Review

Genetic variations in medical research in the past, at present and in the future

By Yoichiro Kamatani*1 and Yusuke Nakamura*2,†

(Edited by Kohei Miyazono, M.J.A.)

Abstract: As we look so different, our genomic sequences vary enormously. The differences in our genome, genetic variations, have played very significant roles in medical research and have contributed to improvement of medical managements in the last 2–3 decades. Genetic variations include germline variations, somatic mutations, and diversities in receptor genes of rearranged immune cells, T cells and B cells. Germline variants are in some cases causative of genetic diseases, are associated with the risk of various diseases, and also affect drug efficacies or adverse events. Some somatic mutations are causative of tumor development. Recent DNA sequencing technologies allow us to perform single-cell analysis or detailed repertoire analysis of B and T cells. It is critically important to investigate temporal changes in immune environment in various anatomical regions in the next one to two decades. In this review article, we would like to introduce the roles of genetic variations in medical fields in the past, at present and in the future.

Keywords: genetic variations, reverse genetics, biobank, genome wide association study (GWAS), pharmacogenomics, immunogenomics

Introduction

We have known for many years that some diseases are genetically determined, and the susceptibility to most of diseases or our phenotypes such as height or blood types are heritable. Genetic variations used/applied in genetic and medical researches are classified into five classes, RFLP (restriction fragment length polymorphism), VNTR (variable number of tandem repeat), STR (short tandem repeat or microsatellite), SNP (single nucleotide polymorphism), and CNV (copy number variation), and other SVs (structural variations) have also been reported.

1. Genetic variations

History of studies in genetic variations is not so long. The first DNA-based genetic variation was reported by Kan and Dozy in 1978.1) They found a polymorphism of DNA sequence adjacent to human β-globin gene. Until the PCR (polymerase chain reaction) method was established, DNA polymorphisms were detected by a combination of DNA restriction enzymes and Southern blotting method. When genetic variations are present in DNA sequences recognized by any restriction endonuclease, the lengths of DNA fragments digested with these enzymes become different individual by individual and can be visualized by Southern blotting and subsequent hybridization with 32P-labeled DNA fragments. However, most of restriction fragment length polymorphism (RFLP) markers were bi-allelic and their heterozygosities was not so high as 10–50%. Before the development of PCR, RFLP was detected by Southern analysis and required a large amount (about 5µg) of non-degraded genomic DNAs. In 1985, Jeffreys et al. reported our genome has the “minisatellite” regions that are highly polymorphic and were used as DNA fingerprinting markers for

---

*1 Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.
*2 Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan.
† Correspondence should be addressed: Y. Nakamura, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, 3-8-31 Ariake, Koto-ku, Tokyo 135-8550, Japan (e-mail: yusuke.nakamura@jfcr.or.jp).

Abbreviations: RFLP: restriction fragment length polymorphism; VNTR: variable number of tandem repeat; STR: short tandem repeat or microsatellite; SNP: single nucleotide polymorphism; CNV: copy number variation; SV: structural variation; PCR: polymerase chain reaction.
forensic purpose such as individual identification and paternity/maternity tests.\(^2\) Since these minisatellite markers were difficult to examine chromosomal segregation, we attempted to isolate single-locus variable number of tandem repeat (VNTR) markers and reported in 1987.\(^3\) The first VNTR marker was reported by Wyman and White in 1980,\(^4\) but systematic VNTR screening was done after the discovery of “minisatellite” loci. Genetic linkage maps with VNTR markers and hundreds of RFLP markers reported by White and his colleagues made it possible to perform reverse genetics approaches for genetic diseases of unknown causes or with no biomarkers. VNTRs markers were also applied in forensic studies\(^5\) and used in the clinic to monitor transplanted cells in recipients having bone marrow transplants.\(^6\)

VNTRs also contributed to discovery of loci of tumor suppressor genes; both alleles of the tumor suppressor genes were inactivated (the two-hit mutation model) and in many cases the inactivation of one allele occurs as the loss of a whole or a part of chromosome (loss of heterozygosity; LOH) including a tumor suppressor gene. The first systematic analysis of loss of all chromosomal arms in colorectal cancer tissues using a set of VNTR markers was performed by Vogelstein and his colleagues.\(^7\) The results led to establish the multi-step carcinogenesis model of human cancers\(^8\) and subsequent analysis of a short arm of chromosome 17 proved the p53 gene as a tumor suppressor gene,\(^9\) which is now shown to be mutated in nearly half of all human cancers.

2. Reverse genetics

The usefulness of RFLP markers in medical research was first suggested by Botstein and colleagues in 1980.\(^10\) RFLP and VNTR show co-dominant Mendelian inheritance and can distinguish the parental alleles (a maternal or paternal origin) of the particular chromosomal loci in our genome. These markers made reverse genetics possible by examining co-segregation of each polymorphic “marker” allele with the inheritance of any genetic diseases with Mendelian inheritance. For example, Gusella et al.\(^11\) identified the genetic locus for Huntington disease to chromosome 4.\(^11\) The development of widely-available DNA polymorphic markers in late 1980s, genetic loci responsible for genetic diseases including cystic fibrosis,\(^12\) multiple endocrine neoplasia type 1 (MEN1),\(^13\) Friedreich’s ataxia,\(^14\) ataxia telangiectasia,\(^15\) familial polyposis coli,\(^16\) neurofibromatosis type 1,\(^17\) and familial breast cancer,\(^18,19\) had been discovered and subsequently their responsible genes were identified although it took a few to several years from the mapping of the disease loci to the discovery of responsible genes. In fact, we identified the locus for familial polyposis coli to a long arm of chromosome 5 in 1987\(^16\) and isolated its responsible APC (adenomatous polyposis coli) gene in 1991.\(^20\) These reverse genetics approach proved that when DNAs from the family members are available, the linkage analysis with polymorphic markers is a very powerful tool to discover responsible genes for genetic diseases in any inheritance models even if no knowledge for the biological or biochemical mechanisms (or abnormalities) is obtained.

