



Review

Genetic Ancestry Inference and Its Application for the Genetic Mapping of Human Diseases

Eva Suarez-Pajes ¹, Ana Díaz-de Usera ², Itahisa Marcelino-Rodríguez ¹, Beatriz Guillen-Guio ^{1,†}
and Carlos Flores ^{1,2,3,*}

¹ Research Unit, Hospital Universitario Nuestra Señora de Candelaria, 38010 Santa Cruz de Tenerife, Spain; eva.sp.95@gmail.com (E.S.-P.); itahisa@gmail.com (I.M.-R.); bguillenguio@gmail.com (B.G.-G.)

² Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), 38600 Santa Cruz de Tenerife, Spain; adiaz@iter.es

³ CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, 28029 Madrid, Spain

* Correspondence: cflores@ull.edu.es

† These authors have contributed equally to this work as senior authors.

Abstract: Admixed populations arise when two or more ancestral populations interbreed. As a result of this admixture, the genome of admixed populations is defined by tracts of variable size inherited from these parental groups and has particular genetic features that provide valuable information about their demographic history. Diverse methods can be used to derive the ancestry apportionment of admixed individuals, and such inferences can be leveraged for the discovery of genetic loci associated with diseases and traits, therefore having important biomedical implications. In this review article, we summarize the most common methods of global and local genetic ancestry estimation and discuss the use of admixture mapping studies in human diseases.

Keywords: admixture mapping; genetic ancestry; ancestry informative markers; next-generation sequencing



Citation: Suarez-Pajes, E.; Díaz-de Usera, A.; Marcelino-Rodríguez, I.; Guillen-Guio, B.; Flores, C. Genetic Ancestry Inference and Its Application for the Genetic Mapping of Human Diseases. *Int. J. Mol. Sci.* **2021**, *22*, 6962. <https://doi.org/10.3390/ijms22136962>

Academic Editors: Akiko Maeda and Maria Adelaide Caligo

Received: 10 June 2021
Accepted: 25 June 2021
Published: 28 June 2021

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



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

1. Genetic Admixture

Admixed populations are the result of gene flow between reproductively isolated groups, owing to events that have occurred throughout human history, including migratory events, the discovery of new territories, or the slave trade. As a result of the intermixture and recombination, over time, the genomes of individuals in the hybrid population will contain a mosaic of ancestries from different population sources in their chromosomes. The length of the chromosome segments inherited from the different ancestral populations will be proportional to the time elapsed since the admixture event. These tracts shorten over the generations by the meiotic recombination process, so that the most recently admixed populations, such as the Canary Islanders in Spain or the Latino populations, would retain longer ancestral tracts, while the populations that mixed more distantly in time, such as the Uyghur in China, would harbor shorter ancestry segments in their chromosomes [1,2].

As such, the admixture proportions and the elapsed time since the admixture event can be inferred based on linkage disequilibrium (LD) [3,4]. When two distant populations interbreed, the admixture linkage disequilibrium (ALD) can be generated among loci with different allelic frequencies in the ancestral populations, leading to a linkage between markers that were previously unlinked. During the first generations since the admixture, ALD is expected to experience a rapid decay between distant loci, while it would be maintained between closer positions and can be detected after generations [5]. Additionally, the ALD dynamics of decay are also influenced by the admixture model. For example, a greater drop of ALD and a faster length decrease in the ancestral chromosomal segments are expected for those populations that have been formed by a single mixing event, compared with admixtures maintained throughout generations [1,6].

Several studies support that genetic ancestry and admixture can partially explain the differences in the prevalence of complex diseases and treatment responses between population groups, due to the unequal distribution of allelic frequencies of the underlying causal variants across populations. As such, and given that the prevalence of many traits and diseases differs between populations, the analysis of genetic ancestry differences between affected and nonaffected subjects, as well as the link of these differences with the pathogenesis, and with evolutionary, environmental, or behavioral factors, plays an important role in biomedical research. Some examples are the high prevalence of multiple sclerosis among Europeans [7], of diabetes in North American and Caribbean regions [8], and of hypertension in African and Asian populations [9].

Many studies also support the existence of genetic variants affecting drug metabolism, transport, and toxicity, which vary widely among populations and ethnic groups [10,11]. In this sense, the response to the anticoagulant warfarin has been studied in different admixed populations in order to explain how genetic ancestries could contribute to the interindividual response to the treatment [12]. In addition, similar studies have developed algorithms to predict warfarin doses and improve the treatment in Hispanic-Caribbeans [13,14]. Furthermore, genetic variants of Cytochrome P450 Family 2 Subfamily C Member 19 (*CYP2C19*) and Paraoxonase 1 (*PON1*) genes, which are highly structured by ancestry, modify the response of the antiplatelet agent clopidogrel in Puerto Ricans [15]. Additionally, among many others, the relationship between genetic ancestry and the response to bronchodilators in patients with asthma [16] or to treatments for acute lymphoblastic leukemia has been investigated [17].

Given the importance of genetic ancestry in medicine, here, we provide an updated review of the most common methods for global and local genetic ancestry estimation, and of their use in admixture mapping approaches, highlighting a few key findings from the recent literature. We also discuss the potential of Next Generation Sequencing (NGS) data for ancestry estimation.

2. Estimation of Genetic Ancestry: Global and Local Ancestry

Global ancestry (GA) is the fraction of genomic ancestry from each admixed individual that can be ascribed to each of the ancestral populations contributing to the recently admixed population (Figure 1A). The estimate of GA can be obtained using different approaches. Some of the most popular methods are based on probabilistic models using genotype data, assuming that populations are in Hardy–Weinberg equilibrium and considering complete linkage equilibrium for all loci considered for the estimation, such as STRUCTURE [18,19] and ADMIXTURE [20,21]. Alternative approaches that allow the estimation of the ancestry proportions are based on principal component decompositions, such as ipPCA [22], and on the study of LD decay curves, such as ALDER [3].

Local ancestry (LA) is a term commonly used to refer to the ancestry in each of the chromosome blocks, also known as ancestral tracks, in recently admixed individuals (Figure 1B). For this, the number of copies derived of each ancestral population, in each genomic position, could be inferred per individual (from zero to two copies). Thus, GA can also be obtained by summarizing LA across the individual genomes. Multiple estimators have been developed to infer LA (Table 1). Briefly, the choice of the most appropriate approach will depend on the number/density of available markers as well as on the evolutionary history of the admixed population under study. Some models are based on haplotype data and require specific reference panels that may not be available for all populations [23–26]. Furthermore, LA inference is complicated in scenarios of admixture between populations with a limited genetic divergence, or of an old admixture event.

Some of the genetic characteristics of recently admixed populations described above (chromosome blocks and ALD distribution) allow the estimation of ancestry with a relatively small number of genetic markers. These markers are usually single nucleotide polymorphisms (SNPs) with a different distribution between populations, known as ancestry informative markers (AIMs). Their number will also depend on the assessed population,

the ancestral groups, and the time since the admixture event [27]. In order to identify those markers that are useful and informative of ancestry, multiple measurements of population differentiation have been proposed [28,29]. Additionally, the widespread use of SNP arrays and NGS technologies with information from hundreds of thousands genetic markers has facilitated LA inferences and the development of AIMs panels for particular admixed populations [30–32].

