

Review

# Human Pangenomics: Promises and Challenges of a Distributed Genomic Reference

Paolo Abondio \* , Elisabetta Cilli  and Donata Luiselli \* 

Laboratory of Ancient DNA, Department of Cultural Heritage, University of Bologna, Via degli Ariani 1, 48121 Ravenna, Italy

\* Correspondence: paolo.abondio2@unibo.it (P.A.); donata.luiselli@unibo.it (D.L.)

**Abstract:** A pangenome is a collection of the common and unique genomes that are present in a given species. It combines the genetic information of all the genomes sampled, resulting in a large and diverse range of genetic material. Pangenomic analysis offers several advantages compared to traditional genomic research. For example, a pangenome is not bound by the physical constraints of a single genome, so it can capture more genetic variability. Thanks to the introduction of the concept of pangenome, it is possible to use exceedingly detailed sequence data to study the evolutionary history of two different species, or how populations within a species differ genetically. In the wake of the Human Pangenome Project, this review aims at discussing the advantages of the pangenome around human genetic variation, which are then framed around how pangenomic data can inform population genetics, phylogenetics, and public health policy by providing insights into the genetic basis of diseases or determining personalized treatments, targeting the specific genetic profile of an individual. Moreover, technical limitations, ethical concerns, and legal considerations are discussed.

**Keywords:** pangenomics; human genomics; pangenome; structural variation; bioinformatics; evolution; selection; phylogenetics; public health; personalized medicine



**Citation:** Abondio, P.; Cilli, E.; Luiselli, D. Human Pangenomics: Promises and Challenges of a Distributed Genomic Reference. *Life* **2023**, *13*, 1360. <https://doi.org/10.3390/life13061360>

Academic Editors: Fan Zhang and Zhiqiang Hu

Received: 15 May 2023

Revised: 2 June 2023

Accepted: 8 June 2023

Published: 9 June 2023



**Copyright:** © 2023 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 concept of pangenome (or “pan-genome”, also called “supragenome”) traces its origin in the molecular study of Prokaryotes, as it mostly refers to the entirety of the transcribed genetic units in all available lineages of a monophyletic (bacterial) group [1–3]. Gathering all the genomes of a phylogenetic clade would provide the result of assembling all the genes for that group in a superior genetic structure that could encompass its complete genetic diversity and repertoire in a theoretical object called a “supergenome” [4,5]. Therefore, the pangenome represents the total compendium of all possible genomic variability for a collection of specimens ideally sharing a common ancestor [6], but does not describe the complete genetic makeup of a species, which could only be achieved by sequencing every single genome of that species. Of course, this definition, although fitting, is influenced by the very nature of the average bacterial genome and the way genetic material can be exchanged among unicellular Prokaryotes. Indeed, the bacterial genome is frequently small, haploid, and almost devoid of introns or intergenic DNA, yet its composition in terms of transcriptional units is extremely flexible, as only a handful of genes are absolutely crucial for the actual functioning of the bacterium, while a wider cluster of accessory genes, mediating survival in specific environmental conditions and other collateral properties (for example, antimicrobial resistance or toxin production), may be integrated, lost, and exchanged through horizontal transfer, or picked up from the environment as fragments of free-floating DNA [7–9]. So, what guarantees rapid adaptation and evolution, as well as extreme plasticity, is the existence of this flexible genomic content that is characteristic of Prokaryotes and that constitutes the “pangenome”.

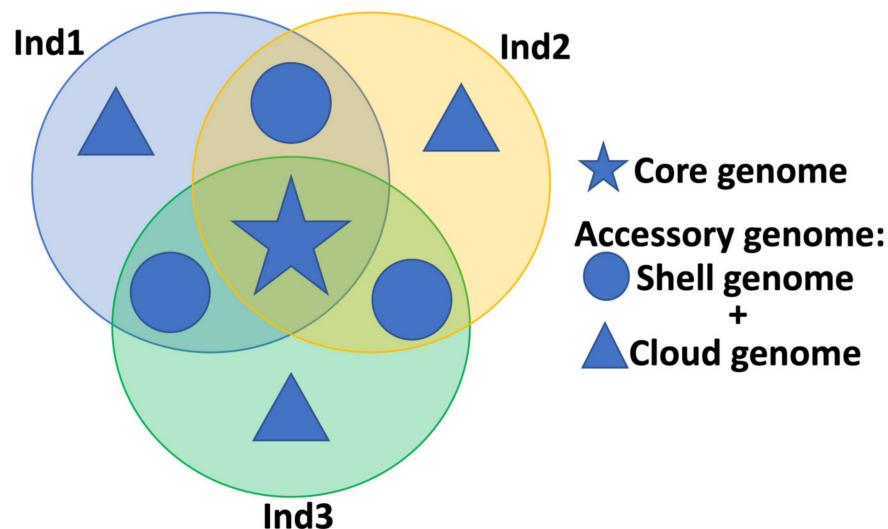
On the other hand, multicellular Eukaryotic organisms usually contain multiple fragments of DNA (chromosomes) as well as multiple copies of the same DNA strand in

each cellular nucleus (polyploidy) [10,11]. These multiple copies carry similar coding information, and virtually all possible genes are always present (although not all actively functioning) in the genetic sequences of each single individual. However, the intricacy of Eukaryotic DNA in complex multicellular organisms lies in it being highly enriched in non-genic material and characterized by a higher degree of structural variation, which moves way beyond individual differences in single nucleotide polymorphisms [12–14]. The notion of pangenome, then, requires the inclusion of novel genetic compositions, whereby its properties are expanded and modified to accommodate a wider and much more complex definition. Indeed, a more general depiction should acknowledge that genomic variation encompasses point mutations but also insertions/deletions, repetitive DNA, mobile genetic elements, inversions, duplications, and gene fusions, especially since gene presence/absence cannot be an informative descriptor of genomic variability in high-order groupings of Eukaryotes [15,16]. Similarly, the “population” under scrutiny can be a taxonomic unit [14,17] but can also be a collection of cells from the same tissue [18] or an ecological community [19,20] depending on what level of biological description one is trying to analyze. Moreover, it must be taken into consideration that many different processes take place having the genetic code on the background, and that these are much more variable among cells and individuals than the simple DNA sequence. Epigenetic patterns [21] and cell-specific chromatin remodeling [22], leading to regulation of transcription and differential translation [23,24], are only some of the phenomena that intervene to determine the flow of information towards the expression of the genetic code, and these can also be considered and studied in a similar way to the pangenome (pan-epigenomics [25,26], pan-transcriptomics [27], pan-proteomics [28], and so on) to encapsulate the overall diversity and general variability of genomic products in a population. Quite recently, the potential of a pangenomic approach (i.e., the identification of gene clusters and definition of relationships between genomes based on gene sharing) has also been applied in the context of metagenomics (i.e., the ability to sequence microbial DNA directly from the environment and define phylogenetic relationships across the spatial dimension of ecological niches) to describe the ecological role of gene clusters linked to niche adaptation and fitness in microbial clades, giving rise to the discipline of pan-metagenomics [29–33].

## 2. Characteristics of the Pangenome

The pangenome concept, as introduced in Section 1, can be applied to any biological population, be it a viral, bacterial, or eukaryotic species. A pangenome can be considered as a collection of all the genes, regulatory entities, and non-genetic segments that are present in different numbers in various lineages of the group under scrutiny. It is also possible to identify which genes are not part of the core collection and which are essential—that is, those that are found in all or most of the lineages (Figure 1). The core genes are conserved, while the accessory genes, present in only some individuals or lineages, are more variable [3,5,34,35]. Moreover, genetic segments that are part of the accessory genome can be further subdivided in the “dispensable” or “shell” genome (i.e., structures that are shared by at least two subjects) and the “unique” or “cloud” genome (i.e., the collection of genomic elements that are specific to a single individual). Sometimes, the definition of “dispensable” genome also encompasses the unique genome, coinciding in this case with the accessory genome [17]. This allows for the characterization of the species’ genetic diversity and for the identification of new alleles and their potential impact on the species’ microevolution [6,36,37]. By comparing the pangenomes of two species, in fact, it is possible to identify which genes are shared and which are unique; this can also provide insights into the evolutionary history of the species and their relationship with each other [38,39]. The pangenome is also an important concept for phylogenetic studies; by comparing the pangenomes of different species, it is possible to identify the evolutionary relationships between them. This can be used to resolve the affiliations between different species and to study the evolution of new genetic structures [4,40,41]. Finally, the pangenome concept can be applied to the study of the genetic basis of adaptation; by comparing the pangenomes

of different species, it is possible to identify the genes that are associated with peculiar phenotypes and to understand how they are related [1,36]. This can provide insights into how particular traits have evolved and how they are maintained across different taxonomic units.



