genes

MDPI

Review

National Genome Initiatives in Europe and the United Kingdom in the Era of Whole-Genome Sequencing: A Comprehensive Review

Jan Smetana [1,*] and Petr Brož [2]

[1] Institute of Food Science and Biotechnology, Faculty of Chemistry, Brno University of Technology, 61200 Brno, Czech Republic
[2] Department of Genetics and Molecular Biology, Institute of Experimental Biology, Faculty of Science, Masaryk University, 61137 Brno, Czech Republic; bbrozpetr@gmail.com
* Correspondence: smetana@fch.vut.cz

Abstract: Identification of genomic variability in population plays an important role in the clinical diagnostics of human genetic diseases. Thanks to rapid technological development in the field of massive parallel sequencing technologies, also known as next-generation sequencing (NGS), complex genomic analyses are now easier and cheaper than ever before, which consequently leads to more effective utilization of these techniques in clinical practice. However, interpretation of data from NGS is still challenging due to several issues caused by natural variability of DNA sequences in human populations. Therefore, development and realization of projects focused on description of genetic variability of local population (often called "national or digital genome") with a NGS technique is one of the best approaches to address this problem. The next step of the process is to share such data via publicly available databases. Such databases are important for the interpretation of variants with unknown significance or (likely) pathogenic variants in rare diseases or cancer or generally for identification of pathological variants in a patient's genome. In this paper, we have compiled an overview of published results of local genome sequencing projects from United Kingdom and Europe together with future plans and perspectives for newly announced ones.

Keywords: national genome project; whole-genome sequencing; population; genetic variability Europe; United Kingdom

check for updates

Citation: Smetana, J.; Brož, P. National Genome Initiatives in Europe and the United Kingdom in the Era of Whole-Genome Sequencing: A Comprehensive Review. *Genes* **2022**, *13*, 556. https://doi.org/10.3390/genes13030556

Academic Editors: Irene Bottillo and Laura Pezzoli

Received: 9 February 2022
Accepted: 18 March 2022
Published: 21 March 2022

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

Copyright: © 2022 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The release of the first human reference genome in 2001 initiated the new era of approach into analyses of human genetic information [1]. Accomplishment of complete whole-genome sequences by pioneers of human genetics James D. Watson and Craig Venter a few years later opened a new path for the utilization of novel massive parallel (next-generation) sequencing [2,3]. These studies showed that 3 billion base pairs encode approximately 26,000 protein coding transcripts and that these coding transcripts represent only 1% of the whole genome. The beginning of the new millennia showed that there is a clear need for construction of reference genomes in order to unravel human genome variability between individuals.

Therefore, several sequencing projects, such as the International HapMap and later 1000 Genomes Project (1KGP), were launched to collect genetic data from various populations. Results from the cohort of 2504 individuals from 26 populations of a stage-three 1KGP project showed a total of nearly 88 million variants and led to the construction of the comprehensive catalogue of structural variants in human genome [4,5]. Data from HapMap and 1KGP were also used as the reference for several studies [6,7] and are still an essential part of several versions of a human reference genome. The current version is known as Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.p13/hg38) [8].

Nowadays, the renaissance of genome-wide sequencing techniques allows us to perform genomic analyses more cheaply and effectively than ever before. The advantages of whole-genome sequencing (WGS) were highlighted in multiple genome-wide association studies (GWAS). The main aim of those studies is the testing of genetic variants across the genomes of many individuals in order to identify genotype–phenotype associations [9]. Data from such studies provided critical data for diagnostics of both rare or cancerous diseases, clinical counseling and decision making for utilization of eligible treatment protocol [10–14]. The vast majority of pathogenic mutations and SVs including deletions, duplications, and inversions were found in exons (coding parts of the genes). In addition, data from recent studies also showed that several disorders, such as Alzheimer's disease or hemophilia A, can be caused by intronic splicing variants, which could have impact on the stability or regulatory function of mRNA [15–17]. Therefore, utilization of WGS technique and analysis of complex genomic data of the patient is considered one of the key steps for personalized medicine [18].

The aim of such analyses is usually to identify pathological or potentially pathological variants through multistep bioinformatic processing of sequencing data. Two major steps are data alignment (comparison between the "standard genome" and genetic information of the patients) and interpretation of genetic variants. Interpretation of genetic variants is commonly based on comparison of variants detected in patients with variants available through online databases such as GnomAD, ClinVar, HGMD, ClinGen or LOVD [19–24]. Comprehensive list of these databases is for example provided by the web pages of the Human Genome Variation Society (www.hgvs.org/locus-specific-mutation-databases). Online databases are broadly considered useful for annotation of genetic variants; nevertheless, there is still the need for better characterization of population-specific genetic variants and their potential significance in patients with rare diseases, neurodevelopmental disorders or in cancer patients treated by targeted therapy [25,26]. Many countries worldwide therefore started their own population-specific initiatives during the last decade. They are typically focused on description of local genetic variability using WGS, creation of online database of identified variants ("digital genome") and possibly generation of population-specific genome assembly. In this review, we bring together basic information about national genome initiatives from the United Kingdom and Europe and discuss the potential novel projects and development in this field.

2. Materials and Methods

Google, PubMed, EGA (The European Genome-Phenome Archive), HGV (Human Genome Variation) and EMBL-EBI data archives were searched in September 2021 to gather information about national genomic initiatives by using the search parameters (<country name> [Title]) and (human genome project) OR (national genome initiative). Criteria for including into analysis were: (1) published results in PubMed, (2) functional website in English language with information about main specifics of the project (e.g., sequencing technology, cohort size), and (3) available information about funding and scientific board of the project. Studies not carried out in a European country were excluded from the analysis. Both authors overviewed the data. Discrepancies and/or inconsistencies were discussed and resolved through mutual agreement.