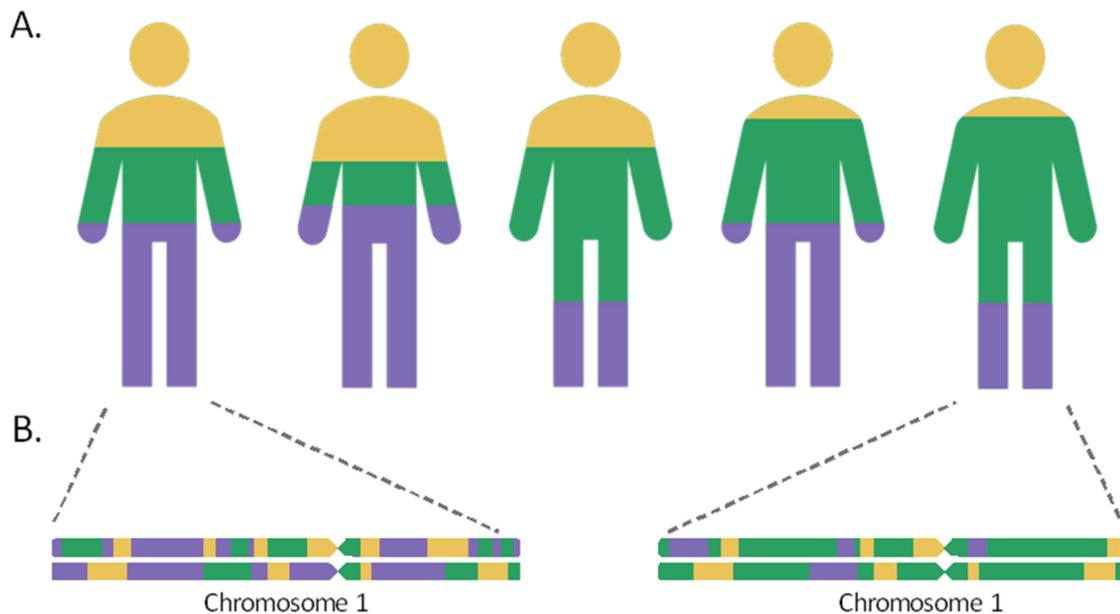


Figure 1. Global (A) and local (B) genetic ancestries in a recently admixed population with three ancestral populations. The proportion of each of the ancestral populations is represented by the colors yellow, blue, and purple.

Table 1. Most common methods to estimate local genetic ancestry.

SOFTWARE	Algorithm	Background LD	Phasing Requirement	Genetic Map	Physical Map	Number of Ancestral Populations	Reference
CHROMOPAINTER	HMM	Yes	Phased	Optional	No	≥ 2	[33]
EILA	k-means	No	Unphased	No	Yes	2 or 3	[34]
ELAI	Two layers HMM	Yes	Phased/Unphased ^a	No	No	≥ 2	[35]
HAPMIX	HMM	Yes	Phased /Unphased ^b	Yes	No	2	[36]
LAMP-LD	HMM	Yes	Phased/Unphased ^b	No	Yes	2, 3 or 5	[37]
Loter	Single layer HMM	No	Phased	No	No	≥ 2	[23]
PCAdmix	HMM and local PCA	No	Phased	Optional	Optional	≥ 2	[25]
RFMIX	CRF	No	Phased	Yes	No	≥ 2	[24]
SABER +	HMM	Yes	Phased	No	No	2–4	[38,39]
SEQMIX	HMM	No	Unphased	Yes	No	2	[40]
SupportMix	SVM	No	Phased	Yes	No	≥ 2	[26]

^a Phased and unphased data are allowed for ancestral and admixed populations. ^b Phased data are needed for the ancestral populations and unphased data for the admixed population. CRF (Conditional Random Field), HMM (Hidden Markov Model), LD (linkage disequilibrium), PCA (Principal Component Analysis), SVM (Support Vector Machines).

The use of genotyping microarrays has also led to the development of improved methods to infer LA, such as LAMP-LD [37], RFMix [24], and HAPMIX [36], among others (Table 1). Compared to the previous methods that were designed to deal with AIMs, these other algorithms rely on denser sets of genetic markers (retaining LD) that allow one to obtain a higher resolution in estimating LA, most of them based on hidden Markov models [35,37]. Additionally, the development of machine learning methods has allowed the use of algorithms such as random forests [24], which have been suggested to provide more accurate estimates [41].

In order to identify the optimal approach for each scenario, benchmarking the different algorithms and reference panels is necessary. Previous reviews have compared the characteristics and effectiveness of local ancestry estimators [42–46], suggesting a few main aspects to consider: (1) the prior requirements of each estimator, and (2) the inherent features of the target population itself. Table 1 shows the main characteristics of the most common methods to estimate local ancestry.

Regarding the necessary requirements for the use of each estimator, it must be considered, for example, whether a phasing step is needed prior to ancestry estimation. This step is crucial for an accurate estimate of ancestry and is closely linked to the density of available markers [47]. Therefore, the use of an algorithm not requiring a prior phasing step (e.g., EILA or SEQMIX) may be less biased. On the other hand, certain tools need specific marker information, such as their intermarker distance (physical maps) and/or the recombination rate (genetic maps). Additionally, other options such as HAPMIX [36], ELAI [35], RFMIX [24], and SUPPORTMIX [26] require imposing a number of generations since the occurrence of the last admixing event. Therefore, uncertainty or the lack of accurate population information can lead to biased estimates.

Moreover, the inherent characteristics of the population under study can also influence software selection. For example, although RFMIX [24] or LAMP-LD [37] have reported accurate local ancestry estimates, their effectiveness is reduced when increasing the number of generations since the admixing event. In the case of ancient admixtures, Loter [23] and ELAI would allow one to obtain better estimates [23,43]. On the other hand, the number of ancestral populations of the target population is also an important factor to be considered in the selection of the estimator (Table 1), as well as the magnitude of the differentiation among the parental populations. For example, ELAI has been shown to better differentiate North African and sub-Saharan ancestry components than has LAMP-LD [2].

Finally, another important aspect that must be considered is the computational requirements of each software, such as the execution time or memory used, which is strongly conditioned by the sample size of the admixed sample under study [43,44]. In this sense, Loter [23] or RFMIX [24] offer the possibility of reducing the computation time via parallel implementation, splitting the process into multiple threads.

3. Admixture Mapping Studies

3.1. Definition

The distribution of allelic frequencies in recently admixed populations is closely related to those frequencies found in their ancestral populations [48,49]. When these ancestral populations have marked differences in the susceptibility to a disease, admixture mapping studies, also known as mapping by admixture linkage disequilibrium (MALD) studies, can be performed to reveal genetic loci harboring variants underlying such differences between population groups [50]. Admixture mapping studies aim to correlate LA with a trait of interest in recently admixed populations in which ALD is still detectable, under the hypothesis that variants associated with increased disease risk will be found in chromosomal fragments inherited from one of the parental populations [51,52]. Thus, an increment (or decrease) in the proportion of the ancestry associated with the trait of interest will be expected in these chromosomal regions (Figure 2, Table 2).

Admixture mapping studies can be performed using case-only or case-control approaches. The case-only design is based on the comparison between observed and average expected ancestry to detect loci enriched in one of the ancestries, while the case-control approach consists of comparing LA pointwise between cases and controls [53,54]. Although case-only studies have shown a greater higher statistical power, case-control analyses are less biased when it comes to verifying that the changes in ancestry are due to the association with the trait of interest and not to other confounding factors [53,54]. Different admixture mapping tools have been developed to perform LA association analyses, both for case-only and for case-control approaches. Classically, the most widely used tools to perform an admixture mapping are ADMIXMAP [55], ANCESTRYMAP [56], or MALDsoft [54].

Additionally, LA estimations could be used as a proxy of the genotype variable within logistic regression models for association study testing on the admixed individuals [57,58]. In the event that the study population presents kinship relationships, it is necessary to use specific methods to correct the population structure, such as PC-Air [59] or EMMAX [60].

