The African Genome Variation Project Shapes Medical Genetics in Africa. *Nature* **2015**, *517*, 327–332, doi:10.1038/nature13997.
28. Mulder, N.; Abimiku, A.; Adebamowo, S.N.; de Vries, J.; Matimba, A.; Olowoyo, P.; Ramsay, M.; Skelton, M.; Stein, D.J. H3Africa: Current Perspectives. *Pharm. Pers. Med.* **2018**, *11*, 59–66, doi:10.2147/PGPM.S141546.
29. Wall, J.D.; Stawiski, E.W.; Ratan, A.; Kim, H.L.; Kim, C.; Gupta, R.; Suryamohan, K.; Gusareva, E.S.; Purbojati, R.W.; Bhangale, T.; et al. The GenomeAsia 100K Project Enables Genetic Discoveries across Asia. *Nature* **2019**, *576*, 106–111, doi:10.1038/s41586-019-1793-z.
30. Ledford, H. AstraZeneca Launches Project to Sequence 2 Million Genomes. *Nat. News* **2016**, *532*, 427, doi:10.1038/nature.2016.19797.
31. Guo, Y.; Dai, Y.; Yu, H.; Zhao, S.; Samuels, D.C.; Shyr, Y. Improvements and Impacts of GRCh38 Human Reference on High Throughput Sequencing Data Analysis. *Genomics* **2017**, *109*, 83–90, doi:10.1016/j.ygeno.2017.01.005.
32. König, I.R.; Loley, C.; Erdmann, J.; Ziegler, A. How to Include Chromosome X in Your Genome-Wide Association Study. *Genet. Epidemiol.* **2014**, *38*, 97–103, doi:10.1002/gepi.21782.
33. Wise, A.L.; Gyi, L.; Manolio, T.A. EXclusion: Toward Integrating the X Chromosome in Genome-Wide Association Analyses. *Am. J. Hum. Genet.* **2013**, *92*, 643–647, doi:10.1016/j.ajhg.2013.03.017.
34. Pesole, G.; Allen, J.F.; Lane, N.; Martin, W.; Rand, D.M.; Schatz, G.; Saccone, C. The Neglected Genome. *EMBO Rep.* **2012**, *13*, 473–474, doi:10.1038/embor.2012.57.

35. Chang, D.; Gao, F.; Slavney, A.; Ma, L.; Waldman, Y.Y.; Sams, A.J.; Billing-Ross, P.; Madar, A.; Spritz, R.; Keinan, A. Accounting for EXentricities: Analysis of the X Chromosome in GWAS Reveals X-Linked Genes Implicated in Autoimmune Diseases. *PLoS ONE* **2014**, *9*, e113684, doi:10.1371/journal.pone.0113684.
36. Case, L.K.; Wall, E.H.; Osmanski, E.E.; Dragon, J.A.; Saligrama, N.; Zachary, J.F.; Lemos, B.; Blankenhorn, E.P.; Teuscher, C. Copy Number Variation in Y Chromosome Multicopy Genes Is Linked to a Paternal Parent-of-Origin Effect on CNS Autoimmune Disease in Female Offspring. *Genome Biol.* **2015**, *16*, 28, doi:10.1186/s13059-015-0591-7.
37. Hudson, G.; Gomez-Duran, A.; Wilson, I.J.; Chinnery, P.F. Recent Mitochondrial DNA Mutations Increase the Risk of Developing Common Late-Onset Human Diseases. *PLoS Genet* **2014**, *10*, e1004369, doi:10.1371/journal.pgen.1004369.
38. Baird, J. YGEN Project. Available online: <https://www.wiki.ed.ac.uk/display/YGEN/Ygen+Home> (accessed on 15 April 2021).
39. Yamamoto, K.; Sakaue, S.; Matsuda, K.; Murakami, Y.; Kamatani, Y.; Ozono, K.; Momozawa, Y.; Okada, Y. Genetic and Phenotypic Landscape of the Mitochondrial Genome in the Japanese Population. *Commun. Biol.* **2020**, *3*, 1–11, doi:10.1038/s42003-020-0812-9.