3. Results

3.1. United Kingdom

The UK has already initiated several genome-wide association studies (GWAS) in large cohorts of individuals. Apart from 1KGP, Wellcome Trust launched the UK10K project aiming at WGS and deep exome sequencing (80×) for identification of both rare and pathogenic SNPs and SVs in the UK population. The WGS cohort, "the UK10K-cohort arm", analyzed data from 3781 individuals. The low pass WGS (median average read depth 7×) found 24 million variants overall, including over 3.5 million indels and 18,739 large deletions (median size 3.7 Kb). The genome of each individual contained on average

3,222,597 SNPs (5073 private), 705,684 indels (295 private) and 215 large deletions (less than 1 private). The dataset from the UK10K project is focused on the genotype/phenotype resource, which will be an order of magnitude deeper than the genetic-only 1000 Genomes Project dataset for Europe [27].

In 2012, National Health Services (NHS) has initialized a new era of genomic medicine with the 100,000 Genomes Project (https://www.genomicsengland.co.uk; accessed on 10 January 2022). Under the auspices of former Prime Minister David Cameron and as the part of the GBP 300 million initiative, NHS-owned company Genomics England is responsible for sequencing of 100,000 genomes from NHS patients with cancer and rare infectious diseases [28]. The project is focused on the better understanding of linkage between diseases and genetic signatures, potential application of genetic information in personal medicine and implementation of WGS into routine medical care [29]. For cancer, 50,000 genomes from 25,000 individuals (germline and tumor pairs) are expected to be collected. The other half of the genomes will involve 15,000 genomes of rare disease patients and 35,000 genomes of their relatives (mainly parents). Sequencing capacity and generation of sequencing data is covered by Illumina, while data analyses and interpretation are realized by several sub-contractors including Iceland's WuXi NextCODE or Wellcome Trust Sanger Institute spin-off Congenica [30]. Nowadays, there are over 92,000 genomes sequenced and pilot studies from the project already showed important genomic data associated with leukemia [31] and rare diseases [32,33]. In addition, a new online database of genetic variants Human Genome Variation Archive (HGVA) was announced [34].

3.2. Iceland

Reykjavik's company deCODE have gathered medical and genotypic data from Iceland's population since 1996. The company researchers have already sequenced a considerable part of Iceland's population. In their recent study, they sequenced genomes of 15,220 Icelanders using Illumina HiSeq platforms with median average read depth of 34x. Overall, they found a total of 31,079,378 SNPs and 7,940,790 indels. Known for its thorough genealogical datasets, they also described the parent of origin of 42,961 de novo mutations [35]. In comparison to other populations, the Icelandic population showed less rare variants and higher frequency of deleterious variants due to the limited population size and geographical isolation leading to higher influence of founder effect [36].

3.3. Sweden

The genetic map of Sweden population was already described in the SweGen study. Based on SNP genotyping of 10,000 individuals, samples from 1000 Swedes reflecting the genetic structure of the Swedish population were carefully selected for WGS [37]. Using a HiSeq X platform, they reached a median average read depth of 36× and found a total of 29.2 million SNPs and 3.8 million indels with 9.9 million of these variants not known in current databases. Furthermore, an average of 7199 individual-specific SNPs and 8645 larger SVs were observed per each sample. In addition, WGS data also showed genetic diversity within Sweden's population (particularly between southernmost and northernmost population of the country) compared to other continental European populations. AS an output of SweGen study, SweFreq online database (https://swefreq.nbis.se/; accessed on 10 Januray 2022) was established, containing whole-genome variant frequencies of all 1000 sequenced Swedish individuals [38].

3.4. Finland

Finland is well-known for its population survey established in 1972 called FINRISK, which collects samples of 6000–8000 individuals every five years to study risk factors of chronic diseases in Finland [39]. Due to Finland's unique population history and advantage resulted from FINRISK clinical data, there are several GWAS studies ongoing in Finland. Sequencing Initiative Suomi in Finland (SISU) compared exome sequence data of 3000 Finns to the same number of non-Finnish Europeans. Results from this recent

SISU study showed that the Finnish gene pool has unique genetic features including fewer variable sites in genome, more low-frequency loss-of-function variants and almost twice as many low-frequency complete gene knockouts [40]. In 2017, the EUR 59 million FinnGen project (https://www.finngen.fi/; accessed on 12 January 2022) was launched as an academic-pharma consortium that involves nine Finnish biobanks, all Finnish University Hospitals and their respective Universities, the Institute of Health and Welfare (THL) and seven large pharmaceutical companies (Abbvie, AstraZeneca, Biogen, Celgene, Genentech, Merck/MSD and Pfizer). The aim is to obtain WGS data from 500,000 Finns, which enables ambitious study designs to improve understanding of the genetic background of diseases and, subsequently, implementation of genome medicine in clinical practice and drug development. There are already 200,000 existing legacy samples, mainly from the THL Biobank, and 300,000 additional prospective samples will be collected by all of the six Finnish hospital biobanks and the Blood Service's biobank [41].

3.5. Denmark

The Denmark population-specific database of SNPs was based on data obtained from the "Danish pan-genome" study, in which authors used WGS for detailed analysis of genomes from 30 trios (parents-offspring). They reported 536,000 novel SNPs and 283,000 novel short indels detected by deep WGS (average read depth of 50×) and they develop a population-wide de novo assembly approach to identify 132,000 novel indels larger than 10 nucleotides with low false discovery rates [42]. Recently, a trio-based approach was utilized to create de novo assemblies of 150 individuals (50 trios) from GenomeDenmark project as a regional reference genome. This approach is unbiased against discovery of SVs and variation in the most complex parts of the genome, and it has the potential to improve the power of future association mapping studies [43].

