eGastroenterology

New horizons of human genetics in digestive diseases

Lanlan Chen , Guoyue Lv

Department of Hepatobiliary and Pancreatic Surgery, General Surgery Center, The First Hospital of Jilin University, Changchun, Jilin, China

Correspondence to
Professor Guoyue Lv;
lvgy@jlu.edu.cn

ABSTRACT

Recent studies have decoded the human Y chromosome sequencing with predominant precision and coverage, offering promising prospects in human genetics and clinical translation. Such an achievement is facilitated by third-generation sequencing technologies including Oxford Nanopore Technology and Pacific Biosciences, which can overcome the limitations of next-generation sequencing. In the context of digestive diseases, these advancements hold significant potential as they can help address the 'missing heritability' problem and detect various genomic variants in genetic association analyses, beyond single nucleotide polymorphisms, hoping to reveal 'major' genes for complex diseases. Besides, the completion of the Y chromosome enables research into sex-specific genetic effects on diseases and this knowledge can lead to sex-specific therapeutic targets and a better understanding of molecular mechanisms behind gender disparities. In summary, the recent decoding of the Y chromosome, coupled with third-generation sequencing, offers new opportunities to address heritability gaps, discover major disease genes and investigate sex-specific effects in digestive diseases, providing valuable insights for clinicians in delivering precise healthcare services.

It has been two decades since the initial assembly of the human Y chromosome sequence was first generated.[1] However, this assembly remains incomplete. Recent groundbreaking developments, as reported in two publications from *Nature*, have now led to the complete decoding of the highly complex Y chromosome, promising a revolution in the field of human genetics. The first paper is led by the T2T (telomere-to-telomere) consortium and presents a complete assembly of the Y chromosome from the HG002 benchmarking genome, which is called T2T-Y, and this assembly has been incorporated into the T2T-CHM13 assembly, generating a new assembly (T2T-CHM13+Y) for mapping XY samples;[2] the second one reveals that the arrangements of repetitive regions, as well as the number of genes, vary dramatically from man to man.[3] The biggest difference between these two studies is that the former uses HG002 cell lines to assemble the genome of chromosome Y, while the latter uses 43 male individuals with genetic diversity from the 1000 Genomes Project.

The first release of the human genome sequence is not a complete work in a strict sense as the sequences of telomere, centromere and sex chromosomes are still incomplete.[4] The significant advances in third-generation sequencing technologies have made deciphering these regions possible, where the ultra-long reads from Oxford Nanopore Technology can create long continuous sequences and HiFi (high-fidelity) reads from Pacific Biosciences can yield reads with high accuracy.[5] If combining ultra-long reads with HiFi reads, the assembly of highly repetitive genomic regions will be no problem, and the two abovementioned papers do combine them in assembling the sequence of chromosome Y. The third-generation sequencing technology conquers the shortcoming of low accuracy in the short reads from the next-generation sequencing (NGS) and should have promise in deciphering complex genomic regions, such as previously inaccessible structural variants, which will unveil the novel mechanism of diseases, especially those rare diseases.[6]

In this editorial, we will mainly discuss how the completion of human chromosome Y and the third-generation sequencing technology should enlighten the research field of digestive diseases. The completion of the human genome will comprehensively create a 'gapless' reference genome data set with high accuracy and coverage. What is more, using T2T-CHM13+Y as the reference might improve the detection of human contamination in genomic databases.[2] Along with the third-generation sequencing, the function of the human genome will be thoroughly decoded accurately and extensively.