3. PCR and microsatellite markers

The development of polymerase chain reaction (PCR) method coupled with isolation of thermostable DNA polymerase revolutionized the speed of DNA analysis. After 1990, scientists introduced microsatellite analysis based on the PCR technology. Microsatellite markers, which are repetitive DNA segments of two to several base-pairs,\(^21\) show high levels of heterozygosity and are present at >100,000 regions covering most of our genome. Because it requires a very small amount (~10 ng) of DNAs and is applicable to the high-throughput PCR system, microsatellite analysis accelerated linkage analysis or population genetics. Furthermore, even if DNAs are degraded at some extent (for example, DNAs from FFPE samples), microsatellite loci can be still analyzed.

One of the advantages of the approach using microsatellite markers is that they can be applied to homozygosity mapping of recessive diseases with very low incidence\(^22\): for example, when patients are born by marriage of relatives or individuals living in the same area for a long period, only several patients are sufficient enough to map a gene responsible to a genetic disease. In fact, we applied this approach and successfully mapped the loci and later identified genes responsible for Fukuyama muscular dystrophy\(^23,24\) and gelatinous drop-like corneal dystrophy.\(^25,26\)

Another advantage of microsatellite markers is for transmission disequilibrium test (TDT) to search genetic loci associated with common diseases. One example of the successes through this approach was shown by Onouchi et al.\(^27\) using 78 Japanese sib pairs with Kawasaki disease; we demonstrated that an SNP in the ITPKC gene was the risk factor for the
disease and associated with response to the intravenous immunoglobulin therapy.

4. SNP and international HapMap project

Around year 2000 when we expected the completion of human genome sequences, many scientists planned to construct a high-density single nucleotide polymorphism (SNP) map with millions of SNP markers covering an entire genome. SNPs are the most common genetic variations (one in 300 base pairs or less) in our genome. On the basis of the hypothesis of “common variant–common disease” association, SNPs were expected to be extremely informative as a tool for population genetics to identify genes (genetic variations) susceptible to various common diseases.

In Japan, the Japanese government announced a 5-year “Japanese Millennium Project” in 2000, which included the construction of Japanese SNP database and establishment of a high-throughput SNP analysis system for genome-wide association study. We have re-sequenced two dozens of Japanese individuals in a total of approximately 154-Mb of exonic segments and found more than 200,000 genetic variations. In addition, we established a high-throughput SNP genotyping system by a combination of multiplex PCR reactions (simultaneous amplification of 96 genomic segments) with Invader assay method. We then constructed the JSNP database including the information of nearly 190,000 SNPs and their allelic frequencies in 768 Japanese individuals in 2002 (the largest SNP database at that time).28

Since multiple large-scale SNP genotyping platforms were developed around 2003, construction of the international consortium for making the SNP (Haplotype) database for three populations had been discussed and finally the international HapMap project including six countries (Canada, China, Japan, Nigeria, U.K., and U.S.A.) was organized in 2003. The aims of this consortium were clearly discussed and finally the international HapMap project including six countries (Canada, China, Japan, Nigeria, U.K., and U.S.A.) was organized in 2003. The aims of this consortium were clearly described in a paper published in Nature in 2003 as follows; “The goal of the International HapMap Project is to determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain. An international consortium is developing a map of these patterns across the genome by determining the genotypes of one million or more sequence variants, their frequencies and the degree of association between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe. The HapMap will allow the discovery of sequence variants that affect common disease, will facilitate development of diagnostic tools and will enhance our ability to choose targets for therapeutic intervention”.29

After the very extensive efforts of participating scientists, the consortium constructed a database consisting of more than one million SNPs in 200530 and reported an extended set in 2007.31 Although the majority of genetic variations were commonly shared among three major populations, a small subset of variations was detected in one particular population. Among them, an SNP in the ABCC11 gene that was uniquely found in the Asian population was later shown to be a determinant of human earwax type.32 This kind of data suggested that if they have any effects on the quality or quantity of the gene product, which evolutionally fits to the certain environmental conditions, SNPs uniquely observed in a certain population might be determinants to characterize them.

5. Biobanks and GWAS

Although there were many criticisms and skepticisms for the “common variation-common diseases” approach for identifying genes (genetic variants) susceptible to common diseases at the beginning of the international HapMap project, many published papers have demonstrated the usefulness of genome wide association study (GWAS) to uncover various genetic factors associated with various diseases including “uncommon diseases”. Although not widely recognized, we started in 2001 whole-genome association studies using nearly 90,000 SNPs that well covered regions containing genes and identified a gene susceptible to myocardial infarction,33 which is the first successful GWAS report in the world.34 We also begun a BioBank Japan project in 2003 and have so far collected DNA samples from nearly 440,000 cases having either of 51 diseases. In combination with advanced technologies and the BioBank Japan samples, we have reported many candidate genes/loci possibly associated with susceptibility to diseases as summarized in Table 1. In other countries, systematic GWAS studies started in 2005 after the international HapMap project offered a large set of SNP information with the development of cheap, commercially-available, and accurate high-throughput SNP typing platforms. These technical and scientific advancement made GWAS a genetic analysis tool in the world, and the number of published GWAS reports exceeded more
Table 1. Genes susceptible to various diseases reported from our centers (Center for Genomic Medicine/Integrative Medical Sciences, RIKEN and Human Genome Center, Institute of Medical Science/Graduate School of Frontier Sciences, The University of Tokyo)