3.6. Norway

The Norwegian 1000 genomes project was founded by the Norwegian Cancer Genomics Consortium (NCGC). While still in the process of collecting samples and processing samples, there is a working database of genetic variants, which already contains 1,547,121 individual variants acquired from 1590 normal chromosomes of cancer patients [44].

3.7. Estonia

The Estonian Genome Center of the University of Tartu (EGCUT) together with Estonian Biobank are collecting samples intended for GWAS studies in Estonian population. This study already consists of 51,535 donors (≥18 years of age), collected to appropriately reflect the age, sex and geographical distribution of the Estonian population (http://www.geenivaramu.ee/for-scientists/data-release/; accessed on 20 December 2021). WGS data are available from 100 individuals together with additional data from SNP arrays (20,000 individuals) and/or NMR metabolome data (11,000 individuals) [45].

3.8. Latvia

Since 2006, The Genome Database of the Latvian Population (LGDB) is collecting and processing health information, data, and biospecimens from representatives of the Latvian population. So far, the LGDB is comprised of samples and associated phenotypic and clinical information from 31,504 participants, constituting approximately 1.5% of the Latvian population [46].

3.9. Lithuania

Genetic data from Lithuanian population come from research project "Genetic diversity of the population of Lithuania and changes of its genetic structure related with evolution and common diseases" (acronym LITGEN). The group previously published data from SNP microarrays describing diversity and distribution of copy number variants (CNVs)

in 286 unrelated individuals from the two main ethnolinguistic groups (Aukštaičiai and Žemaičiai) of the Lithuanian population [47]. Recently, first 96 exomes from healthy Lithuanian individuals were sequenced. An average of 42,139 SNPs and 2306 short indels were found in each individual exome together with five pathogenic genomic variants that were inherited in an autosomal recessive pattern and that statistically significantly differed from the European population data from 1KGP [48].

3.10. Spain

There are several smaller projects focused on genetic variability and rare diseases in the Spanish population, such as the Medical Genome Project (MGP) [49] or The Genoma 1000 Navarra Research Project (NAGEN 1000) [50]; however, a national genome project is missing. On the other hand, the CSVS (Collaborative Spanish Variability Server), a crowdsourcing database of the Spanish population genetic variability currently aggregates more than 2000 genomes and exomes of unrelated Spanish individual. Based on the collected data so far, CSVS produced the first version of the Spanish Genome Reference Panel (SGRP1.0) [51].

3.11. France

The French National Alliance for Life Sciences and Health (Aviesan) started in 2015 national plan: the EUR 670 million "2025 France Genomic Medicine Initiative (PFMG2025), responsible for introducing precision medicine into the care pathway and developing a national framework for "big-genomic data" medicine [52]. Technological aspects of the project are secured via France Genomique, an infrastructure which joins together the four main French public research organizations: CEA, CNRS, INRA and INSERM.

3.12. Netherlands

Genome of the Netherlands (GoNL) is a Dutch reference genome project in which whole genomes of 250 Dutch trios (750 individuals) were sequenced (average read depth of 13×) [53]. In 1990, the Netherlands also established a population-based cohort study called the Rotterdam study. Recently, 2628 DNA samples from this study were used for exome sequencing (average read depth of 53×) and this dataset was denoted "Rotterdam Study Exome Sequencing set 2" (RSX2). The authors of the projects have stated that from the 439,633 coding variants, 120,109 were absent from six other public population databases including ExAC2.0, ESP6500, 1 KG, Icelandic deCODE, GoNL and UK10K. The smallest overlap was seen with the Icelandic population, which is in line with previous statements. In general, each dataset showed contained variants not present in any of the other datasets. The results suggested that both smaller population-specific datasets as well as large aggregation datasets contributed information and each one of them contributed variants that were not seen yet [54].

3.13. Italy

In Italy, several GWAS studies characterized the genetic variability of local populations, including Sardinians [55] or Lombards in the North Italian region [56], using SNP-genotyping. In 2015, Italian National Commission for genomics was established in order to set up a national plan for the use of genomic knowledge and technologies in healthcare known as the Italian Genome project (IGP). Based on the sequencing data, a new Italian Genome Reference Panel (IGRP1.0) was defined. Pivotal results also extended the knowledge of genetic variability in the Italian population, including variants not known in the previous datasets such as β thalassemia-related variant GRCh37 chr11:g.5248004G>A (rs11549407), distribution of deleterious variants and incidence of human knockouts, and overall confirmed the necessity of distinct genome references for the Italian population [57].

3.14. Germany

A similar situation to France is found in Germany, where nowadays plenty of GWAS projects are running with the idea of building a human genome database; for example, a consortium consisting of Kühne-Stiftung, University Hospital Zurich (UHZ) and UKE Hamburg. This consortium plans to obtain whole-genome sequencing data from over 9000 people in the German-language area with funding of 12.5 million euros [58]. GWAS-suitable infrastructure is already in place in Germany, such as the German Human Genome-Phenome Archive (CHGA), which is available to scientific community through the German Cancer Research Center DKFZ, Heidelberg. In addition, The German Ministry of Health announced the foundation of the German Genomics Initiative (genomDE). GenomDE should entail a legal and ethical framework for organization, data infrastructure and reimbursement as well as a communication campaign aimed on both the public and healthcare professionals in Germany in the field of population genomics

3.15. Czech Republic

The National Center for Medical Genomics (NCGM) recently launched the project "Analysis of Czech Genomes for Theranostics" (ACTG), which is focused on WGS analysis of 1500 genomes from Czech population by the 2022. So far, the genomic database consists of 1055 analyzed genomes [59].