**Figure 1.** Example of core and accessory genome from three sequenced samples.

The pangenome concept can be formalized by focusing on its different components, which aim to provide a better understanding of the structure and dynamics of the gene content of a population. It can be tied to notions of static and dynamic genome composition and evolution, each of which has its own advantages and limitations. Static formulations of the pangenome can analyze the gene content of a population at a given time or location and are typically used to describe the core and accessory components of the pangenome [42,43]. A focus on core genes, for instance, considers only those that are shared by all members of the population and disregards the variability of the accessory genome. This model is useful for framing the stability and resilience of a population in changing environmental conditions by identifying shared or conserved genomic elements and can be used to pinpoint common druggable targets for antibiotic therapy [42,44,45]. An expanded model, on the other hand, may consider the variability of the accessory genome and provide a more accurate description of the genetic repertoire of a species, highlighting the emergence of novel genes by selection and adaptation [46]. The pangenome, on the other hand, can also be considered in its dynamic structure, by analyzing the temporal changes in the gene content of a population [8]. A “gene birth-and-death” model [47], for example, could consider the acquisition and loss of genes along a phylogeny and over time in order to predict the emergence of novel genetic components and, consequently, the evolutionary dynamics of a population [33,48]. However, they also have their own limitations, such as the fact that they may not consider stochastic changes in gene frequencies over time, the effects of gene interactions, and the complex regulatory networks that control the expression of genes in a population.

### 3. The Paradigm of Human Cell Types as Species

The concept of pangenome has recently been extended to higher-order organisms, and is being applied most notably to the human species [49–52]. By extension, the notion of a pangenome is based on the idea that different cell types can be regarded as “species” in their own right, each with a unique constitution of genetic material [18,53]. Indeed, human cells differ from each other in terms of their transcriptomes, proteomes, and metabolomes, as well as their genomic content [21,54,55]. This means that a single human tissue can contain several different genetic programmes, and this heterogeneity is reflected in the different collections of cell types that exist, as in the case of immune T cells studied at single-cell

level [54,56,57]. This view allows researchers to study the genetic composition of the same individual at different levels of detail, from the single cell to the entire organism [58,59]. Pangenome analysis of human cells can therefore be used to identify genetic differences between cell types, as well as to identify genes that are associated with particular cellular phenotypes. Such analysis can help us to understand the function of individual genes, as well as to identify novel gene products that are associated with specific cell types. Additionally, pangenome analysis can also be used to identify genetic elements that are associated with diseases, such as cancer, and to develop therapeutic approaches that target these objects [60,61]. Indeed, although some cancers present a non-negligible familiarity due to germline mutations in specific target genes [62–64], most of them are somatic in nature, and a pangenomic approach to their cellular makeup has become a primary way of exploring the molecular identity and phenotypic variability of this relatively common pathology [53,60,61,65,66].

So, the pangenome provides a powerful tool for understanding the complexity of human cells and for developing novel therapies for the treatment of different human diseases. This concept is particularly relevant to the field of personalized medicine, as it allows for a much more targeted approach to diagnosis and treatment. By considering the individual's genetic composition at different levels, a better assessment could be provided for the likelihood of a patient developing a particular condition, or how well a particular treatment will work for them. In addition, the concept of pangenome has important implications for the field of biotechnology. By understanding the pangenomic composition of different cell types, it is possible to develop new technologies that are tailored to work with a particular set of genetic material. For example, gene-editing techniques such as CRISPR-Cas9 can be used to modify the genetic material of a particular cell type in order to perform lab screening, treat a particular condition, or to create a new type of cell with desired characteristics [67,68]. This has the potential to revolutionize the way targeted treatments are designed and applied in a clinical setting.

#### 4. Describing the Repertoire of Structural Variation

Structural variation (SV) refers to the large-scale changes in the structure of the genome, such as deletions, duplications, inversions, and translocations [69]. These changes can occur in both coding and non-coding regions and can lead to significant changes in gene expression, function, and underlying phenotype [70–73]. SVs are a major contributor to the diversity of the human genome, and they can cause disease when they alter important regulatory elements or disrupt essential genes [74,75]. The vast majority of SVs are not found in the reference genome, so their detection and characterization are challenging [76–79]. The study of structural variation has been greatly facilitated by the development of high-throughput sequencing technologies, as these allow for the detection of such events at a much higher resolution and in larger numbers than previously possible [76,80]. The repertoire of SVs that can be found in a given population is vast, and it is constantly evolving; SVs are dynamic and can be used by a variety of processes, such as the unequal crossing over between homologous chromosomes, non-homologous end joining, and replication errors; they can also be inherited from parent to offspring or can be acquired *de novo*; they can also occur in any region of the genome, from the smallest single nucleotide changes to the largest chromosomal rearrangements [81–83]. The study of SVs provides important insights into the evolution of the human genome, as well as the genetic basis of diseases, as they can affect gene expression, leading to changes in phenotype, and can influence the efficacy of certain treatments [82,83]. In addition, SVs can provide important clues to the evolutionary history of a species, as they can provide evidence of recombination events between closely related lineages or between populations [79,83,84]. As such, SVs are an important aspect of the pan-genome, providing a dynamic view of the evolution of a species over time.

## 5. Increasing SNP Discovery, Mappability, and Association

Recent advances in the field of genomics have facilitated the identification of single nucleotide polymorphisms (SNPs), which are the most common type of genetic variation in the human genome [85–87]. SNPs are small genetic changes that can occur at a single base pair and, when they occur within a gene or a regulatory region of the genome, they can affect its function or the regulation of its expression [88–90]. While the discovery of SNPs is relatively straightforward, the challenge lies in their mapping and in determining how they contribute to the phenotype of an individual [91]. The ability to map SNPs has been greatly enhanced by the development of high-throughput sequencing technologies such as next-generation sequencing (NGS), which allows for rapid and cost-effective whole-genome sequencing. The NGS data generated can then be used to identify SNPs and map them to specific locations in the genome. These data can also be used to identify gene expression patterns associated with SNPs and to determine how they contribute to phenotypic variation. In addition to advances in technology, the development of databases and bioinformatics tools has also enabled the search for SNPs in large datasets. The use of bioinformatics tools to integrate SNP data with other types of data, such as transcriptomic and proteomic data, has also enabled researchers to identify correlations between SNPs and phenotypic variation [92–94]. This has been particularly useful in the field of personalized medicine, where the identification of SNPs associated with disease can help to develop more precise treatments and therapies [95,96]. The pangenome has enabled a much better understanding of genetic variation and diversity among organisms, allowing for an unprecedented level of SNP discovery and mappability [97,98]. Indeed, similarly to structural variants, by studying the entire genomic content of a population rather than individual genomes, a much wider range of SNPs can be identified. The pangenome can be used to improve the accuracy of existing genetic maps, as it can be used to infer population structure and identify recombination hotspots between different lineages [99,100]. This can then be used to refine existing maps and identify areas of higher recombination, which is especially important for genome-wide association analyses. Conversely, the pangenome can also be used to identify areas of conserved genetic variation, which can aid in the identification of regions of interest in evolutionary studies [98]. Finally, the pangenome can also be used to create more accurate and comprehensive databases of SNPs, which can then be used to facilitate the development of more powerful bioinformatics tools and data mining techniques [101–103]. So, by utilizing the pangenome, researchers can compare the entire genome of an individual to the various genetic compositions present within a population and identify SNPs that are specific to an individual. This allows for a more comprehensive analysis of the effects of an SNP on the genetic code of an individual, as well as providing a more detailed understanding of the population-specific genetic makeup.

In the context of SNP-phenotype relationship elucidation, genome-wide association studies (GWAS) have been increasingly important for deciphering the complex relationships between genes and physical characteristics [104–107]. GWAS is a type of statistical approach that seeks to identify correlations between genetic variants in large populations starting from given phenotypes [108]. Initially undertaken with the application of single nucleotide polymorphisms (SNPs), GWAS approaches have since been adapted for use with other genetic variants such as insertions and deletions [108]. An important consideration in the effectiveness of GWAS is the sample size used, as the larger the sample, the more accurate and comprehensive the results become [109–111]. The use of pangenomics as part of GWAS has grown in recent years as a means of improving the accuracy of estimated associations (pan-GWAS) as well as novel discovery [14,112–114]. Indeed, the collective term “pangenomics” includes different types of data and information that are taken from multiple sources of genetic material through whole-genome sequencing and genotyping. From these datasets, an extended set of SNPs and other variants can be identified, yielding more genetic information than is available with single-source datasets. Therefore, using pangenomics-based genetic data in GWAS, researchers can obtain more precise and comprehensive understanding of the genetic variants associated with diseases

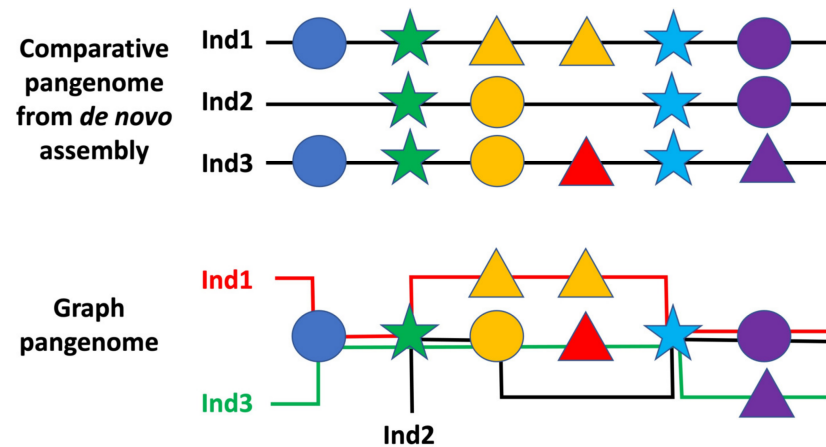
and other phenotypes at a higher level of taxonomic description. Interestingly, the use of pangenomics has also extended to the study of the roles of gene-gene (or eQTL) interactions in regulating gene expression levels and the resulting phenotypes [114,115]. In this way, pangenomics can directly enhance GWAS results, as the genome-wide scan can be augmented with additional data on eQTLs and protein-protein interactions to more accurately identify potentially causative genetic variants which may be involved in complex multifactorial diseases such as diabetes or dementia. Another useful tool in pangenomic-driven association studies is cell-line (or organoid [116–118])-based phenotypic screening [119,120]. This technique involves using cell-based models, such as a lab-created in vitro cell cultures, to screen for a specific phenotype change in response to a genetic variant. By introducing an engineered genetic variant detected through pan-GWAS into a cell, researchers can use high-throughput assays to assess the effect of the variant on a particular phenotype. This approach, supplementing genome-wide scans, provides results which can be directly attributed to a genetic variant.

