genes

MDPI

Review

The Y Chromosome: A Complex Locus for Genetic Analyses of Complex Human Traits

Katherine Parker [1], A. Mesut Erzurumluoglu [2] and Santiago Rodriguez [1,3,*]

[1]   Bristol Medical School, University of Bristol, Bristol BS8 1QU, UK; kp16445@bristol.ac.uk
[2]   Medical Research Council (MRC) Epidemiology Unit, University of Cambridge, Cambridge CB2 0SL, UK;
      mesut.erzurumluoglu@mrc-epid.cam.ac.uk
[3]   MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, Bristol Medical School,
      University of Bristol, Bristol BS8 1QU, UK
*   Correspondence: santi.rodriguez@bristol.ac.uk; Tel.: +44-(0)117-3310133

check for updates

Abstract: The Human Y chromosome (ChrY) has been demonstrated to be a powerful tool for phylogenetics, population genetics, genetic genealogy and forensics. However, the importance of ChrY genetic variation in relation to human complex traits is less clear. In this review, we summarise existing evidence about the inherent complexities of ChrY variation and their use in association studies of human complex traits. We present and discuss the specific particularities of ChrY genetic variation, including Y chromosomal haplogroups, that need to be considered in the design and interpretation of genetic epidemiological studies involving ChrY.

Keywords: Y chromosome; Y haplogroups; complex locus; genetic association analyses; genetic epidemiology; complex human traits

1. Introduction

During the last decade, genome-wide association studies (GWASs) have identified large numbers of loci associated with common complex human traits, enabling a better understanding of the genetic architecture of human disease [1]. This progress is documented by the NHGRI-EBI GWAS Catalogue [2], which includes all published GWAS hits to date. The latest version (accessed 11/10/20) of the widely used diagram, "SNP-trait associations, mapped onto the human genome by chromosomal location and displayed on the human karyotype", provided by the catalogue (Figure 1), shows evidence of many SNP–trait associations on all 22 autosomes and the X chromosome. Notably, however, none of the GWAS hits found to date reside on the Y chromosome (ChrY).

Results from studies of various designs not limited to GWASs (e.g., candidate gene association analyses) also provide evidence that minimises the role of ChrY common genetic variation in relation to human complex traits. It is important to consider alternative explanations for this lack of association, such as the inherent genetic complexity of ChrY, which differs from the remainder of human chromosomes in terms of structure, function and population history. These differences make it more complex to analyse than other human chromosomes, which consequently leads to the removal or misanalysis of ChrY variants from association studies.

This review addresses specific features of human ChrY relevant for the design and interpretation of genetic association studies of complex human traits and covers different study designs applied in genetic epidemiological studies of ChrY, discussing their strengths and weaknesses, as well as potential avenues for future research.
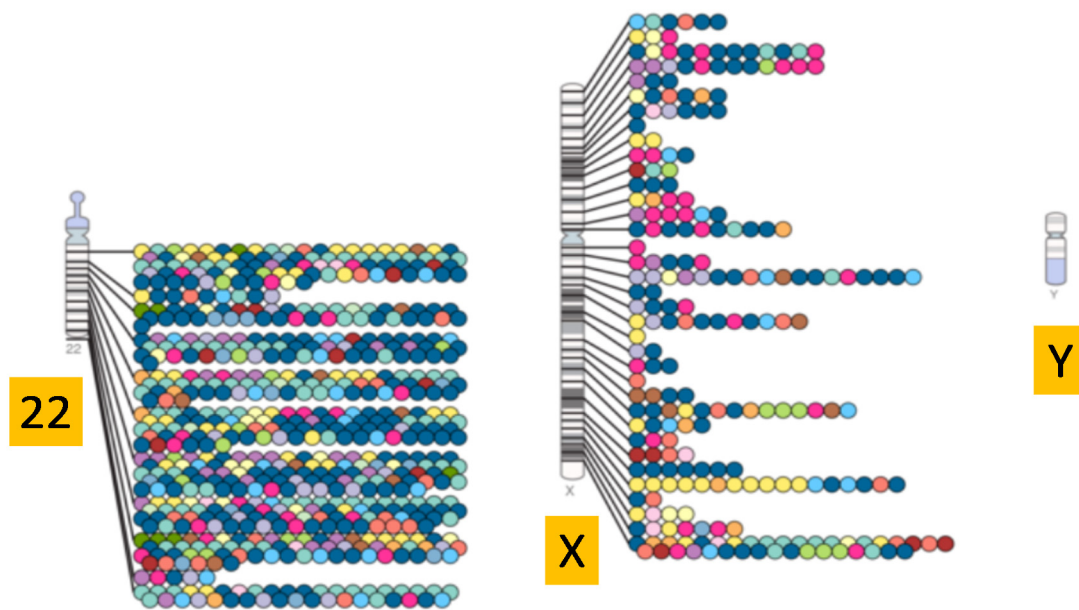
Figure 1. GWASs can be used to look for genetic associations with human disease traits across the human genome. This figure is a modified version of the GWAS diagram from the September 2019 version of the GWAS Catalog, a directory which contains more than 157,000 associations from 4220 publications [2]. The diagram shows hits identified on chromosomes 22 and X, with different categories of disease indicated by different coloured markers. There are no hits which correspond to ChrY. Legend has not been included for clarity, but the original diagram and legend can be found at https://www.ebi.ac.uk/gwas/diagram.

2. The Human Y Chromosome, a Complex Locus for Complex Trait Analysis

Since the discovery of ChrY in 1921 [3], specific particularities related to its structure, function and population history have been discovered. A comprehensive list can be found in Table 1.

Table 1. Chronological order of the discovery of the particularities of the Y chromosome from the discovery of the male-determining region to exploration of gene function.

Year	Description	References
1985	ChrY utilised for evolutionary studies 'genetic distance' genealogy	[4]
1986	First evidence that the male-determining region is located on the short arm of the Y chromosome	[5]
1989	ChrY polymorphisms utilised in phylogenetics	[6]
1990	First gene, sex-determining region Y (SRY), mapped onto ChrY	[7]
1990	Animal models used to investigate the influence of ChrY on hypertension	[8]
1991	Sry used in animal models to form transgenic mice	[9]
1997	ChrY utilised for forensic science and paternity testing	[10]
2000	ChrY utilised for phylogenetics	[11]
2000 onwards	Association between hind III restriction fragment polymorphism and cardiovascular disease detected	[12–15]
2002	Y-chromosomal haplogroups established	[16]
2003	MSY first sequenced	[17]
2005 onwards	Unconvincing evidence relating the MSY to cardiovascular disease risk is reported	[18–21]
2009 onwards	Y haplotypes utilised to investigate association between ChrY and other complex traits	[22–24]
2016	Exploration of the function of genes identified on within MSY	[25]

In summary, ChrY is the sex-determining chromosome in humans and is passed strictly from fathers to sons. Current understanding suggests that the function of ChrY is limited with regards to complex disease causal genes. Here we provide a review of the literature related to ChrY and complex diseases and also make suggestions for future studies.

2.1. Structure of Human Chromosome Y and Recombination

In mammals, gender is dictated by either the presence or absence of ChrY. This chromosome is structurally unusual in relation to autosomes and, in humans, bears an extremely high quantity of repetitive content. ChrY also contains heterochromatic, X-transposed, X-degenerate, ampliconic and pseudoautosomal regions (PAR) [17], formed over time, as the previously autologous pair underwent a sequence of large scale inversions and deletion events [26].

ChrY shows similarity with its counterpart, the X chromosome, at the pseudoautosomal regions present at the telomeres (Figure 2). As a result, crossovers involving ChrY can only occur at these regions. The remainder of ChrY (~95%) is referred to as the male-specific region of ChrY (MSY) [17], previously known as the non-recombining region of Y (NRY) [11]. Variation of the MSY over generations is contributed to by *de novo* mutations and both X-to-Y and Y-to-Y gene conversion [17,27–29]. One study suggests an average of 600 nucleotides per ChrY in a newborn male have undergone Y-to-Y gene conversion in recent human evolution [27]. Whilst significant, this amount of variation is relatively small when compared with autosomes and, when accompanied by uniparental inheritance and lack of crossing over, means that Y chromosome lineages are relatively easier to trace back through time than autosomal lineages.
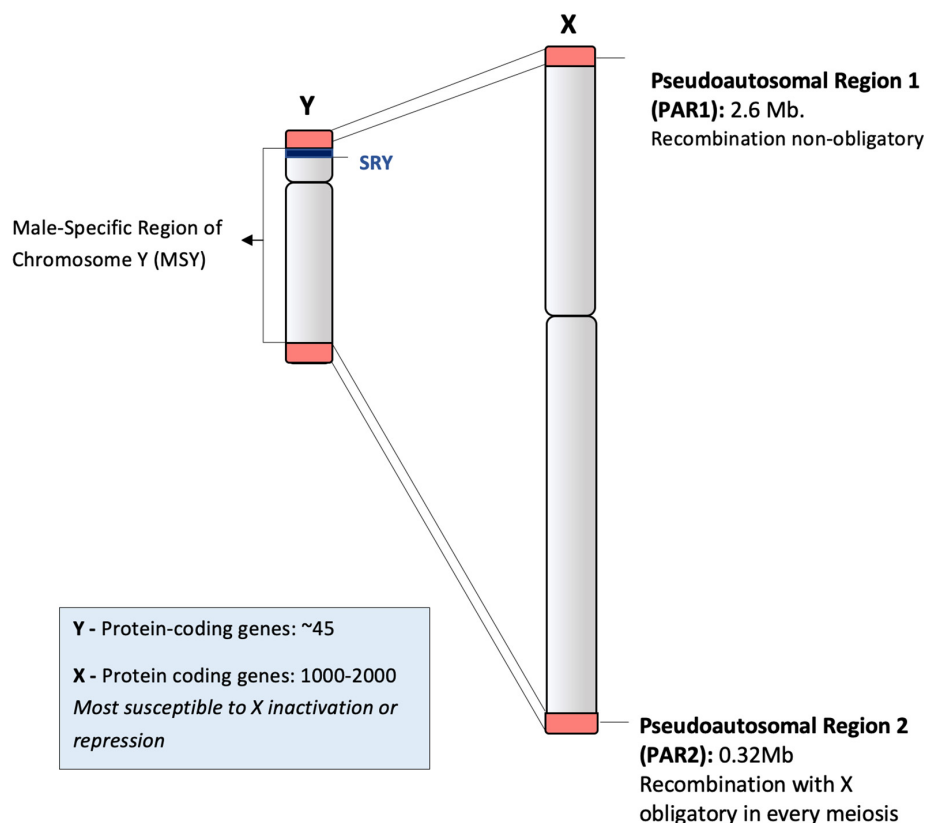
Figure 2. Schematic representation of the Y and X chromosomes. Pseudoautosomal regions 1 and 2, the MSY and the location of SRY are indicated.