40. Dilthey, A.T.; Moutsianas, L.; Leslie, S.; McVean, G. HLA*IMP—An Integrated Framework for Imputing Classical HLA Alleles from SNP Genotypes. *Bioinforma. Oxf. Engl.* **2011**, *27*, 968–972, doi:10.1093/bioinformatics/btr061.
41. Vukcevic, D.; Traherne, J.A.; Naess, S.; Ellinghaus, E.; Kamatani, Y.; Dilthey, A.; Lathrop, M.; Karlsen, T.H.; Franke, A.; Moffatt, M.; et al. Imputation of KIR Types from SNP Variation Data. *Am. J. Hum. Genet.* **2015**, *97*, 593–607, doi:10.1016/j.ajhg.2015.09.005.
42. Hindorf, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9362–9367, doi:10.1073/pnas.0903103106.
43. Cooper, G.M.; Shendure, J. Needles in Stacks of Needles: Finding Disease-Causal Variants in a Wealth of Genomic Data. *Nat. Rev. Genet.* **2011**, *12*, 628–640, doi:10.1038/nrg3046.
44. Kircher, M.; Witten, D.M.; Jain, P.; O’Roak, B.J.; Cooper, G.M.; Shendure, J. A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants. *Nat. Genet.* **2014**, *46*, 310–315, doi:10.1038/ng.2892.
45. Ionita-Laza, I.; McCallum, K.; Xu, B.; Buxbaum, J.D. A Spectral Approach Integrating Functional Genomic Annotations for Coding and Noncoding Variants. *Nat. Genet.* **2016**, *48*, 214, doi:10.1038/ng.3477.
46. Chen, W.; Larrabee, B.R.; Ovsyannikova, I.G.; Kennedy, R.B.; Haralambieva, I.H.; Poland, G.A.; Schaid, D.J. Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics. *Genetics* **2015**, *200*, 719–736, doi:10.1534/genetics.115.176107.
47. Farh, K.K.-H.; Marson, A.; Zhu, J.; Kleinewietfeld, M.; Housley, W.J.; Beik, S.; Shores, N.; Whitton, H.; Ryan, R.J.H.; Shishkin, A.A.; et al. Genetic and Epigenetic Fine Mapping of Causal Autoimmune Disease Variants. *Nature* **2015**, *518*, 337–343, doi:10.1038/nature13835.
48. Asimit, J.L.; Hatzikotoulas, K.; McCarthy, M.; Morris, A.P.; Zeggini, E. Trans-Ethnic Study Design Approaches for Fine-Mapping. *Eur. J. Hum. Genet.* **2016**, *24*, 1330–1336, doi:10.1038/ejhg.2016.1.
49. Tachmazidou, I.; Dedoussis, G.; Southam, L.; Farmaki, A.-E.; Ritchie, G.R.S.; Xifara, D.K.; Matchan, A.; Hatzikotoulas, K.; Rayner, N.W.; Chen, Y.; et al. A Rare Functional Cardioprotective APOC3 Variant Has Risen in Frequency in Distinct Population Isolates. *Nat. Commun.* **2013**, *4*, 2872, doi:10.1038/ncomms3872.
50. Sulem, P.; Gudbjartsson, D.F.; Walters, G.B.; Helgadóttir, H.T.; Helgason, A.; Gudjonsson, S.A.; Zanon, C.; Besenbacher, S.; Bjornsdóttir, G.; Magnusson, O.T.; et al. Identification of Low-Frequency Variants Associated with Gout and Serum Uric Acid Levels. *Nat. Genet.* **2011**, *43*, 1127–1130, doi:10.1038/ng.972.
51. Schwarze, K.; Buchanan, J.; Fermont, J.M.; Drea, H.; Tilley, M.W.; Taylor, J.M.; Antoniou, P.; Knight, S.J.L.; Camps, C.; Pentony, M.M.; et al. The Complete Costs of Genome Sequencing: A Microcosting Study in Cancer and Rare Diseases from a Single Center in the United Kingdom. *Genet. Med.* **2020**, *22*, 85–94, doi:10.1038/s41436-019-0618-7.