Overall, the pangenome concept represents a powerful tool for increasing the discovery and resolution of single nucleotide polymorphisms (SNPs), as well as structural variants, in the human genome. By providing a more comprehensive analysis of the genetic code of an individual, as well as providing insights into the population-specific genetic makeup, the pangenome is a valuable tool for researchers studying the genetic diversity of human populations.

## 6. Pangenomic Non-Linearity and Larger Structural Variations

The concept of the pangenome has opened up the possibility of looking into the genome as a non-linear object (Figure 2). This is because the genomes of individuals are not necessarily identical, and different individuals may have different genomic compositions [121]. In fact, there are various factors that influence the genetic makeup of individuals, such as genetic recombination, mutations, and even environmental factors [122,123], and this means that the genetic composition of individuals can vary, even though they are members of the same species. As such, the study of the human pangenome has the opportunity to become a powerful tool for further understanding the complexity of genome diversity of humans, i.e., their structural variation between individuals, aiming for better understand the genetic basis of common diseases. Additionally, the study of the human pangenome can also help us to better disentangle the evolution of the human genome, and how different genetic components interact with one another. The implications of the human pangenome go beyond just the study of disease, as it can also be used to explore the functional roles of different genes and gene clusters in various biological processes and how they interact with one another in various ways [124–126]. Thinking in terms of population genetics, the fact that the pangenome can be interpreted as a non-linear object implies that the genetic potential of a group of individuals is not restricted to a linear arrangement of nucleotides, but instead involves a much more complex repertoire of structural variation that may be shaped at the individual level by various epigenetic, regulatory, and metabolic patterns. Furthermore, it is now known that processes such as gene expression and chromatin remodelling are largely responsible for drastically altering the quality, quantity, and even the type of gene products generated from the same sequence of DNA. This means that the same gene in the pangenome can give rise to different outcomes, depending on how it is regulated and expressed, implying that the linearity of a single individual's DNA is not enough to determine the final outcome of a particular gene. In addition, pangenomic non-linearity also allows for the detection of novel trait-related structural gene variants that may not have been possible to identify before, similarly to the accessory genes that favour the development of drug resistance in bacteria or the emergence of new morphological features in plants. Ideally, the non-linearity of the pangenome through structural variation is an essential factor that may contribute to the evolution of a species and the development of novel traits. It is also relevant to the study of human health and disease, as it can help to

explain the diversity of clinical manifestations seen in various genetic disorders and can assist in the development of better therapies and treatments.



**Figure 2.** Example of a comparative and a graph interpretation of the pangenome based on three sequenced individuals. Symbols correspond to genomic elements belonging to the core, shell or cloud genome as shown in Figure 1. Colors represent elements in the same position. Individual 1 (Ind1) shows a duplication of the third genomic element, which is different than the one carried by the other individuals and is, therefore, part of the cloud genome.

## 7. Technical, Ethical, and Legal Considerations

The prospect of a human pangenome holds tremendous potential for applications ranging from personalized medicine to forensics and population genetics. However, the technical implications of such a project should not be overlooked [49,50]. Indeed, the sheer amount of data that would need to be generated and stored to map the human pangenome is immense and could prove to be an obstacle for its implementation. For example, the number of individuals that would need to be genotyped in order to obtain a reliable and accurate representation of the human pangenome could be so large that it is not necessarily feasible to do it with the current technology. However, novel bioinformatic methods have been developed in recent years to allow the analysis of such mass of data from microbes, plants, and animals [98,127–129]. Furthermore, the data generated from individuals involved in such projects would not only be massive, but also highly sensitive and personal, as they would contain the complete and exact sequence of an individual’s genetic material [130–134]. This raises ethical issues related to the right to privacy and data protection, as well as potential misuse, as the data could be used to identify subjects or to discriminate against certain populations or individuals [135–139]. Human genome analysis is already being used to identify new ways to diagnose and treat medical conditions, as well as to better understand the underlying genetic basis of disease. By understanding a person’s full genetic makeup, doctors can more accurately diagnose diseases, as well as tailor treatments to target specific genetic mutations. So, the pangenome can be used to develop personalized medicine, where treatments are tailored to an individual’s unique genetic profile. Overall, the pangenome has the potential to revolutionize genetics-informed medical practice and provide valuable insight into the genetic basis of diseases. On the other hand, the use of the pangenome could also lead to further ethical dilemmas. For example, it could potentially be used to discriminate against individuals or communities based on their genetic makeup. This could lead to people being denied access to healthcare, education, or employment, due to their genetic profile [140,141]. It is also important to consider the potential repercussions of data privacy and sharing around the human pangenome in order to ensure that it is used responsibly and ethically. For example, who owns and has the right to access the data that is generated? Who has the right to determine how the data should be used? How will the data be used and how will they be protected from misuse? What legal framework can be put in place to ensure that the data are used in a responsible and appropriate way [142–145]? From

a public health perspective, it is possible that the data generated could reveal important information about certain diseases or health conditions or could be used to predict certain health outcomes. As such, it is important to consider how this information should be used, as well as the potential implications for public health policy [146–148]. If a person's entire genetic make-up were to be known, then it would be possible to identify and compare them to others with similar genetic profiles, with legal issues concerning, in particular, privacy and the unauthorized use of genetic data; an individual's genetic profile could be used to determine whether they may be predisposed to certain diseases or conditions, and this information could be used to deny them rights or opportunities. This could be especially concerning if employers or insurance companies were to gain access to an individual's genetic profile and use it to deprive them of access to a job or health insurance [149]. In addition, the privacy of an individual's genetic information could be compromised if it were to be shared without their knowledge or consent. Furthermore, the use of the human pangenome could also lead to further legal issues in the field of intellectual property. For instance, if a gene or genes were to be identified as being responsible for a particular trait, then it might be possible to patent those genes [150–153]. This could potentially lead to disagreements over who has the rights to the gene and its uses. Furthermore, the infringement of such patents could also potentially lead to further legal disputes. Finally, the use of the pangenome (both human and microbial) could also lead to further issues in the field of criminal justice [154,155]. For instance, it could potentially be used for suspect identification, and this could possibly lead to legal debates over the accuracy and reliability of the evidence. In addition, the use of genetic information could also potentially lead to arguments over the ethical and moral implications of using such evidence in criminal trials. Overall, the legal implications of the human pangenome are complex and far-reaching. Although the prospective benefits of utilizing this information are great, the potential risks and implications must also be taken into consideration. The concerns emerging from the use, production, and discovery of genetic data are, of course, already known at the level of the single individual; therefore, it is essential that any legal framework surrounding the use of this information be carefully considered and that appropriate measures are taken to ensure the privacy, safety, and ethical use of such data at a population level as well. Indeed, legal and technical safeguards are already in place to protect not only the physical entity that is the DNA but also the flow of all derived genomic data information, as well as to uphold anonymity and confidentiality, such as via cryptography, access control, and data perturbation [156–158]. However, in a world where at-home genetic testing is easily accessible for a small fee, and consumer genetics companies are allowed to sell their data to pharmaceutical corporations for drug development [159–161] or provide them to law enforcement [162,163], moving through the legal and ethical maze of what constitutes explicit consent to data sharing is still very unsteady territory and must be thoroughly scrutinized [164–166]. As subject privacy and data sharing cannot be mutually exclusive in the setting of a truly democratic science, encrypting the individual's genetic data but providing access to the pangenome, which does not inform on any single person's specific characteristics, may seem like a reasonable trade-off.