3.16. Poland

In Poland during 2014–2020, the European Centre for Bioinformatics and Genomics (ECBiG) consortium planned the sequencing of whole human genomes of about 5000 inhabitants from all over the country [60]. Pilot results from a cohort of 1079 individuals showed total of 31.24 million SNPs and 5.63 million small indels. On average, 4.48 million small variants per individual were found, of which 16,473 were private variants [61].

3.17. Slovenia

The Slovenian Genome project (SGP) was announced as support for the cooperation in +1MG. The consortium of University of Ljubljana, Institute of Oncology Ljubljana and Institute Service of Slovenia for Transfusion Medicine is focused on the creation of an environment for the collection of genetic, health and environmental risk factors and development of personalized medicine in Slovenia. The pilot project is based on sequencing of 300 Slovene genomes, which will represent a foundation of data analysis platform for the project [62].

3.18. Greece

The genomic initiative "Genome of Greece" (GoGreece) was launched in 2010 through The Laboratory of Pharmacogenomics and Individualized Therapy of the University of Patras. The project is based on WGS of >100,000 Greek individuals in order to delineate the genetic etiology of the underlying clinical phenotype of patients suffering from monogenic and multifactorial diseases and to determine the genetic variability of the Hellenic population [63]. Results from this cohort showed incidence of novel *FTO* and *TBC1D1* genetic variants associated with Amyotrophic lateral sclerosis (ALS) in the Greek population [64], or six genomic variants (*SLC9A4* c.1919G>A, *KIAA1109* c.2933T>C and c.4268_4269delCCinsTA, *HoxB6* c.668C>A, *HoxD12* c.418G>A, and *NCK2* c.745_746delAAinsG) with the potential of celiac disease predisposition in the Greek population [65].

3.19. Cyprus

The Cyprus genome initiative was funded as a part of EUR 38 million Horizont 2020 Biobanking and the Cyprus Human Genome Project, which is based on an existing Biobank of the University of Cyprus and its transformation to a Center of Excellence in Biobanking and Biomedical Research. As a part of the University of Cyprus, this core facility

will collect over 16,500 donors and together with other partners (Medical University of Graz, Austria, Biobanking and BioMolecular Resources Research Infrastructure-European Research Infrastructure Consortium/BBMRI-ERIC, Austria; RTD TALOS Limited Cyprus) it will be responsible for completion of the Cyprus human genome project [66].

3.20. Malta

The Maltese Genome Project was founded as a part of the Malta BioBank, which is a member of BBMRI-ERIC. The main aim of the genome project is to obtain genome data from 1% of the Maltese population in relation to origins, mobility, epidemiology, pharmacogenomics and immunogenetics for gene discovery research [67].

3.21. Russia

The extreme diversity of the Russian population is one of the main causes of under-representation of genetic information in large worldwide datasets such as HapMap or 1KGP. The Genome Russia project (http://genomerussia.spbu.ru; accessed on 1 February 2022), launched by St. Petersburg State University and Dobzhansky Center for Genome Bioinformatics, is focusing on collection of samples from at least 3000 individuals from different parts of the Russian federation, whose ancestors are indigenous to the region for several generations. The sequencing data from this trio-design study will allow creating the database of medically relevant genomic variants characteristic to the Russian population, which would be the basis for developing the principles of the future personalized medicine [68]. A pilot study of the project used the WGS approach for analysis of genetic variability in a cohort of 264 samples obtained from 52 isolated populations across the Russian federation. The variant calling showed 8 million SNPs and 2 million indels per population and 4% of these SNPs were classified as novel when compared to dbSNP [69].

4. Discussion

Recent development of massive parallel sequencing technology launched plenty of both international and local GWAS studies, leading to better characterization of genetic variability among human populations. Current version of human genome assembly GRCh38/hg38 and its predecessors covered genetic variability of the local European population only briefly, and thus there is a strong need for further genomic data from local or native populations such as Finland, Iceland, Baltic and south European countries. From this point of view, it is not surprising that basically every further published national genome project from European country showed additional data not included in this assembly. In addition, those studies showed novel information about other aspects such as biobanking, GDPR, library preparation, sequencing workflow, utilization of novel data processing and data mining algorithms. Information about projects with published scientific results is summarized in Table 1. Further development of local genome projects in Europe should also bring ´1+ Million Genomes (+1MG)' European initiative. The goal of 22 signatory EU countries is to obtain sequenced genomes from more than 1 million individuals by 2022 in order to create a framework that will cover analysis of genomic and health data both inside and across national boundaries in Europe. The methodical basis of the initiative is based on the Horizon 2020 project Beyond 1 Million Genomes (B1MG; https://b1mg-project.eu/), which is focused on infrastructure setup, legal and technical guidance, data standards and best practices to enable data access [70]. Another project associated with the +1MG initiative is the multi-country project called Genome of Europe. Together with B1MG it is focused on building a robust and high-quality European network of national genomic reference cohorts, representative of the European population. Connected via +1MG initiative, those individual datasets from the EU population will create a world-class European reference database for research and innovation of healthcare [71].

Table 1. Overview of basic characterizes of UK or European national genome projects with published scientific results.