52. Shafin, K.; Pesout, T.; Lorig-Roach, R.; Haukness, M.; Olsen, H.E.; Bosworth, C.; Armstrong, J.; Tigyi, K.; Maurer, N.; Koren, S.; et al. Nanopore Sequencing and the Shasta Toolkit Enable Efficient de Novo Assembly of Eleven Human Genomes. *Nat. Biotechnol.* **2020**, *38*, 1044–1053, doi:10.1038/s41587-020-0503-6.
53. Layer, R.M.; Kindlon, N.; Karczewski, K.J.; Exome Aggregation Consortium; Quinlan, A.R. Efficient Genotype Compression and Analysis of Large Genetic-Variation Data Sets. *Nat. Methods* **2016**, *13*, 63–65, doi:10.1038/nmeth.3654.
54. Kumar, V.; Muratani, M.; Rayan, N.A.; Kraus, P.; Lufkin, T.; Ng, H.H.; Prabhakar, S. Uniform, Optimal Signal Processing of Mapped Deep-Sequencing Data. *Nat. Biotechnol.* **2013**, *31*, 615–622, doi:10.1038/nbt.2596.
55. Libbrecht, M.W.; Noble, W.S. Machine Learning Applications in Genetics and Genomics. *Nat. Rev. Genet.* **2015**, *16*, 321–332, doi:10.1038/nrg3920.
56. Zhao, S.; Agafonov, O.; Azab, A.; Stokowy, T.; Hovig, E. Accuracy and Efficiency of Germline Variant Calling Pipelines for Human Genome Data. *Sci. Rep.* **2020**, *10*, 20222, doi:10.1038/s41598-020-77218-4.
57. Children’s Hospital of Philadelphia. Edico Set World Record for Secondary Analysis Speed. Available online: <https://www.bioitworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed> (accessed on 14 April 2021).
58. Hail Team. Scalable Genomic Data Analysis. Available online: <https://github.com/hail-is/hail> (accessed on 15 April 2021).
59. Siretskiy, A.; Sundqvist, T.; Voznesenskiy, M.; Spjuth, O. A Quantitative Assessment of the Hadoop Framework for Analyzing Massively Parallel DNA Sequencing Data. *GigaScience* **2015**, *4*, 26, doi:10.1186/s13742-015-0058-5.

60. The ENCODE Project Consortium an Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **2012**, *489*, 57–74, doi:10.1038/nature11247.
61. Bernstein, B.E.; Stamatoyannopoulos, J.A.; Costello, J.F.; Ren, B.; Milosavljevic, A.; Meissner, A.; Kellis, M.; Marra, M.A.; Beaudet, A.L.; Ecker, J.R.; et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **2010**, *28*, 1045–1048, doi:10.1038/nbt1010-1045.
62. Martens, J.H.A.; Stunnenberg, H.G. BLUEPRINT: Mapping Human Blood Cell Epigenomes. *Haematologica* **2013**, *98*, 1487–1489, doi:10.3324/haematol.2013.094243.
63. Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C.H.; Corthals, G.L.; Costello, C.E.; et al. The Human Proteome Project: Current State and Future Direction. *Mol. Cell. Proteomics MCP* **2011**, *10*, doi:10.1074/mcp.M111.009993.
64. Regev, A.; Teichmann, S.A.; Lander, E.S.; Amit, I.; Benoist, C.; Birney, E.; Bodenmiller, B.; Campbell, P.; Carninci, P.; Clatworthy, M.; et al. The Human Cell Atlas. *eLife* **2017**, *6*, doi:10.7554/eLife.27041.
65. Streeter, I.; Harrison, P.W.; Faulconbridge, A.; Flicek, P.; Parkinson, H.; Clarke, L. The Human-Induced Pluripotent Stem Cell Initiative—Data Resources for Cellular Genetics. *Nucleic Acids Res.* **2017**, *45*, D691–D697, doi:10.1093/nar/gkw928.