**Author Contributions:** Conceptualization, P.A. and D.L.; methodology, P.A. and E.C.; investigation, P.A. and E.C.; resources, D.L.; data curation, P.A.; writing—original draft preparation, P.A. and E.C.; writing—review and editing, P.A., E.C. and D.L.; supervision, D.L.; project administration, E.C. and D.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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



## References

1. Brockhurst, M.A.; Harrison, E.; Hall, J.P.J.; Richards, T.; McNally, A.; MacLean, C. The Ecology and Evolution of Pangenomes. *Curr. Biol.* **2019**, *29*, R1094–R1103. [[CrossRef](#)] [[PubMed](#)]
2. Golicz, A.A.; Bayer, P.E.; Bhalla, P.L.; Batley, J.; Edwards, D. Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. *Trends Genet.* **2020**, *36*, 132–145. [[CrossRef](#)] [[PubMed](#)]
3. Medini, D.; Donati, C.; Rappuoli, R.; Tettelin, H. The Pangenome: A Data-Driven Discovery in Biology. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; Springer: Cham, Switzerland, 2020; ISBN 978-3-030-38280-3.
4. Moldovan, M.A.; Gelfand, M.S. Pangenomic Definition of Prokaryotic Species and the Phylogenetic Structure of *Prochlorococcus* spp. *Front. Microbiol.* **2018**, *9*, 428. [[CrossRef](#)] [[PubMed](#)]
5. Bobay, L.-M. The Prokaryotic Species Concept and Challenges. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; Springer: Cham, Switzerland, 2020; ISBN 978-3-030-38280-3.
6. McInerney, J.O.; Whelan, F.J.; Domingo-Sananes, M.R.; McNally, A.; O'Connell, M.J. Pangenomes and Selection: The Public Goods Hypothesis. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; Springer: Cham, Switzerland, 2020; ISBN 978-3-030-38280-3.
7. Sela, I.; Wolf, Y.I.; Koonin, E.V. Theory of prokaryotic genome evolution. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11399–11407. [[CrossRef](#)]
8. Cummins, E.A.; Hall, R.J.; McInerney, J.O.; McNally, A. Prokaryote pangenomes are dynamic entities. *Curr. Opin. Microbiol.* **2022**, *66*, 73–78. [[CrossRef](#)]
9. McInerney, J.O. Prokaryotic Pangenomes Act as Evolving Ecosystems. *Mol. Biol. Evol.* **2023**, *40*, msac232. [[CrossRef](#)]
10. Vellai, T.; Vida, G. The origin of eukaryotes: The difference between prokaryotic and eukaryotic cells. *Proc. Biol. Sci.* **1999**, *266*, 1571–1577. [[CrossRef](#)]
11. Gabaldón, T. Origin and Early Evolution of the Eukaryotic Cell. *Annu. Rev. Microbiol.* **2021**, *75*, 631–647. [[CrossRef](#)]
12. Elliott, T.A.; Gregory, T.R. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, *370*, 20140331. [[CrossRef](#)]
13. Šatovič-Vukšić, E.; Plohl, M. Satellite DNAs-From Localized to Highly Dispersed Genome Components. *Genes* **2023**, *14*, 742. [[CrossRef](#)]
14. Li, N.; He, Q.; Wang, J.; Wang, B.; Zhao, J.; Huang, S.; Yang, T.; Tang, Y.; Yang, S.; Aisimutuola, P.; et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* **2023**, *55*, 852–860. [[CrossRef](#)]
15. Brooks, S.A.; Palermo, K.M.; Kahn, A.; Hein, J. Impact of white-spotting alleles, including W20, on phenotype in the American Paint Horse. *Anim. Genet.* **2020**, *51*, 707–715. [[CrossRef](#)]
16. Sibbald, S.J.; Eme, L.; Archibald, J.M.; Roger, A.J. Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes. *Trends Parasitol.* **2020**, *36*, 927–941. [[CrossRef](#)]
17. Aggarwal, S.K.; Singh, A.; Choudhary, M.; Kumar, A.; Rakshit, S.; Kumar, P.; Bohra, A.; Varshney, R.K. Pangenomics in Microbial and Crop Research: Progress, Applications, and Perspectives. *Genes* **2022**, *13*, 598. [[CrossRef](#)]
18. Apicella, C.; Ruano, C.S.M.; Thilaganathan, B.; Khalil, A.; Giorgione, V.; Gascoin, G.; Marcellin, L.; Gaspar, C.; Jacques, S.; Murdoch, C.E.; et al. Pan-Genomic Regulation of Gene Expression in Normal and Pathological Human Placentas. *Cells* **2023**, *12*, 578. [[CrossRef](#)]
19. Boeuf, D.; Eppley, J.M.; Mende, D.R.; Malmstrom, R.R.; Woyke, T.; DeLong, E.F. Metapangenomics reveals depth-dependent shifts in metabolic potential for the ubiquitous marine bacterial SAR324 lineage. *Microbiome* **2021**, *9*, 172. [[CrossRef](#)]
20. Singh, S.; Aghdam, S.A.; Lahowetz, R.M.; Brown, A.M.V. Metapangenomics of wild and cultivated banana microbiome reveals a plethora of host-associated protective functions. *Environ. Microbiome* **2023**, *18*, 36. [[CrossRef](#)]
21. Wang, J.; Shi, A.; Lyu, J. A comprehensive atlas of epigenetic regulators reveals tissue-specific epigenetic regulation patterns. *Epigenetics* **2023**, *18*, 2139067. [[CrossRef](#)]
22. Mawla, A.M.; van der Meulen, T.; Huising, M.O. Chromatin accessibility differences between alpha, beta, and delta cells identifies common and cell type-specific enhancers. *BMC Genom.* **2023**, *24*, 202. [[CrossRef](#)]
23. Schaeffer, M.; Nollmann, M. Contributions of 3D chromatin structure to cell-type-specific gene regulation. *Curr. Opin. Genet. Dev.* **2023**, *79*, 102032. [[CrossRef](#)]
24. Yaschenko, A.E.; Fenech, M.; Mazzoni-Putman, S.; Alonso, J.M.; Stepanova, A.N. Deciphering the molecular basis of tissue-specific gene expression in plants: Can synthetic biology help? *Curr. Opin. Plant Biol.* **2022**, *68*, 102241. [[CrossRef](#)]
25. El-Zein, M.; Cheishvili, D.; Gotlieb, W.; Gilbert, L.; Hemmings, R.; Behr, M.A.; Szyf, M.; Franco, E.L.; MARKER Study Group. Genome-wide DNA methylation profiling identifies two novel genes in cervical neoplasia. *Int. J. Cancer* **2020**, *147*, 1264–1274. [[CrossRef](#)] [[PubMed](#)]
26. Chen, P.; Bando, D.J.D.; Weimer, B.C. Bacterial Epigenomics: Epigenetics in the Age of Population Genomics. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; Springer: Cham, Switzerland, 2020; ISBN 978-3-030-38280-3.
27. Cui, W.-J.; Zhang, B.; Zhao, R.; Liu, L.-X.; Jiao, J.; Zhang, Z.; Tian, C.-F. Lineage-Specific Rewiring of Core Pathways Predating Innovation of Legume Nodules Shapes Symbiotic Efficiency. *mSystems* **2021**, *6*, e01299-20. [[CrossRef](#)] [[PubMed](#)]