2.2. The Functional Role of the Human Y Chromosome

It has been frequently suggested that the functional significance of ChrY is minimal. ChrY would therefore represent a 'functional wasteland' and it would be destined to disappear from the nuclear genome [30]. Although absence of evidence does not imply evidence of absence for ChrY [31], this has contributed to the historical consideration of ChrY as a human chromosome with an insignificant functional role.

2.2.1. Deterioration of the Y Chromosome

Current evidence suggests that deterioration of ChrY was initiated by the acquisition of a male sex-determining function early in mammalian evolution and subsequently perpetuated by repression of recombination. This lack of recombination is also partly responsible for the loss of much of the ancestral gene content, by way of a phenomenon known as Muller's ratchet [32]. As a consequence, ChrY has acquired genes which relate to male-specific functions.

There is controversy about the consequences of ChrY deterioration. Some authors have argued that it will inevitably lead to continued degradation and eventual disappearance of the human ChrY [33]. This argument sites evidence from parallel systems of sex determination (ZW) and comparisons with other mammals [30,34,35]. In contrast, other authors site gene conversion within palindromes [17] and a relative slowing of rate of gene decay in ChrY over recent millennia as evidence supporting continuing evolution in current human populations [26,36].

Disentangling which of these opposed hypotheses better explains the recent evolution of ChrY is an active area of research [37,38]. Unanswered questions to date include (a) to what extent is direct comparison of equivalent but different genetic sex-determining systems, namely XY and ZW, appropriate? (b) Does the comparison of ChrY from long-diverged species overlook the complexities of its evolution and bring about misleading conclusions? (c) Would understanding the extent of genetic deterioration of ChrY be relevant and informative for genetic association analyses?

The degradation of ChrY may represent a biological explanation for the reduced number of functional loci on ChrY and hence a reduced number of associations found by association mapping approaches. If this degradation continues, then one would expect further loss of functional elements over time. Alternatively, it could be possible that degradation of ChrY has reached an equilibrium implying that current functional elements are important for males and will be maintained in the population over time.

2.2.2. Genetic Content of the Y Chromosome

The entirety of the Chr Y was first sequenced in 2003 by Skaletsky et al. who identified 78 protein-coding genes, encoding 27 distinct proteins or protein families [17]. However, current transcript- and protein-level evidence suggests the presence of 64 protein-coding genes including genes on the PAR (Ensembl BioMart; HGNC; CCDS, all accessed 11/10/20) [39–41]. When the MSY is considered in isolation (i.e., the genes on PAR1 and PAR2 excluded), the region has 45 protein-coding genes (Ensembl BioMart; neXtProt v2020-07-17; both accessed 11/10/20) excluding PRYP3 and TSPY9P which are now thought to be pseudogenes (Table S1) (Human Protein Atlas, accessed 11/10/20) [42]. This number is considerably smaller than the number of genes harboured within the X chromosome (n = 852) and all other autosomes (range n = 234–2059), making ChrY the chromosome with the least number of protein-coding genes in humans (Ensembl BioMart, accessed 11/10/20).

In contrast to its lack of protein-coding genes, ChrY presents an unusually high level of structural polymorphisms. ChrY is enriched with various types of repeat polymorphisms including SINEs, endogenous retroviruses and segmental duplications [32].

2.2.3. Gene Function of the Y Chromosome

The first genes identified on the MSY were found to be instrumental in initiating and maintaining the male phenotype [5,7,9] and conferring male fertility [43,44]. These findings likely contributed to the belief that the principal function of the MSY was to facilitate development and maintenance of the male phenotype, and further, that it held little consequence for health in general. Although this has been the consensus for some time, confidence in this hypothesis is dwindling. The functions of genes encoded by ChrY have been summarised by Bellott et al. [45]. These functions range from histone lysine demethylation, regulation of stem-cell self-renewal and translation, and protein modification (such as the deubiquitinase USP9Y) [45,46]. Although our understanding of MSY and the role of these genes is far from complete, it seems plausible that variation within these loci could influence fitness.

These findings have promoted a shift in thinking that has given way to an explosion of research into ChrY and numerous studies have now been published that search for associations between the MSY and male traits.

2.3. Specific Particularities of the MSY in Relation to Human Population History

The MSY is a powerful tool with which to investigate human population history [47–52]. In order to design appropriate genetic association studies of human complex traits, one must possess an appreciation of the specific particularities of the MSY. This fundamental requirement may also lead to the discovery of novel applications of the MSY, promoting a deepening of our knowledge of human population history and structure in relation to genetic architecture of human traits.

Firstly, the haploid nature of ChrY means that it has a smaller effective population size when compared with either autosomes or the X chromosome, which have four times and three times greater effective population sizes, respectively. As a result, ChrY is much more vulnerable to the effects of drift (already a key dictator of the evolution of ChrY due to lack of recombination), which results in the accentuation of the differentiation between populations from one generation to the next [53,54].

Secondly, the perspective offered by studying ChrY is representative of only males within a population. This is comparable to reasoning applied when using mitochondrial DNA (mtDNA), which is (thought to be) predominantly inherited through the female line [55–57]. A plethora of demographic and sociocultural practices between males and females (for example patrilocality versus matrilocality, polygamy and disparity in life expectancy [58]) have the potential to influence gene flow in the population. If ChrY is considered in isolation, intricacies of population dynamics may be missed [59]. Combining two data types (that of MSY and mtDNA) is considered more efficacious. This contextual approach to the analysis of the MSY and its variation is in itself another application for the MSY, which has allowed collation of more information about the variation in movement between males and females over time [60–62].

Thirdly and importantly, there are key assumptions upon which many of the investigations mentioned above are based—including that these markers are neutral. This assumption is being tested in recent research which looks at whether variation in the MSY could confer significant differences in fitness or affect complex traits [15,63]. In addition, Y chromosome data can be used to estimate the time to most recent common ancestor (TMRCA) and, due to its inherently smaller population size, the estimation given by using the Y chromosome should be proportionally more recent than the time indicated by neutral markers found on autosomes or the X chromosome [54]. TMRCA estimates based on ChrY are revised as the knowledge in the field increases. For example, it has been suggested that TMRCA estimates given by ChrY data are even more recent than would be predicted and thus provide possible evidence of natural selection acting on ChrY [54,64]. In fact, there is recent evidence suggesting that natural selection is involved in the maintenance of genetic variation involving copy number variants [65]. This is an area where some controversy remains, following the discovery of a new lineage (haplogroup A00) in 2013, revisions of the estimated TMRCA have been suggested [66–69].

3. Y Haplogroups

A key particularity of the MSY in relation to human population history is the existence of Y haplogroups. This source of genetic variation, absent on autosomes, is created by the absence of recombination on the MSY. Y haplogroups are defined as groups of similar Y haplotypes (genetic markers that are inherited together) that share a common ancestor with a particular SNP mutation (Figure 3).
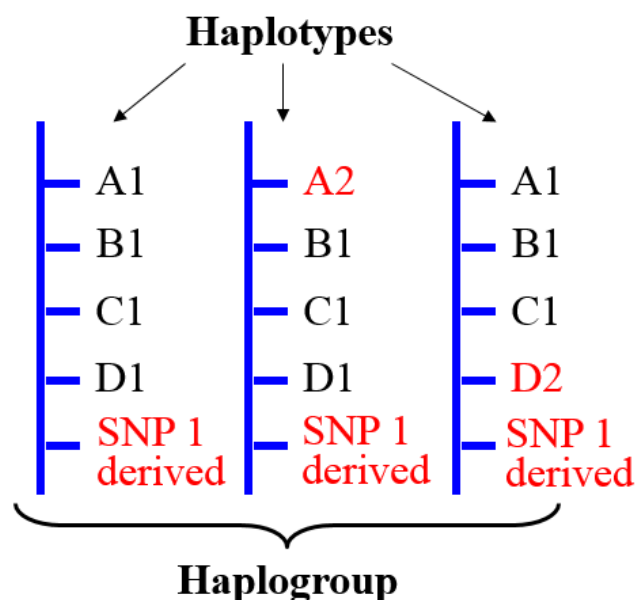
Figure 3. Haplogroups are groups of haplotypes that share a common ancestor. Individuals belonging to a specific haplogroup share the same derived allele for one informative SNP. A derived allele is a new variant in a locus, different from the original non-mutated allele (which is known as the ancestral allele). This figure shows three haplotypes. All of them share the derived allele for SNP1. Variation in other SNPs (in red) creates different haplotypes within this haplogroup.