66. Morgan, X.C.; Tickle, T.L.; Sokol, H.; Gevers, D.; Devaney, K.L.; Ward, D.V.; Reyes, J.A.; Shah, S.A.; LeLeiko, N.; Snapper, S.B.; et al. Dysfunction of the Intestinal Microbiome in Inflammatory Bowel Disease and Treatment. *Genome Biol.* **2012**, *13*, R79, doi:10.1186/gb-2012-13-9-r79.
67. Castro-Nallar, E.; Bendall, M.L.; Pérez-Losada, M.; Sabuncyan, S.; Severance, E.G.; Dickerson, F.B.; Schroeder, J.R.; Yolken, R.H.; Crandall, K.A. Composition, Taxonomy and Functional Diversity of the Oropharynx Microbiome in Individuals with Schizophrenia and Controls. *PeerJ* **2015**, *3*, e1140, doi:10.7717/peerj.1140.
68. Wang, Z.; Roberts, A.B.; Buffa, J.A.; Levison, B.S.; Zhu, W.; Org, E.; Gu, X.; Huang, Y.; Zamanian-Daryoush, M.; Culley, M.K.; et al. Non-Lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. *Cell* **2015**, *163*, 1585–1595, doi:10.1016/j.cell.2015.11.055.
69. Blekhman, R.; Goodrich, J.K.; Huang, K.; Sun, Q.; Bukowski, R.; Bell, J.T.; Spector, T.D.; Keinan, A.; Ley, R.E.; Gevers, D.; et al. Host Genetic Variation Impacts Microbiome Composition across Human Body Sites. *Genome Biol.* **2015**, *16*, 191, doi:10.1186/s13059-015-0759-1.
70. Bai, W.; Shi, W.; de Marvao, A.; Dawes, T.J.W.; O'Regan, D.P.; Cook, S.A.; Rueckert, D. A Bi-Ventricular Cardiac Atlas Built from 1000+ High Resolution MR Images of Healthy Subjects and an Analysis of Shape and Motion. *Med. Image Anal.* **2015**, *26*, 133–145, doi:10.1016/j.media.2015.08.009.
71. Thompson, P.M.; Stein, J.L.; Medland, S.E.; Hibar, D.P.; Vasquez, A.A.; Renteria, M.E.; Toro, R.; Jahanshad, N.; Schumann, G.; Franke, B.; et al. The ENIGMA Consortium: Large-Scale Collaborative Analyses of Neuroimaging and Genetic Data. *Brain Imaging Behav.* **2014**, *8*, 153–182, doi:10.1007/s11682-013-9269-5.
72. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of Integrating Data to Uncover Genotype-Phenotype Interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97, doi:10.1038/nrg3868.
73. Ernst, J.; Kellis, M. Large-Scale Imputation of Epigenomic Datasets for Systematic Annotation of Diverse Human Tissues. *Nat. Biotechnol.* **2015**, *33*, 364–376, doi:10.1038/nbt.3157.
74. Gamazon, E.R.; Wheeler, H.E.; Shah, K.P.; Mozaffari, S.V.; Aquino-Michaels, K.; Carroll, R.J.; Eyler, A.E.; Denny, J.C.; Nicolae, D.L.; GTEx Consortium; et al. A Gene-Based Association Method for Mapping Traits Using Reference Transcriptome Data. *Nat. Genet.* **2015**, *47*, 1091–1098, doi:10.1038/ng.3367.
75. Trynka, G.; Sandor, C.; Han, B.; Xu, H.; Stranger, B.E.; Liu, X.S.; Raychaudhuri, S. Chromatin Marks Identify Critical Cell Types for Fine Mapping Complex Trait Variants. *Nat. Genet.* **2013**, *45*, 124–130, doi:10.1038/ng.2504.