28. Broadbent, J.A.; Broszczak, D.A.; Tennakoon, I.U.K.; Huygens, F. Pan-proteomics, a concept for unifying quantitative proteome measurements when comparing closely-related bacterial strains. *Expert Rev. Proteom.* **2016**, *13*, 355–365. [[CrossRef](#)] [[PubMed](#)]
29. Ma, B.; France, M.; Ravel, J. Meta-Pangenome: At the Crossroad of Pangenomics and Metagenomics. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; Springer: Cham, Switzerland, 2020; ISBN 978-3-030-38280-3.
30. Zhong, C.; Chen, C.; Wang, L.; Ning, K. Integrating pan-genome with metagenome for microbial community profiling. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1458–1466. [[CrossRef](#)] [[PubMed](#)]
31. Li, T.; Yin, Y. Critical assessment of pan-genomic analysis of metagenome-assembled genomes. *Brief. Bioinform.* **2022**, *23*, bbac413. [[CrossRef](#)]
32. Zhai, Y.; Wei, C. Open pangenome of *Lactococcus lactis* generated by a combination of metagenome-assembled genomes and isolate genomes. *Front. Microbiol.* **2022**, *13*, 948138. [[CrossRef](#)]
33. Romero Picazo, D.; Werner, A.; Dagan, T.; Kupczok, A. Pangenome Evolution in Environmentally Transmitted Symbionts of Deep-Sea Mussels Is Governed by Vertical Inheritance. *Genome Biol. Evol.* **2022**, *14*, evac098. [[CrossRef](#)]
34. Jaiswal, A.K.; Tiwari, S.; Tavares, G.C.; Da Silva, W.M.; De Castro Oliveira, L.; Ibraim, I.C.; Guimarães, L.C.; Gomide, A.C.P.; Jamal, S.B.; Pantoja, Y.; et al. Pan-omics focused to Crick's central dogma. In *Pan-genomics: Applications, Challenges, and Future Prospects*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 1–41. ISBN 978-0-12-817076-2.
35. Innamorati, K.A.; Earl, J.P.; Aggarwal, S.D.; Ehrlich, G.D.; Hiller, N.L. The Bacterial Guide to Designing a Diversified Gene Portfolio. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; Springer: Cham, Switzerland, 2020; ISBN 978-3-030-38280-3.
36. Tiwary, B.K. Evolutionary pan-genomics and applications. In *Pan-genomics: Applications, Challenges, and Future Prospects*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 65–80. ISBN 978-0-12-817076-2.
37. Douglas, G.M.; Shapiro, B.J. Genic Selection Within Prokaryotic Pangenomes. *Genome Biol. Evol.* **2021**, *13*, evab234. [[CrossRef](#)]
38. Hyun, J.C.; Monk, J.M.; Palsson, B.O. Comparative pangenomics: Analysis of 12 microbial pathogen pangenomes reveals conserved global structures of genetic and functional diversity. *BMC Genom.* **2022**, *23*, 7. [[CrossRef](#)]
39. Anderson, B.D.; Bisanz, J.E. Challenges and opportunities of strain diversity in gut microbiome research. *Front. Microbiol.* **2023**, *14*, 1117122. [[CrossRef](#)]
40. Maistrenko, O.M.; Mende, D.R.; Luetge, M.; Hildebrand, F.; Schmidt, T.S.B.; Li, S.S.; Rodrigues, J.F.M.; von Mering, C.; Pedro Coelho, L.; Huerta-Cepas, J.; et al. Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. *ISME J.* **2020**, *14*, 1247–1259. [[CrossRef](#)]
41. Köstlbacher, S.; Collingro, A.; Halter, T.; Schulz, F.; Jungbluth, S.P.; Horn, M. Pangenomics reveals alternative environmental lifestyles among chlamydiae. *Nat. Commun.* **2021**, *12*, 4021. [[CrossRef](#)]
42. de Korne-Elenbaas, J.; Bruisten, S.M.; van Dam, A.P.; Maiden, M.C.J.; Harrison, O.B. The *Neisseria gonorrhoeae* Accessory Genome and Its Association with the Core Genome and Antimicrobial Resistance. *Microbiol. Spectr.* **2022**, *10*, e0265421. [[CrossRef](#)]
43. Mesa, V.; Monot, M.; Ferraris, L.; Popoff, M.; Mazuet, C.; Barbut, F.; Delannoy, J.; Dupuy, B.; Butel, M.-J.; Aires, J. Core-, pan- and accessory genome analyses of *Clostridium neonatale*: Insights into genetic diversity. *Microb. Genom.* **2022**, *8*, mgen000813. [[CrossRef](#)]
44. Zakham, F.; Sironen, T.; Vapalahti, O.; Kant, R. Pan and Core Genome Analysis of 183 *Mycobacterium tuberculosis* Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains. *Antibiotics* **2021**, *10*, 500. [[CrossRef](#)]
45. Golchha, N.C.; Nighojkar, A.; Nighojkar, S. Redefining genomic view of *Clostridioides difficile* through pangenome analysis and identification of drug targets from its core genome. *Drug Target Insights* **2022**, *16*, 17–24. [[CrossRef](#)]
46. Whelan, F.J.; Hall, R.J.; McInerney, J.O. Evidence for Selection in the Abundant Accessory Gene Content of a Prokaryote Pangenome. *Mol. Biol. Evol.* **2021**, *38*, 3697–3708. [[CrossRef](#)]
47. Zarebski, A.E.; Du Plessis, L.; Parag, K.V.; Pybus, O.G. A computationally tractable birth-death model that combines phylogenetic and epidemiological data. *PLoS Comput. Biol.* **2022**, *18*, e1009805. [[CrossRef](#)]
48. Tranchant-Dubreuil, C.; Rouard, M.; Sabot, F. Plant Pangenome: Impacts on Phenotypes and Evolution. In *Annual Plant Reviews Online*; Roberts, J.A., Ed.; Wiley: Hoboken, NJ, USA, 2019; pp. 453–478. ISBN 978-1-119-31299-4.
49. Miga, K.H.; Wang, T. The Need for a Human Pangenome Reference Sequence. *Annu. Rev. Genomics Hum. Genet.* **2021**, *22*, 81–102. [[CrossRef](#)]
50. Wang, T.; Antonacci-Fulton, L.; Howe, K.; Lawson, H.A.; Lucas, J.K.; Phillippy, A.M.; Popejoy, A.B.; Asri, M.; Carson, C.; Chaisson, M.J.P.; et al. The Human Pangenome Project: A global resource to map genomic diversity. *Nature* **2022**, *604*, 437–446. [[CrossRef](#)] [[PubMed](#)]
51. Singh, V.; Pandey, S.; Bhardwaj, A. From the reference human genome to human pangenome: Premise, promise and challenge. *Front. Genet.* **2022**, *13*, 1042550. [[CrossRef](#)] [[PubMed](#)]
52. Liao, W.-W.; Asri, M.; Ebler, J.; Doerr, D.; Haukness, M.; Hickey, G.; Lu, S.; Lucas, J.K.; Monlong, J.; Abel, H.J.; et al. A draft human pangenome reference. *Nature* **2023**, *617*, 312–324. [[CrossRef](#)] [[PubMed](#)]
53. Neou, M.; Villa, C.; Armignacco, R.; Jouinot, A.; Raffin-Sanson, M.-L.; Septier, A.; Letourneur, F.; Diry, S.; Diedisheim, M.; Izac, B.; et al. Pangenomic Classification of Pituitary Neuroendocrine Tumors. *Cancer Cell* **2020**, *37*, 123–134.e5. [[CrossRef](#)]
54. Pace, L. Temporal and Epigenetic Control of Plasticity and Fate Decision during CD8+ T-Cell Memory Differentiation. *Cold Spring Harb. Perspect. Biol.* **2021**, *13*, a037754. [[CrossRef](#)]