Y haplogroups have been identified and organised into a phylogenetic tree which is added to and amended as new ChrY SNPs are discovered [16,70]. Currently, more than 300 haplogroups have been identified throughout the world and have been studied as a means to investigate and understand several aspects related to Genetics. Their use in genetic genealogy and forensics has been discussed at length in recent reviews [71,72]. We briefly discuss here the use of Y haplogroups in phylogenetics and population genetics.

The use of Y haplogroups as a tool for informing investigation into early human and indeed primate population history has been widely studied. By looking at common Y haplogroups and the variation which they contain, it is possible to estimate the time of origin and hypothesise about the movement of their carriers across the globe. To this end, climatic and archaeological data have been analysed in conjunction with MSY insertion and deletion events [51,73,74], microsatellites [75] and minisatellites [76]. Analysis of the MSY has provided solid evidence about the origins and early migration of *Homo sapiens* [47,77,78]; the identification of candidate founder haplotypes for numerous populations [79] and the elucidation of intricacies of inter and intrapopulation dynamics across the globe, including suggestions of partition and coalescence events and previously unknown genetic bottlenecks [74,80].

In addition to the investigation of human population history, the MSY has been used to gather information about more recent human demographics. Analysis has allowed the identification of Y haplotype clusters and their distribution throughout Europe [81–85]. These subdivisions within the

population are a potential source of variation in disease risk and further studies incorporating this aspect of variation could have important implications for using haplogroup data in the study of disease.

4. The Y Chromosome in Genetic Epidemiological Studies of Human Complex Traits

Genetic epidemiological studies involving ChrY aim to analyse the role of ChrY and environmental factors in determining or influencing health and disease in families and in populations. To date, a variety of human complex traits including behaviour and psychiatric traits [24], cancer risk [86,87], autoimmunity [88,89] and HIV progression [23] have been interrogated to this end, but the relationship between the human ChrY and cardiovascular risk is one of the best studied examples. There is a recent review [90] covering many aspects of this relationship, and therefore, we will not cover details here. Instead, we will present key aspects for consideration when utilising ChrY in genetic epidemiological studies using studies that explore cardiovascular risk as examples. We will discuss gender differences in relation to disease incidence and progression, study design and the use of Y haplogroups in association mapping.

4.1. Gender Differences in Relation to Disease

The fact that females lack all ChrY-specific genes implies that ChrY does not contain genes essential for survival. However, gender differences in disease incidence have been documented for many conditions [91–93]. For example, females are more likely to suffer from certain systemic autoimmune conditions such as systemic lupus erythematosus [94] and rheumatoid arthritis [95]. Conversely, men are more likely to suffer from schizophrenia [92] and cardiovascular disease [93]. There are also significant gender differences in progression of certain diseases [96,97]. It is likely that an interplay of genetic and environmental factors explains these differences. Variables that have been identified thus far include gender differences in lifestyle patterns (e.g., alcohol consumption, smoking and dietary habits) [98], variation in hormonal levels during development [99], hormonal influence during adult life [100] and many others such as pregnancy [101].

4.2. Study Design

Study design is key for the success of association mapping analyses [102]. The association mapping study designs utilised to date to explore the relationship between genetic variation on ChrY and human complex traits include linkage studies, candidate gene studies, animal models and genome-wide association studies (GWASs). Table 2 summarises different aspects of each of these study designs and examples where they have been applied to disentangle the genetic contribution of the MSY in relation to human complex traits. Table 2 also shows the implicit assumptions of each study design, their strengths and limitations and examples of published studies to help contextualise the contribution of these association mapping approaches to current understanding of the relationship between MSY and complex traits.

Table 2. Basis of different study designs used to investigate the association between ChrY and complex traits, with strengths and limitations for consideration when analysing findings.

| Study Design	Basis	Strengths	Limitations	Reference
Linkage analysis	First-degree relatives are compared in order to ascertain the potential for a genetic component of disease susceptibility	1- Successful in the identification of highly penetrant genetic variants related to Mendelian traits or monogenic disorders	1- Limited application for complex traits due to the use of individuals that share similar genetic and environmental constituents, establishing true effects in multifactorial traits is limited	[44]
Candidate gene studies	A particular gene is studied based on biological plausibility. Variation at this gene is investigated in genetic association studies	1- Highly specific for genetic variation which focuses on the MSY	1- Locus selected in absence of understanding of its function and potential effects 2- Highly vulnerable to chance making conflicting evidence more likely	[12–14,19–21,103]
Animal models	Animal studies are used as a framework for looking at human disease	1- Allows careful control and manipulation of both genetic and external environment to isolate the effects of the MSY	1- Application of animal models to human disease makes two key assumptions that may be incorrect: - -That disease process studied is the same in humans as in other animals - -That the MSY in other animals and that in humans is equivalent 2- Application of studies using consomic strains is limited due to confounding factors	[104–107]
GWAS	Case-control study design used to look for common genetic variants more frequently identified in those with particular diseases	1- Allows genotype-first analysis of the MSY for which understanding of genetic content is limited	1- Sex chromosomes are routinely excluded from these study types as the entirety of the MSY is effectively in linkage disequilibrium 2- Haplotypes and haplogroups used as a basis for this type of analysis may be flawed 3- Frequency of haplotypes within haplogroups in different cohorts may be sufficiently different to dilute or exaggerate relationships seen in other populations	[15,18,22,23,108–111]

4.2.1. Genome-Wide Association Studies (GWASs)

Table 3 shows a summary of issues relevant to genetic association studies carried out using MSY SNPs (including GWASs). These include statistical power, linkage disequilibrium patterns, population stratification, presence of complex loci, colocalisation, pleiotropy, gene–gene interactions, gene/protein expression and replication.

Although these aspects are also relevant to genetic association studies involving SNPs located on autosomes, studies involving ChrY need to account for additional issues that are specific to the genetic complexities of the MSY. These specific characteristics of ChrY open additional effects and possible solutions to account for them in the design and interpretation of GWASs involving ChrY SNPs. Table 3 summarises the implications of each of them on the interpretation of genetic association studies. We also point to possible solutions to minimise their impact.

The current GWAS literature does not offer conclusive evidence for a role for the MSY in disease susceptibility. Various potential reasons for this exist. For example, ChrY is excluded from most GWASs. In fact, ChrY SNPs are excluded from some commonly used genotyping arrays [112]. Another reason is that some authors opt for to exclude ChrY SNPs from GWAS analyses. Of note however, is that ChrY is routinely excluded from GWASs as all common variants in the MSY are effectively in linkage disequilibrium (Table 3). Although more recently X has been more likely to be included [113], this practice has left a dearth of evidence with respect to ChrY. For example, out of more than 50 GWASs conducted over the last 10 years on chronic kidney disease, only one reported results of ChrY analyses [112].
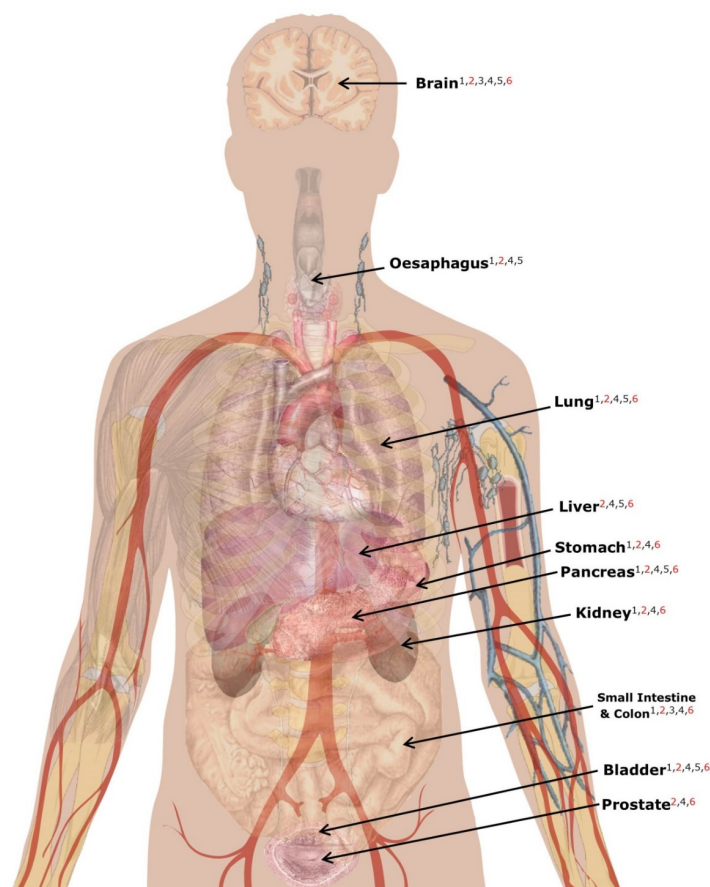
Figure 4. Protein-coding Y-chromosomal genes expressed in non-gonadal tissues (not comprehensive) as detected by the Human Protein Atlas. Protein labelled 1: DDX3Y; 2: EIF1AY; 3: NLGN4Y; 4: RPS4Y1; 5: UTY; 6: ZFY. Proteins highlighted in red have a low quality of supporting evidence. Image source: commons.wikimedia.org/wiki/Human_body_diagrams (Wikimedia CC0 licence).