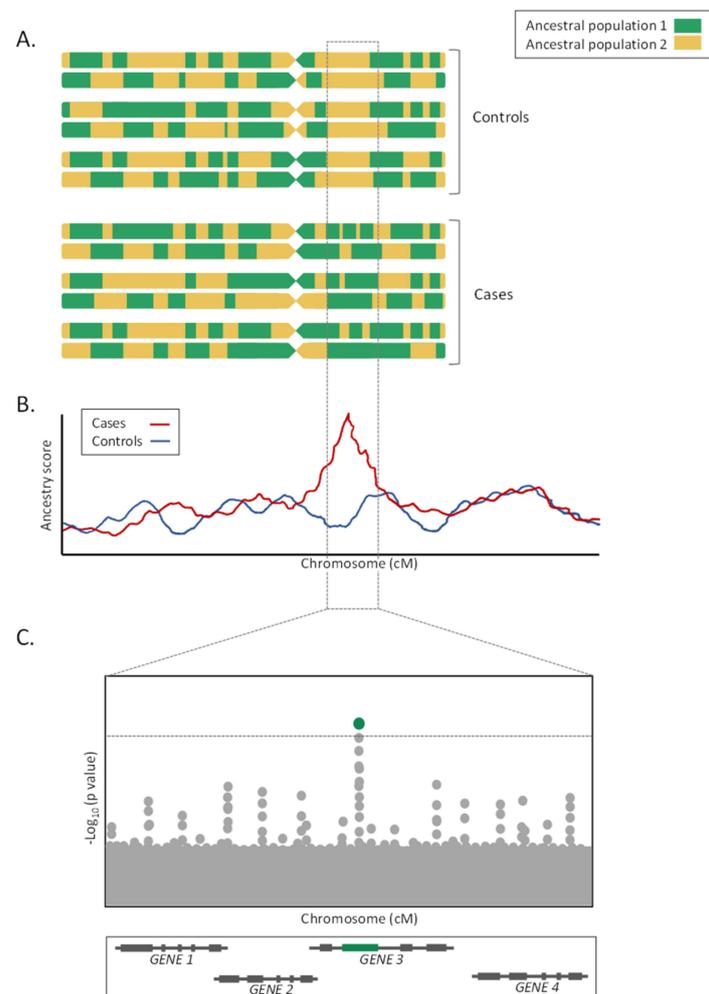


Figure 2. Scheme of an admixture mapping study. (A) LA estimates in cases and controls individuals from a recently admixed population. (B) Comparison of local ancestry scores of all chromosomal regions between cases and controls. (C) Fine mapping study on genomic regions where genetic ancestry is associated with a trait.

Table 2. Definition of the main concepts.

Concept	Definition
Ancestry informative marker (AIM)	Genetic variants, usually SNPs, that show large frequency differences between the parental populations and that are, thus, highly informative for ancestry estimation in admixed populations.
Admixture model	A simple model to describe how gene flow between ancestral populations could have occurred. Admixed populations can be the result of a mixture between individuals from two or more populations and that can be maintained in various generations (gradual admixture) or be a result of a single event (hybrid isolation).
Ancestry estimation	In admixed populations, this allows the determination of the proportion of each of the ancestries for a given admixture model.
Global ancestry (GA)	Estimated ancestry proportion with which each parental population contributes on average to the genome of an admixed individual for a given admixture model.
Local ancestry (LA)	Estimated ancestry proportion with which each parental population contributes to each locus of the genome of an admixed individual for a given admixture model.
Admixture mapping	Method that allows detecting if the genetic ancestry of a particular section of the genome in a mixed population tends to be inherited with a particular trait.

3.2. Advantages and Disadvantages of Admixture Mapping Studies

Since 2005, genome-wide association studies (GWAS) have allowed the identification of thousands of genetic loci associated with many traits, including complex diseases [61]. However, as of June 2020, nearly 90% of GWAS reported in the GWAS Catalog were obtained from Europeans [62]. Thus, a significant part of the genetic variability exclusive to or better represented in other ethnic groups remains largely unexplored, which has obvious consequences for the generalized implementation of precision medicine. Assuming that the effect of the risk variants can vary depending on the surrounding genetic variation, some results could be affected by the hidden stratification caused by the LA or if the direction of the effect is opposite in the distinct ancestral populations [58,63]. In this sense, admixture mapping studies overcome these limitations, offering a more efficient alternative to contribute to the disentangling of the genetic architecture of diseases and traits in recently admixed populations. As a major advantage, given that LA tracks are usually large, often measured in the megabase-scale, the significance penalty of these studies is much lower than for GWAS, therefore increasing the statistical power for a given sample size [48]. Furthermore, these studies are less affected by allelic heterogeneity than GWAS, because they are based on LA and not on SNP alleles directly.

On the contrary, admixture mapping studies lose efficiency when allele frequencies are similarly distributed among ancestral populations and when the LD in the parental populations is unknown [64,65]. Furthermore, for the LA estimations to be accurate, the choice of the LA estimator, as well as the number and density of genotyped markers, is critical [65]. Additionally, given the megabase size of the loci detected by the admixture mapping approaches, fine mapping studies or candidate gene association studies focused on the prioritized genomic regions must follow for the study to be fully completed (Figure 2C). Finally, as in the GWAS, admixture mapping approaches only allow the detection of the genetic risks associated with the trait of interest, and do not consider the environmental, cultural, or socioeconomic factors [64].

3.3. Applications of Admixture Mapping Studies in Biomedical Research

Based on admixed populations, several studies have implemented admixture mapping approaches to reveal novel risk factors associated with complex diseases, including cancer, hypertension, and autoimmune, respiratory, and infectious diseases. Primarily, these studies have been widely applied in African Americans and Hispanic/Latino populations. The genetic contribution of each of the parental populations has been estimated for African Americans from four different states, providing an estimate of average ancestry corresponding to 76.4% African, 20.9% European, and 2.7% Native American [66]. A similar approach was carried out in the Hispanic population of Southern Colorado, for whom the composition was reported to correspond to 62.7% European, 34.1% Native American, and 3.2% African [67]. However, these estimates differ widely between the states [68]. In this sense, recent studies in African American populations have found genomic regions with increased proportions of African ancestry associated with risk of prostate and breast cancer [69–71]. Additionally, Schwartz and colleagues conducted admixture mapping studies using both case-only and case-control approaches in 1,812 African Americans, and revealed two loci linked to lung cancer susceptibility, one of them at 1q42 with an excess of European ancestry, and another at 3q25 enriched in African ancestry [72]. Furthermore, Yang and colleagues reported genomic loci enriched in Native American ancestry that could influence relapse and a poorer prognosis for lymphoblastic leukemia in a heterogeneous cohort of self-reported ancestries [17].

Zhu and colleagues performed an admixture mapping in African Americans, identifying an excess of African ancestry at 6p24 and 21p21 associated with hypertension [73]. Gignoux and colleagues detected a genomic region in 18q12 in the Latino population, where an increase in Native American ancestry was associated with a greater asthma risk, while European ancestry was associated with asthma protection [74]. Furthermore, an increased risk of multiple sclerosis was linked to European genetic risk factors in African

American and Hispanic individuals [75]. Admixture mapping studies have also been used to explain why some populations are more susceptible to suffer infections caused by microbes. For instance, admixture mapping studies allowed explanations for the susceptibility to tuberculosis caused by *Mycobacterium tuberculosis* in the South African Colored population [76] and the incidence of *Staphylococcus aureus* in African Americans [77], allowing the detection of genomic regions that could be involved in the pathogenesis. Furthermore, these approaches have also been applied to investigate the treatment response to identify the genetic differences that modify the response to drugs. Recent studies suggest that genetic variants related to drug metabolism, transport, and toxicity may vary between ethnic groups [10,11]. Therefore, admixture mapping approaches are necessary to study the implication of the genetic ancestry in the response to a treatment. In this sense, Spear and colleagues found an African region at 8p11 with a suggestive association to bronchodilator response in African Americans patients with asthma [78].

Additionally, although less frequently, admixture mapping studies have also been performed in admixed populations other than Latinos or African Americans. For instance, Sun and colleagues recently performed an admixture mapping approach in the Hawaiian population in order to determine the relation of the genetic admixture with different cardiovascular traits. Interestingly, this admixed population (78% Native Hawaiian, 11.5% European, and 7.8% Asian [79]) has a high burden of cardiovascular diseases and diabetes [80,81]. Despite that, it has been a poorly studied minority. The study revealed an excess of native Hawaiian ancestry on chromosome 6, and further single variant association tests focused on this region identified a variant in the 5'UTR region of the eyes shut homolog (*EYS*) gene significantly associated with type 2 diabetes [82]. Moreover, this strategy has been applied to recently admixed western European populations, specifically in the Spanish population of the Canary Islands in the Macaronesia, where the use of SNP arrays and whole-genome sequencing revealed that average proportions of ancestry correspond to 75% European, 22% North African, and 3% South Saharan African [2]. There is only one admixture mapping study in the European population published to date. This was performed by Guillen-Guio and colleagues, who performed an admixture mapping study of asthma in the Canary Island population [83]. This study allowed the detection of a novel locus within chromosome 16 (16q23.3) enriched in North African ancestry that was associated with asthma risk. Subsequent whole-exome sequencing analyses revealed that the phospholipase C gamma 2 (*PLCG2*) gene, located in that region, was enriched of deleterious variants among asthma cases [83]. Likewise, other populations that could be excellent targets for admixture mapping studies are the Cape Verde population, also from the Macaronesia, since they are a result of recent gene flow between Africans and Europeans [84], and the Uyghur population, which can be modelled by a two-way admixture between Europeans and East Asians [31]. Genetic ancestry studies could assist in unraveling the genetics underlying prevalent diseases in these populations, helping to raise the representation of diversity in the genetic architecture of diseases, and thus result in a better transference of knowledge to personalized medicine.