55. Melby, J.A.; Brown, K.A.; Gregorich, Z.R.; Roberts, D.S.; Chapman, E.A.; Ehlers, L.E.; Gao, Z.; Larson, E.J.; Jin, Y.; Lopez, J.R.; et al. High sensitivity top-down proteomics captures single muscle cell heterogeneity in large proteoforms. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2222081120. [[CrossRef](#)]
56. Pace, L.; Amigorena, S. Epigenetics of T cell fate decision. *Curr. Opin. Immunol.* **2020**, *63*, 43–50. [[CrossRef](#)]
57. Pace, L.; Goudot, C.; Zueva, E.; Gueguen, P.; Burgdorf, N.; Waterfall, J.J.; Quivy, J.-P.; Almouzni, G.; Amigorena, S. The epigenetic control of stemness in CD8<sup>+</sup> T cell fate commitment. *Science* **2018**, *359*, 177–186. [[CrossRef](#)]
58. Sreenivasan, V.K.A.; Balachandran, S.; Spielmann, M. The role of single-cell genomics in human genetics. *J. Med. Genet.* **2022**, *59*, 827–839. [[CrossRef](#)]
59. Vodovotz, Y. Towards systems immunology of critical illness at scale: From single cell 'omics to digital twins. *Trends Immunol.* **2023**, *44*, 345–355. [[CrossRef](#)]
60. Stenman, A.; Yang, M.; Paulsson, J.O.; Zedenius, J.; Paulsson, K.; Juhlin, C.C. Pan-Genomic Sequencing Reveals Actionable CDKN2A/2B Deletions and Kataegis in Anaplastic Thyroid Carcinoma. *Cancers* **2021**, *13*, 6340. [[CrossRef](#)]
61. Yu, Y.; Zhang, Z.; Dong, X.; Yang, R.; Duan, Z.; Xiang, Z.; Li, J.; Li, G.; Yan, F.; Xue, H.; et al. Pangenomic analysis of Chinese gastric cancer. *Nat. Commun.* **2022**, *13*, 5412. [[CrossRef](#)]
62. Hemminki, K.; Li, X.; Försti, A.; Eng, C. Are population level familial risks and germline genetics meeting each other? *Hered. Cancer Clin. Pract.* **2023**, *21*, 3. [[CrossRef](#)]
63. Hart, S.N.; Polley, E.C.; Yussuf, A.; Yadav, S.; Goldgar, D.E.; Hu, C.; LaDuca, H.; Smith, L.P.; Fujimoto, J.; Li, S.; et al. Mutation prevalence tables for hereditary cancer derived from multigene panel testing. *Hum. Mutat.* **2020**, *41*, e1–e6. [[CrossRef](#)]
64. Tomlinson-Hansen, S.; Beaston, M. Hereditary Cancer Genes and Related Risks. *Rhode Isl. Med. J.* **2023**, *106*, 12–17.
65. Stenman, A.; Backman, S.; Johansson, K.; Paulsson, J.O.; Stålberg, P.; Zedenius, J.; Juhlin, C.C. Pan-genomic characterization of high-risk pediatric papillary thyroid carcinoma. *Endocr. Relat. Cancer* **2021**, *28*, 337–351. [[CrossRef](#)]
66. López-Carrasco, A.; Berbegall, A.P.; Martín-Vañó, S.; Blanquer-Maceiras, M.; Castel, V.; Navarro, S.; Noguera, R. Intra-Tumour Genetic Heterogeneity and Prognosis in High-Risk Neuroblastoma. *Cancers* **2021**, *13*, 5173. [[CrossRef](#)]
67. Tay Fernandez, C.G.; Nestor, B.J.; Danilevicz, M.F.; Marsh, J.I.; Petereit, J.; Bayer, P.E.; Batley, J.; Edwards, D. Expanding Gene-Editing Potential in Crop Improvement with Pangenomes. *Int. J. Mol. Sci.* **2022**, *23*, 2276. [[CrossRef](#)]
68. Gao, Y.; Guitton-Sert, L.; Dessapt, J.; Coulombe, Y.; Rodrigue, A.; Milano, L.; Blondeau, A.; Larsen, N.B.; Duxin, J.P.; Hussein, S.; et al. A CRISPR-Cas9 screen identifies EXO1 as a formaldehyde resistance gene. *Nat. Commun.* **2023**, *14*, 381. [[CrossRef](#)]
69. Weckselblatt, B.; Rudd, M.K. Human Structural Variation: Mechanisms of Chromosome Rearrangements. *Trends Genet.* **2015**, *31*, 587–599. [[CrossRef](#)]
70. Boyling, A.; Perez-Siles, G.; Kennerson, M.L. Structural Variation at a Disease Mutation Hotspot: Strategies to Investigate Gene Regulation and the 3D Genome. *Front. Genet.* **2022**, *13*, 842860. [[CrossRef](#)] [[PubMed](#)]
71. Zedan, H.T.; Ali, F.H.; Zayed, H. The spectrum of chromosomal translocations in the Arab world: Ethnic-specific chromosomal translocations and their relevance to diseases. *Chromosoma* **2022**, *131*, 127–146. [[CrossRef](#)] [[PubMed](#)]
72. Acemel, R.D.; Lupiáñez, D.G. Evolution of 3D chromatin organization at different scales. *Curr. Opin. Genet. Dev.* **2023**, *78*, 102019. [[CrossRef](#)]
73. Gavril, E.-C.; Nucă, I.; Pânzaru, M.-C.; Ivanov, A.V.; Mihai, C.-T.; Antoci, L.-M.; Ciobanu, C.-G.; Rusu, C.; Popescu, R. Genotype-Phenotype Correlations in 2q37-Deletion Syndrome: An Update of the Clinical Spectrum and Literature Review. *Genes* **2023**, *14*, 465. [[CrossRef](#)] [[PubMed](#)]
74. Macciardi, F.; Giulia Bacalini, M.; Miramontes, R.; Boattini, A.; Taccioli, C.; Modenini, G.; Malhas, R.; Anderlucci, L.; Gusev, Y.; Gross, T.J.; et al. A retrotransposon storm marks clinical phenoconversion to late-onset Alzheimer's disease. *GeroScience* **2022**, *44*, 1525–1550. [[CrossRef](#)]
75. Modenini, G.; Abondio, P.; Guffanti, G.; Boattini, A.; Macciardi, F. Evolutionarily recent retrotransposons contribute to schizophrenia. *Transl. Psychiatry* **2023**, *13*, 181. [[CrossRef](#)]
76. Mahmoud, M.; Gobet, N.; Cruz-Dávalos, D.I.; Mounier, N.; Dessimoz, C.; Sedlazeck, F.J. Structural variant calling: The long and the short of it. *Genome Biol.* **2019**, *20*, 246. [[CrossRef](#)]
77. Zanini, S.F.; Bayer, P.E.; Wells, R.; Snowdon, R.J.; Batley, J.; Varshney, R.K.; Nguyen, H.T.; Edwards, D.; Golicz, A.A. Pangenomics in crop improvement—from coding structural variations to finding regulatory variants with pangenome graphs. *Plant Genome* **2022**, *15*, e20177. [[CrossRef](#)]
78. Liu, Z.; Roberts, R.; Mercer, T.R.; Xu, J.; Sedlazeck, F.J.; Tong, W. Towards accurate and reliable resolution of structural variants for clinical diagnosis. *Genome Biol.* **2022**, *23*, 68. [[CrossRef](#)]
79. Pokrovac, I.; Pezer, Ž. Recent advances and current challenges in population genomics of structural variation in animals and plants. *Front. Genet.* **2022**, *13*, 1060898. [[CrossRef](#)]
80. Warburton, P.E.; Sebra, R.P. Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annu. Rev. Genomics Hum. Genet.* **2023**, *24*. [[CrossRef](#)]
81. Currall, B.B.; Chiang, C.; Talkowski, M.E.; Morton, C.C. Mechanisms for Structural Variation in the Human Genome. *Curr. Genet. Med. Rep.* **2013**, *1*, 81–90. [[CrossRef](#)]
82. Carvalho, C.M.B.; Lupski, J.R. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* **2016**, *17*, 224–238. [[CrossRef](#)]

83. Collins, R.L.; Brand, H.; Karczewski, K.J.; Zhao, X.; Alföldi, J.; Francioli, L.C.; Khera, A.V.; Lowther, C.; Gauthier, L.D.; Wang, H.; et al. A structural variation reference for medical and population genetics. *Nature* **2020**, *581*, 444–451. [[CrossRef](#)]
84. Quan, C.; Lu, H.; Lu, Y.; Zhou, G. Population-scale genotyping of structural variation in the era of long-read sequencing. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 2639–2647. [[CrossRef](#)]
85. Jiang, Z.; Wang, H.; Michal, J.J.; Zhou, X.; Liu, B.; Woods, L.C.S.; Fuchs, R.A. Genome Wide Sampling Sequencing for SNP Genotyping: Methods, Challenges and Future Development. *Int. J. Biol. Sci.* **2016**, *12*, 100–108. [[CrossRef](#)]
86. Van Asselt, A.J.; Ehli, E.A. Whole-Genome Genotyping Using DNA Microarrays for Population Genetics. *Methods Mol. Biol.* **2022**, *2418*, 269–287. [[CrossRef](#)]
87. Kockum, I.; Huang, J.; Stridh, P. Overview of Genotyping Technologies and Methods. *Curr. Protoc.* **2023**, *3*, e727. [[CrossRef](#)]
88. Robert, F.; Pelletier, J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. *Front. Genet.* **2018**, *9*, 507. [[CrossRef](#)]
89. Shatoff, E.; Bundschuh, R. Single nucleotide polymorphisms affect RNA-protein interactions at a distance through modulation of RNA secondary structures. *PLoS Comput. Biol.* **2020**, *16*, e1007852. [[CrossRef](#)]
90. Ye, Y.; Zhang, Z.; Liu, Y.; Diao, L.; Han, L. A Multi-Omics Perspective of Quantitative Trait Loci in Precision Medicine. *Trends Genet.* **2020**, *36*, 318–336. [[CrossRef](#)] [[PubMed](#)]
91. Zheng, H.; Zhao, X.; Wang, H.; Ding, Y.; Lu, X.; Zhang, G.; Yang, J.; Wang, L.; Zhang, H.; Bai, Y.; et al. Location deviations of DNA functional elements affected SNP mapping in the published databases and references. *Brief. Bioinform.* **2020**, *21*, 1293–1301. [[CrossRef](#)] [[PubMed](#)]
92. Manzoni, C.; Kia, D.A.; Vandrovcova, J.; Hardy, J.; Wood, N.W.; Lewis, P.A.; Ferrari, R. Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Brief. Bioinform.* **2018**, *19*, 286–302. [[CrossRef](#)] [[PubMed](#)]
93. Pinu, F.R.; Beale, D.J.; Paten, A.M.; Kouremenos, K.; Swarup, S.; Schirra, H.J.; Wishart, D. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites* **2019**, *9*, 76. [[CrossRef](#)] [[PubMed](#)]
94. Balagué-Dobón, L.; Cáceres, A.; González, J.R. Fully exploiting SNP arrays: A systematic review on the tools to extract underlying genomic structure. *Brief. Bioinform.* **2022**, *23*, bbac043. [[CrossRef](#)]
95. Strianese, O.; Rizzo, F.; Ciccarelli, M.; Galasso, G.; D’Agostino, Y.; Salvati, A.; Del Giudice, C.; Tesorio, P.; Rusciano, M.R. Precision and Personalized Medicine: How Genomic Approach Improves the Management of Cardiovascular and Neurodegenerative Disease. *Genes* **2020**, *11*, 747. [[CrossRef](#)]
96. Johansson, Å.; Andreassen, O.A.; Brunak, S.; Franks, P.W.; Hedman, H.; Loos, R.J.F.; Meder, B.; Melén, E.; Wheelock, C.E.; Jacobsson, B. Precision medicine in complex diseases—Molecular subgrouping for improved prediction and treatment stratification. *J. Intern. Med.* **2023**. *early view*. [[CrossRef](#)]
97. Hurgobin, B.; Edwards, D. SNP Discovery Using a Pangenome: Has the Single Reference Approach Become Obsolete? *Biology* **2017**, *6*, 21. [[CrossRef](#)]
98. Gong, Y.; Li, Y.; Liu, X.; Ma, Y.; Jiang, L. A review of the pangenome: How it affects our understanding of genomic variation, selection and breeding in domestic animals? *J. Anim. Sci. Biotechnol.* **2023**, *14*, 73. [[CrossRef](#)]
99. MacAlasdair, N.; Pesonen, M.; Brynildsrud, O.; Eldholm, V.; Kristiansen, P.A.; Corander, J.; Caugant, D.A.; Bentley, S.D. The effect of recombination on the evolution of a population of *Neisseria meningitidis*. *Genome Res.* **2021**, *31*, 1258–1268. [[CrossRef](#)]
100. Preska Steinberg, A.; Lin, M.; Kussell, E. Core genes can have higher recombination rates than accessory genes within global microbial populations. *eLife* **2022**, *11*, e78533. [[CrossRef](#)]
101. Sirén, J.; Monlong, J.; Chang, X.; Novak, A.M.; Eizenga, J.M.; Markello, C.; Sibbesen, J.A.; Hickey, G.; Chang, P.-C.; Carroll, A.; et al. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* **2021**, *374*, abg8871. [[CrossRef](#)]
102. Ebler, J.; Ebert, P.; Clarke, W.E.; Rausch, T.; Audano, P.A.; Houwaart, T.; Mao, Y.; Korbel, J.O.; Eichler, E.E.; Zody, M.C.; et al. Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes. *Nat. Genet.* **2022**, *54*, 518–525. [[CrossRef](#)]
103. Mun, T.; Vaddadi, N.S.K.; Langmead, B. Pangenomic genotyping with the marker array. *Algorithms Mol. Biol.* **2023**, *18*, 2. [[CrossRef](#)]
104. Alonso-Hearn, M.; Badia-Bingué, G.; Canive, M. Genome-wide association studies for the identification of cattle susceptible and resilient to paratuberculosis. *Front. Vet. Sci.* **2022**, *9*, 935133. [[CrossRef](#)]
105. Narumi, S. Genome-wide association studies for thyroid physiology and diseases. *Endocr. J.* **2023**, *70*, 9–17. [[CrossRef](#)]
106. Walsh, R.; Jurgens, S.J.; Erdmann, J.; Bezzina, C.R. Genome-wide association studies of cardiovascular disease. *Physiol. Rev.* **2023**, *103*, 2039–2055. [[CrossRef](#)]
107. Vallarino, J.G.; Jun, H.; Wang, S.; Wang, X.; Sade, N.; Orf, I.; Zhang, D.; Shi, J.; Shen, S.; Cuadros-Inostroza, Á.; et al. Limitations and advantages of using metabolite-based genome-wide association studies: Focus on fruit quality traits. *Plant Sci. Int. J. Exp. Plant Biol.* **2023**, *333*, 111748. [[CrossRef](#)]
108. Lemay, M.-A.; Malle, S. A Practical Guide to Using Structural Variants for Genome-Wide Association Studies. *Methods Mol. Biol.* **2022**, *2481*, 161–172. [[CrossRef](#)]
109. Nishino, J.; Ochi, H.; Kochi, Y.; Tsunoda, T.; Matsui, S. Sample Size for Successful Genome-Wide Association Study of Major Depressive Disorder. *Front. Genet.* **2018**, *9*, 227. [[CrossRef](#)]
110. Moore, C.M.; Jacobson, S.A.; Fingerlin, T.E. Power and Sample Size Calculations for Genetic Association Studies in the Presence of Genetic Model Misspecification. *Hum. Hered.* **2019**, *84*, 256–271. [[CrossRef](#)] [[PubMed](#)]