Project	Country	Cohort Size	Year of Publishing	Library Preparation	Sequencing Technology	Website	Reference
UK10k	United Kingdom	3781	2013	Illumina pair-end (BGI, Sanger)	Illumina (BGI, Sanger)	https://www.uk10k.org/	[27]
deCODE Genetics	Iceland	2636	2015	TruSeq SBS	HiSeq, GAIIx	https://www.decode.com/research/	[36]
SweGen	Sweden	1000	2017	TruSeq PCR-free 2.0	HiSeq X	https://swefreq.nbis.se/dataset/SweGen	[38]
Sequencing Initiative Suomi in Finland (SISU)	Finnland	3000	2014	Agilent, Illumina, Roche	NA	http://www.sisuproject.fi/	[40]
Genome Denmark	Denmark	150	2017	Illumina	Illumina HiSeq2000	https://genome.au.dk/	[42]
Genome of Netherlands (GoNL)	Netherlands	769	2014	TruSeq 2.0, Nextera	HiSeq 2000	http://www.nlgenome.nl/	[53]
Italian Genome Reference Panel (IGRP1.0)	Italy	947	2020	Illumina	Illumina	https://www.iigm.it/site/	[57]
The Thousand Polish Genomes Project	Poland	1079	2021	TruSeq DNA PCR-free kit	NovaSeq 6000	https://www.genompolski.pl/	[61]
Genome Russia	Russia	3000	2018	TruSeq PCR free	HiSeq X, NovaSeq	http://genomerussia.spbu.ru/	[69]

Taken together, the need for the completion of the genetic diversity map of human populations makes obvious that further local sequencing projects are still needed. Improvement of datasets and references introduced by previous large-scale sequencing initiatives such as 1KGP or HapMap is now more effective thanks to broader availability of WGS techniques. Genomic data provided by various countries world-wide from local sequencing projects should therefore lead to rapid improvement in the area of precision and/or personalized medicine and thus bring another important tool to the clinical diagnostics of the diseases.

Author Contributions: J.S.—Conceptualization, methodology, writing—review and editing; P.B.—writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Health of the Czech Republic, grant nr. NU20-07-00145.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are thankful to their colleagues for their support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Collins, F.S. Contemplating the end of the beginning. *Genome Res.* **2001**, *11*, 641–643. [CrossRef]
2. Levy, S.; Sutton, G.; Ng, P.C.; Feuk, L.; Halpern, A.L.; Walenz, B.P.; Axelrod, N.; Huang, J.; Kirkness, E.F.; Denisov, G.; et al. The diploid genome sequence of an individual human. *PLoS Biol.* **2007**, *5*, e254. [CrossRef] [PubMed]
3. Wheeler, D.A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y.J.; Makhijani, V.; Roth, G.T.; et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **2008**, *452*, 872–876. [CrossRef]
4. The Genomes Project, C.; Auton, A.; Abecasis, G.R.; Altshuler, D.M.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68. Available online: https://www.nature.com/articles/nature15393#supplementary-information (accessed on 5 May 2021). [CrossRef]
5. Sudmant, P.H.; Rausch, T.; Gardner, E.J.; Handsaker, R.E.; Abyzov, A.; Huddleston, J.; Zhang, Y.; Ye, K.; Jun, G.; Hsi-Yang Fritz, M.; et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **2015**, *526*, 75. Available online: https://www.nature.com/articles/nature15394#supplementary-information (accessed on 5 May 2021). [CrossRef]
6. Nothnagel, M.; Lu, T.T.; Kayser, M.; Krawczak, M. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum. Mol. Genet.* **2010**, *19*, 2927–2935. [CrossRef]