One problem in genetics is the 'missing heritability'.[7] The 'heritability' is a statistic that refers to the proportion of the variation in a given phenotype within a certain population that can be explained by genetic variation. For example, the heritability of liver enzymes, which was estimated in a broad sense within twin families, ranged from 22% to 60%, but the heritability calculated based on single

nucleotide polymorphisms (SNPs) could usually be <2%.[8] A similar situation was observed in non-alcoholic fatty liver disease (NAFLD), where the broad-sense heritability of NAFLD was >20% but the SNP-based heritability was <10%.[9] Usually, the SNP-based heritability can be <10%, which is much smaller than the broad-sense heritability. The so-called 'missing heritability' can be accounted for by two aspects: (1) only SNPs cannot represent the genetic heritability as the human genome consists of other variants, such as copy number variants (CNV) and structural variants (SV); and (2) a series of biological processes are involved in transforming genotype to phenotype, such as transcriptome, proteome, epigenome and metabolome, and combining these omics information should improve heritability estimates. Therein, the first aspect can be settled down by the expansion of genetic markers (CNV, SV and other variants) in the genome-wide association study (GWAS).[10] Undoubtedly, third-generation sequencing will help find the 'missing heritability' using the complete human genome as the reference in further GWAS analysis.

Another important aspect is the discovery of novel mechanisms for complex diseases, especially for those without 'major' genes currently. In a recent publication, Mukamel et al[10] performed an extensive and comprehensive association analysis to estimate the association between the variable number of tandem repeats (VNTRs) with a wide range of human traits in the UK Biobank. This study unveiled that repeat polymorphism at *EIF3H* displayed a larger effect size than common SNPs in colorectal cancer, suggesting other genomic variants should play an important role in GWAS analysis as well. Although Mukamel et al[10] applied the statistical method they developed to quantify the VNTRs using the NGS data, the results can be improved with high accuracy if using third-generation sequencing and taking the complete human genome as the reference. A previous large-scale, exome-wide analysis failed to discover any 'major' gene for type 2 diabetes, and rare protein-coding variants explain much less phenotypic variance than common variants.[11 12] Thus, future GWAS analysis should not be limited to SNP–trait associations. It is probable that the 'major' genes for the complex disease should be discovered if including other types of genomic variants in the GWAS analysis, such as CNV and SV.

The last but most important aspect should be the sex-specific genetic effects on diseases. Now, with the completion of human chromosome Y, we can answer to what extent the sex differences of a disease are determined by genetics. In traditional SNP-based GWAS analysis, the genotype can be denoted as 0, 1 and 2 from the autosome, and either linear or logistic regression model will be applied to test the SNP–trait associations. However, analysing SNPs from the sex chromosomes is not the case, and it is the most difficult aspect because of the statistical challenges caused by X-inactivation uncertainty.[13] Thus, sex chromosomes are usually omitted in GWAS analysis, and it was reported that only 25%

of GWAS reported results for chromosome X and only 3% provided results for chromosome Y in the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog (https://www.ebi.ac.uk/gwas/).[14] The current study recommended estimating genetic effects in a sex-stratified fashion to construct a sex-specific polygenic risk score (PRS), similar to the population-specific PRS.[14] Besides, other statistical methods in genetics should be updated and developed to fit with the analysis of sex chromosomes, such as linkage disequilibrium score regression, colocalisation and Mendelian randomisation. If the genetic analysis is performed with sex stratification, the genetic effects of sex-specific genes will be more accurate.

Gender differences have been observed in many digestive diseases, and one of the most prominent diseases is autoimmune liver diseases, which consist of autoimmune hepatitis (AIH), primary biliary cholangitis (PBC) and primary sclerosing cholangitis (PSC). Therein, the proportion of women in AIH, PBC and PSC is 80%, 91% and 67%, respectively.[15] Currently, evidence suggests that the epigenetic changes of X chromosome-related genes, X chromosome instability and inactivation are important sex-related factors.[15] Besides, gender differences have been observed in NAFLD,[16] gastric cancer[17] and colorectal cancer.[18] However, the role of Y chromosome in gender differences is largely unknown. The completion of Y chromosome sequencing may help us to discover these sex-related molecular mechanisms in the future.