Table 3. Considerations and possible solutions for genetic association studies carried out using SNPs in the MSY.

GWAS Features	Y Chromosome vs. Autosomes	Implication	Possible Solution(s)	References
Statistical power	Only men inherit a Y chromosome	~50% reduction in sample size	1- Include the Y chromosome in GWASs—could be mandated by funders 2- Share GWAS summary results (e.g., via LD Hub, GWAS Catalogue) 3- Increase sample sizes	[2,112,115]
LD structure	All the common variants in the MSY are in LD	Identifying the causal variant is very difficult	1- Larger sample sizes 2- Fine mapping by (i) sequencing the MSY and the associated region, (ii) carrying out transethnic studies and/or (iii) Y-DNA haplogroup association studies 3- Functional analyses (e.g., single SNP editing)	[15,24,116]
Population stratification	Principal components calculated using autosomal SNPs are not applicable	Potential overadjustment and loss of statistical power	1- Sensitivity analyses with and without genetic principal components 2- Sensitivity analyses with and without Y-DNA haplogroup information as a covariate in (male-only) GWASs and looking for SNPs with significant differences in effect sizes	[117–119]
Complex loci	Many highly variable regions and repetitive sequences	Variant calling may not be accurate	Only include variants called with high confidence	[17]
Colocalisation of eQTL and GWAS signals	No eQTLs identified for the Y chromosome	GWAS-eQTL colocalisation analysis cannot be carried out at present	Initiate trans-ethnic eeGWASs and/or study rare variants on the Y chromosome	[114]
Pleiotropy	No conclusive GWAS signal identified	Not much information to link potential findings with other biological pathways	1- Initiate a consortium to identify associations on the MSY 2- Carry out PheWASs using all Y chromosomal SNPs (incl. rare variants)	[2]
Gene–gene interactions	Very few examples identified in autosomes. None with SNPs on the Y chromosome	Almost no statistical power to detect small effects	Hypothesis driven approaches (e.g., between SNPs associated with obesity/CVD and SNPs in/near *UTY*—a gene expressed in non-gonadal tissues)	[120]
Gene/protein Expression	Many MSY genes/proteins are not expressed at detectable levels in non-gonadal tissues	Identified associations are likely to be biologically implausible if not expressed in disease-relevant tissue	Query the Human Protein Atlas to check whether the putatively causal gene/protein is active in a relevant tissue (Figure 4)	[121]

4.2.2. Expression Quantitative Trait Loci (eQTLs) and Protein Quantitative Trait Loci (pQTLs)

Currently, there are 45 protein-coding genes specific to the Y chromosome (i.e., excl. genes in the pseudoautosomal regions; Ensembl BioMart, accessed: 11/10/20). These 45 genes code for 27 distinct proteins/protein complexes [114].

If ChrY variants are included in a GWAS, any identified associations will need to be biologically plausible. To aid the search for prime candidates for GWASs of common complex diseases, we queried the Human Protein Atlas (v19.2; Ensembl v92) for all 45 genes [121] (using aliases from Ensembl BioMart database [122] to make sure we captured all) in all tissues except the testis and female-specific tissues (i.e., vagina, ovary and fallopian tubes). We also removed results marked with "not detected"—indicating that the levels of expression, if any, were below the limit of detection. As a positive control, we queried to confirm whether the proteins were detected in the testis; 27 distinct proteins were detected (validated by transcript-level evidence): DDX3Y, ZFY, RPS4Y1, VCY1B, VCY, CDY2B, TSPY2, RBMY1F, CDY1, CDY1B, CDY2A, UTY, DAZ3, DAZ1, EIF1AY, DAZ4, DAZ2, RBMY1J, TSPY3, TSPY8, TSPY4, RBMY1A1, TSPY10, RBMY1E, RBMY1B, RBMY1D, and TSPY1.

The only proteins that were detectable outside of the testes by the Human Protein Atlas were DDX3Y, ZFY, RPS4Y1, NLGN4Y, UTY, and EIF1AY, which would make them and eQTLs/pQTLs affecting their expression prime candidates for non-fertility related diseases/traits (Figure 4). There are three more proteins—KDM5D, USP9Y, TMSB4Y—which are putatively ubiquitously expressed as determined by RNA-seq carried out by GTEx [114] but are not yet validated by the Human Protein Atlas' immunohistochemistry assays – and would thus also be candidates for non-fertility related diseases/traits [63,114].

We further searched the literature (e.g., GTEx [114], BRAINEAC [123], and Blood eQTL server [124]) to identify eQTLs on the Y chromosome. However, we could not find any—reflecting the need for further research in this area (Table 3).

When interpreting these findings, one must consider the limitations of the analysis carried out by the Human Protein Atlas, which uses an antibody-detection method, inevitably dependent on the accuracy of antibodies used [125]. In some cases, antibodies are not able to distinguish between proteins in the same family [125]. For example, antibodies may fail to distinguish between RBMY1A1 and RBMY1B, two members of the same family, both located on ChrY; similarly, between E1F1AY and E1F1AX which are located on ChrY and ChrX, respectively. However, the Human Protein Atlas uses other lines of evidence for antibody verification, including transcript mRNA validation (https://www.proteinatlas.org/about/antibody+validation). Secondly, investigation here is limited to those genes that are ubiquitously expressed, there are other protein-coding genes whose proteins have been detected outside the gonads but not ubiquitously which may be plausible targets for complex trait analysis. For example, PCDH11Y, predominantly expressed in the brain [63], which could perhaps influence brain related diseases or behavioural traits. Thirdly, as protein expression varies with development, there may be ChrY gene expression which affects complex trait development, which is undetectable by currently used methods because they are not expressed in adult tissues [114].

4.3. Use of Y Haplogroups in Genetic Association Studies of Common Complex Traits

Y chromosome (Y-DNA) haplogroups are more widely used in population genetics than in genetic epidemiology, although associations between Y-DNA haplogroups and several traits, including cardio-metabolic traits and psychiatric traits have been reported [20,24,110].

In contrast, non-recombining genetic variation, such as Y chromosomal (Y-DNA) haplogroups, has rarely been considered in the design and interpretation of genetic association studies [82]. We have recently studied weather hidden stratification and/or differential phenotypic effects by Y-DNA haplogroups could exist [118]. Y haplogroups can be used to stratify individuals in genetic association analyses. To this end, we followed a two-stage approach. Firstly, we stratified individuals according to their Y haplogroups and then tested for association between 32 autosomal BMI-related SNPs from Speliotes et al. [126] and BMI within each haplogroup after adjusting for the top 10 genetic principal

components. Although we found evidence suggestive of an interaction between haplogroup I and *FTO* SNPs in ALSPAC, we could not fully replicate these findings in the 1958 Birth Cohort. Further studies are needed.

Consideration of haplogroups in the design and interpretation of genetic association studies of ChrY, could be a crucial and commonly omitted source of complexity. Potential caveats and limitations are in addition to those listed on Table 3 and relate to novel issues on genetic complexity, population structure and statistical power.

4.3.1. Genetic Complexity Inherent to Y Haplogroups

Y haplogroups are complex genetic variants that combine haplotypes sharing a common ancestor [70]. As a consequence, individuals sharing the same haplogroup are genetically heterogeneous in relation to single genetic common variation, including SNPs. This may have direct implications for genetic association studies. A SNP signal identified in GWASs could be diluted or more difficult to identify when studying Y haplogroups. Similarly, comparing individuals from different cohorts that share the same haplogroup could represent an additional layer of heterogeneity. It is likely that such individuals differ genotypically in relation to a number of SNPs, despite belonging to the same haplogroup. This heterogeneity needs to be taken into account in the interpretation of results observed from different populations, traits and haplogroups.

Alternatively, haplogroups could be well placed to capture the effect of combinations of two or more SNPs. In fact, haplotypes can define functional units of genes [127]. Advantages of analysing haplogroups in association studies would include the fact that genetic variation in populations is organised into haplotypes and that combining SNPs into haplogroups reduces the dimension of association tests and may augment statistical power.

4.3.2. Statistical Power Issues in Studies Involving Y Haplogroups

The increase in power resulting from combining SNPs into haplogroups has a differential effect on different haplogroups. An illustration table from [24] shows the frequency of different haplogroups observed from a European cohort. 72.1% of individuals belong to haplogroup R, the commonest haplogroup in Europe, 19% to haplogroup I. The remaining 9% of individuals belong to 10 other haplogroups. As a result, there is considerable within-study power heterogeneity for different haplogroups. This is a common feature of most studies involving Y haplogroups.

