Journal of Orthopaedic Science xxx (2017) 1–5

Contents lists available at ScienceDirect

Journal of Orthopaedic Science

journal homepage: http://www.elsevier.com/locate/jos

Review Article

Human genetics contributes to the understanding of disease pathophysiology and drug discovery☆

Saori Sakaue [a, b], Yukinori Okada [a, *]

[a] Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan
[b] Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan

ARTICLE INFO

Article history:
Received 23 January 2017
Received in revised form
9 July 2017
Accepted 29 July 2017
Available online xxx

ABSTRACT

Background: Today, sequencing technology has markedly reduced the cost and time needed to read the human genome than ever before. Genome-wide association studies have successfully identified a number of disease risk genes.
Contribution to understanding of disease pathophysiology: Recent advancements in genomic technology have substantially furthered our understanding of the pathophysiology of many diseases, such as rheumatoid arthritis.
Toward drug discovery and future direction: Accumulating genomic information is now expected to accelerate the discovery of novel drugs. Rapidly growing multi-dimensional information in life sciences would make human genetics significantly important in the near future.

1. Introduction

To our surprise, today complete set of human genome from an individual can be read on only a thousand dollars [1], thanks to the recent technological advance in genomic science. The amount of information being compiled in life science fields is rapidly growing. Together with computational technology, we are expecting the upcoming of the Precision Medicine, genomic drug discovery or other application by using these genomic "Big Data." In this review, we summarize the basics of statistical genetics and highlight recent updates in the genomic research on rheumatoid arthritis (RA) and their future application in translational research.

2. Basics of statistical genomics

The human genome is the complete set of genetic information, preserved as DNA sequences in 23 chromosome pairs. DNA strands are composed of units called nucleotides, each containing one of the following four nucleobases: cytosine (C), guanine (G), adenine (A), and thymine (T). Approximately 0.01% of variations exist between the genomes of human individuals, and single-nucleotide polymorphism (SNP) is one of the well-studied variations. Statistical genetics converts the genomic information of each individual, including SNPs, into a digital code and analyzes its effect on a variety of phenotypes of each individual. A statistical association study in general has limitations such as an ambiguous temporal ordering of cause-and-effect relationships, whereas genomic information (cause) always precedes the phenotype (effect) in statistical genetics, which leads to reproducible results and high applicability.

The Human Genome Project was declared complete in 2003, and several international projects such as the HapMap Project [2] and 1000 Genomes Project [3–5] were then undertaken to accumulate information on vast amounts of reference human genomes from all around the world. Japan also has its own biobank that combines medical and genome information [6,7]. The genome-wide association study (GWAS)—a thorough analysis of the association of a specific disease or trait with tens of millions of genome-wide SNPs—has successfully identified a number of disease risk genes, and has led to the better understanding of the pathophysiology of many diseases. Recent development of human genome sequencing technologies, such as high-density SNP microarrays and next generation sequencers, have substantially improved the cost and time needed to read genome sequences. Statistical genetics is expected to be increasingly important not only because there are rapidly growing human genomic data thanks to the evolution of technology, but also because we are reaching a point where the

☆ This Review Article was presented at the 31st Annual Research Meeting of the Japanese Orthopaedic Association, Fukuoka, Oct. 13, 2016.

* Corresponding author. Fax: +81 6 6879 3975.
E-mail address: yokada@sg.med.osaka-u.ac.jp (Y. Okada).

http://dx.doi.org/10.1016/j.jos.2017.07.009
0949-2658/

analytical results can contribute to the clinical application. The driving force of genomic research is now making a transition from "how to read the sequence rapidly, economically, and accurately" to "how to meaningfully interpret genomic information both biologically and clinically."

3. GWAS elucidates disease pathophysiology

The goal of the genetic studies is not only to identify disease risk genes, but to clarify the underlying disease biology and to apply it to clinical field. Although comprehensive catalogs of genetic risk loci on various human traits have been identified, little is known about the means to integrate the findings of genetics research with diverse biological knowledge.

In RA, genetic factors have long been considered to contribute to disease pathogenesis, and an epidemiological study revealed that half of RA risk might be related to hereditary factors [8]. In 2014, Okada et al. performed the largest scale of GWAS on RA, analyzing more than hundreds of thousands of samples from various ethnic backgrounds and tens of millions of variants, which led to the identification of over 100 disease susceptibility loci, of which 42 were novel [9]. This discovery enabled exhaustive research on RA pathophysiology, involving the use of various biological database and resources. Many of the RA susceptibility loci were found to be common with type 1 diabetes, inflammatory bowel disease, and hematological malignancy, indicating that these diseases might share pathogenic backgrounds. The RA susceptibility loci were also analyzed together with epigenetic chromatin marks, and a significant enrichment of RA risk alleles with trimethylation of histone H3 at lysine 4 (H3K4me3) peaks in primary CD4+ regulatory T cells was observed. A pathway analysis database showed that RA susceptibility genes aggregated in T cells, B cells and cytokine-related pathway. These results implicated the contribution of regulatory T cells, B cells, and a specific type of cytokines in RA pathogenesis.