Currently, there are numerous GWAS on digestive diseases, but few of them reported the genetic associations of SNPs in chromosomes X and Y. This creates a gap in sex-specific effects between results from genetic analyses and those from traditional epidemiological studies. As can be seen, the genetically sex-specific effects usually shed light on sex-specific therapeutic targets and give possible explanations of the molecular mechanisms on the observed sex differences. For instance, many promising preventive therapies have been proposed for primary liver cancer, but the gender disparity should not be neglected as the activation of the sex-determining region on Y chromosome (*SRY*) can promote male-specific hepatocarcinogenesis.[19 20] Future GWAS analyses should reveal the associations of *SRY* region with hepatocellular carcinoma in male individuals. Thus, the sex-specific effects can be well explained if taking sex chromosomes into consideration in GWAS analyses.

Generally, together with third-generation sequencing, the completion of human chromosomes offers us unprecedented opportunities to fill the 'missing heritability', discover the 'major' genes for complex diseases and appraise the sex-specific effects in disease initiation, progression and prognosis. As clinicians, we are encouraged to grasp these new opportunities to answer scientific questions about digestive diseases and provide healthcare service with precision.

ORCID iD
Lanlan Chen http://orcid.org/0000-0002-1075-5592

REFERENCES

1 Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, *et al*. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003;423:825–37.
2 Rhie A, Nurk S, Cechova M, *et al*. The complete sequence of a human Y chromosome. *Nature* 2023;621:344–54.
3 Hallast P, Ebert P, Loftus M, *et al*. Assembly of 43 human Y chromosomes reveals extensive complexity and variation. *Nature* 2023;621:355–64.
4 Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. *Science* 2001;291:1304–51.
5 Marx V. Long road to long-read assembly. *Nat Methods* 2021;18:125–9.
6 Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet* 2020;21:597–614.
7 Yang J, Lee SH, Goddard ME, *et al*. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.
8 van Beek JHDA, Lubke GH, de Moor MHM, *et al*. Heritability of liver enzyme levels estimated from genome-wide SNP data. *Eur J Hum Genet* 2015;23:1223–8.
9 Sookoian S, Pirola CJ. Genetic predisposition in nonalcoholic fatty liver disease. *Clin Mol Hepatol* 2017;23:1–12.
10 Mukamel RE, Handsaker RE, Sherman MA, *et al*. Repeat polymorphisms underlie top genetic risk loci for glaucoma and colorectal cancer. *Cell* 2023;186:3659–73.
11 Flannick J, Mercader JM, Fuchsberger C, *et al*. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* 2019;570:71–6.
12 Weiner DJ, Nadig A, Jagadeesh KA, *et al*. Polygenic architecture of rare coding variation across 394,783 exomes. *Nature* 2023;614:492–9.
13 Keur N, Ricaño-Ponce I, Kumar V, *et al*. A systematic review of analytical methods used in genetic association analysis of the X-chromosome. *Brief Bioinform* 2022;23:bbac287.
14 Sun L, Wang Z, Lu T, *et al*. eXclusionarY: 10 years later, where are the sex chromosomes in GWASs? *Am J Hum Genet* 2023;110:903–12.
15 Schwinge D, Schramm C. Sex-related factors in autoimmune liver diseases. *Semin Immunopathol* 2019;41:165–75.
16 Lonardo A, Nascimbeni F, Ballestri S, *et al*. Sex differences in nonalcoholic fatty liver disease: state of the art and identification of research gaps. *Hepatology* 2019;70:1457–69.
17 Janulaityte-Günther D, Kupcinskas L, Pavilonis A, *et al*. Helicobacter pylori antibodies and gastric cancer: a gender-related difference. *FEMS Immunol Med Microbiol* 2005;44:191–5.
18 Choi Y, Kim N. Sex difference of colon adenoma pathway and colorectal carcinogenesis. *World J Mens Health* August 29, 2023.
19 Guillot A, Tacke F. Spatial dimension of macrophage heterogeneity in liver diseases. *Egastro* 2023;1:e000003.
20 Liu C, Ren Y-F, Dong J, *et al*. Activation of SRY accounts for male-specific hepatocarcinogenesis: implication in gender disparity of hepatocellular carcinoma. *Cancer Lett* 2017;410:20–31.