This has implications for study design and interpretation. Some analyses combine these heterogeneous haplogroups into a new group, whereas other simply remove them from the analyses. Collapsing different haplogroups is not an optimal solution, since different mutations are likely to be associated with a specific haplotypic background.

On occasion, studies report suggestive associations involving low-frequency haplogroups. The inherent low power of these associations together with the aforementioned heterogeneity among populations make replication of these signals difficult. This also applies to subgroups of common haplogroups. Refining associations within a haplogroup by considering different subgroups (e.g., R1a1, R1b1, R1b1b2, R1b1b2g, R1b1b2h . . .) mirrors the power issue that we have previously shown.

4.3.3. Population Structure Particularities Specific to Y Haplogroups

Population samples of Y haplogroups present specific idiosyncrasies, different from population samples of SNPs, that need to be taken into consideration in the design and interpretation of genetic association studies.

Extensive population evidence shows that Y haplogroups (and specifically, genetic trees that generated them) describe ChrY correctly from a population point of view. Evolutionary history of Y haplogroups correlates with geography. The question, still unresolved, is whether this also creates an association to biology, specifically, to genetic association with human complex traits. Interestingly, a recent study which investigated the pattern of methylation amongst some Y haplogroups [128] has

identified haplogroup-specific methylation sites accompanied by SNPs, irrespective of geographical origin. These results also prompt one to question whether evolutionary history of ChrY could correlate with human complex trait variation independent of geography.

5. Conclusions

To date, there is little evidence supporting a role of human ChrY variation on human complex traits. A possible reason for this is that ChrY variation is actually irrelevant in relation to human complex traits. Alternatively, one could argue that the lack of evidence could be related to study design and interpretation issues. Here, we have reviewed and presented here evidence that highlights the complexities of this issue and disentangles relevant variables and their possible role.

References

1. Visscher, P.M.; Wray, N.R.; Zhang, Q.; Sklar, P.; McCarthy, M.I.; Brown, M.A.; Yang, J. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Human Genet.* **2017**, *101*, 5–22. [CrossRef] [PubMed]
2. 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] [PubMed]
3. Painter, T.S. The y-chromosome in mammals. *Science* **1921**, *53*, 503–504. [CrossRef] [PubMed]
4. Casanova, M.; Leroy, P.; Boucekkine, C.; Weissenbach, J.; Bishop, C.; Fellous, M.; Purrello, M.; Fiori, G.; Siniscalco, M. A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* **1985**, *230*, 1403–1406. [CrossRef] [PubMed]
5. Disteche, C.M.; Casanova, M.; Saal, H.; Friedman, C.; Sybert, V.; Graham, J.; Thuline, H.; Page, D.C.; Fellous, M. Small deletions of the short arm of the Y chromosome in 46,XY females. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 7841–7844. [CrossRef]
6. Lucotte, G.; Guérin, P.; Hallé, L.; Loirat, F.; Hazout, S. Y chromosome DNA polymorphisms in two African populations. *Am. J. Human Genet.* **1989**, *45*, 16–20.
7. Sinclair, A.H.; Berta, P.; Palmer, M.S.; Hawkins, J.R.; Griffiths, B.L.; Smith, M.J.; Foster, J.W.; Frischauf, A.M.; Lovell-Badge, R.; Goodfellow, P.N. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **1990**, *346*, 240–244. [CrossRef] [PubMed]
8. Ely, D.L.; Turner, M.E. Hypertension in the spontaneously hypertensive rat is linked to the Y chromosome. *Hypertension* **1990**, *16*, 277–281. [CrossRef]
9. Koopman, P.; Gubbay, J.; Vivian, N.; Goodfellow, P.; Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. *Nature* **1991**, *351*, 117–121. [CrossRef]
10. Jobling, M.A.; Pandya, A.; Tyler-Smith, C. The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal Med.* **1997**, *110*, 118–124. [CrossRef] [PubMed]
11. Underhill, P.A.; Shen, P.; Lin, A.A.; Jin, L.; Passarino, G.; Yang, W.H.; Kauffman, E.; Bonné-Tamir, B.; Bertranpetit, J.; Francalacci, P.; et al. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **2000**, *26*, 358–361. [CrossRef]
12. Ellis, J.A.; Stebbing, M.; Harrap, S.B. Association of the human Y chromosome with high blood pressure in the general population. *Hypertension* **2000**, *36*, 731–733. [CrossRef] [PubMed]
13. Charchar, F.J.; Tomaszewski, M.; Padmanabhan, S.; Lacka, B.; Upton, M.N.; Inglis, G.C.; Anderson, N.H.; McConnachie, A.; Zukowska-Szczechowska, E.; Grzeszczak, W.; et al. The Y chromosome effect on blood pressure in two European populations. *Hypertension* **2002**, *39*, 353–356. [CrossRef] [PubMed]

14. Charchar, F.J.; Tomaszewski, M.; Lacka, B.; Zakrzewski, J.; Zukowska-Szczechowska, E.; Grzeszczak, W.; Dominiczak, A.F. Association of the human Y chromosome with cholesterol levels in the general population. *Arterioscler. Thromb. Vasc. Biol.* **2004**, *24*, 308–312. [CrossRef] [PubMed]

15. Charchar, F.J.; Bloomer, L.D.; Barnes, T.A.; Cowley, M.J.; Nelson, C.P.; Wang, Y.; Denniff, M.; Debiec, R.; Christofidou, P.; Nankervis, S.; et al. Inheritance of coronary artery disease in men: An analysis of the role of the Y chromosome. *Lancet* **2012**, *379*, 915–922. [CrossRef]

16. Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **2002**, *12*, 339–348. [CrossRef]

17. Skaletsky, H.; Kuroda-Kawaguchi, T.; Minx, P.J.; Cordum, H.S.; Hillier, L.; Brown, L.G.; Repping, S.; Pyntikova, T.; Ali, J.; Bieri, T.; et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **2003**, *423*, 825–837. [CrossRef]

18. Haitjema, S.; van Setten, J.; Eales, J.; van der Laan, S.W.; Gandin, I.; de Vries, J.P.; de Borst, G.J.; Pasterkamp, G.; Asselbergs, F.W.; Charchar, F.J.; et al. Genetic variation within the Y chromosome is not associated with histological characteristics of the atherosclerotic carotid artery or aneurysmal wall. *Atherosclerosis* **2017**, *259*, 114–119. [CrossRef]

19. Kostrzewa, G.; Broda, G.; Konarzewska, M.; Krajewki, P.; Płoski, R. Genetic polymorphism of human Y chromosome and risk factors for cardiovascular diseases: A study in WOBASZ cohort. *PLoS ONE* **2013**, *8*, e68155. [CrossRef]

20. Rodríguez, S.; Chen, X.H.; Miller, G.J.; Day, I.N. Non-recombining chromosome Y haplogroups and centromeric HindIII RFLP in relation to blood pressure in 2743 middle-aged Caucasian men from the UK. *Human Genet.* **2005**, *116*, 311–318. [CrossRef]

21. Russo, P.; Venezia, A.; Lauria, F.; Strazzullo, P.; Cappuccio, F.P.; Iacoviello, L.; Barba, G.; Siani, A. HindIII(+/-) polymorphism of the Y chromosome, blood pressure, and serum lipids: No evidence of association in three white populations. *Am. J. Hypertens* **2006**, *19*, 331–338. [CrossRef]

22. Lu, C.; Wen, Y.; Hu, W.; Lu, F.; Qin, Y.; Wang, Y.; Li, S.; Yang, S.; Lin, Y.; Wang, C.; et al. Y chromosome haplogroups based genome-wide association study pinpoints revelation for interactions on non-obstructive azoospermia. *Sci. Rep.* **2016**, *6*, 33363. [CrossRef] [PubMed]

23. Sezgin, E.; Lind, J.M.; Shrestha, S.; Hendrickson, S.; Goedert, J.J.; Donfield, S.; Kirk, G.D.; Phair, J.P.; Troyer, J.L.; O'Brien, S.J.; et al. Association of Y chromosome haplogroup I with HIV progression, and HAART outcome. *Human Genet.* **2009**, *125*, 281–294. [CrossRef] [PubMed]

24. Howe, L.J.; Erzurumluoglu, A.M.; Davey Smith, G.; Rodriguez, S.; Stergiakouli, E. Y Chromosome, Mitochondrial DNA and Childhood Behavioural Traits. *Sci. Rep.* **2017**, *7*, 11655. [CrossRef] [PubMed]

25. Aken, B.L.; Ayling, S.; Barrell, D.; Clarke, L.; Curwen, V.; Fairley, S.; Fernandez Banet, J.; Billis, K.; García Girón, C.; Hourlier, T.; et al. The Ensembl gene annotation system. *Database (Oxford)* **2016**, *2016*. [CrossRef]

26. Hughes, J.F.; Skaletsky, H.; Brown, L.G.; Pyntikova, T.; Graves, T.; Fulton, R.S.; Dugan, S.; Ding, Y.; Buhay, C.J.; Kremitzki, C.; et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **2012**, *483*, 82–86. [CrossRef]

27. Rozen, S.; Skaletsky, H.; Marszalek, J.D.; Minx, P.J.; Cordum, H.S.; Waterston, R.H.; Wilson, R.K.; Page, D.C. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **2003**, *423*, 873–876. [CrossRef]

28. Cruciani, F.; Trombetta, B.; Macaulay, V.; Scozzari, R. About the X-to-Y gene conversion rate. *Am. J. Human Genet.* **2010**, *86*, 495–497; author reply 497–498. [CrossRef]

29. Trombetta, B.; Cruciani, F. Y chromosome palindromes and gene conversion. *Human Genet.* **2017**, *136*, 605–619. [CrossRef]

30. Aitken, R.J.; Marshall Graves, J.A. The future of sex. *Nature* **2002**, *415*, 963. [CrossRef]

31. Hawley, R.S. The human Y chromosome: Rumors of its death have been greatly exaggerated. *Cell* **2003**, *113*, 825–828. [CrossRef]

32. Jobling, M.; Hollox, E.; Hurles, M.; Kivisild, T.; Tyler-Smith, C. *Human Evolutionary Genetics*, 2nd ed.; Garland Science: New York, NY, USA, 2013.