76. Fu, J.; Wolfs, M.G.M.; Deelen, P.; Westra, H.-J.; Fehrmann, R.S.N.; Te Meerman, G.J.; Buurman, W.A.; Rensen, S.S.M.; Groen, H.J.M.; Weersma, R.K.; et al. Unraveling the Regulatory Mechanisms Underlying Tissue-Dependent Genetic Variation of Gene Expression. *PLoS Genet.* **2012**, *8*, e1002431, doi:10.1371/journal.pgen.1002431.
77. Brown, C.D.; Mangravite, L.M.; Engelhardt, B.E. Integrative Modeling of EQTLs and Cis-Regulatory Elements Suggests Mechanisms Underlying Cell Type Specificity of EQTLs. *PLoS Genet.* **2013**, *9*, e1003649, doi:10.1371/journal.pgen.1003649.
78. Weinshilboum, R.; Wang, L. Pharmacogenomics: Precision Medicine and Drug Response. *Mayo Clin. Proc.* **2017**, *92*, 1711–1722, doi:10.1016/j.mayocp.2017.09.001.
79. Visscher, P.M.; Wray, N.R.; Zhang, Q.; Sklar, P.; McCarthy, M.I.; Brown, M.A.; Yang, J. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **2017**, *101*, 5–22, doi:10.1016/j.ajhg.2017.06.005.
80. Gallagher, M.D.; Chen-Plotkin, A.S. The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* **2018**, *102*, 717–730, doi:10.1016/j.ajhg.2018.04.002.
81. Thomas, M.; Sakoda, L.C.; Hoffmeister, M.; Rosenthal, E.A.; Lee, J.K.; van Duijnhoven, F.J.B.; Platz, E.A.; Wu, A.H.; Dampier, C.H.; de la Chapelle, A.; et al. Genome-Wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am. J. Hum. Genet.* **2020**, *107*, 432–444, doi:10.1016/j.ajhg.2020.07.006.
82. Forgetta, V.; Keller-Baruch, J.; Forest, M.; Durand, A.; Bhatnagar, S.; Kemp, J.P.; Nethander, M.; Evans, D.; Morris, J.A.; Kiel, D.P.; et al. Development of a Polygenic Risk Score to Improve Screening for Fracture Risk: A Genetic Risk Prediction Study. *PLoS Med.* **2020**, *17*, e1003152, doi:10.1371/journal.pmed.1003152.

83. Inouye, M.; Abraham, G.; Nelson, C.P.; Wood, A.M.; Sweeting, M.J.; Dudbridge, F.; Lai, F.Y.; Kaptoge, S.; Brozynska, M.; Wang, T.; et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J. Am. Coll. Cardiol.* **2018**, *72*, 1883–1893, doi:10.1016/j.jacc.2018.07.079.
84. Burgess, S.; Freitag, D.F.; Khan, H.; Gorman, D.N.; Thompson, S.G. Using Multivariable Mendelian Randomization to Disentangle the Causal Effects of Lipid Fractions. *PLoS ONE* **2014**, *9*, e108891, doi:10.1371/journal.pone.0108891.
85. Navarese, E.P.; Kolodziejczak, M.; Schulze, V.; Gurbel, P.A.; Tantry, U.; Lin, Y.; Brockmeyer, M.; Kandzari, D.E.; Kubica, J.M.; D'Agostino, R.B.; et al. Effects of Proprotein Convertase Subtilisin/Kexin Type 9 Antibodies in Adults With Hypercholesterolemia: A Systematic Review and Meta-Analysis. *Ann. Intern. Med.* **2015**, *163*, 40–51, doi:10.7326/M14-2957.
86. Sirota, M.; Schaub, M.A.; Batzoglou, S.; Robinson, W.H.; Butte, A.J. Autoimmune Disease Classification by Inverse Association with SNP Alleles. *PLoS Genet.* **2009**, *5*, e1000792, doi:10.1371/journal.pgen.1000792.