7. Nelis, M.; Esko, T.; Magi, R.; Zimprich, F.; Zimprich, A.; Toncheva, D.; Karachanak, S.; Piskackova, T.; Balascak, I.; Peltonen, L.; et al. Genetic structure of Europeans: A view from the North-East. *PLoS ONE* **2009**, *4*, e5472. [CrossRef] [PubMed]
8. Pan, B.; Kusko, R.; Xiao, W.; Zheng, Y.; Liu, Z.; Xiao, C.; Sakkiah, S.; Guo, W.; Gong, P.; Zhang, C.; et al. Similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinform.* **2019**, *20*, 101. [CrossRef]
9. Tam, V.; Patel, N.; Turcotte, M.; Bossé, Y.; Paré, G.; Meyre, D. Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **2019**, *20*, 467–484. [CrossRef]
10. Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **2009**, *461*, 272–276. [CrossRef] [PubMed]
11. Carss, K.J.; Arno, G.; Erwood, M.; Stephens, J.; Sanchis-Juan, A.; Hull, S.; Megy, K.; Grozeva, D.; Dewhurst, E.; Malka, S.; et al. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am. J. Hum. Genet.* **2017**, *100*, 75–90. [CrossRef]
12. Wang, J.; Skoog, T.; Einarsdottir, E.; Kaartokallio, T.; Laivuori, H.; Grauers, A.; Gerdhem, P.; Hytonen, M.; Lohi, H.; Kere, J.; et al. Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. *Sci. Rep.* **2016**, *6*, 33256. [CrossRef] [PubMed]
13. Chapman, M.A.; Lawrence, M.S.; Keats, J.J.; Cibulskis, K.; Sougnez, C.; Schinzel, A.C.; Harview, C.L.; Brunet, J.P.; Ahmann, G.J.; Adli, M.; et al. Initial genome sequencing and analysis of multiple myeloma. *Nature* **2011**, *471*, 467–472. [CrossRef] [PubMed]
14. Alioto, T.; Buchhalter, I.; Derdak, S.; Hutter, B.; Eldridge, M.D.; Hovig, E.; Heisler, L.E.; Beck, T.A.; Simpson, J.T.; Tonon, L.; et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat. Commun.* **2015**, *6*, 10001. [CrossRef]
15. Bach, J.E.; Oldenburg, J.; Muller, C.R.; Rost, S. Mutational spectrum and deep intronic variants in the factor VIII gene of haemophilia A patients. Identification by next generation sequencing. *Hamostaseologie* **2016**, *36*, S25–S28. [PubMed]
16. Bucossi, S.; Polimanti, R.; Ventriglia, M.; Mariani, S.; Siotto, M.; Ursini, F.; Trotta, L.; Scrascia, F.; Callea, A.; Vernieri, F.; et al. Intronic rs2147363 variant in ATP7B transcription factor-binding site associated with Alzheimer's disease. *J. Alzheimer's Dis. JAD* **2013**, *37*, 453–459. [CrossRef] [PubMed]
17. Gelfman, S.; Wang, Q.; McSweeney, K.M.; Ren, Z.; La Carpia, F.; Halvorsen, M.; Schoch, K.; Ratzon, F.; Heinzen, E.L.; Boland, M.J.; et al. Annotating pathogenic non-coding variants in genic regions. *Nat. Commun.* **2017**, *8*, 236. [CrossRef]
18. Tremblay, J.; Hamet, P. Role of genomics on the path to personalized medicine. *Metab. Clin. Exp.* **2013**, *62* (Suppl. 1), S2–S5. [CrossRef]
19. Riggs, E.R.; Andersen, E.F.; Cherry, A.M.; Kantarci, S.; Kearney, H.; Patel, A.; Raca, G.; Ritter, D.I.; South, S.T.; Thorland, E.C.; et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2020**, *22*, 245–257. [CrossRef]
20. Gudmundsson, S.; Singer-Berk, M.; Watts, N.A.; Phu, W.; Goodrich, J.K.; Solomonson, M.; Rehm, H.L.; MacArthur, D.G.; O'Donnell-Luria, A. Variant interpretation using population databases: Lessons from gnomAD. *Hum. Mutat.* **2021**. *Early View*. [CrossRef]
21. MacDonald, J.R.; Ziman, R.; Yuen, R.K.; Feuk, L.; Scherer, S.W. The Database of Genomic Variants: A curated collection of structural variation in the human genome. *Nucleic Acids Res.* **2014**, *42*, D986–D992. [CrossRef] [PubMed]