Another discovery using GWAS and Big Data analysis led to the revision of the "shared epitope hypothesis." Since the 1980s, *HLA-DRB1* has long been thought to be one of the strongest disease susceptibility genes in RA, and conserved amino acid sequences at positions 70–74 were thought to be essential to antigen recognition and pathogenesis of RA, thus termed as "shared epitope hypothesis." Recently, together with improved power of GWAS, a method called "human leukocyte antigen (HLA) imputation" was developed. This method computationally estimates HLA alleles of individuals by using reference panels of GWAS data from each ethnic background, thereby enabling the assessment of variants in HLA loci comprehensively and highly accurately without additional costs. Application of the HLA imputation method to seropositive RA revealed that the majority of the disease risk at MHC was explained by amino acid sequence polymorphisms at 11 and 13 positions of the HLA-DRβ1 molecule in multiple populations [10–13], which were different from those of shared epitope hypothesis. In addition to *HLA-DRB1*, this method identified disease risk loci in *HLA-A*, *HLA-B*, and *HLA-DPB1*, all of which are located in the peptide-binding grooves and might contribute to RA pathophysiology [10,11] (Fig. 1). A recent study also revealed that ACPA-positive RA was independently associated with a synonymous mutation in the non-classical HLA gene of *HLA-DOA*, which led to an expression quantitative locus (eQTL) effect on the *HLA-DOA* mRNA expressions, suggesting the dosage contribution of non-classical HLA genes to the disease biology for the first time [14].

4. Contribution of statistical genetics in drug discovery

In addition to elucidating disease pathogenesis, using genetic information in novel drug discovery was attempted for the first time in the abovementioned example of RA GWAS [9]. Okada et al. comprehensively analyzed important RA susceptibility genes using databases on existing drugs and information on their therapeutic targets. They were found to be strongly related to the targets of known approved RA medications, such as TNF-inhibitors or other biologics and also oral drugs, through protein–protein interactions (Fig. 2). This methodology also identified a non-RA drug, CDK4/6 inhibitor, which is currently approved for breast cancer, as a candidate for the treatment of RA. This was further supported by an experiment that demonstrated the effectiveness of a CDK4/6 inhibitor on animal models for RA [15,16]. Pharmaceutical companies are presently developing CDK4/6 inhibitors as a potential indication for RA.

Another example that exploits GWAS results for a novel drug development is a method called phenome-wide association study (PheWAS), which analyzes the association between genomic and phenotype information that is preserved in an electronic medical records [17,18]. PheWAS can help predict unexpected adverse effects of an investigatory drug that would otherwise remain unknown before clinical trials. Diogo et al. conducted PheWAS on a population with a loss of function on *TYK2* gene, which is a target of tofacitinib, an approved drug for RA [19]. PheWAS successfully showed that people with variants on *TYK2* gene were unlikely to have autoimmune diseases, such as RA, systemic lupus erythematosus, and inflammatory bowel disease, and likely to be susceptible to pneumonia and increased LDL-cholesterol levels. This was in concordance with the post-marketing surveillance of tofacitinib that revealed its therapeutic effect on patients with RA and showed an increased risk of respiratory infection and increased LDL-cholesterol levels.

5. Future direction

Although drug discovery is indispensable to future medical progress, pharmaceutical companies are facing a productivity crisis because of increased cost and decreased success rate. In general, $1.8 billion and over 10 years of research are necessary for developing a single novel drug. Unexpected side effects after clinical trials and marketing can lead to the withdrawal of the drug and

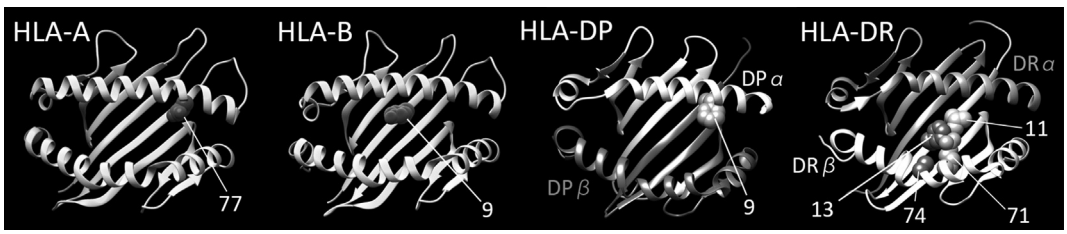

Fig. 1. Three-dimensional representation of RA risk amino acid positions. Key amino acid positions are highlighted as spheres. We used Protein Data Bank entries 1x7q (HLA-A), 2bvp (HLA-B), 3lqz (HLA-DP) and 3pdo (HLA-DR) with UCSF Chimera to prepare the figure.

Please cite this article in press as: Sakaue S, Okada Y, Human genetics contributes to the understanding of disease pathophysiology and drug discovery, Journal of Orthopaedic Science (2017), http://dx.doi.org/10.1016/j.jos.2017.07.009