| Trait                                           | Association                                                   | Reference                                      |
|-------------------------------------------------|--------------------------------------------------------------|-----------------------------------------------|
| Adolescent idiopathic scoliosis                 |                  | Takahashi. Nat. Genet. 2011;43(12):1237–1240.               |
|                                                  | LBX1, GPR126, BNC2  | Kon. Nat. Genet. 2013;45(6):670–679.                        |
| Age-related macular degeneration                |                  | Ogura. Am. J. Hum. Genet. 2015;97(2):337–342.                |
|                                                  | TNFRSF10A, 4q12  | Arakawa. Nat. Genet. 2011;43(10):1001–1004.                  |
| Atopic dermatitis                               |                  | Hirota. Nat. Genet. 2012;44(11):1222–1226.                   |
| Atrial fibrillation                             |                  | Low. Nat. Genet. 2017;49(6):953–958.                         |
| Body mass index                                 |                  | Okada. Nat. Genet. 2012;44(3):302–306.                       |
|                                                  | CDKAL1, KLF9, 112 new loci                                   | Aktyama. Nat. Genet. 2017;49(10):1458–1467.               |
| Bronchial asthma                                |                  | Hirita. Nat. Genet. 2011;43(9):893–896.                      |
| Cataract                                        |                  | Ishigaki. Nat. Genet. 2020;52(7):669–679.                    |
| Cerebral aneurysm                               |                  | Ishigaki. Nat. Genet. 2020;52(7):669–679.                    |
| Chronic kidney disease                         |                  | Ishigaki. Nat. Genet. 2020;52(7):669–679.                    |
| Chronic obstructive pulmonary disease (COPD)    |                  | Ishigaki. Nat. Genet. 2020;52(7):669–679.                    |
| Duodenal ulcer                                  |                  | Tanikawa. Nat. Genet. 2012;44(4):430–434, S1–S2.             |
| Endometriosis                                   |                  | Uno. Nat. Genet. 2010;42(8):707–710.                         |
| Gastric cancer                                  |                  | Ishigaki. Nat. Genet. 2020;52(7):669–679.                    |
| Graves’ disease                                 |                  | Ishigaki. Nat. Genet. 2020;52(7):669–679.                    |
| HCV-positive hepatocellular carcinoma           |                  | Kumar. Nat. Genet. 2011;43(5):455–458.                       |
| Hepatitis B                                      | HLA, 6q27        | Kamatani. Nat. Genet. 2009;41(5):591–595.                    |
| Hepatocellular carcinoma                       | DEPDC5, IFNL2-IFNL3 | Miki. Nat. Genet. 2011;43(8):797–800.                       |
| Interstitial lung disease                       |                  | Ishigaki. Nat. Genet. 2020;52(7):669–679.                    |
| Kawasaki disease                                |                  | Onouchi. Nat. Genet. 2008;40(1):35–42.                       |
| Keloid                                          |                  | Onouchi. Nat. Genet. 2012;44(5):517–521.                     |

Continued on next page.
Continued.

**Osteoarthritis**

- *GDF5*  
  Miyamoto. Nat. Genet. 2007;39(4):529–533.

- *DVWA*  
  Miyamoto. Nat. Genet. 2008;40(8):994–998.

**Lumbar disc herniation**

- *CILP*  
  Seki. Nat. Genet. 2005;37(6):607–612.

- *COL11A1*  
  Mio. Am. J. Hum. Genet. 2007;81(6):1271–1277.

- *THBS2*  
  Hirose. Am. J. Hum. Genet. 2008;82(5):1122–1129.

**Lung adenocarcinoma**

- *TERT, TP63*  
  Miki. Nat. Genet. 2010;42(10):893–896.

**Lung cancer**

- *POT1*  
  Ishigaki. Nat. Genet. 2020;52(7):669–679.

**Myocardial infarction/Coronary artery disease**

- *LTA*  
  Ozaki. Nat. Genet. 2002;32(4):650–654.

- *LGALS2*  
  Ozaki. Nature. 2004;429(6987):72–75.

- *PSMA6*  
  Ozaki. Nat. Genet. 2006;38(8):921–925.

- *BRAP*  
  Ozaki. Nat. Genet. 2009;41(3):329–333.

- 35 new loci  
  Koyama. Nat. Genet. 2020;52(11):1169–1177.

**Ossification of the posterior longitudinal ligament of the spine (OPLL)**

- 5 new loci  
  Nakajima. Nat. Genet. 2014;46(9):1012–1016.

**Osteoporosis**

- *STK39*  
  Ishigaki. Nat. Genet. 2020;52(7):669–679.

**Prostate cancer**

- 5p15, *GPRC6A-RFX6, 13q22, FOXP4*  
  Takata. Nat. Genet. 2010;42(9):751–754.

- 11q12, 10q26, 3p11.2  
  Akamatsu. Nat. Genet. 2012;44(4):426–429, S1.

- 12 new loci  
  Takata. Nat. Commun. 2019;10(1):4422.

**Rheumatoid arthritis**

- *PADI4*  
  Suzuki. Nat. Genet. 2003;34(4):395–402.

- *SLC22A4*  
  Tokuhiro. Nat. Genet. 2003;35(4):341–348.