33. Graves, J.A. Sex chromosome specialization and degeneration in mammals. *Cell* **2006**, *124*, 901–914. [CrossRef] [PubMed]

34. Marshall Graves, J.A. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu Rev. Genet.* **2008**, *42*, 565–586. [CrossRef]

35. Arakawa, Y.; Nishida-Umehara, C.; Matsuda, Y.; Sutou, S.; Suzuki, H. X-chromosomal localization of mammalian Y-linked genes in two XO species of the Ryukyu spiny rat. *Cytogenet. Genome Res.* **2002**, *99*, 303–309. [CrossRef] [PubMed]

36. Hughes, J.F.; Skaletsky, H.; Pyntikova, T.; Minx, P.J.; Graves, T.; Rozen, S.; Wilson, R.K.; Page, D.C. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **2005**, *437*, 100–103. [CrossRef]

37. Charlesworth, B.; Charlesworth, D. The degeneration of Y chromosomes. *Philos Trans. R Soc. Lond B Biol. Sci.* **2000**, *355*, 1563–1572. [CrossRef]

38. Skov, L.; Schierup, M.H.; Consortium, D.P.G. Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. *PLoS Genet.* **2017**, *13*, e1006834. [CrossRef] [PubMed]

39. Smedley, D.; Haider, S.; Durinck, S.; Pandini, L.; Provero, P.; Allen, J.; Arnaiz, O.; Awedh, M.H.; Baldock, R.; Barbiera, G.; et al. The BioMart community portal: An innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **2015**, *43*, W589–W598. [CrossRef]

40. Bruford, E.A.; Braschi, B.; Denny, P.; Jones, T.E.M.; Seal, R.L.; Tweedie, S. Guidelines for human gene nomenclature. *Nat. Genet.* **2020**, *52*, 754–758. [CrossRef]

41. Pruitt, K.D.; Harrow, J.; Harte, R.A.; Wallin, C.; Diekhans, M.; Maglott, D.R.; Searle, S.; Farrell, C.M.; Loveland, J.E.; Ruef, B.J.; et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* **2009**, *19*, 1316–1323. [CrossRef]

42. Gaudet, P.; Michel, P.A.; Zahn-Zabal, M.; Britan, A.; Cusin, I.; Domagalski, M.; Duek, P.D.; Gateau, A.; Gleizes, A.; Hinard, V.; et al. The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D177–D182. [CrossRef] [PubMed]

43. Girardi, S.K.; Mielnik, A.; Schlegel, P.N. Submicroscopic deletions in the Y chromosome of infertile men. *Human Reprod.* **1997**, *12*, 1635–1641. [CrossRef] [PubMed]

44. Vogt, P.H.; Edelmann, A.; Kirsch, S.; Henegariu, O.; Hirschmann, P.; Kiesewetter, F.; Köhn, F.M.; Schill, W.B.; Farah, S.; Ramos, C.; et al. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Mol. Genet.* **1996**, *5*, 933–943. [CrossRef]

45. Bellott, D.W.; Hughes, J.F.; Skaletsky, H.; Brown, L.G.; Pyntikova, T.; Cho, T.J.; Koutseva, N.; Zaghlul, S.; Graves, T.; Rock, S.; et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **2014**, *508*, 494–499. [CrossRef]

46. Meyfour, A.; Pooyan, P.; Pahlavan, S.; Rezaei-Tavirani, M.; Gourabi, H.; Baharvand, H.; Salekdeh, G.H. Chromosome-Centric Human Proteome Project Allies with Developmental Biology: A Case Study of the Role of Y Chromosome Genes in Organ Development. *J. Proteome Res.* **2017**, *16*, 4259–4272. [CrossRef]

47. Underhill, P.A.; Kivisild, T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev. Genet.* **2007**, *41*, 539–564. [CrossRef]

48. Hammer, M.F.; Spurdle, A.B.; Karafet, T.; Bonner, M.R.; Wood, E.T.; Novelletto, A.; Malaspina, P.; Mitchell, R.J.; Horai, S.; Jenkins, T.; et al. The geographic distribution of human Y chromosome variation. *Genetics* **1997**, *145*, 787–805.

49. Watanabe, Y.; Naka, I.; Khor, S.S.; Sawai, H.; Hitomi, Y.; Tokunaga, K.; Ohashi, J. Analysis of whole Y-chromosome sequences reveals the Japanese population history in the Jomon period. *Sci. Rep.* **2019**, *9*, 8556. [CrossRef] [PubMed]

50. Dupanloup, I.; Pereira, L.; Bertorelle, G.; Calafell, F.; Prata, M.J.; Amorim, A.; Barbujani, G. A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *J. Mol. Evol.* **2003**, *57*, 85–97. [CrossRef]

51. Hammer, M.F.; Redd, A.J.; Wood, E.T.; Bonner, M.R.; Jarjanazi, H.; Karafet, T.; Santachiara-Benerecetti, S.; Oppenheim, A.; Jobling, M.A.; Jenkins, T.; et al. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6769–6774. [CrossRef]

52. Karmin, M.; Saag, L.; Vicente, M.; Wilson Sayres, M.A.; Järve, M.; Talas, U.G.; Rootsi, S.; Ilumäe, A.M.; Mägi, R.; Mitt, M.; et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* **2015**, *25*, 459–466. [CrossRef] [PubMed]

53. Wright, S. The Distribution of Gene Frequencies in Populations of Polyploids. *Proc. Natl. Acad. Sci. USA* **1938**, *24*, 372–377. [CrossRef]

54. Jobling, M.A.; Tyler-Smith, C. The human Y chromosome: An evolutionary marker comes of age. *Nat. Rev. Genet.* **2003**, *4*, 598–612. [CrossRef]

55. Giles, R.E.; Blanc, H.; Cann, H.M.; Wallace, D.C. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 6715–6719. [CrossRef] [PubMed]

56. Sato, M.; Sato, K. Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochim. Biophys. Acta* **2013**, *1833*, 1979–1984. [CrossRef] [PubMed]

57. Chinnery, P.F. Inheritance of mitochondrial disorders. *Mitochondrion* **2002**, *2*, 149–155. [CrossRef] [PubMed]

58. Cavalli-Sforza, L.L.; Feldman, M.W. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* **2003**, *33*, 266–275. [CrossRef]

59. Seielstad, M.T.; Minch, E.; Cavalli-Sforza, L.L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **1998**, *20*, 278–280. [CrossRef]

60. Bolnick, D.A.; Bolnick, D.I.; Smith, D.G. Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Mol. Biol. Evol.* **2006**, *23*, 2161–2174. [CrossRef]

61. Oota, H.; Settheetham-Ishida, W.; Tiwawech, D.; Ishida, T.; Stoneking, M. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat. Genet.* **2001**, *29*, 20–21. [CrossRef]

62. Lippold, S.; Xu, H.; Ko, A.; Li, M.; Renaud, G.; Butthof, A.; Schröder, R.; Stoneking, M. Human paternal and maternal demographic histories: Insights from high-resolution Y chromosome and mtDNA sequences. *Investig. Genet.* **2014**, *5*, 13. [CrossRef]

63. Maan, A.A.; Eales, J.; Akbarov, A.; Rowland, J.; Xu, X.; Jobling, M.A.; Charchar, F.J.; Tomaszewski, M. The Y chromosome: A blueprint for men's health? *Eur. J. Hum. Genet.* **2017**, *25*, 1181–1188. [CrossRef] [PubMed]

64. Pritchard, J.K.; Seielstad, M.T.; Perez-Lezaun, A.; Feldman, M.W. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol. Biol. Evol.* **1999**, *16*, 1791–1798. [CrossRef] [PubMed]

65. Teitz, L.S.; Pyntikova, T.; Skaletsky, H.; Page, D.C. Selection Has Countered High Mutability to Preserve the Ancestral Copy Number of Y Chromosome Amplicons in Diverse Human Lineages. *Am. J. Hum. Genet.* **2018**, *103*, 261–275. [CrossRef] [PubMed]