4. NGS and Genetic Ancestry Estimation

The technological development of NGS over the years, together with the reduction in the sequencing costs, offers a great opportunity for genetic ancestry studies to develop further. Among the major advantages of this technology compared to microarrays (Table 3) is its high-throughput capacity, resulting in thousands of DNA fragments being sequenced simultaneously, offering the possibility of covering a larger fraction of the genome. Therefore, the use of NGS, especially of whole-genome sequencing (WGS), allows an increase in the number of markers tested to infer LA, and the possibility to find optimal ancestry-specific genetic markers. This permits one to obtain reference panels of ancestral populations and then design panels of more restricted AIMs, as done by Li-Ju Wang and colleagues, who proposed a specific panel of AIMs to infer three-way genetic admixture (European, East Asian, and African) by using whole-exome sequencing (WES) data [85]. Furthermore, the

NGS technology allows the detection of information from the entire spectrum of allelic frequencies, from common variants to low-frequency and rare variants, which, by definition, are expected to be more structured among populations compared to the common variation that is typically covered by most SNP genotyping microarrays. This leads to better detection of population-specific variants and, therefore, improved LA estimation [86]. Additionally, another advantage offered by the NGS is that detected SNPs are not affected by ascertainment bias, which is induced by an incorrect or nonrepresentative selection of markers. In this sense, Maróti and colleagues assessed the used of WGS, WES, and SNP genotyping microarray data in population genetic analyses [87]. Their results suggested that SNP genotyping data may be more prone to biasing the results, as they are related to significantly higher cross-validation error values and an overestimation of the admixture proportions than are WES or WGS data. Accordingly, Lachance and Tishkoff suggested that the use of biased markers from genotyping arrays may misestimate LD and overestimate population differences [88]. Since these aspects are important for LA inference and, consequently, for the proper performance of an admixture mapping study, we anticipate that the use of NGS will lead to more accurate estimates.

Table 3. Advantages and disadvantages of using NGS for LA estimation.

Advantages	Disadvantages
<ul style="list-style-type: none"> • Larger fraction of the genome covered. • Detection of low-frequency and population-specific variants. • More accurate LA estimate. • SNPs not affected by ascertainment bias. 	<ul style="list-style-type: none"> • Lack of specific algorithms and software. • Accuracy depends on sequencing coverage. • WES covers a small portion of the genome. • Higher computational and economic costs.

LA (Local ancestry), WES (Whole-Exome Sequence), SNP (Single Nucleotide Polymorphism).

However, although the application of NGS implies important advances, its use to estimate LA still has important limitations (Table 3). First, the large amount of data generated by NGS approaches implies a large computational requirement and high economic costs. Second, most existing algorithms have been designed for SNP genotyping microarray data (e.g., LAMP-LD limits the use to datasets of markers up to 500,000 SNPs [37]). Additionally, the analysis of WGS data using methods not optimized for NGS can lead to decreased accuracy of the LA estimation [86]. To cope with this limitation, specific NGS software is being developed for ancestry estimation, mainly aimed at WES analysis [40,86]. Third, the LA estimation is conditioned by the depth of coverage of the sequencing data, as a low depth of coverage increases the likelihood of introducing false genotype calls that could lead to biases in the estimation of the ancestry proportions. Different tools have recently been developed to estimate LA addressing this problem. For example, SEQMIX, which has been optimized to use low depth of coverage data from WES or targeted sequencing, allows one to obtain accurate estimations by combining data from off-target and on-target reads [40]. Lanc-CSV uses continent-specific variants (CSV) (i.e., variants that are private to one of the continental populations) and preprocessing steps to guarantee a minimal depth of coverage in CSV for them to be considered further downstream for LA estimation [86]. Finally, considering that WES covers only 1–2% of the genome, the inference of ancestry blocks using WES data is reduced to this portion of the genome. In this sense, new methods to infer the variation of uncovered regions have been developed and could be applied to improve LA inference [89].

Although the blooming use of NGS represents a great opportunity for population studies improvement, only two NGS-based admixture mapping studies have been published to date [90,91]. Liu and colleagues performed an admixture mapping study of blood pressure phenotypes using WES data from African American individuals to understand their higher prevalence of hypertension, revealing four regions enriched in African ancestry linked to diastolic blood pressure, two of them also overlapping with regions that were

significantly associated with mean arterial pressure [90]. Additionally, Lin and colleagues accomplished a multi-ethnic study using WGS data to study glomerular phenotypes. Although no significant associations were detected in the admixture mapping study, using WGS data allowed them to identify three rare variants associated with estimated glomerular filtration rate [91]. Therefore, although the ancestry inference using sequencing data continues to be a challenge, the results of these two studies support the huge potential for the use of NGS in admixture mapping approaches.

5. Concluding Remarks

Genetic ancestry studies and admixture mapping approaches have expanded genetic knowledge in biomedical research, revealing new loci associated with traits and diseases that could not have been detected by conventional association studies. Despite this, the available genomic resources need to be improved to obtain more accurate ancestry inferences. For instance, WGS data for African or Middle Eastern populations are still limited. Interestingly, the genetic features of Arab populations could be a valuable opportunity, especially in the study of recessive traits given the high rate of consanguinity that could be overlooked in other population groups [62]. In this context, additional deep characterizations of the genetic variation on these populations are needed to further improve estimations and reduce biases or false associations caused by the lack of proper reference datasets. Therefore, initiatives such as The Human Heredity and Health in Africa (H3Africa) consortium have been developed in order to promote genetic and environmental base studies of human diseases in Africans and their clinical application [92,93]. On the other hand, the National Arab Genome Project in the United Arab Emirates (UAE) aims to achieve a greater representation of Arab genomes through NGS technologies [94].

Consequently, a more equitable representation of the ancestral groups in association studies will improve the development of personalized medicine. The unequal proportion in European-based GWAS is known to generate a bias in the findings that could lead to misinterpretations and hamper the risk prediction, prognosis, or response to drugs in under-represented populations [95]. In this regard, Manrai and colleagues detected that patients of African ancestry with hypertrophic cardiomyopathy received reports with misclassified risk variants [96]. This highlights the need to expand the catalog of genetic variation across diverse populations, as well as to promote studies in non-European populations in order to improve the prediction and treatment of diseases for individuals, irrespective of their ancestry.

There are other approaches currently used in biomedical research that would benefit from taking genetic ancestry into account. Gene expression studies, which seek to explain how the expression of certain genes influences the development or severity of a trait, are an example. Based on the evidence that gene expression levels also differ between populations, in recent years, there has been a need to incorporate ancestry in expression studies and generate ancestry-dependent transcriptomic profiles [97–99]. Additionally, polygenic risk score (PRS) modeling, which uses a composite of the individual risk variant effects, estimated from association studies, to predict the overall genetic risk of developing a particular trait, have relied primarily on European GWAS data. Therefore, the transferability of the PRS scores to other population groups is limited and, as a consequence, a worse performance has been described in African populations [100]. Therefore, there is an urgent need to include other admixed populations and to consider genetic ancestry in these models, to optimize the PRS, and improve the efficiency of this approach [101,102]. Finally, recent studies show that rare or low-frequency variants are more likely to have a larger effect on complex traits [103,104]. Most studies that have focused on rare variants have been designed for genetically homogeneous populations and do not consider the effect of local genetic ancestry. Recently, Qin and colleagues have developed an approach to identify chromosomal blocks that harbor rare variants using the local ancestry, which has shown more powerful results than other methods for the case of admixed populations [105]. In

this sense, increasing the knowledge in this field will allow for a better understanding of how these uncommon variants influence diseases and traits.