- *FCRL3*  
  Kochi. Nat. Genet. 2005;37(5):478–485.

- *CD244*  
  Suzuki. Nat. Genet. 2008;40(10):1224–1229.

- *CCR6*  
  Kochi. Nat. Genet. 2010;42(6):515–519.

- 9 new loci  
  Okada. Nat. Genet. 2012;44(5):511–516.

- 42 new loci  
  Okada. Nature. 2014;506(7488):376–381.

**Stroke**

- *PRKCH*  
  Kubo. Nat. Genet. 2007;39(2):212–217.

**Type 2 diabetes**

- *WNT5B*  
  Kanazawa. Am. J. Hum. Genet. 2004;75(5):832–843.

- *KCNQ1*  
  Unoki. Nat. Genet. 2008;40(9):1098–1102.

- 7 new loci  
  Imamura. Nat. Commun. 2016;7(1):10531.

- 28 new loci  
  Suzuki. Nat. Genet. 2019;51(3):379–386.

- 7 new loci  
  Ishigaki. Nat. Genet. 2020;52(7):669–679.

**Ulcerative colitis**

- *FCGR2A, 13q12, SLC26A3*  
  Asano. Nat. Genet. 2009;41(12):1325–1329.

**Urolithiasis**

- *STIM1*  
  Ishigaki. Nat. Genet. 2020;52(7):669–679.

**Uterine fibrosis**

- 3 new loci  
  Cha. Nat. Genet. 2011;43(5):447–450.
than 500 by the end of 2009 and reached 3567 in 2018.\textsuperscript{35}

After the initial wave of GWAS, the importance of bigger genetic data was soon recognized, and the international collaborative consortia were organized to combine internationally existing samples. Because linkage disequilibrium (LD) structures are different among populations, this kind of meta-analysis can be simply done only if the ethnicity is similar among the participated groups. However, more sophisticated analysis to combine trans-ancestry samples could contribute to an increase of the statistical power for discovery and finer causative variant mapping.\textsuperscript{36}

For example, trans-ancestry analyses conducted by AFGen consortium for atrial fibrillation\textsuperscript{37} or MEGASTROKE consortium for stroke\textsuperscript{38} took this strategy, and identified 97 atrial fibrillation loci or 32 stroke loci, respectively. The BioBank Japan project substantially contributed to both studies as the largest single non-European sample set.

These efforts aggregately led to the great accomplishments for understanding the complex disease-related genomics. Firstly, GWAS identified more than hundreds or thousands of susceptibility loci; for example, we found 1,407 loci for medical tests\textsuperscript{39} and 276 loci for diseases\textsuperscript{40} using BioBank Japan data set, which has supported the validity of polygenic model hypothesis proposed about 100 years ago.\textsuperscript{41} Secondly, many polygenic susceptibility variants were shown to be enriched for non-coding regulatory DNA elements,\textsuperscript{42–44} along with experimental evidence showing that GWAS variants alter the regulatory role of enhancers,\textsuperscript{45,46} indicating that the genetics of complex traits is mainly governed by the diversity in gene regulation system among different individuals. This further enables cell-type specificity analysis using only genomic data; we found the enrichment of obesity-related loci in a B-cell specific active enhancer region.\textsuperscript{47} Lastly, although the effect sizes of individual variants are modest or small, the cumulative effect of thousands of variants (polygenic score; PGS) was shown to have the predictive ability equivalent to a monogenic driver variant.\textsuperscript{48}

However, there is a serious bottleneck in utilizing GWAS results into clinic. PGS will not predict so accurately when the ethnicities of target populations used for GWAS are different.\textsuperscript{49} This predictivity bias may happen even when the geographical regions within Japan are different.\textsuperscript{50,51} The fact that \textasciitilde97% of GWAS participants are from European ancestry while they make up only 16% of the global population hampers the clinical utility of PGS. To alleviate this Eurocentric bias, non-European large genomic data sets are imperative. Ongoing world-wide efforts including the BioBank Japan and other biobanks are trying to address this problem (see Koyama \textit{et al.}\textsuperscript{52} for example).

6. Pharmacogenomics

It is now widely recognized that germline variations/somatic mutations in genes involved in drug metabolisms, drug transports, drug receptors, and drug target/downstream molecules, influence the effectiveness and adverse reaction risks of various drugs.

Adverse drug reactions (ADRs) can be classified simply into two groups, one that can be explained by the mode of therapeutic drug actions and the other for which little or no information was available for their underlying mechanisms. Typical examples in the former cases are leukocytopenia, nausea/vomiting, and hair loss caused by cytotoxic anti-cancer drugs as well as bleeding in brain or intestines caused by anticoagulants. Representative examples for the latter one are severe drug eruption and drug-induced liver injury caused by various drugs although risk factors for some of them have been identified in the last two decades. Millions of patients suffer from severe ADRs worldwide and a small subset of patients lose their lives. Since genetic variations are known to be one of the important factors among multiple factors involved in the etiology of ADRs, we are able to apply such genetic factors that increase risks of ADRs to improve medical management for patients having the high risk for ADRs. The improved management should also contribute to significant reduction of the unnecessary medical costs (one example for warfarin in https://core.ac.uk/display/6665518).

In the last decade, extensive research challenges have been taken to uncover genetic factors underlying ADRs and have successfully identified the genetic variations that increase risks of ADRs. Genetic biomarkers that have been so far clarified are listed in “Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels” in U.S. Food and Drug Administration homepage (https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling). Proteins encoded by listed genes in the websites are classified mainly into three major groups.