87. Lipsky, P.E.; van der Heijde, D.M.; St Clair, E.W.; Furst, D.E.; Breedveld, F.C.; Kalden, J.R.; Smolen, J.S.; Weisman, M.; Emery, P.; Feldmann, M.; et al. Infliximab and Methotrexate in the Treatment of Rheumatoid Arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. *N. Engl. J. Med.* **2000**, *343*, 1594–1602, doi:10.1056/NEJM200011303432202.
88. Grainger, R.; Harrison, A.A. Infliximab in the Treatment of Ankylosing Spondylitis. *Biol. Targets Ther.* **2007**, *1*, 163–171.
89. Lin, J.; Ziring, D.; Desai, S.; Kim, S.; Wong, M.; Korin, Y.; Braun, J.; Reed, E.; Gjertson, D.; Singh, R.R. TNF α Blockade in Human Diseases: An Overview of Efficacy and Safety. *Clin. Immunol. Orlando Fla* **2008**, *126*, 13–30, doi:10.1016/j.clim.2007.08.012.
90. Singer, J.B.; Lewitzky, S.; Leroy, E.; Yang, F.; Zhao, X.; Klickstein, L.; Wright, T.M.; Meyer, J.; Paulding, C.A. A Genome-Wide Study Identifies HLA Alleles Associated with Lumiracoxib-Related Liver Injury. *Nat. Genet.* **2010**, *42*, 711–714, doi:10.1038/ng.632.
91. Heap, G.A.; Weedon, M.N.; Bewshea, C.M.; Singh, A.; Chen, M.; Satchwell, J.B.; Vivian, J.P.; So, K.; Dubois, P.C.; Andrews, J.M.; et al. HLA-DQA1-HLA-DRB1 Variants Confer Susceptibility to Pancreatitis Induced by Thiopurine Immunosuppressants. *Nat. Genet.* **2014**, *46*, 1131–1134, doi:10.1038/ng.3093.
92. Yang, S.-K.; Hong, M.; Baek, J.; Choi, H.; Zhao, W.; Jung, Y.; Haritunians, T.; Ye, B.D.; Kim, K.-J.; Park, S.H.; et al. A Common Missense Variant in NUDT15 Confers Susceptibility to Thiopurine-Induced Leukopenia. *Nat. Genet.* **2014**, *46*, 1017–1020, doi:10.1038/ng.3060.
93. Wang, Y.; Tong, C.; Wang, Z.; Wang, Z.; Mauger, D.; Tantisira, K.G.; Israel, E.; Szeffler, S.J.; Chinchilli, V.M.; Boushey, H.A.; et al. Pharmacodynamic Genome-Wide Association Study Identifies New Responsive Loci for Glucocorticoid Intervention in Asthma. *Pharm. J.* **2015**, *15*, 422–429, doi:10.1038/tpj.2014.83.
94. Rader, D.J. New Therapies for Coronary Artery Disease: Genetics Provides a Blueprint. *Sci. Transl. Med.* **2014**, *6*, 239ps4–239ps4, doi:10.1126/scitranslmed.3008535.
95. Russ, A.P.; Lampel, S. The Druggable Genome: An Update. *Drug Discov. Today* **2005**, *10*, 1607–1610, doi:10.1016/S1359-6446(05)03666-4.
96. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672, doi:10.1093/nar/gkj067.
97. Chen, X.; Ji, Z.L.; Chen, Y.Z. TTD: Therapeutic Target Database. *Nucleic Acids Res.* **2002**, *30*, 412–415.
98. Hewett, M.; Oliver, D.E.; Rubin, D.L.; Easton, K.L.; Stuart, J.M.; Altman, R.B.; Klein, T.E. PharmGKB: The Pharmacogenetics Knowledge Base. *Nucleic Acids Res.* **2002**, *30*, 163–165.
99. Fang, H.; Wolf, H.D.; Knezevic, B.; Burnham, K.L.; Osgood, J.; Sanniti, A.; Lara, A.L.; Kasela, S.; de Cesco, S.; Egner, J.K.W.; et al. A Genetics-Led Approach Defines the Drug Target Landscape of 30 Immune-Related Traits. *Nat. Genet.* **2019**, *51*, 1082–1091, doi:10.1038/s41588-019-0456-1.