In summary, promoting genetic studies in admixed populations, and the use of admixture mapping studies, combined with the alternative approaches described, promise the identification of novel disease associations and a better understanding of complex trait genetics. Eventually, these results will translate into a more equitable representation of the catalogs of genetic variation across populations.

Author Contributions: E.S.-P. and B.G.-G. wrote the first draft of the manuscript; E.S.-P., A.D.-d.U., I.M.-R., B.G.-G., and C.F. revised the draft and prepared the figures and tables; C.F. obtained funding; C.F. and B.G.-G. designed the structure of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Instituto de Salud Carlos III (PI14/00844, PI17/00610, CD19/00231, PI20/00876), co-funded by ERD Funds, “A way of making Europe” from the EU; ITER agreement OA17/008. A.D.-d.U. was supported by a fellowship from the Spanish Ministry of Education and Vocational Education (Grant No. FPU16/01435).

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. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Jin, W.; Li, R.; Zhou, Y.; Xu, S. Distribution of ancestral chromosomal segments in admixed genomes and its implications for inferring population history and admixture mapping. *Eur. J. Hum. Genet.* **2014**, *22*, 930–937. [[CrossRef](#)] [[PubMed](#)]
- Guillen-Guio, B.; Lorenzo-Salazar, J.M.; Gonzalez-Montelongo, R.; Díaz-de Usera, A.; Marcelino-Rodríguez, I.; Corrales, A.; de Leon, A.C.; Alonso, S.; Flores, C. Genomic analyses of human European diversity at the southwestern edge: Isolation, African influence and disease associations in the Canary Islands. *Mol. Biol. Evol.* **2018**, *35*, 3010–3026. [[CrossRef](#)] [[PubMed](#)]
- Loh, P.R.; Lipson, M.; Patterson, N.; Moorjani, P.; Pickrell, J.K.; Reich, D.; Berger, B. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **2013**, *193*, 1233–1254. [[CrossRef](#)] [[PubMed](#)]
- Zhou, Y.; Qiu, H.; Xu, S. Modeling Continuous Admixture Using Admixture-Induced Linkage Disequilibrium. *Sci. Rep.* **2017**, *7*, 43054. [[CrossRef](#)] [[PubMed](#)]
- Chakraborty, R.; Weiss, K.M. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 9119–9123. [[CrossRef](#)]
- Pfaff, C.L.; Parra, E.J.; Bonilla, C.; Hiester, K.; McKeigue, P.M.; Kamboh, M.I.; Hutchinson, R.G.; Ferrell, R.E.; Boerwinkle, E.; Shriver, M.D. Population structure in admixed populations: Effect of admixture dynamics on the pattern of linkage disequilibrium. *Am. J. Hum. Genet.* **2001**, *68*, 198–207. [[CrossRef](#)]
- Compston, A.; Coles, A. Multiple sclerosis. *Lancet* **2008**, *372*, 1502–1517. [[CrossRef](#)]
- Cho, N.H.; Shaw, J.E.; Karuranga, S.; Huang, Y.; da Rocha Fernandes, J.D.; Ohlrogge, A.W.; Malanda, B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res. Clin. Pract.* **2018**, *138*, 271–281. [[CrossRef](#)] [[PubMed](#)]
- Mills, K.T.; Stefanescu, A.; He, J. The global epidemiology of hypertension. *Nat. Rev. Nephrol.* **2020**, *16*, 223–237. [[CrossRef](#)]
- Ortega, V.E.; Meyers, D.A. Pharmacogenetics: Implications of race and ethnicity on defining genetic profiles for personalized medicine. *J. Allergy Clin. Immunol.* **2014**, *133*, 16–26. [[CrossRef](#)]
- Suarez-Kurtz, G. Pharmacogenomics in admixed populations. *Trends Pharmacol. Sci.* **2005**, *26*, 196–201. [[CrossRef](#)]
- Villagra, D.; Duconge, J.; Windemuth, A.; Cadilla, C.L.; Kocherla, M.; Gorowski, K.; Bogaard, K.; Renta, J.Y.; Cruz, I.A.; Mirabal, S.; et al. CYP2C9 and VKORC1 genotypes in Puerto Ricans: A case for admixture-matching in clinical pharmacogenetic studies. *Clin. Chim. Acta* **2010**, *411*, 1306–1311. [[CrossRef](#)]
- Duconge, J.; Ramos, A.S.; Claudio-Campos, K.; Rivera-Miranda, G.; Bermúdez-Bosch, L.; Renta, J.Y.; Cadilla, C.L.; Cruz, I.; Feliu, J.F.; Vergara, C.; et al. A novel admixture-based pharmacogenetic approach to refine warfarin dosing in caribbean hispanics. *PLoS ONE* **2016**, *11*, e0145480. [[CrossRef](#)]
- Roche-Lima, A.; Roman-Santiago, A.; Feliu-Maldonado, R.; Rodríguez-Maldonado, J.; Nieves-Rodríguez, B.G.; Carrasquillo-Carrion, K.; Ramos, C.M.; Da Luz Sant’Ana, I.; Massey, S.E.; Duconge, J. Machine learning algorithm for predicting warfarin dose in caribbean hispanics using pharmacogenetic data. *Front. Pharmacol.* **2020**, *10*, 1–8. [[CrossRef](#)]