66. Mendez, F.L.; Krahn, T.; Schrack, B.; Krahn, A.M.; Veeramah, K.R.; Woerner, A.E.; Fomine, F.L.; Bradman, N.; Thomas, M.G.; Karafet, T.M.; et al. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am. J. Hum. Genet.* **2013**, *92*, 454–459. [CrossRef]

67. Elhaik, E.; Tatarinova, T.V.; Klyosov, A.A.; Graur, D. The 'extremely ancient' chromosome that isn't: A forensic bioinformatic investigation of Albert Perry's X-degenerate portion of the Y chromosome. *Eur. J. Human Genet.* **2014**, *22*, 1111–1116. [CrossRef]

68. Mendez, F.L.; Veeramah, K.R.; Thomas, M.G.; Karafet, T.M.; Hammer, M.F. Reply to 'The 'extremely ancient' chromosome that isn't' by Elhaik et al. *Eur. J. Human Genet.* **2015**, *23*, 564–567. [CrossRef]

69. Elhaik, E.; Tatarinova, T.V.; Klyosov, A.A.; Graur, D. Reply to Mendez et al: The 'extremely ancient' chromosome that still isn't. *Eur J. Human Genet.* **2015**, *23*, 567–568. [CrossRef]

70. Karafet, T.M.; Mendez, F.L.; Meilerman, M.B.; Underhill, P.A.; Zegura, S.L.; Hammer, M.F. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* **2008**, *18*, 830–838. [CrossRef]

71. Calafell, F.; Larmuseau, M.H.D. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. *Human Genet.* **2017**, *136*, 559–573. [CrossRef]

72. Kayser, M. Forensic use of Y-chromosome DNA: A general overview. *Human Genet.* **2017**, *136*, 621–635. [CrossRef]

73. Underhill, P.A.; Jin, L.; Zemans, R.; Oefner, P.J.; Cavalli-Sforza, L.L. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 196–200. [CrossRef] [PubMed]

74. Underhill, P.A.; Jin, L.; Lin, A.A.; Mehdi, S.Q.; Jenkins, T.; Vollrath, D.; Davis, R.W.; Cavalli-Sforza, L.L.; Oefner, P.J. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **1997**, *7*, 996–1005. [CrossRef]

75. de Knijff, P.; Kayser, M.; Caglià, A.; Corach, D.; Fretwell, N.; Gehrig, C.; Graziosi, G.; Heidorn, F.; Herrmann, S.; Herzog, B.; et al. Chromosome Y microsatellites: Population genetic and evolutionary aspects. *Int. J. Legal Med.* **1997**, *110*, 134–149. [CrossRef]

76. Jobling, M.A.; Bouzekri, N.; Taylor, P.G. Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). *Human Mol. Genet.* **1998**, *7*, 643–653. [CrossRef]

77. Seielstad, M.; Bekele, E.; Ibrahim, M.; Touré, A.; Traoré, M. A view of modern human origins from Y chromosome microsatellite variation. *Genome Res.* **1999**, *9*, 558–567.

78. Wells, R.S.; Yuldasheva, N.; Ruzibakiev, R.; Underhill, P.A.; Evseeva, I.; Blue-Smith, J.; Jin, L.; Su, B.; Pitchappan, R.; Shanmugalakshmi, S.; et al. The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10244–10249. [CrossRef] [PubMed]

79. Pena, S.D.; Santos, F.R.; Bianchi, N.O.; Bravi, C.M.; Carnese, F.R.; Rothhammer, F.; Gerelsaikhan, T.; Munkhtuja, B.; Oyunsuren, T. A major founder Y-chromosome haplotype in Amerindians. *Nat. Genet.* **1995**, *11*, 15–16. [CrossRef]

80. Hammer, M.F.; Karafet, T.; Rasanayagam, A.; Wood, E.T.; Altheide, T.K.; Jenkins, T.; Griffiths, R.C.; Templeton, A.R.; Zegura, S.L. Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **1998**, *15*, 427–441. [CrossRef] [PubMed]

81. Roewer, L.; Croucher, P.J.; Willuweit, S.; Lu, T.T.; Kayser, M.; Lessig, R.; de Knijff, P.; Jobling, M.A.; Tyler-Smith, C.; Krawczak, M. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Human Genet.* **2005**, *116*, 279–291. [CrossRef]

82. Chen, X.H.; Rodríguez, S.; Hawe, E.; Talmud, P.J.; Miller, G.J.; Underhill, P.; Humphries, S.E.; Day, I.N. Evidence of admixture from haplotyping in an epidemiological study of UK Caucasian males: Implications for association analyses. *Human Hered* **2004**, *57*, 142–155. [CrossRef]

83. Rootsi, S.; Behar, D.M.; Järve, M.; Lin, A.A.; Myres, N.M.; Passarelli, B.; Poznik, G.D.; Tzur, S.; Sahakyan, H.; Pathak, A.K.; et al. Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. *Nat. Commun.* **2013**, *4*, 2928. [CrossRef] [PubMed]

84. Kayser, M.; Lao, O.; Anslinger, K.; Augustin, C.; Bargel, G.; Edelmann, J.; Elias, S.; Heinrich, M.; Henke, J.; Henke, L.; et al. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Human Genet.* **2005**, *117*, 428–443. [CrossRef]

85. Lao, O.; Lu, T.T.; Nothnagel, M.; Junge, O.; Freitag-Wolf, S.; Caliebe, A.; Balascakova, M.; Bertranpetit, J.; Bindoff, L.A.; Comas, D.; et al. Correlation between genetic and geographic structure in Europe. *Curr. Biol.* **2008**, *18*, 1241–1248. [CrossRef]

86. Patel, R.; Khalifa, A.O.; Isali, I.; Shukla, S. Prostate cancer susceptibility and growth linked to Y chromosome genes. *Front. Biosci. (Elite Ed.)* **2018**, *10*, 423–436. [CrossRef]

87. Kido, T.; Lau, Y.F. Roles of the Y chromosome genes in human cancers. *Asian J. Androl* **2015**, *17*, 373–380. [CrossRef]

88. Lleo, A.; Oertelt-Prigione, S.; Bianchi, I.; Caliari, L.; Finelli, P.; Miozzo, M.; Lazzari, R.; Floreani, A.; Donato, F.; Colombo, M.; et al. Y chromosome loss in male patients with primary biliary cirrhosis. *J. Autoimmun.* **2013**, *41*, 87–91. [CrossRef]

89. Persani, L.; Bonomi, M.; Lleo, A.; Pasini, S.; Civardi, F.; Bianchi, I.; Campi, I.; Finelli, P.; Miozzo, M.; Castronovo, C.; et al. Increased loss of the Y chromosome in peripheral blood cells in male patients with autoimmune thyroiditis. *J. Autoimmun.* **2012**, *38*, J193–J196. [CrossRef]

90. Khan, S.I.; Andrews, K.L.; Jennings, G.L.; Sampson, A.K.; Chin-Dusting, J.P.F. Y Chromosome, Hypertension and Cardiovascular Disease: Is Inflammation the Answer? *Int. J. Mol. Sci.* **2019**, *20*. [CrossRef]

91. Institute of Medicine (US) Committee on Understanding the Biology of Sex and Gender Differences; Wizemann, T.M.; Pardue, M.L. *Exploring the Biological Contributions to Human Health: Does Sex Matter?*; The National Acadamies Press (US): Washington, DC, USA, 2001.

92. McGrath, J.; Saha, S.; Chant, D.; Welham, J. Schizophrenia: A concise overview of incidence, prevalence, and mortality. *Epidemiol. Rev.* **2008**, *30*, 67–76. [CrossRef]

93. Mozaffarian, D.; Benjamin, E.J.; Go, A.S.; Arnett, D.K.; Blaha, M.J.; Cushman, M.; Das, S.R.; de Ferranti, S.; Després, J.P.; Fullerton, H.J.; et al. Executive Summary: Heart Disease and Stroke Statistics—2016 Update: A Report From the American Heart Association. *Circulation* **2016**, *133*, 447–454. [CrossRef]

94. Somers, E.C.; Thomas, S.L.; Smeeth, L.; Schoonen, W.M.; Hall, A.J. Incidence of systemic lupus erythematosus in the United Kingdom, 1990–1999. *Arthritis Care Rheum.* **2007**, *57*, 612–618. [CrossRef] [PubMed]

95. Ngo, S.T.; Steyn, F.J.; McCombe, P.A. Gender differences in autoimmune disease. *Front. Neuroendocrinol.* **2014**, *35*, 347–369. [CrossRef] [PubMed]

96. Whitacre, C.C.; Reingold, S.C.; O'Looney, P.A. A gender gap in autoimmunity. *Science* **1999**, *283*, 1277–1278. [CrossRef]

97. Voskuhl, R.R.; Sawalha, A.H.; Itoh, Y. Sex chromosome contributions to sex differences in multiple sclerosis susceptibility and progression. *Mult. Scler.* **2018**, *24*, 22–31. [CrossRef]

98. Anand, S.S.; Islam, S.; Rosengren, A.; Franzosi, M.G.; Steyn, K.; Yusufali, A.H.; Keltai, M.; Diaz, R.; Rangarajan, S.; Yusuf, S.; et al. Risk factors for myocardial infarction in women and men: Insights from the INTERHEART study. *Eur. Heart J.* **2008**, *29*, 932–940. [CrossRef]