111. Politi, C.; Roumeliotis, S.; Tripepi, G.; Spoto, B. Sample Size Calculation in Genetic Association Studies: A Practical Approach. *Life* **2023**, *13*, 235. [[CrossRef](#)] [[PubMed](#)]
112. Jin, S.; Han, Z.; Hu, Y.; Si, Z.; Dai, F.; He, L.; Cheng, Y.; Li, Y.; Zhao, T.; Fang, L.; et al. Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. *Mol. Plant* **2023**, *16*, 678–693. [[CrossRef](#)] [[PubMed](#)]
113. Gupta, P.K. GWAS for genetics of complex quantitative traits: Genome to pangenome and SNPs to SVs and k-mers. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **2021**, *43*, e2100109. [[CrossRef](#)]
114. Zhou, Y.; Zhang, Z.; Bao, Z.; Li, H.; Lyu, Y.; Zan, Y.; Wu, Y.; Cheng, L.; Fang, Y.; Wu, K.; et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **2022**, *606*, 527–534. [[CrossRef](#)]
115. Li, J.; Yuan, D.; Wang, P.; Wang, Q.; Sun, M.; Liu, Z.; Si, H.; Xu, Z.; Ma, Y.; Zhang, B.; et al. Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biol.* **2021**, *22*, 119. [[CrossRef](#)]
116. Magré, L.; Verstegen, M.M.A.; Buschow, S.; van der Laan, L.J.W.; Peppelenbosch, M.; Desai, J. Emerging organoid-immune co-culture models for cancer research: From oncimmunology to personalized immunotherapies. *J. Immunother. Cancer* **2023**, *11*, e006290. [[CrossRef](#)]
117. Neal, J.T.; Li, X.; Zhu, J.; Giangarra, V.; Grzeskowiak, C.L.; Ju, J.; Liu, I.H.; Chiou, S.-H.; Salahudeen, A.A.; Smith, A.R.; et al. Organoid Modeling of the Tumor Immune Microenvironment. *Cell* **2018**, *175*, 1972–1988.e16. [[CrossRef](#)]
118. Zhou, J.-Q.; Zeng, L.-H.; Li, C.-T.; He, D.-H.; Zhao, H.-D.; Xu, Y.-N.; Jin, Z.-T.; Gao, C. Brain organoids are new tool for drug screening of neurological diseases. *Neural Regen. Res.* **2023**, *18*, 1884–1889. [[CrossRef](#)]
119. Moffat, J.G.; Vincent, F.; Lee, J.A.; Eder, J.; Prunotto, M. Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nat. Rev. Drug Discov.* **2017**, *16*, 531–543. [[CrossRef](#)]
120. Heinrich, L.; Kumbier, K.; Li, L.; Altschuler, S.J.; Wu, L.F. Selection of Optimal Cell Lines for High-Content Phenotypic Screening. *ACS Chem. Biol.* **2023**, *18*, 679–685. [[CrossRef](#)]
121. Eichler, E.E. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *N. Engl. J. Med.* **2019**, *381*, 64–74. [[CrossRef](#)]
122. Mathieson, I. Human adaptation over the past 40,000 years. *Curr. Opin. Genet. Dev.* **2020**, *62*, 97–104. [[CrossRef](#)]
123. Soto, D.C.; Uribe-Salazar, J.M.; Shew, C.J.; Sekar, A.; McGinty, S.P.; Dennis, M.Y. Genomic structural variation: A complex but important driver of human evolution. *Am. J. Biol. Anthropol.* **2023**. *early view*. [[CrossRef](#)]
124. Simonti, C.N.; Capra, J.A. The evolution of the human genome. *Curr. Opin. Genet. Dev.* **2015**, *35*, 9–15. [[CrossRef](#)]
125. Torres, R.; Szpiech, Z.A.; Hernandez, R.D. Human demographic history has amplified the effects of background selection across the genome. *PLoS Genet.* **2018**, *14*, e1007387. [[CrossRef](#)]
126. Werren, E.A.; Garcia, O.; Biggam, A.W. Identifying adaptive alleles in the human genome: From selection mapping to functional validation. *Hum. Genet.* **2021**, *140*, 241–276. [[CrossRef](#)]
127. Fagorzi, C.; Checcucci, A. A Compendium of Bioinformatic Tools for Bacterial Pangenomics to Be Used by Wet-Lab Scientists. *Methods Mol. Biol.* **2021**, *2242*, 233–243. [[CrossRef](#)]
128. Baaijens, J.A.; Bonizzoni, P.; Boucher, C.; Della Vedova, G.; Pirola, Y.; Rizzi, R.; Sirén, J. Computational graph pangenomics: A tutorial on data structures and their applications. *Nat. Comput.* **2022**, *21*, 81–108. [[CrossRef](#)]
129. Eisenstein, M. Every base everywhere all at once: Pangenomics comes of age. *Nature* **2023**, *616*, 618–620. [[CrossRef](#)]
130. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The complete sequence of a human genome. *Science* **2022**, *376*, 44–53. [[CrossRef](#)] [[PubMed](#)]
131. Aganezov, S.; Yan, S.M.; Soto, D.C.; Kirsche, M.; Zarate, S.; Avdeyev, P.; Taylor, D.J.; Shafin, K.; Shumate, A.; Xiao, C.; et al. A complete reference genome improves analysis of human genetic variation. *Science* **2022**, *376*, eabl3533. [[CrossRef](#)] [[PubMed](#)]
132. Mao, Y.; Zhang, G. A complete, telomere-to-telomere human genome sequence presents new opportunities for evolutionary genomics. *Nat. Methods* **2022**, *19*, 635–638. [[CrossRef](#)] [[PubMed](#)]
133. Kille, B.; Balaji, A.; Sedlazeck, F.J.; Nute, M.; Treangen, T.J. Multiple genome alignment in the telomere-to-telomere assembly era. *Genome Biol.* **2022**, *23*, 182. [[CrossRef](#)]
134. Kim, D.S.; Wiel, L.; Ashley, E.A. Mind the Gap: The Complete Human Genome Unlocks Benefits for Clinical Genomics. *Clin. Chem.* **2023**, *69*, 6–8. [[CrossRef](#)]
135. Joly, Y.; Dupras, C.; Pinkesz, M.; Tovino, S.A.; Rothstein, M.A. Looking Beyond GINA: Policy Approaches to Address Genetic Discrimination. *Annu. Rev. Genomics Hum. Genet.* **2020**, *21*, 491–507. [[CrossRef](#)]
136. Kim, H.; Ho, C.W.L.; Ho, C.-H.; Athira, P.S.; Kato, K.; De Castro, L.; Kang, H.; Huxtable, R.; Zwart, H.; Ives, J.; et al. Genetic discrimination: Introducing the Asian perspective to the debate. *NPJ Genomic Med.* **2021**, *6*, 54. [[CrossRef](#)]
137. Joly, Y.; Huerne, K.; Arych, M.; Bombard, Y.; De Paor, A.; Dove, E.S.; Granados Moreno, P.; Ho, C.W.L.; Ho, C.-H.; Van Hoyweghen, I.; et al. The Genetic Discrimination Observatory: Confronting novel issues in genetic discrimination. *Trends Genet.* **2021**, *37*, 951–954. [[CrossRef](#)]
138. Tiller, J.; Delatycki, M.B. Genetic discrimination in life insurance: A human rights issue. *J. Med. Ethics* **2021**, *47*, 484–485. [[CrossRef](#)]
139. Joly, Y.; Dalpe, G. Genetic discrimination still casts a large shadow in 2022. *Eur. J. Hum. Genet.* **2022**, *30*, 1320–1322. [[CrossRef](#)]
140. Chapman, C.R.; Mehta, K.S.; Parent, B.; Caplan, A.L. Genetic discrimination: Emerging ethical challenges in the context of advancing technology. *J. Law Biosci.* **2020**, *7*, lsz016. [[CrossRef](#)]
141. Clayton, E.W.; Evans, B.J.; Hazel, J.W.; Rothstein, M.A. The law of genetic privacy: Applications, implications, and limitations. *J. Law Biosci.* **2019**, *6*, 1–36. [[CrossRef](#)]