15. Duconge, J.; Escalera, O.; Korchela, M.; Ruao, G. Clinical Implications of Genetic Admixture in Hispanic Puerto Ricans: Impact on the Pharmacogenetics of CYP2C19 and PON1. In *Clinical Applications of Pharmacogenetics*; IntechOpen: Rijeka, Croatia, 2012; Volume 19, pp. 151–163.
16. Corvol, H.; De Giacomo, A.; Eng, C.; Seibold, M.; Ziv, E.; Chapela, R.; Rodriguez-Santana, J.R.; Rodriguez-Cintron, W.; Thyne, S.; Watson, H.G.; et al. Genetic ancestry modifies pharmacogenetic gene-gene interaction for asthma. *Pharmacogenet. Genom.* **2009**, *19*, 489–496. [[CrossRef](#)]
17. Yang, J.J.; Cheng, C.; Devidas, M.; Cao, X.; Fan, Y.; Campana, D.; Yang, W.; Neale, G.; Cox, N.J.; Scheet, P.; et al. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nat. Genet.* **2011**, *43*, 237–241. [[CrossRef](#)] [[PubMed](#)]
18. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes* **2007**, *7*, 574–578. [[CrossRef](#)]
19. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [[CrossRef](#)] [[PubMed](#)]
20. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)] [[PubMed](#)]
21. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **2011**, *12*, 246. [[CrossRef](#)] [[PubMed](#)]
22. Limpiti, T.; Intarapanich, A.; Assawamakin, A.; Shaw, P.J.; Wangkumhang, P.; Piriyaopongsa, J.; Ngamphiw, C.; Tongsimma, S. Study of large and highly stratified population datasets by combining iterative pruning principal component analysis and structure. *BMC Bioinform.* **2011**, *12*, 255. [[CrossRef](#)] [[PubMed](#)]
23. Dias-Alves, T.; Mairal, J.; Blum, M.G.B. Loter: A software package to infer local ancestry for a wide range of species. *Mol. Biol. Evol.* **2018**, *35*, 2318–2326. [[CrossRef](#)]
24. Maples, B.K.; Gravel, S.; Kenny, E.E.; Bustamante, C.D. RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **2013**, *93*, 278–288. [[CrossRef](#)]
25. Brisbin, A.; Bryc, K.; Byrnes, J.; Zakharia, F.; Omberg, L.; Degenhardt, J.; Reynolds, A.; Ostrer, H.; Mezey, J.G.; Bustamante, C.D. Pcadmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* **2012**, *84*, 343–364. [[CrossRef](#)]
26. Omberg, L.; Salit, J.; Hackett, N.; Fuller, J.; Matthew, R.; Chouchane, L.; Rodriguez-Flores, J.L.; Bustamante, C.; Crystal, R.G.; Mezey, J.G. Inferring genome-wide patterns of admixture in Qataris using fifty-five ancestral populations. *BMC Genet.* **2012**, *13*, 49. [[CrossRef](#)]
27. Winkler, C.A.; Nelson, G.W.; Smith, M.W. Admixture Mapping Comes of Age. *Annu. Rev. Genom. Hum. Genet.* **2010**, *11*, 65–89. [[CrossRef](#)]
28. Rosenberg, N.A.; Li, L.M.; Ward, R.; Pritchard, J.K. Informativeness of Genetic Markers for Inference of Ancestry. *Am. J. Hum. Genet.* **2003**, *73*, 1402–1422. [[CrossRef](#)] [[PubMed](#)]
29. Ding, L.; Wiener, H.; Abebe, T.; Altaye, M.; Go, R.C.P.; Kercksmar, C.; Grabowski, G.; Martin, L.J.; Khurana Hershey, G.K.; Chakorborty, R.; et al. Comparison of measures of marker informativeness for ancestry and admixture mapping. *BMC Genom.* **2011**, *12*, 622. [[CrossRef](#)] [[PubMed](#)]
30. Chen, G.; Shriner, D.; Zhou, J.; Doumatey, A.; Huang, H.; Gerry, N.P.; Herbert, A.; Christman, M.F.; Chen, Y.; Dunston, G.M.; et al. Development of admixture mapping panels for African Americans from commercial high-density SNP arrays. *BMC Genom.* **2010**, *11*, 417. [[CrossRef](#)] [[PubMed](#)]
31. Xu, S.; Jin, L. A Genome-wide Analysis of Admixture in Uyghurs and a High-Density Admixture Map for Disease-Gene Discovery. *Am. J. Hum. Genet.* **2008**, *83*, 322–336. [[CrossRef](#)] [[PubMed](#)]
32. Mao, X.; Bigham, A.W.; Mei, R.; Gutierrez, G.; Weiss, K.M.; Brutsaert, T.D.; Leon-Velarde, F.; Moore, L.G.; Vargas, E.; McKeigue, P.M.; et al. A genomewide admixture mapping panel for hispanic/latino populations. *Am. J. Hum. Genet.* **2007**, *80*, 1171–1178. [[CrossRef](#)] [[PubMed](#)]
33. Lawson, D.J.; Hellenthal, G.; Myers, S.; Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **2012**, *8*, e1002453. [[CrossRef](#)]
34. Yang, J.J.; Li, J.; Buu, A.; Williams, L.K. Efficient inference of local ancestry. *Bioinformatics* **2013**, *29*, 2750–2756. [[CrossRef](#)]
35. Guan, Y. Detecting structure of haplotypes and local ancestry. *Genetics* **2014**, *196*, 625–642. [[CrossRef](#)]
36. Price, A.L.; Tandon, A.; Patterson, N.; Barnes, K.C.; Rafaels, N.; Ruczinski, I.; Beaty, T.H.; Mathias, R.; Reich, D.; Myers, S. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **2009**, *5*, e1000519. [[CrossRef](#)]
37. Baran, Y.; Pasaniuc, B.; Sankararaman, S.; Torgerson, D.G.; Gignoux, C.; Eng, C.; Rodriguez-Cintron, W.; Chapela, R.; Ford, J.G.; Avila, P.C.; et al. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* **2012**, *28*, 1359–1367. [[CrossRef](#)]
38. Tang, H.; Coram, M.; Wang, P.; Zhu, X.; Risch, N. Reconstructing genetic ancestry blocks in admixed individuals. *Am. J. Hum. Genet.* **2006**, *79*, 1–12. [[CrossRef](#)] [[PubMed](#)]
39. Johnson, N.A.; Coram, M.A.; Shriver, M.D.; Romieu, I.; Barsh, G.S.; London, S.J.; Tang, H. Ancestral Components of Admixed Genomes in a Mexican Cohort. *PLoS Genet.* **2011**, *7*, e1002410. [[CrossRef](#)]
40. Hu, Y.; Willer, C.; Zhan, X.; Kang, H.M.; Abecasis, G.R. Accurate Local-Ancestry Inference in Exome-Sequenced Admixed Individuals via Off-Target Sequence Reads. *Am. J. Hum. Genet.* **2013**, *93*, 891–899. [[CrossRef](#)]

41. Uren, C.; Hoal, E.G.; Möller, M. Putting RFMix and ADMIXTURE to the test in a complex admixed population. *BMC Genet.* **2020**, *21*, 40. [[CrossRef](#)]
42. Geza, E.; Mugo, J.; Mulder, N.J.; Wonkam, A.; Chimusa, E.R.; Mazandu, G.K. A comprehensive survey of models for dissecting local ancestry deconvolution in human genome. *Brief. Bioinform.* **2019**, *20*, 1709–1724. [[CrossRef](#)]
43. Schubert, R.; Andaleon, A.; Wheeler, H.E. Comparing local ancestry inference models in populations of two- And three-way admixture. *PeerJ* **2020**, *8*, 1–19. [[CrossRef](#)] [[PubMed](#)]
44. Hui, D.; Fang, Z.; Lin, J.; Duan, Q.; Li, Y.; Hu, M.; Chen, W. LAIT: A local ancestry inference toolkit. *BMC Genet.* **2017**, *18*, 83. [[CrossRef](#)] [[PubMed](#)]
45. Yuan, K.; Zhou, Y.; Ni, X.; Wang, Y.; Liu, C.; Xu, S. Models, methods and tools for ancestry inference and admixture analysis. *Quant. Biol.* **2017**, *5*, 236–250. [[CrossRef](#)]
46. Thornton, T.A.; Bermejo, J.L. Local and global ancestry inference and applications to genetic association analysis for admixed Populations. *Genet. Epidemiol.* **2014**, *38*, S5–S12. [[CrossRef](#)]
47. Browning, S.R.; Browning, B.L. Haplotype phasing: Existing methods and new developments. *Nat. Rev. Genet.* **2011**, *12*, 703–714. [[CrossRef](#)]
48. Shriner, D.; Adeyemo, A.; Ramos, E.; Chen, G.; Rotimi, C.N. Mapping of disease-associated variants in admixed populations. *Genome Biol.* **2011**, *12*, 223. [[CrossRef](#)]
49. Auton, A.; Abecasis, G.R.; Altshuler, D.M.; Durbin, R.M.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; Eichler, E.E.; Flück, P.; et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74. [[CrossRef](#)]
50. Claiborne Stephens, J.; Briscoe, D.; O'Brien, S.J. Mapping by admixture linkage disequilibrium in human populations: Limits and guidelines. *Am. J. Hum. Genet.* **1994**, *55*, 809–824. [[PubMed](#)]
51. McKeigue, P.M. Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am. J. Hum. Genet.* **1997**, *60*, 188–196. [[PubMed](#)]
52. McKeigue, P.M. Mapping genes that underlie ethnic differences in disease risk: Methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am. J. Hum. Genet.* **1998**, *63*, 241–251. [[CrossRef](#)]
53. Hoggart, C.J.; Shriver, M.D.; Kittles, R.A.; Clayton, D.G.; McKeigue, P.M. Design and Analysis of Admixture Mapping Studies. *Am. J. Hum. Genet.* **2004**, *74*, 965–978. [[CrossRef](#)]
54. Montana, G.; Pritchard, J.K. Statistical tests for admixture mapping with case-control and cases-only data. *Am. J. Hum. Genet.* **2004**, *75*, 771–789. [[CrossRef](#)]
55. Hoggart, C.J.; Parra, E.J.; Shriver, M.D.; Bonilla, C.; Kittles, R.A.; Clayton, D.G.; McKeigue, P.M. Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet.* **2003**, *72*, 1492–1504. [[CrossRef](#)]
56. Patterson, N.; Hattangadi, N.; Lane, B.; Lohmueller, K.E.; Hafler, D.A.; Oksenberg, J.R.; Hauser, S.L.; Smith, M.W.; O'Brien, S.J.; Altshuler, D.; et al. Methods for High-Density Admixture Mapping of Disease Genes. *Am. J. Hum. Genet.* **2004**, *74*, 979–1000. [[CrossRef](#)]
57. Atkinson, E.G.; Maihofer, A.X.; Kanai, M.; Martin, A.R.; Karczewski, K.J.; Santoro, M.L.; Ulirsch, J.C.; Kamatani, Y.; Okada, Y.; Finucane, H.K.; et al. Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **2021**. [[CrossRef](#)]
58. Wang, X.; Zhu, X.; Qin, H.; Cooper, R.S.; Ewens, W.J.; Li, C.; Li, M. Adjustment for local ancestry in genetic association analysis of admixed populations. *Bioinformatics* **2011**, *27*, 670–677. [[CrossRef](#)]
59. Conomos, M.P.; Miller, M.; Thornton, T. Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness. *Genet. Epidemiol.* **2015**, *39*, 276–293. [[CrossRef](#)]
60. Kang, H.M.; Sul, J.H.; Service, S.K.; Zaitlen, N.A.; Kong, S.; Freimer, N.B.; Sabatti, C.; Eskin, E. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **2010**, *42*, 348–354. [[CrossRef](#)]
61. Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **2019**, *47*, D1005–D1012. [[CrossRef](#)]
62. Abou Tayoun, A.N.; Rehm, H.L. Genetic variation in the Middle East—An opportunity to advance the human genetics field. *Genome Med.* **2020**, *12*, 12–15. [[CrossRef](#)]
63. Derks, E.M.; Zwinderman, A.H.; Gamazon, E.R. The Relation Between Inflation in Type-I and Type-II Error Rate and Population Divergence in Genome-Wide Association Analysis of Multi-Ethnic Populations. *Behav. Genet.* **2017**, *47*, 360–368. [[CrossRef](#)]
64. Shriner, D. Overview of Admixture Mapping. *Curr. Protoc. Hum. Genet.* **2017**, *94*, 1.23.1–1.23.8. [[CrossRef](#)]
65. Smith, M.W.; O'Brien, S.J. Mapping by admixture linkage disequilibrium: Advances, limitations and guidelines. *Nat. Rev. Genet.* **2005**, *6*, 623–632. [[CrossRef](#)]
66. Reiner, A.P.; Ziv, E.; Lind, D.L.; Nievergelt, C.M.; Schork, N.J.; Cummings, S.R.; Phong, A.; Burchard, E.G.; Harris, T.B.; Psaty, B.M.; et al. Population structure, admixture, and aging-related phenotypes in African American adults: The cardiovascular health study. *Am. J. Hum. Genet.* **2005**, *76*, 463–477. [[CrossRef](#)]
67. Bonilla, C.; Parra, E.J.; Pfaff, C.L.; Dios, S.; Marshall, J.A.; Hamman, R.F.; Ferrell, R.E.; Hoggart, C.L.; McKeigue, P.M.; Shriver, M.D. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann. Hum. Genet.* **2004**, *68*, 139–153. [[CrossRef](#)]