22. Rehm, H.L.; Berg, J.S.; Brooks, L.D.; Bustamante, C.D.; Evans, J.P.; Landrum, M.J.; Ledbetter, D.H.; Maglott, D.R.; Martin, C.L.; Nussbaum, R.L.; et al. ClinGen–the Clinical Genome Resource. *N. Engl. J. Med.* **2015**, *372*, 2235–2242. [CrossRef]
23. Stenson, P.D.; Mort, M.; Ball, E.V.; Chapman, M.; Evans, K.; Azevedo, L.; Hayden, M.; Heywood, S.; Millar, D.S.; Phillips, A.D.; et al. The Human Gene Mutation Database (HGMD(®)): Optimizing its use in a clinical diagnostic or research setting. *Hum. Genet.* **2020**, *139*, 1197–1207. [CrossRef] [PubMed]
24. Fokkema, I.F.; den Dunnen, J.T.; Taschner, P.E. LOVD: Easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. *Hum. Mutat.* **2005**, *26*, 63–68. [CrossRef]
25. Hehir-Kwa, J.Y.; Marschall, T.; Kloosterman, W.P.; Francioli, L.C.; Baaijens, J.A.; Dijkstra, L.J.; Abdellaoui, A.; Koval, V.; Thung, D.T.; Wardenaar, R.; et al. A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. *Nat. Commun.* **2016**, *7*, 12989. [CrossRef]
26. Levy-Sakin, M.; Pastor, S.; Mostovoy, Y.; Li, L.; Leung, A.K.Y.; McCaffrey, J.; Young, E.; Lam, E.T.; Hastie, A.R.; Wong, K.H.Y.; et al. Genome maps across 26 human populations reveal population-specific patterns of structural variation. *Nat. Commun.* **2019**, *10*, 1025. [CrossRef] [PubMed]
27. Walter, K.; Min, J.L.; Huang, J.; Crooks, L.; Memari, Y.; McCarthy, S.; Perry, J.R.; Xu, C.; Futema, M.; Lawson, D.; et al. The UK10K project identifies rare variants in health and disease. *Nature* **2015**, *526*, 82–90. [CrossRef] [PubMed]
28. Marx, V. The DNA of a nation. *Nature* **2015**, *524*, 503. [CrossRef]
29. Mark, C.; Jim, D.; Martin, D.; Leila, E.; Tom, F.; Sue, H.; Tim, H.; Luke, J.; Nick, M.; Jeanna, M.-P.; et al. The 100,000 Genomes Project Protocol. Available online: https://figshare.com/articles/journal_contribution/GenomicEnglandProtocol_pdf/4530893/4 (accessed on 24 October 2021).
30. Turnbull, C.; Scott, R.H.; Thomas, E.; Jones, L.; Murugaesu, N.; Pretty, F.B.; Halai, D.; Baple, E.; Craig, C.; Hamblin, A.; et al. The 100 000 Genomes Project: Bringing whole genome sequencing to the NHS. *BMJ* **2018**, *361*, k1687. [CrossRef]
31. Klintman, J.; Barmpouti, K.; Knight, S.J.L.; Robbe, P.; Dreau, H.; Clifford, R.; Ridout, K.; Burns, A.; Timbs, A.; Bruce, D.; et al. Clinical-grade validation of whole genome sequencing reveals robust detection of low-frequency variants and copy number alterations in CLL. *Br. J. Haematol.* **2018**, *182*, 412–417. [CrossRef]
32. Radziwon, A.; Arno, G.; Wheaton, D.; McDonagh, E.M.; Baple, E.L.; Webb-Jones, K.; Webster, A.R.; MacDonald, I.M. Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum. Mutat.* **2017**, *38*, 704–715. [CrossRef] [PubMed]
33. Gräf, S.; Haimel, M.; Bleda, M.; Hadinnapola, C.; Southgate, L.; Li, W.; Hodgson, J.; Liu, B.; Salmon, R.M.; Southwood, M.; et al. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat. Commun.* **2018**, *9*, 1416. [CrossRef]
34. Lopez, J.; Coll, J.; Haimel, M.; Kandasamy, S.; Tarraga, J.; Furio-Tari, P.; Bari, W.; Bleda, M.; Rueda, A.; Gräf, S.; et al. HGVA: The Human Genome Variation Archive. *Nucleic Acids Res.* **2017**, *45*, W189–W194. [CrossRef] [PubMed]
35. Jonsson, H.; Sulem, P.; Kehr, B.; Kristmundsdottir, S.; Zink, F.; Hjartarson, E.; Hardarson, M.T.; Hjorleifsson, K.E.; Eggertsson, H.P.; Gudjonsson, S.A.; et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **2017**, *549*, 519–522. [CrossRef] [PubMed]
36. Gudbjartsson, D.F.; Helgason, H.; Gudjonsson, S.A.; Zink, F.; Oddson, A.; Gylfason, A.; Besenbacher, S.; Magnusson, G.; Halldorsson, B.V.; Hjartarson, E.; et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **2015**, *47*, 435–444. [CrossRef] [PubMed]
37. Humphreys, K.; Grankvist, A.; Leu, M.; Hall, P.; Liu, J.; Ripatti, S.; Rehnstrom, K.; Groop, L.; Klareskog, L.; Ding, B.; et al. The genetic structure of the Swedish population. *PLoS ONE* **2011**, *6*, e22547. [CrossRef] [PubMed]
38. Ameur, A.; Dahlberg, J.; Olason, P.; Vezzi, F.; Karlsson, R.; Martin, M.; Viklund, J.; Kahari, A.K.; Lundin, P.; Che, H.; et al. SweGen: A whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur. J. Hum. Genet. EJHG* **2017**, *25*, 1253–1260. [CrossRef]
39. Borodulin, K.; Tolonen, H.; Jousilahti, P.; Jula, A.; Juolevi, A.; Koskinen, S.; Kuulasmaa, K.; Laatikainen, T.; Mannisto, S.; Peltonen, M.; et al. Cohort Profile: The National FINRISK Study. *Int. J. Epidemiol.* **2017**, *47*, 696–696i. [CrossRef]
40. Lim, E.T.; Wurtz, P.; Havulinna, A.S.; Palta, P.; Tukiainen, T.; Rehnstrom, K.; Esko, T.; Magi, R.; Inouye, M.; Lappalainen, T.; et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **2014**, *10*, e1004494. [CrossRef]
41. FinnGen. Available online: https://www.finngen.fi/en/about (accessed on 15 December 2021).
42. Besenbacher, S.; Liu, S.; Izarzugaza, J.M.G.; Grove, J.; Belling, K.; Bork-Jensen, J.; Huang, S.; Als, T.D.; Li, S.; Yadav, R.; et al. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat. Commun.* **2015**, *6*, 5969. Available online: https://www.nature.com/articles/ncomms6969#supplementary-information (accessed on 19 September 2021). [CrossRef]
43. Maretty, L.; Jensen, J.M.; Petersen, B.; Sibbesen, J.A.; Liu, S.; Villesen, P.; Skov, L.; Belling, K.; Theil Have, C.; Izarzugaza, J.M.G.; et al. Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. *Nature* **2017**, *548*, 87. Available online: https://www.nature.com/articles/nature23264#supplementary-information (accessed on 19 September 2021). [CrossRef] [PubMed]
44. NCGC. Conditions for use of the 1000 Genomes. Available online: https://kreftgenomikk.no/en/vilkar-for-bruk-av-1000genomes-no/ (accessed on 27 November 2021).

