

## REVIEW

# DNA variations in human and medical genetics: 25 years of my experience

Yusuke Nakamura<sup>1,2,3</sup>

DNA variations have contributed enormously to the fields of medical and forensic science, especially through their use in studies on genes responsible or susceptible to various diseases and those on screening of chromosomal abnormalities in tumors. The types of genetic variations used in these studies have changed in the past 25 years and can be classified into five major 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). Genetic linkage analysis using these tools helped to map and discover genes responsible for hundreds of hereditary diseases. Furthermore, construction of the international SNP database and recent development of high-throughput SNP typing platforms enabled us to perform genome-wide association studies, which have identified genes (or genetic variations) susceptible to common diseases or those associated with drug responses. Genome-wide sequencing of individual DNAs is gaining immense scope. Here, I summarize the history of polymorphic DNA markers and their contribution to the genetic analysis of both rare hereditary diseases and common diseases, as well as recent advances in pharmacogenetics, including our contribution to these areas.

*Journal of Human Genetics* (2009) 54, 1–8; doi:10.1038/jhg.2008.6; published online 9 January 2009

**Keywords:** CNV; DNA polymorphism; GWAS; microsatellite; RFLP; SNP; VNTR

### HISTORY OF DNA VARIATIONS

Just as we all look different, our genomic sequences vary significantly. These differences in our genome are known as genetic variations, and are caused by various types of mechanisms, such as nucleotide substitutions, insertion/deletion of nucleotides, differences in the number of tandem-repeat sequences, differences in the copy number of genomic segments and the combinations of these changes. A subset of these genetic variations is defined as genetic polymorphisms when they are observed at a frequency of  $\geq 1\%$  in a certain population.

The DNA polymorphisms are classified into six classes as shown in Table 1. The importance of restriction fragment length polymorphism (RFLP) in medical research was first suggested by Botstein *et al.*<sup>1</sup> RFLP is generated by differences in the size of the DNA fragment digested by a certain restriction endonuclease due to the base substitutions at the site recognized by the endonuclease. Before the discovery of polymerase chain reaction (PCR) RFLP was detected by Southern analysis and required a large amount of non-degraded genomic DNAs. RFLP patterns show co-dominant Mendelian inheritance and can help distinguish between the parental alleles (whether the allele is of maternal or of paternal origin) of the particular loci in our genome. The variable number of tandem repeats (VNTRs) was first reported as one type of RFLPs by Nakamura *et al.*,<sup>2</sup> but they are highly poly-

morphic, with high heterozygosity in given populations because of a wide variety of the copy number of tandem-repeat DNA sequences. Isolation of VNTR markers was done by an extension of the report of hypervariable 'minisatellite' sequences by Jeffreys *et al.*<sup>3</sup> Minisatellite probes identified very variable multiple loci simultaneously, but each VNTR marker (also called a single-locus minisatellite marker) identified a single highly polymorphic locus in our genome. Because of their highly polymorphic nature, both VNTRs and minisatellite markers were applied in forensic studies,<sup>4</sup> and also used in clinics to monitor recipients of bone marrow transplants.<sup>5</sup>

VNTRs also contributed to the studies of tumor suppressor genes, most of which were characterized by inactivation of both alleles in tumors (the two-hit mutation model) and were very useful for detection of loss of a chromosome or a part of a chromosome (loss of heterozygosity) by the disappearance of a paternal or maternal allele on Southern analysis. The first systematic analysis of loss of all chromosomal arms (allelotype analysis) was reported in colorectal cancer by Vogelstein *et al.*<sup>6</sup> using a set of VNTR markers. The results became one of the important clues for establishing multi-step carcinogenesis model of colorectal cancers.<sup>7</sup> Subsequent detailed analysis of the short arm of chromosome 17 led us to find the mutations of the p53 gene and prove it to be a tumor suppressor gene.<sup>8</sup>

<sup>1</sup>Japan Society of Human Genetics, Tokyo, Japan; <sup>2</sup>Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan and <sup>3</sup>Center for Genomic Medicine, RIKEN, Japan

Correspondence: Dr Y Nakamura, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.

E-mail: yusuke@ims.u-tokyo.ac.jp

Received 28 October 2008; revised 5 November 2008; accepted 6 November 2008; published online 9 January 2009

**Table 1** Types of genetic variation

	Type	Established (or widely used)
First generation	RFLP (restriction fragment length polymorphism)	1980
Second generation	VNTR (variable number of tandem repeat) or minisatellite	1985
Third generation	Microsatellite or STRP (short tandem repeat polymorphism)	1989
Fourth generation	SNP (single nucleotide polymorphism)	2000
Fifth generation	CNV (copy number variation)	2005
Sixth generation	Whole genome sequencing	2012(?)