68. Bryc, K.; Durand, E.Y.; Macpherson, J.M.; Reich, D.; Mountain, J.L. The genetic ancestry of african americans, latinos, and european Americans across the United States. *Am. J. Hum. Genet.* **2015**, *96*, 37–53. [[CrossRef](#)]
69. Freedman, M.L.; Haiman, C.A.; Patterson, N.; McDonald, G.J.; Tandon, A.; Waliszewska, A.; Penney, K.; Steen, R.G.; Ardlie, K.; John, E.M.; et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14068–14073. [[CrossRef](#)]
70. Bock, C.H.; Schwartz, A.G.; Ruterbusch, J.J.; Levin, A.M.; Neslund-Dudas, C.; Land, S.J.; Wenzlaff, A.S.; Reich, D.; McKeigue, P.; Chen, W.; et al. Results from a prostate cancer admixture mapping study in African-American men. *Hum. Genet.* **2009**, *126*, 637–642. [[CrossRef](#)]
71. Ruiz-Narváez, E.A.; Sucheston-Campbell, L.; Bensen, J.T.; Yao, S.; Haddad, S.; Haiman, C.A.; Bandera, E.V.; John, E.M.; Bernstein, L.; Hu, J.J.; et al. Admixture mapping of African-American women in the AMBER Consortium identifies new loci for breast cancer and estrogen-receptor subtypes. *Front. Genet.* **2016**, *7*, 1–10. [[CrossRef](#)]
72. Schwartz, A.G.; Wenzlaff, A.S.; Bock, C.H.; Ruterbusch, J.J.; Chen, W.; Cote, M.L.; Artis, A.S.; van Dyke, A.L.; Land, S.J.; Harris, C.C.; et al. Admixture mapping of lung cancer in 1812 African-Americans. *Carcinogenesis* **2011**, *32*, 312–317. [[CrossRef](#)]
73. Zhu, X.; Luke, A.; Cooper, R.S.; Quertermous, T.; Hanis, C.; Mosley, T.; Gu, C.C.; Tang, H.; Rao, D.C.; Risch, N.; et al. Admixture mapping for hypertension loci with genome-scan markers. *Nat. Genet.* **2005**, *37*, 177–181. [[CrossRef](#)]
74. Gignoux, C.R.; Torgerson, D.G.; Pino-Yanes, M.; Uricchio, L.H.; Galanter, J.; Roth, L.A.; Eng, C.; Hu, D.; Nguyen, E.A.; Huntsman, S.; et al. An admixture mapping meta-analysis implicates genetic variation at 18q21 with asthma susceptibility in Latinos. *J. Allergy Clin. Immunol.* **2019**, *143*, 957–969. [[CrossRef](#)]
75. Chi, C.; Shao, X.; Rhead, B.; Gonzales, E.; Smith, J.B.; Xiang, A.H.; Graves, J.; Waldman, A.; Lotze, T.; Schreiner, T.; et al. Admixture mapping reveals evidence of differential multiple sclerosis risk by genetic ancestry. *PLoS Genet.* **2019**, *15*, e1007808. [[CrossRef](#)]
76. Daya, M.; van der Merwe, L.; Gignoux, C.R.; van Helden, P.D.; Möller, M.; Hoal, E.G. Using multi-way admixture mapping to elucidate TB susceptibility in the South African Coloured population. *BMC Genom.* **2014**, *15*, 1021. [[CrossRef](#)]
77. Cyr, D.D.; Allen, A.S.; Du, G.; Ruf, F.; Adams, C.; Thaden, J.T.; Maskarinec, S.A.; Souli, M.; Guo, S.; Dykxhoorn, D.M.; et al. Evaluating genetic susceptibility to Staphylococcus aureus bacteremia in African Americans using admixture mapping. *Genes Immun.* **2017**, *18*, 95–99. [[CrossRef](#)]
78. Spear, M.L.; Hu, D.; Pino-Yanes, M.; Huntsman, S.; Eng, C.; Levin, A.M.; Ortega, V.E.; White, M.J.; McGarry, M.E.; Thakur, N.; et al. A genome-wide association and admixture mapping study of bronchodilator drug response in African Americans with asthma. *Pharm. J.* **2019**, *19*, 249–259. [[CrossRef](#)]
79. Kim, S.K.; Gignoux, C.R.; Wall, J.D.; Lum-Jones, A.; Wang, H.; Haiman, C.A.; Chen, G.K.; Henderson, B.E.; Kolonel, L.N.; Le Marchand, L.; et al. Population Genetic Structure and Origins of Native Hawaiians in the Multiethnic Cohort Study. *PLoS ONE* **2012**, *7*, e47881. [[CrossRef](#)]
80. Mau, M.K.; Sinclair, K.; Saito, E.P.; Baumhofer, K.N.; Kaholokula, J.K.A. Cardiometabolic health disparities in native hawaiians and other pacific islanders. *Epidemiol. Rev.* **2009**, *31*, 113–129. [[CrossRef](#)]
81. Maskarinec, G.; Erber, E.; Grandinetti, A.; Verheus, M.; Oum, R.; Hopping, B.N.; Schmidt, M.M.; Uchida, A.; Juarez, D.T.; Hodges, K.; et al. Diabetes incidence based on linkages with health plans: The multiethnic cohort. *Diabetes* **2009**, *58*, 1732–1738. [[CrossRef](#)]
82. Sun, H.; Lin, M.; Russell, E.M.; Minster, R.L.; Chan, T.F.; Dinh, B.L.; Naseri, T.; Reupena, M.S.; Lum-Jones, A.; Cheng, I.; et al. The impact of global and local Polynesian genetic ancestry on complex traits in Native Hawaiians. *PLoS Genet.* **2021**, *17*, e1009273. [[CrossRef](#)] [[PubMed](#)]
83. Guillen-Guio, B.; Hernández-Beeftink, T.; Marcelino-Rodríguez, I.; Rodríguez-Pérez, H.; Lorenzo-Salazar, J.M.; Espinilla-Peña, M.; Corrales, A.; Pino-Yanes, M.; Callero, A.; Perez-Rodríguez, E.; et al. Admixture mapping of asthma in southwestern Europeans with North African ancestry influences. *Am. J. Physiol. Cell. Mol. Physiol.* **2020**, *318*, 965–975. [[CrossRef](#)] [[PubMed](#)]
84. Beleza, S.; Campos, J.; Lopes, J.; Araújo, I.I.; Hoppfer Almada, A.; Correia e Silva, A.; Parra, E.J.; Rocha, J. The Admixture Structure and Genetic Variation of the Archipelago of Cape Verde and Its Implications for Admixture Mapping Studies. *PLoS ONE* **2012**, *7*, e51103. [[CrossRef](#)]
85. Wang, L.J.; Zhang, C.W.; Su, S.C.; Chen, H.I.H.; Chiu, Y.C.; Lai, Z.; Bouamar, H.; Ramirez, A.G.; Cigarroa, F.G.; Sun, L.Z.; et al. An ancestry informative marker panel design for individual ancestry estimation of Hispanic population using whole exome sequencing data. *BMC Genom.* **2019**, *20*, 1007. [[CrossRef](#)]
86. Brown, R.; Pasaniuc, B. Enhanced Methods for Local Ancestry Assignment in Sequenced Admixed Individuals. *PLoS Comput. Biol.* **2014**, *10*, e1003555. [[CrossRef](#)]
87. Maróti, Z.; Boldogkői, Z.; Tombácz, D.; Snyder, M.; Kalmár, T. Evaluation of whole exome sequencing as an alternative to BeadChip and whole genome sequencing in human population genetic analysis. *BMC Genom.* **2018**, *19*, 778. [[CrossRef](#)]
88. Lachance, J.; Tishkoff, S.A. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays* **2013**, *35*, 780–786. [[CrossRef](#)]
89. Díaz-de Usera, A.; Lorenzo-Salazar, J.M.; Rubio-Rodríguez, L.A.; Muñoz-Barrera, A.; Guillen-Guio, B.; Marcelino-Rodríguez, I.; García-Olivares, V.; Mendoza-Alvarez, A.; Corrales, A.; Íñigo-Campos, A.; et al. Evaluation of Whole-Exome Enrichment Solutions: Lessons from the High-End of the Short-Read Sequencing Scale. *J. Clin. Med.* **2020**, *9*, 3656. [[CrossRef](#)]