45. Leitsalu, L.; Haller, T.; Esko, T.; Tammesoo, M.L.; Alavere, H.; Snieder, H.; Perola, M.; Ng, P.C.; Magi, R.; Milani, L.; et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **2015**, *44*, 1137–1147. [CrossRef] [PubMed]
46. Rovite, V.; Wolff-Sagi, Y.; Zaharenko, L.; Nikitina-Zake, L.; Grens, E.; Klovins, J. Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. *J. Epidemiol.* **2018**, *28*, 353–360. [CrossRef]
47. Urnikyte, A.; Domarkiene, I.; Stoma, S.; Ambrozaityte, L.; Uktveryte, I.; Meskiene, R.; Kasiulevicius, V.; Burokiene, N.; Kucinskas, V. CNV analysis in the Lithuanian population. *BMC Genet.* **2016**, *17*, 64. [CrossRef] [PubMed]
48. Rancelis, T.; Arasimavicius, J.; Ambrozaityte, L.; Kavaliauskiene, I.; Domarkiene, I.; Karciauskaite, D.; Kucinskiene, Z.A.; Kucinskas, V. Analysis of pathogenic variants from the ClinVar database in healthy people using next-generation sequencing. *Genet. Res.* **2017**, *99*, e6. [CrossRef]
49. Dopazo, J.; Amadoz, A.; Bleda, M.; Garcia-Alonso, L.; Alemán, A.; García-García, F.; Rodriguez, J.A.; Daub, J.T.; Muntané, G.; Rueda, A.; et al. 267 Spanish Exomes Reveal Population-Specific Differences in Disease-Related Genetic Variation. *Mol. Biol. Evol.* **2016**, *33*, 1205–1218. [CrossRef]
50. NAGEN: Proyecto Genoma 1000 Navarra (NAGEN 1000). Available online: https://www.nagen1000navarra.es/en (accessed on 13 November 2021).
51. Peña-Chilet, M.; Roldán, G.; Perez-Florido, J.; Ortuño, F.M.; Carmona, R.; Aquino, V.; Lopez-Lopez, D.; Loucera, C.; Fernandez-Rueda, J.L.; Gallego, A.; et al. CSVS, a crowdsourcing database of the Spanish population genetic variability. *Nucleic Acids Res.* **2021**, *49*, D1130–D1137. [CrossRef]
52. Lévy, Y. Genomic medicine 2025: France in the race for precision medicine. *Lancet* **2016**, *388*, 2872. [CrossRef]
53. Boomsma, D.I.; Wijmenga, C.; Slagboom, E.P.; Swertz, M.A.; Karssen, L.C.; Abdellaoui, A.; Ye, K.; Guryev, V.; Vermaat, M.; van Dijk, F.; et al. The Genome of the Netherlands: Design, and project goals. *Eur. J. Hum. Genet. EJHG* **2014**, *22*, 221–227. [CrossRef]
54. Francioli, L.C.; Menelaou, A.; Pulit, S.L.; van Dijk, F.; Palamara, P.F.; Elbers, C.C.; Neerincx, P.B.T.; Ye, K.; Guryev, V.; Kloosterman, W.P.; et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **2014**, *46*, 818–825. [CrossRef]
55. Di Gaetano, C.; Fiorito, G.; Ortu, M.F.; Rosa, F.; Guarrera, S.; Pardini, B.; Cusi, D.; Frau, F.; Barlassina, C.; Troffa, C.; et al. Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection. *PLoS ONE* **2014**, *9*, e91237. [CrossRef] [PubMed]
56. Vai, S.; Ghirotto, S.; Pilli, E.; Tassi, F.; Lari, M.; Rizzi, E.; Matas-Lalueza, L.; Ramirez, O.; Lalueza-Fox, C.; Achilli, A.; et al. Genealogical relationships between early medieval and modern inhabitants of Piedmont. *PLoS ONE* **2015**, *10*, e0116801. [CrossRef] [PubMed]
57. Cocca, M.; Barbieri, C.; Concas, M.P.; Robino, A.; Brumat, M.; Gandin, I.; Trudu, M.; Sala, C.F.; Vuckovic, D.; Girotto, G.; et al. A bird's-eye view of Italian genomic variation through whole-genome sequencing. *Eur. J. Hum. Genet.* **2020**, *28*, 435–444. [CrossRef] [PubMed]
58. Lemm, S. Germany's Largest Research Programme for Genome Sequencing Launched. Available online: https://idw-online.de/de/news725422 (accessed on 25 September 2021).
59. Analysis of Czech Genomes for Theranostics (ACGT). Available online: https://www.acgt.cz/en/ (accessed on 19 November 2021).
60. Genomic Map of Poland. Available online: http://ecbig.pl/page/genomic-map-of-poland/ (accessed on 12 October 2021).
61. Kaja, E.; Lejman, A.; Sielski, D.; Sypniewski, M.; Gambin, T.; Suchocki, T.; Dawidziuk, M.; Golik, P.; Wojtaszewska, M.; Stępień, M.; et al. 'The Thousand Polish Genomes Project'—A national database of Polish variant allele frequencies. *bioRxiv* **2021**, 451425. [CrossRef]
62. SGP. Slovenski Genomski Projekt. Available online: http://genom.si/en/index.html (accessed on 19 November 2021).
63. Patrinos, G.P.; Pasparakis, E.; Koiliari, E.; Pereira, A.C.; Hünemeier, T.; Pereira, L.V.; Mitropoulou, C. Roadmap for Establishing Large-Scale Genomic Medicine Initiatives in Low- and Middle-Income Countries. *Am. J. Hum. Genet.* **2020**, *107*, 589–595. [CrossRef]
64. Mitropoulos, K.; Merkouri Papadima, E.; Xiromerisiou, G.; Balasopoulou, A.; Charalampidou, K.; Galani, V.; Zafeiri, K.V.; Dardiotis, E.; Ralli, S.; Deretzi, G.; et al. Genomic variants in the FTO gene are associated with sporadic amyotrophic lateral sclerosis in Greek patients. *Hum Genom.* **2017**, *11*, 30. [CrossRef]
65. Balasopoulou, A.; Stanković, B.; Panagiotara, A.; Nikčevic, G.; Peters, B.A.; John, A.; Mendrinou, E.; Stratopoulos, A.; Legaki, A.I.; Stathakopoulou, V.; et al. Novel genetic risk variants for pediatric celiac disease. *Hum Genom.* **2016**, *10*, 34. [CrossRef]
66. CY-BIOBANK. Center of Excellence—Biobanking and the Cyprus Human Genome Project. Available online: https://www.ucy.ac.cy/mmrc/en/cybiobank (accessed on 1 September 2021).
67. Borg, J. Malta Human Genome Project. Available online: https://www.researchgate.net/publication/327954831_Malta_Human_Genome_Project?channel=doi&linkId=5baf348592851ca9ed2e8197&showFulltext=true (accessed on 24 October 2021).
68. Oleksyk, T.K.; Brukhin, V.; O'Brien, S.J. The Genome Russia project: Closing the largest remaining omission on the world Genome map. *Gigascience* **2015**, *4*, 53. [CrossRef]
69. Zhernakova, D.V.; Kliver, S.; Cherkasov, N.; Tamazian, G.; Rotkevich, M.; Krasheninnikova, K.; Evsyukov, I.; Sidorov, S.; Dobrynin, P.; Yurchenko, A.A.; et al. Analytical "bake-off" of whole genome sequencing quality for the Genome Russia project using a small cohort for autoimmune hepatitis. *PLoS ONE* **2018**, *13*, e0200423. [CrossRef]

70. The Beyond 1 Million Genomes (B1MG). Available online: https://b1mg-project.eu/ (accessed on 2 October 2021).
71. Saunders, G.; Baudis, M.; Becker, R.; Beltran, S.; Béroud, 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. [CrossRef]