100. Kirchhof, P.; Sipido, K.R.; Cowie, M.R.; Eschenhagen, T.; Fox, K.A.A.; Katus, H.; Schroeder, S.; Schunkert, H.; Priori, S.; ESC CRT R&D and European Affairs Work Shop on Personalized Medicine the Continuum of Personalized Cardiovascular Medicine: A Position Paper of the European Society of Cardiology. *Eur. Heart J.* **2014**, *35*, 3250–3257, doi:10.1093/eurheartj/ehu312.
101. Enroth, S.; Johansson, A.; Enroth, S.B.; Gyllenstein, U. Strong Effects of Genetic and Lifestyle Factors on Biomarker Variation and Use of Personalized Cutoffs. *Nat. Commun.* **2014**, *5*, 4684, doi:10.1038/ncomms5684.
102. Morgan, P.; Brown, D.G.; Lennard, S.; Anderton, M.J.; Barrett, J.C.; Eriksson, U.; Fidock, M.; Hamrén, B.; Johnson, A.; March, R.E.; et al. Impact of a Five-Dimensional Framework on R&D Productivity at AstraZeneca. *Nat. Rev. Drug Discov.* **2018**, *17*, 167–181, doi:10.1038/nrd.2017.244.
103. Weitzel, K.W.; Smith, D.M.; Elsey, A.R.; Duong, B.Q.; Burkley, B.; Clare-Salzler, M.; Gong, Y.; Higgins, T.A.; Kong, B.; Langae, T.; et al. Implementation of Standardized Clinical Processes for TPMT Testing in a Diverse Multidisciplinary Population: Challenges and Lessons Learned. *Clin. Transl. Sci.* **2018**, *11*, 175–181, doi:10.1111/cts.12533.
104. Cheng, L.; Lopez-Beltran, A.; Massari, F.; MacLennan, G.T.; Montironi, R. Molecular Testing for BRAF Mutations to Inform Melanoma Treatment Decisions: A Move toward Precision Medicine. *Mod. Pathol.* **2018**, *31*, 24–38, doi:10.1038/mod-pathol.2017.104.
105. In Deutschland zugelassene Arzneimittel für die Personalisierte Medizin. Available online: vfa.de/personalisiert (accessed on 22 November 2019).
106. Stark, Z.; Dolman, L.; Manolio, T.A.; Ozenberger, B.; Hill, S.L.; Caulfield, M.J.; Levy, Y.; Glazer, D.; Wilson, J.; Lawler, M.; et al. Integrating Genomics into Healthcare: A Global Responsibility. *Am. J. Hum. Genet.* **2019**, *104*, 13–20, doi:10.1016/j.ajhg.2018.11.014.

107. Human Genes Research Act—Riigi Teataja. Available online: <https://www.riigiteataja.ee/en/eli/531102013003/consolide> (accessed on 14 April 2021).
108. Leitsalu, L. Communicating Genomic Research Results to Population-Based Biobank Participants. Ph.D. Thesis, University of Tartu, Tartu, Estonia, 2016.
109. Sepper, R.; Ross, P.; Tiik, M. Nationwide Health Data Management System: A Novel Approach for Integrating Biomarker Measurements with Comprehensive Health Records in Large Populations Studies. *J. Proteome Res.* **2011**, *10*, 97–100, doi:10.1021/pr1007784.
110. Priisalu, J.; Ottis, R. Personal Control of Privacy and Data: Estonian Experience. *Health Technol.* **2017**, *7*, 441–451, doi:10.1007/s12553-017-0195-1.
111. Leitsalu, L.; Alavere, H.; Tammesoo, M.-L.; Leego, E.; Metspalu, A. Linking a Population Biobank with National Health Registries—The Estonian Experience. *J. Pers. Med.* **2015**, *5*, 96–106, doi:10.3390/jpm5020096.