Group 1: Drug metabolizing enzymes and drug transporters related to pharmacokinetics.
Group 2: Proteins related to pharmacodynamics.

Group 3: Human leukocyte antigens (HLAs).

It is obvious that poor clearance of drugs from our body results in elevation of drug concentration, leading to the toxic events; for example, genetic variants in UDP glucuronosyltransferase 1A1 (UGT1A1) that metabolizes irinotecan were shown to an association with myelotoxicity.53) Similarly, inactivating variants of CYP2C9, which metabolizes warfarin, increase the risk of bleeding.54) Furthermore, as an example of group 2, genetic variations in VKORC1 involved in vitamin K recycling influence pharmacodynamics of warfarin and define the necessary dose of warfarin in individual patients.54)

For life-threatening drug skin adverse reactions of unknown causes, drug-HLA interactions have been proven to be triggers for severe immune reactions attacking to skin cells; firstly, an HLA-B*1502 allele was shown to be very strongly linked to serious dermatologic reactions such as toxic epidermal necrolysis (TEN) and Stevens-Johnson syndrome (SJS) in patients taking carbamazepine.55) Following their discovery, we found a strong association of an HLA-B*3505 allele with nevirapine-induced skin adverse drug reactions in Thai patients.56) Although the exact molecular mechanisms how HLA molecules can be a trigger to induce such serious immune reactions are unclear, we and others suspect that a certain drug (or drug metabolite) binds to a certain HLA (or an HLA-peptide complex), is recognized as the presence of a non-self molecule, activates cytotoxic CD8+ T cells that can recognize the HLA-antigen complex, and then cause deleterious reactions.

In addition, some cases that influence the effectiveness of drugs were reported. A good example (although still in a big debate) is genotypes of cytochrome P450 2D6 (CYP2D6) and clinical outcome of breast cancer patients treated with tamoxifen. It is in an agreement that CYP2D6 metabolizes tamoxifen to its anti-estrogenic metabolites, 4-hydroxytamoxifen and endoxifen, and patients with some CYP2D6 variants show the significantly lower plasma concentration of endoxifen. Higher incidences of recurrence after breast cancer surgical treatment were observed in patients with null or low-activity genotypes in Caucasian and Japanese populations when patients received the tamoxifen monotherapy.57),58) However, some groups, which included the cases treated with hormone therapy in combination of chemotherapy, insisted to conclude no association between CYP2D6 genotypes and the outcome of tamoxifen treatment.59) Since patients need to take tamoxifen for 5–10 years, we are confident that the genetic diagnosis for tamoxifen treatment should improve the prognosis of breast cancer patients.

Although a large amount of information in the pharmacogenomics field has been accumulated, it is still limited in effective uses in clinical setting. Avoidance of ADRs certainly makes the quality of patient’s quality of life better and reduce the unnecessary medical costs in any countries. We need to construct an international network to further accelerate the application of useful pharmacogenomics information.

7. Immunogenomics and immunopharmacogenomics

As mentioned above, pharmacogenomics is a research area that combine pharmacology with genomics to examine the role of genetic variations (including somatic mutations) in drug responses. “Immunogenomics” is defined as a research filed which examines genetic variations in human leukocyte antigen (HLA) molecules or expression levels of immune-related genes in certain disease conditions. Due to the very rapid advancement in genomic analysis, we have been proposing the research area of “immunopharmacogenomics (IPG)” which further investigate the detailed immune responses including immunorepertoire changes at genomic levels in combination with pharmacological responses, as defined in a book entitled “Immunopharmacogenomics”.60) Particularly, T and B lymphocytes, which have receptors, the T cell receptor (TCR) and B cell receptor (BCR), respectively, are well known to play essential roles in the adaptive immune system and are critically important for inducing the various immunological reactions. T and B lymphocytes need to prepare for responding to exposures to pathogens and chemicals with a huge diversity. Considering the extremely-high complexity of the immune cells and responses, a comprehensive approach to fully characterize chronological changes (also may be at anatomical locations) of the repertoires of T-cell and B-cell receptors according to the disease progression/remission are urgently and essentially required for better understanding of roles of immune cells and their chronological changes in drug responses as well as various disease or allergic conditions including anaphylaxis caused by vaccinations.

During the differentiation of lymphocytes, the genes encoding TCRs and BCRs, which include
variable (V), diversity (D) (not for TCR α), and joining (J) exon segments, need to go through the essential biological process called “rearrangement” to generate functional receptors for huge numbers of pathogens. Extremely-high diversity of TCR and BCR repertoires can be generated through combinations of each of V, (D), and J exon segments with dozens of distinct segments as well as through the template-independent insertion and deletion of nucleotides at rearranged junctions. The high variable complementary determining region 3 (CDR3) regions, which are critically important to determine the specificity and affinity for the antigen recognition by T cells, was generated by the combination of V-(D)-J rearrangement and insertion/deletion at the junction sites. The complexity of the TCR repertoire is still not fully clarified, but it is estimated to be approximately $10^{15–18}$ as a heterodimer of the TCR α and β chains in human. In addition, the higher diversity of the BCR repertoire is also well known because of the much higher somatic hyper-mutation processes in the BCR gene.