90. Liu, Z.; Shriner, D.; Hansen, N.F.; Rotimi, C.N.; Mullikin, J.C.; Barnabas, B.B.; Black, S.; Bouffard, G.G.; Brooks, S.Y.; Coleman, H.; et al. Admixture mapping identifies genetic regions associated with blood pressure phenotypes in African Americans. *PLoS ONE* **2020**, *15*, e0232048. [[CrossRef](#)]
91. Lin, B.M.; Grinde, K.E.; Brody, J.A.; Breeze, C.E.; Raffield, L.M.; Mychaleckyj, J.C.; Thornton, T.A.; Perry, J.A.; Baier, L.J.; de las Fuentes, L.; et al. Whole genome sequence analyses of eGFR in 23,732 people representing multiple ancestries in the NHLBI trans-omics for precision medicine (TOPMed) consortium. *EBioMedicine* **2021**, *63*, 103157. [[CrossRef](#)]
92. Mulder, N.; Abimiku, A.; Adebamowo, S.N.; de Vries, J.; Matimba, A.; Olowoyo, P.; Ramsay, M.; Skelton, M.; Stein, D.J. H3Africa: Current perspectives. *Pharmacogenomics Pers. Med.* **2018**, *11*, 59–66. [[CrossRef](#)]
93. De Vries, J.; Tindana, P.; Littler, K.; Ramsay, M.; Rotimi, C.; Abayomi, A.; Mulder, N.; Mayosi, B.M.; Europe PMC Funders Group. The H3Africa policy framework: Negotiating fairness in genomics. *Trends Genet.* **2015**, *31*, 117–119. [[CrossRef](#)] [[PubMed](#)]
94. Al-Ali, M.; Osman, W.; Tay, G.K.; Alsafar, H.S. A 1000 Arab genome project to study the Emirati population. *J. Hum. Genet.* **2018**, *63*, 533–536. [[CrossRef](#)]
95. Sirugo, G.; Williams, S.M.; Tishkoff, S.A. The Missing Diversity in Human Genetic Studies. *Cell* **2019**, *177*, 26–31. [[CrossRef](#)]
96. Manrai, A.K.; Funke, B.H.; Rehm, H.L.; Olesen, M.S.; Maron, B.A.; Szolovits, P.; Margulies, D.M.; Loscalzo, J.; Kohane, I.S. Genetic Misdiagnoses and the Potential for Health Disparities. *N. Engl. J. Med.* **2016**, *375*, 655–665. [[CrossRef](#)]
97. Li, X.; Battle, A.; Karczewski, K.J.; Zappala, Z.; Knowles, D.A.; Smith, K.S.; Kukurba, K.R.; Wu, E.; Simon, N.; Montgomery, S.B. Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. *Am. J. Hum. Genet.* **2014**, *95*, 245–256. [[CrossRef](#)] [[PubMed](#)]
98. Bhattacharya, A.; García-Closas, M.; Olshan, A.F.; Perou, C.M.; Troester, M.A.; Love, M.I. A framework for transcriptome-wide association studies in breast cancer in diverse study populations. *Genome Biol.* **2020**, *21*, 42. [[CrossRef](#)]
99. Roelands, J.; Mall, R.; Almeer, H.; Thomas, R.; Mohamed, M.G.; Bedri, S.; Al-Bader, S.B.; Junejo, K.; Ziv, E.; Sayaman, R.W.; et al. Ancestry-associated transcriptomic profiles of breast cancer in patients of African, Arab, and European ancestry. *npj Breast Cancer* **2021**, *7*, 10. [[CrossRef](#)]
100. Duncan, L.; Shen, H.; Gelaye, B.; Meijssen, J.; Ressler, K.; Feldman, M.; Peterson, R.; Domingue, B. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **2019**, *10*, 3328. [[CrossRef](#)]
101. Cavazos, T.B.; Witte, J.S. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *Hum. Genet. Genomics Adv.* **2021**, *2*, 100017. [[CrossRef](#)]
102. Bitarello, B.D.; Mathieson, I. Polygenic Scores for Height in Admixed Populations. *G3 Genes Genomes Genet.* **2020**, *10*, 4027–4036. [[CrossRef](#)] [[PubMed](#)]
103. Momozawa, Y.; Mizukami, K. Unique roles of rare variants in the genetics of complex diseases in humans. *J. Hum. Genet.* **2021**, *66*, 11–23. [[CrossRef](#)] [[PubMed](#)]
104. Kosmicki, J.A.; Churchhouse, C.L.; Rivas, M.A.; Neale, B.M. Discovery of rare variants for complex phenotypes. *Hum. Genet.* **2016**, *135*, 625–634. [[CrossRef](#)] [[PubMed](#)]
105. Qin, H.; Zhao, J.; Zhu, X. Identifying Rare Variant Associations in Admixed Populations. *Sci. Rep.* **2019**, *9*, 5458. [[CrossRef](#)]