142. Arias, J.J.; Pham-Kanter, G.; Gonzalez, R.; Campbell, E.G. Trust, vulnerable populations, and genetic data sharing. *J. Law Biosci.* **2015**, *2*, 747–753. [[CrossRef](#)]
143. Borry, P.; Bentzen, H.B.; Budin-Ljøsne, I.; Cornel, M.C.; Howard, H.C.; Feeney, O.; Jackson, L.; Mascalconi, D.; Mendes, Á.; Peterlin, B.; et al. The challenges of the expanded availability of genomic information: An agenda-setting paper. *J. Community Genet.* **2018**, *9*, 103–116. [[CrossRef](#)]
144. Scollen, S.; Page, A.; Wilson, J. From the Data on Many, Precision Medicine for “One”: The Case for Widespread Genomic Data Sharing. *Biomed. Hub* **2017**, *2*, 104–110. [[CrossRef](#)]
145. Rehm, H.L.; Page, A.J.H.; Smith, L.; Adams, J.B.; Alterovitz, G.; Babb, L.J.; Barkley, M.P.; Baudis, M.; Beauvais, M.J.S.; Beck, T.; et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. *Cell Genom.* **2021**, *1*, 100029. [[CrossRef](#)]
146. Pollack Porter, K.M.; Rutkow, L.; McGinty, E.E. The Importance of Policy Change for Addressing Public Health Problems. *Public Health Rep.* **2018**, *133*, 9S–14S. [[CrossRef](#)]
147. Molster, C.M.; Bowman, F.L.; Bilkey, G.A.; Cho, A.S.; Burns, B.L.; Nowak, K.J.; Dawkins, H.J.S. The Evolution of Public Health Genomics: Exploring Its Past, Present, and Future. *Front. Public Health* **2018**, *6*, 247. [[CrossRef](#)]
148. Oliver, K.; Lorenc, T.; Tinkler, J.; Bonell, C. Understanding the unintended consequences of public health policies: The views of policymakers and evaluators. *BMC Public Health* **2019**, *19*, 1057. [[CrossRef](#)]
149. Bélisle-Pipon, J.-C.; Vayena, E.; Green, R.C.; Cohen, I.G. Genetic testing, insurance discrimination and medical research: What the United States can learn from peer countries. *Nat. Med.* **2019**, *25*, 1198–1204. [[CrossRef](#)]
150. Sherkow, J.S.; Greely, H.T. The History of Patenting Genetic Material. *Annu. Rev. Genet.* **2015**, *49*, 161–182. [[CrossRef](#)] [[PubMed](#)]
151. Nicol, D.; Dreyfuss, R.C.; Gold, E.R.; Li, W.; Liddicoat, J.; Van Overwalle, G. International Divergence in Gene Patenting. *Annu. Rev. Genom. Hum. Genet.* **2019**, *20*, 519–541. [[CrossRef](#)] [[PubMed](#)]
152. Liddicoat, J.; Liddell, K.; McCarthy, A.H.; Hogarth, S.; Aboy, M.; Nicol, D.; Patton, S.; Hopkins, M.M. Continental drift? Do European clinical genetic testing laboratories have a patent problem? *Eur. J. Hum. Genet.* **2019**, *27*, 997–1007. [[CrossRef](#)] [[PubMed](#)]
153. Du, L.; Lin, S.; Kamenova, K. Framing Ethical Concerns and Attitudes towards Human Gene Patents in the Chinese Press. *Asian Bioeth. Rev.* **2020**, *12*, 307–323. [[CrossRef](#)]
154. Larregue, J. The long hard road to the doability of interdisciplinary research projects: The case of biosocial criminology. *New Genet. Soc.* **2018**, *37*, 21–43. [[CrossRef](#)]
155. Tozzo, P.; D’Angiolla, G.; Brun, P.; Castagliuolo, I.; Gino, S.; Caenazzo, L. Skin Microbiome Analysis for Forensic Human Identification: What Do We Know So Far? *Microorganisms* **2020**, *8*, 873. [[CrossRef](#)]
156. Bonomi, L.; Huang, Y.; Ohno-Machado, L. Privacy challenges and research opportunities for genomic data sharing. *Nat. Genet.* **2020**, *52*, 646–654. [[CrossRef](#)]
157. Dhirani, L.L.; Mukhtiar, N.; Chowdhry, B.S.; Newe, T. Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review. *Sensors* **2023**, *23*, 1151. [[CrossRef](#)]
158. Wan, Z.; Hazel, J.W.; Clayton, E.W.; Vorobeychik, Y.; Kantarcioglu, M.; Malin, B.A. Sociotechnical safeguards for genomic data privacy. *Nat. Rev. Genet.* **2022**, *23*, 429–445. [[CrossRef](#)]
159. Schaper, M.; Schick Tanz, S. Medicine, market and communication: Ethical considerations in regard to persuasive communication in direct-to-consumer genetic testing services. *BMC Med. Ethics* **2018**, *19*, 56. [[CrossRef](#)]
160. Koplin, J.J.; Skeggs, J.; Gyngell, C. Ethics of Buying DNA. *J. Bioethical Inq.* **2022**, *19*, 395–406. [[CrossRef](#)]
161. Martins, M.F.; Murry, L.T.; Telford, L.; Moriarty, F. Direct-to-consumer genetic testing: An updated systematic review of healthcare professionals’ knowledge and views, and ethical and legal concerns. *Eur. J. Hum. Genet.* **2022**, *30*, 1331–1343. [[CrossRef](#)]
162. de Groot, N.F.; van Beers, B.C.; Meynen, G. Commercial DNA tests and police investigations: A broad bioethical perspective. *J. Med. Ethics* **2021**, *47*, 788–795. [[CrossRef](#)]
163. Skeva, S.; Larmuseau, M.H.; Shabani, M. Review of policies of companies and databases regarding access to customers’ genealogy data for law enforcement purposes. *Pers. Med.* **2020**, *17*, 141–153. [[CrossRef](#)]
164. Horton, R.; Lucassen, A. Consent and Autonomy in the Genomics Era. *Curr. Genet. Med. Rep.* **2019**, *7*, 85–91. [[CrossRef](#)]
165. Rego, S.; Grove, M.E.; Cho, M.K.; Ormond, K.E. Informed Consent in the Genomics Era. *Cold Spring Harb. Perspect. Med.* **2020**, *10*, a036582. [[CrossRef](#)]
166. Koplin, J.J.; Gyngell, C.; Savulescu, J.; Vears, D.F. Moving from ‘fully’ to ‘appropriately’ informed consent in genomics: The PROMICE framework. *Bioethics* **2022**, *36*, 655–665. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.