Considering the importance of immunogenomics analysis in coming years, we developed a systematic, accurate and unbiased analysis of TCR and BCR transcripts using the next generation sequencing technology and proposing their application in medical research areas as shown in Fig. 1. Firstly, further genome-wide and immunogenomics analyses of drug-induced adverse reactions (skin hypersensitivity and liver injuries) can find the risk genetic factors to cause these reactions. In fact, we reported the two genetic factors to increase the risk of severe skin hypersensitivity; an HLA-B*3505 allele for nevirapine-induced skin adverse drug reactions in Thai HIV-infected patients as well as an HLA-A*3101 allele for carbamazepine-induced skin adverse drug reactions in Japanese population. Drug-induced liver injuries were also indicated their associations with HLA molecules. Secondly, the immunoreper-
toire analysis may help better understanding of the acute and chronic graft-versus-host disease (GVHD) and graft-versus-leukemia (GVL) effect of bone marrow (or cord blood) transplantation or tissue rejection after organ transplantation, leading to development of novel therapies. We reported that T cell repertoire analysis may predict the occurrence of GVHD in weeks earlier than when patients show the clinical symptoms.\textsuperscript{63} Thirdly, it is obvious that we can analyze how individual immune cells play key roles in autoimmune diseases and food allergy (or allergy to chemicals). We have analyzed the BCR repertoires in Kawasaki disease patients before and after globulin treatment\textsuperscript{64} as well as peanut allergy patients who received de-sensitization,\textsuperscript{65} and found the drastic changes of BCR repertoires during the treatment. Fourthly, immunogenomics analysis (immune signature analysis) in combination with somatic mutation analysis in tumors may predict the efficacy of not only immunotherapy but also chemotherapy or radiation therapy.\textsuperscript{66} In addition, through the BCR analysis we interestingly identified clonal expansion of plasma cells in a pseudo-progression tumor in a patient who received the immune checkpoint inhibitors.\textsuperscript{67} Fifthly, immune repertoire drastically changes during the aging and should be important to understand various aging-related changes such as the severity of COVID-19 infection. Lastly, it is undoubtedly essential for investigating immune responses after vaccination, particularly anaphylactic reactions that occurred immediately after the vaccination.

8. Precision medicine

In 2015, Barack Obama, the former president of U.S., announced a new research initiative called “Precision Medicine Initiative (PMI)”.\textsuperscript{68} The main aim was to take individual differences into account to improve health; the concept of “an appropriate dose of a right drug to a right patient” was not new and had been called as “personalized medicine”. The increasing availability of the advanced technologies including genomics, proteomics, metabolomics and other areas will encourage the realization of not only cancer precision medicine, but also of the whole range of health and medical cares including prevention of most of life style-related diseases.

It is not surprising to think that one of the scientific bases of PMI should be the genetic variation-aware personalized health care, and the utilization of PGS should be the promising implementation. However, the lack of portability of PGS among different populations needs to be addressed first. At present, the implementation of PGS will offer a far better precision to the subjects from European ancestry than those from the other populations, but scientists in the world need to consider this issue seriously as we previously had the good spirit to deliver the SNP information to all three populations in the international HapMap project. There would be several possibilities to overcome it as the ongoing international efforts of whole genome sequencing of the world-wide population would be crucial to find out the clues and resolve this issue. We should explore the road to the sustainable implementation of precision medicine based on genetic variations to all the people on the earth.

9. Conclusion

We here described briefly the history of human genetic variations as well as their significance in life science, particularly in medical genetics, pharmacogenetics and immunogenomics. One of the major goals in life science is to improve our quality of lives anywhere in the world. We believe that understanding of the molecular mechanisms causing diseases is essential and critical for the achievement of our goals. Genetic variations have played and will play very important roles in our medical and health cares. However, we must stress that we should pay much more efforts to avoid genetic discrimination. We need to teach that as genetic variations tell us, we are different, but we are equal and should respect the differences each other.

References

1) Kan, Y.W. and Dozy, A.M. (1978) Polymorphism of DNA sequence adjacent to human \(\beta\)-globin structural gene: Relationship to sickle mutation. Proc. Natl. Acad. Sci. U.S.A. \textbf{75}, 5631–5635.
2) Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hypervariable ‘minisatellite’ regions in human DNA. Nature \textbf{314}, 67–73.
3) Nakamura, Y., Leppert, M., O’Connell, P., Wolff, R., Holm, T., Culver, M. \textit{et al.} (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science \textbf{235}, 1616–1622.
4) Wyman, A.R. and White, R. (1980) A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. U.S.A. \textbf{77}, 6754–6758.
5) Odelberg, S.J., Plaetke, R., Eldridge, J.R., Ballard, L., O’Connell, P., Nakamura, Y. \textit{et al.} (1989) Characterization of eight VNTR loci by agarose gel electrophoresis. Genomics \textbf{5}, 915–924.
6) Gatti, R.A., Nakamura, Y., Nussmeier, M., Susi, E., Shan, W. and Grody, W.W. (1989) Informativeness of VNTR genetic markers for detecting
chimerism after bone marrow transplantation. Dis. Markers 7, 105–112.

7) Vogelstein, B., Fearon, E.R., Kern, S.E., Hamilton, S.R., Preisinger, A.C., Nakamura, Y. et al. (1989) Allelotype of colorectal carcinomas. Science 244, 207–211.

8) Vogelstein, B., Fearon, E.R., Botstein, D., Hamilton, S.R., Preisinger, A.C., Jessup, J.M. et al. (1989) Genetic alterations during colorectal-tumor development. N. Engl. J. Med. 319, 527–532.

9) Baker, S.J., Fearon, E.R., Nigro, J.M., Hamilton, S.R., Preisinger, A.C., Leppert, M. et al. (1989) Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science 244, 217–221.

10) Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314–331.

11) Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E. et al. (1983) A polymorphic DNA marker genetically linked to Huntington’s disease. Nature 306, 234–238.

12) Tsui, L.C., Buchwald, M., Barker, D., Braman, J.C., Knowlton, R., Schum, J.W. et al. (1985) Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. Science 230, 1054–1057.

13) Larsson, C., Skogseid, B., Öberg, K., Nakamura, Y. and Nordenskjöld, M. (1988) Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. Nature 332, 85–87.

14) Chamberlain, S., Shaw, J., Rowland, A., Wallis, J., South, S., Nakamura, Y. et al. (1988) Mapping of mutation causing Friedreich’s ataxia to human chromosome 9. Nature 334, 248–250.

15) Gatti, R.A., Berkel, I., Boder, E., Braedt, G., Charmley, P., Conconnon, P. et al. (1988) Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. Nature 336, 577–580.

16) Leppert, M., Dobbs, M., Scambler, P., O’Connell, P., Nakamura, Y., Stauffer, D. et al. (1987) The gene for familial polyposis coli maps to the long arm of chromosome 5. Science 238, 1411–1413.

17) Barker, D., Wright, E., Nguyen, K., Cannon, L., Fain, P., Goldgar, D. et al. (1987) Gene for von Recklinghausen neurofibromatosis is in the pericentric region of chromosome 17. Science 236, 1100–1102.

18) Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A., Huey, B. et al. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250, 1684–1689.

19) Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N. et al. (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 265, 2088–2090.

20) Nishida, I., Nakamura, Y., Miki, Y., Ando, H., Horii, A. et al. (1991) Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 253, 665–669.

21) Litt, M. and Luty, J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44, 397–401.

22) Lander, E.S. and Botstein, D. (1987) Homozygosity mapping: A way to map human recessive traits with the DNA of inbred children. Science 236, 1567–1570.

23) Toda, T., Segawa, M., Nomura, Y., Nonaka, I., Masuda, K., Ishihara, T. et al. (1993) Localization of a gene for Fukuyama type congenital muscular dystrophy (FCMD) to chromosome 9q31-33. Nat. Genet. 5, 283–286.

24) Kobayashi, K., Nakahori, Y., Miyake, M., Nomura, Y., Yoshioka, M., Saito, K. et al. (1998) An ancient retrotransposon insertion causes Fukuyama-type congenital muscular dystrophy (FCMD). Nature 394, 388–392.

25) Tsujikawa, M., Kurahashi, H., Tanaka, T., Okada, M., Yamamoto, S., Maeda, N. et al. (1998) Homozygosity mapping of a gene responsible for gelatinous drop-like corneal dystrophy (GDDL) to chromosome 1p. Am. J. Hum. Genet. 63, 1073–1077.

26) Tsujikawa, M., Kurahashi, H., Tanaka, T., Nishida, K., Shimomura, Y., Tano, Y. et al. (1999) Identification of the gene responsible for gelatinous drop-like corneal dystrophy. Nat. Genet. 21, 420–423.

27) Onouchi, Y., Gunji, T., Burns, J.C., Shimizu, C., Newburger, J.W., Yashiro, M. et al. (2008) ITPKC functional polymorphism associated with Kawasaki disease susceptibility and formation of coronary artery aneurysms. Nat. Genet. 40, 35–42.

28) Hirakawa, M., Tanaka, T., Hashimoto, Y., Kuroda, M., Takagi, T. and Nakamura, Y. (2002) JSNP: A database of common gene variations in the Japanese population. Nucleic Acids Res. 30, 158–162.

29) The International HapMap Consortium, Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F. et al. (2003) The International HapMap Project. Nature 426, 789–796.

30) The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437, 1299–1302.

31) Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A. et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861.

32) Yoshihara, K.I., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T. et al. (2006) A SNP in the ABCC11 gene is the determinant of human earwax type. Nat. Genet. 38, 324–330.

33) Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsumoda, T. et al. (2002) Functional SNPs in the lymphotoxin-zeta gene that are associated with susceptibility to myocardial infarction. Nat. Genet. 31, 324–327.

34) Shendure, J., Findlay, G.M. and Snyder, M.W. (2019) Genomic medicine—Progress, pitfalls, and promise. Cell 177, 45–57.

35) Buniello, A., Macarthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C. et al. (2019)
The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005–D1012.

Mägi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N. et al. (2017) Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. Hum. Mol. Genet. 26, 3639–3650.

Roselli, C., Chaffin, M.D., Weng, L.-C., Aeschbacher, S., Ahlberg, G., Albert, C.M. et al. (2018) Multi-ethnic genome-wide association study for atrial fibrillation. Nat. Genet. 50, 1225–1233.

Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A. et al. (2018) Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat. Genet. 50, 524–537.

Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M. et al. (2018) Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat. Genet. 50, 390–400.

Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H. et al. (2020) Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat. Genet. 52, 669–679.

Fisher, R.A. (1919) The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52, 399–433.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H. et al. (2012) Systematic localization of common disease-associate variation in regulatory DNA. Science 337, 1190–1195.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M. et al. (2014) An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461.

Roadmap Epigenetics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A. et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330.

Clausnitzer, M., Dankel, S.N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C. et al. (2015) FTO obesity variant circuitry and adipocyte browning in humans. N. Engl. J. Med. 373, 895–907.

Gupta, R.M., Hadayo, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D. et al. (2017) A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. Cell 170, 522–533.e15.

Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M. et al. (2017) Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat. Genet. 49, 1458–1467.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H. et al. (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet. 50, 1219–1224.

Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M. and Daly, M.J. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat. Genet. 51, 584–591.