112. Leitsalu, L.; Alavere, H.; Jacquemont, S.; Kolk, A.; Maillard, A.M.; Reigo, A.; Nõukas, M.; Reymond, A.; Männik, K.; Ng, P.C.; et al. Reporting Incidental Findings of Genomic Disorder-Associated Copy Number Variants to Unselected Biobank Participants. *Pers. Med.* **2016**, *13*, 303–314.
113. Alver, M.; Palover, M.; Saar, A.; Läll, K.; Zekavat, S.M.; Tõnisson, N.; Leitsalu, L.; Reigo, A.; Nikopensus, T.; Ainla, T.; et al. Recall by Genotype and Cascade Screening for Familial Hypercholesterolemia in a Population-Based Biobank from Estonia. *Genet. Med.* **2018**, *1*, doi:10.1038/s41436-018-0311-2.
114. Reisberg, S.; Krebs, K.; Lepamets, M.; Kals, M.; Mägi, R.; Metsalu, K.; Lauschke, V.M.; Vilo, J.; Milani, L. Translating Genotype Data of 44,000 Biobank Participants into Clinical Pharmacogenetic Recommendations: Challenges and Solutions. *Genet. Med.* **2018**, *1*, doi:10.1038/s41436-018-0337-5.
115. Ministry of Social Affairs. Preliminary Study on the Personal Medicine Project. Available online: <http://www.sm.ee/et/personaalmeditsiini-juhtprojekti-eeluuring> (accessed on 11 April 2021).
116. Läll, K.; Lepamets, M.; Palover, M.; Esko, T.; Metspalu, A.; Tõnisson, N.; Padrik, P.; Mägi, R.; Fischer, K. Polygenic Prediction of Breast Cancer: Comparison of Genetic Predictors and Implications for Screening. *bioRxiv* **2018**, 448597, doi:10.1101/448597.
117. Läll, K.; Mägi, R.; Morris, A.; Metspalu, A.; Fischer, K. Personalized Risk Prediction for Type 2 Diabetes: The Potential of Genetic Risk Scores. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2017**, *19*, 322–329, doi:10.1038/gim.2016.103.
118. Gibson, G. On the Utilization of Polygenic Risk Scores for Therapeutic Targeting. *PLoS Genet.* **2019**, *15*, e1008060, doi:10.1371/journal.pgen.1008060.
119. Garner, C. Upward Bias in Odds Ratio Estimates from Genome-Wide Association Studies. *Genet. Epidemiol.* **2007**, *31*, 288–295, doi:10.1002/gepi.20209.
120. Wauters, A.; Van Hoyweghen, I. Global Trends on Fears and Concerns of Genetic Discrimination: A Systematic Literature Review. *J. Hum. Genet.* **2016**, *61*, 275–282, doi:10.1038/jhg.2015.151.
121. Genetic Tests Can Hurt Your Chances of Getting Some Types of Insurance. Available online: <https://www.npr.org/sections/health-shots/2018/08/07/636026264/genetic-tests-can-hurt-your-chances-of-getting-some-types-of-insurance> (accessed on 15 April 2021).
122. e-Estonia briefing centre Electronic Health Records (e-Health Records). Available online: <https://e-estonia.com/solutions/healthcare/e-health-record/> (accessed on 19 April 2021).
123. Leitsalu, L.; Hercher, L.; Metspalu, A. Giving and Withholding of Information Following Genomic Screening: Challenges Identified in a Study of Primary Care Physicians in Estonia. *J. Genet. Couns.* **2012**, *21*, 591–604, doi:10.1007/s10897-011-9424-3.
124. Saunders, G.; Baudis, M.; Becker, R.; Beltran, S.; Bérout, C.; Birney, E.; Brooksbank, C.; Brunak, S.; Van den Bulcke, M.; Drysdale, R.; et al. Leveraging European Infrastructures to Access 1 Million Human Genomes by 2022. *Nat. Rev. Genet.* **2019**, *20*, 693–701, doi:10.1038/s41576-019-0156-9.