99. Bao, A.M.; Swaab, D.F. Sexual differentiation of the human brain: Relation to gender identity, sexual orientation and neuropsychiatric disorders. *Front. Neuroendocrinol.* **2011**, *32*, 214–226. [CrossRef]

100. Ding, E.L.; Song, Y.; Malik, V.S.; Liu, S. Sex differences of endogenous sex hormones and risk of type 2 diabetes: A systematic review and meta-analysis. *JAMA* **2006**, *295*, 1288–1299. [CrossRef]

101. Adams Waldorf, K.M.; Nelson, J.L. Autoimmune disease during pregnancy and the microchimerism legacy of pregnancy. *Immunol. Invest.* **2008**, *37*, 631–644. [CrossRef]

102. Terwilliger, J.D.; Göring, H.H. Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis, and experimental design. 2000. *Human Biol.* **2009**, *81*, 663–728. [CrossRef] [PubMed]

103. García, E.C.; González, P.; Castro, M.G.; Alvarez, R.; Reguero, J.R.; Batalla, A.; Cortina, A.; Alvarez, V. Association between genetic variation in the Y chromosome and hypertension in myocardial infarction patients. *Am. J. Med. Genet. A* **2003**, *122A*, 234–237. [CrossRef] [PubMed]

104. Suto, J.; Satou, K. Effect of the Y chromosome on plasma high-density lipoprotein-cholesterol levels in Y-chromosome-consomic mouse strains. *BMC Res. Notes* **2014**, *7*, 393. [CrossRef] [PubMed]

105. Kren, V.; Qi, N.; Krenova, D.; Zidek, V.; Sladká, M.; Jáchymová, M.; Míková, B.; Horky, K.; Bonne, A.; Van Lith, H.A.; et al. Y-chromosome transfer induces changes in blood pressure and blood lipids in SHR. *Hypertension* **2001**, *37*, 1147–1152. [CrossRef]

106. Dickey, C.; Toot, J.; Terwilliger, M.; Payne, R.; Turner, M.; Ely, D. The SHR Y chromosome increases cardiovascular, endocrine, and behavioral responses to stress compared to the WKY Y chromosome. *Physiol. Behav.* **2012**, *106*, 101–108. [CrossRef]

107. Prokop, J.W.; Tsaih, S.W.; Faber, A.B.; Boehme, S.; Underwood, A.C.; Troyer, S.; Playl, L.; Milsted, A.; Turner, M.E.; Ely, D.; et al. The phenotypic impact of the male-specific region of chromosome-Y in inbred mating: The role of genetic variants and gene duplications in multiple inbred rat strains. *Biol. Sex. Differ.* **2016**, *7*, 10. [CrossRef] [PubMed]

108. Voskarides, K.; Hadjipanagi, D.; Papazachariou, L.; Griffin, M.; Panayiotou, A.G. Evidence for contribution of the y chromosome in atherosclerotic plaque occurrence in men. *Genet. Test. Mol. Biomarkers* **2014**, *18*, 552–556. [CrossRef]

109. Bloomer, L.D.; Nelson, C.P.; Eales, J.; Denniff, M.; Christofidou, P.; Debiec, R.; Moore, J.; Zukowska-Szczechowska, E.; Goodall, A.H.; Thompson, J.; et al. Male-specific region of the Y chromosome and cardiovascular risk: Phylogenetic analysis and gene expression studies. *Arterioscler. Thromb. Vasc. Biol.* **2013**, *33*, 1722–1727. [CrossRef]

110. O'Keeffe, L.M.; Howe, L.D.; Fraser, A.; Hughes, A.D.; Wade, K.H.; Anderson, E.L.; Lawlor, D.A.; Erzurumluoglu, A.M.; Davey-Smith, G.; Rodriguez, S.; et al. Associations of Y chromosomal haplogroups with cardiometabolic risk factors and subclinical vascular measures in males during childhood and adolescence. *Atherosclerosis* **2018**, *274*, 94–103. [CrossRef]

111. Sato, Y.; Shinka, T.; Nozawa, S.; Yoshiike, M.; Koh, E.; Kanaya, J.; Namiki, M.; Matsumiya, K.; Tsujimura, A.; Komatsu, K.; et al. Y chromosome haplogroup D2a1 is significantly associated with high levels of luteinizing hormone in Japanese men. *Andrology* **2015**, *3*, 520–525. [CrossRef]

112. Anderson, K.; Cañadas-Garre, M.; Chambers, R.; Maxwell, A.P.; McKnight, A.J. The Challenges of Chromosome Y Analysis and the Implications for Chronic Kidney Disease. *Front. Genet.* **2019**, *10*, 781. [CrossRef]

113. Winham, S.J.; de Andrade, M.; Miller, V.M. Genetics of cardiovascular disease: Importance of sex and ethnicity. *Atherosclerosis* **2015**, *241*, 219–228. [CrossRef] [PubMed]

114. Battle, A.; Brown, C.D.; Engelhardt, B.E.; Montgomery, S.B.; GTEx Consortium; Laboratory, Data Analysis & Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; et al. Genetic effects on gene expression across human tissues. *Nature* **2017**, *550*, 204–213. [CrossRef] [PubMed]

115. Zheng, J.; Erzurumluoglu, A.M.; Elsworth, B.L.; Kemp, J.P.; Howe, L.; Haycock, P.C.; Hemani, G.; Tansey, K.; Laurin, C.; Pourcain, B.S.; et al. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **2017**, *33*, 272–279. [CrossRef] [PubMed]

116. Riley, V.; Erzurumluoglu, A.M.; Rodriguez, S.; Bonilla, C. Mitochondrial DNA Haplogroups and Breast Cancer Risk Factors in the Avon Longitudinal Study of Parents and Children (ALSPAC). *Genes (Basel)* **2018**, *9*. [CrossRef]

117. Lawson, D.J.; Davies, N.M.; Haworth, S.; Ashraf, B.; Howe, L.; Crawford, A.; Hemani, G.; Davey Smith, G.; Timpson, N.J. Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genet.* **2020**, *139*, 23–41. [CrossRef]

118. Erzurumluoglu, A.M.; Baird, D.; Richardson, T.G.; Timpson, N.J.; Rodriguez, S. Using Y-Chromosomal Haplogroups in Genetic Association Studies and Suggested Implications. *Genes (Basel)* **2018**, *9*. [CrossRef]

119. Makowsky, R.; Yan, Q.; Wiener, H.W.; Sandel, M.; Aissani, B.; Tiwari, H.K.; Shrestha, S. The utility of mitochondrial and y chromosome phylogenetic data to improve correction for population stratification. *Front. Genet.* **2012**, *3*, 301. [CrossRef]

120. Hemani, G.; Shakhbazov, K.; Westra, H.J.; Esko, T.; Henders, A.K.; McRae, A.F.; Yang, J.; Gibson, G.; Martin, N.G.; Metspalu, A.; et al. Detection and replication of epistasis influencing transcription in humans. *Nature* **2014**, *508*, 249–253. [CrossRef]

121. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef]

122. Hunt, S.E.; McLaren, W.; Gil, L.; Thormann, A.; Schuilenburg, H.; Sheppard, D.; Parton, A.; Armean, I.M.; Trevanion, S.J.; Flicek, P.; et al. Ensembl variation resources. *Database (Oxford)* **2018**, *2018*. [CrossRef]

123. Ramasamy, A.; Trabzuni, D.; Guelfi, S.; Varghese, V.; Smith, C.; Walker, R.; De, T.; Coin, L.; de Silva, R.; Cookson, M.R.; et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **2014**, *17*, 1418–1428. [CrossRef] [PubMed]

124. Westra, H.J.; Peters, M.J.; Esko, T.; Yaghootkar, H.; Schurmann, C.; Kettunen, J.; Christiansen, M.W.; Fairfax, B.P.; Schramm, K.; Powell, J.E.; et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **2013**, *45*, 1238–1243. [CrossRef] [PubMed]

125. Uhlen, M.; Bandrowski, A.; Carr, S.; Edwards, A.; Ellenberg, J.; Lundberg, E.; Rimm, D.L.; Rodriguez, H.; Hiltke, T.; Snyder, M.; et al. A proposal for validation of antibodies. *Nat. Methods* **2016**, *13*, 823–827. [CrossRef]

126. Speliotes, E.K.; Willer, C.J.; Berndt, S.I.; Monda, K.L.; Thorleifsson, G.; Jackson, A.U.; Lango Allen, H.; Lindgren, C.M.; Luan, J.; Magi, R.; et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **2010**, *42*, 937–948. [CrossRef]

127. Clark, A.G. The role of haplotypes in candidate gene studies. *Genet. Epidemiol.* **2004**, *27*, 321–333. [CrossRef] [PubMed]

128. Zhang, M.; Wang, C.C.; Yang, C.; Meng, H.; Agbagwa, I.O.; Wang, L.X.; Wang, Y.; Yan, S.; Ren, S.; Sun, Y.; et al. Epigenetic Pattern on the Human Y Chromosome Is Evolutionarily Conserved. *PLoS ONE* **2016**, *11*, e0146402. [CrossRef] [PubMed]