Kerminen, S., Martin, A.R., Koskela, J., Ruotsalainen, S.E., Havulinna, A.S., Surakka, I. et al. (2019) Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. Am. J. Hum. Genet. 104, 1169–1181.

Sakaue, S., Hirata, J., Kanai, M., Suzuki, K., Akiyama, M., Lai Too, C. et al. (2020) Dimensionality reduction reveals fine-structure scale in the Japanese population with consequences for polygenic risk prediction. Nat. Commun. 11, 1569.

Koyama, S., Ito, K., Terao, C., Akiyama, M., Horikoshi, M., Momozawa, Y. et al. (2020) Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. Nat. Genet. 52, 1169–1177.

Iyer, L., Das, S., Janisch, L., Wen, M., Ramírez, J., Karrison, T. et al. (2002) UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. Pharmacogenomics J. 2, 43–47.

The International Warfarin Pharmacogenetics Consortium (2009) Estimation of the warfarin dose prediction accuracy in warfarin dose prediction of current polygenic risk scores may exacerbate health disparities. Nat. Genet. 51, 113–121.

Kiyotani, K., Yoshihara, T., Sassa, M., Bando, Y., Sumitomo, I., Hosono, N. et al. (2008) Impact of CYP2D6*10 on recurrence-free survival in breast cancer patients receiving adjuvant tamoxifen therapy. Cancer Sci. 99, 995–999.

Kiyotani, K., Yoshihara, T., Hosono, N., Tsunoda, T., Kubo, M., Aki, F. et al. (2010) Lessons for pharmacogenomics studies: Association study between CYP2D6 genotype and tamoxifen response. Pharmacogenet. Genomics 20, 565–568.

Nakamura, Y. (ed.) (2015) Immunopharmacogenomics. Springer, Tokyo.

Ozeki, T., Yoshihara, T., Yowang, A., Takahashi, A., Kubo, M., Shirakata, Y. et al. (2011) Genome-wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in...
Japanese population. Hum. Mol. Genet. 20, 1034–1041.

62) Daly, A.K., Donaldson, P.T., Bhatnagar, P., Shen, Y., Pe’er, I., Floratos, A. et al. (2009) HLA-B5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. Nat. Genet. 41, 816–819.

63) Yew, P.Y., Alachkar, H., Yamaguchi, R., Kiyotani, K., Fang, H., Yap, K.L. et al. (2015) Quantitative characterization of T-cell repertoire in allogeneic hematopoietic stem cell transplant recipients. Bone Marrow Transplant. 50, 1227–1234.

64) Ko, T.M., Kiyotani, K., Chang, J.S., Park, J.H., Yew, P.Y., Chen, Y.T. et al. (2018) Immunoglobulin profiling identifies unique signatures in patients with Kawasaki disease during intravenous immunoglobulin treatment. Hum. Mol. Genet. 27, 2671–2677.

65) Kiyotani, K., Mai, T.H., Yamaguchi, R., Yew, P.Y., Kulis, M., Orgel, K. et al. (2018) Characterization of the B-cell receptor repertoires in peanut allergic subjects undergoing oral immunotherapy. J. Hum. Genet. 63, 239–248.

66) Inoue, H., Park, J.H., Kiyotani, K., Zewde, M., Miyashita, A., Jinnin, M. et al. (2016) Intratumoral expression levels of PD-L1, GZMA, and HLA-A along with oligoclonal T cell expansion associate with response to nivolumab in metastatic melanoma. Oncoimmunology 5, e1204507.

67) Santa-Maria, C.A., Kato, T., Park, J.H., Kiyotani, K., Rademaker, A., Shah, A.N. et al. (2018) A pilot study of durvalumab and tremelimumab and immunogenomic dynamics in metastatic breast cancer. Oncotarget 9, 18985–18996.

68) Collins, F.S. and Varmus, H. (2015) A new initiative on Precision Medicine. N. Engl. J. Med. 372, 793–795.

Profile

Yusuke Nakamura was born in Osaka Prefecture in 1952 and graduated from Osaka University School of Medicine in 1977. After graduation, he majored in Surgery, especially in surgical oncology and received Ph.D. degree in 1984. Then he spent five years as a postdoctoral fellow and a faculty member at Howard Hughes Medical Institute, the University of Utah to pursue an advanced study in genetics. In 1989, he became Head of Biochemistry Department at Cancer Institute, Tokyo. In 1994, he was appointed as Professor at Institute of Medical Science, The University of Tokyo and became Director of Human Genome Center in 1995. In 2000, he also became a leader for the SNP typing group at RIKEN SNP Research Center and subsequently served as director of the RIKEN Center between 2005 to 2010. In 2011, he was appointed as Special Advisor to the Japanese Cabinet. In 2012, he moved to Department of Medicine at the University of Chicago as Professor and Deputy Director of Center for Personalized Therapeutics. In 2018, he returned to Japan as Director of Cross-ministerial Strategic Innovation Promotion Program of Japanese Cabinet Office as well as Director at the Cancer Precision Center, Japanese Foundation for Cancer Research. He is an internationally well-known pioneer of genetic variation studies including VNTR (variable number of tandem repeat) markers and their applications in medical fields including cancer genomics. He isolated and characterized a number of genes involved in carcinogenesis including APC that is responsible for familial polyposis coli. For his accomplishment, he received many awards that include Medal with Purple Ribbon, Keio Medical Science Prize, Takeda Medical Prize and Clarivate Citation Laureate.

(Received Mar. 26, 2021; accepted Apr. 14, 2021)