### FIRST-GENERATION GENETIC LINKAGE MAPS OF HUMAN CHROMOSOMES AND GENETIC LINKAGE ANALYSIS

Using these polymorphic DNA markers, White and his co-workers constructed genetic linkage maps of all human chromosomes and also made these DNA markers freely available to scientific communities (for an example, see Nakamura *et al.*<sup>9</sup>). The study by Botstein *et al.*<sup>1</sup> indicated that genetic loci responsible for genetic diseases, for which no information for speculating the causes of the disease mechanism was available, could be mapped by linkage analysis if an appropriate number of families with certain genetic diseases was available. As polymorphic DNA markers allow us to distinguish a chromosome as of maternal or paternal origin, we can examine whether each polymorphic ‘marker’ allele co-segregates with the inheritance of a disease. In fact, Gusella *et al.*<sup>10</sup> determined the genetic locus for Huntington’s disease to the short arm of chromosome 4. Subsequent to the increase of available DNA polymorphic markers in the late 1980s, many genes for relatively common genetic diseases, such as cystic fibrosis, familial polyposis coli, neurofibromatosis type 1, multiple endocrine neoplasia type 1 and familial breast cancer, were discovered and their responsible genes identified a few years later. These studies proved that linkage analysis with DNA polymorphic markers using families with genetic diseases in any inheritance model was a powerful tool to discover responsible genes even without any knowledge about the biological or biochemical mechanisms (or such abnormalities). This approach is known as the ‘reverse genetics’ method.

### SECOND-GENERATION GENETIC LINKAGE MAP OF HUMAN CHROMOSOMES

It is widely accepted that the PCR method revolutionized the size and speed of DNA analysis. DNA polymorphic markers were also drastically influenced by the development of PCR systems. After 1990, scientists switched from RFLP analysis based on Southern technology to microsatellite analysis based on PCR technology. Microsatellite markers were first described by Weber and May,<sup>11</sup> and are short segments of two or more base pairs repeated tandemly in tracts. Unlike VNTR loci, which number a few thousand in our genome, microsatellite loci are present at >100 000 regions that cover most of the genomic regions. Hence, microsatellite markers have been successfully used for linkage analysis or population genetics because they can be easily adapted to the high-throughput system and require a very small amount of DNA. Even if DNA is degraded to some extent, microsatellite loci can be analyzed. Owing to their high levels of heterozygosity, they can distinguish paternal alleles with high probability and are very informative in linkage analysis. Through the development of very detailed second-generation genetic

linkage maps of human chromosomes, a large number of genetic diseases were mapped and, subsequently, their responsible genes were isolated. Rare mutations have so far been reported in 2000 Mendelian diseases.

One of the advantages of the approach using microsatellite markers is the possibility of applying linkage analysis (also known as homozygosity mapping) to recessive diseases with very low incidence for which only a very small number of patients can be collected for study.<sup>12</sup> However, in the case of patients born by consanguineous marriage (or marriage between individuals living in the same area where residents stay for a long period), only a few patients are sufficient to map a region including a gene responsible for a genetic disease. In fact, we applied this approach and determined the loci for Fukuyama muscular dystrophy and gelatinous drop-like corneal dystrophy, and isolated the responsible genes.<sup>13,14</sup>

Another advantage of microsatellite markers is their application in a sib-pair analysis or transmission disequilibrium test, both of which are useful for searching genetic loci associated with common diseases. Although these types of genetic approaches were used widely, they were not so successful in identifying genes or loci associated with common diseases because a huge number of siblings or families were required to determine the genetic factors with very modest effects that increase the risk of diseases (limitation of the statistical power). One example of the success of this type of approach was shown by Onouchi *et al.*<sup>15</sup> using 78 Japanese sibling pairs with Kawasaki disease, the leading cause of acquired heart disease among children in developed countries; we showed that an SNP in the ITPKC gene increased the risk of the disease and might be associated with a response to intravenous immunoglobulin therapy.

### HapMap PROJECT AND GWAS (GENOME-WIDE ASSOCIATION STUDY) USING SNPs (SINGLE-NUCLEOTIDE POLYMORPHISMS)

Around 2000, scientists planned to develop millions of SNP markers covering an entire genome and construct a high-density SNP (also haplotype) map. SNPs are the genetic variations that are most commonly present in our genome. They are expected to be present in one in every 500–1000 bp on average, but their distribution is likely to be higher (say, one in every 300 bp). On the basis of the ‘common variant–common disease’ association hypothesis, SNPs were considered to be very useful as a tool for population genetics, particularly for identification of genes (or genetic variations) susceptible to various diseases. The Japanese government announced a 5-year ‘Japanese Millennium Project’ in 2000 ([http://www.nature.com/nbt/journal/v18/n2/full/nbt0200\\_142.html](http://www.nature.com/nbt/journal/v18/n2/full/nbt0200_142.html)), which included (1) construction of an SNP database for the Japanese population and (2) establishment of a high-throughput SNPTyping system for the genome-wide association study, as one of the five arms of the project. We resequenced 24 Japanese individuals in a total of about 154-Mb genomic segments, mainly in regions containing genes, and found nearly 200 000 genetic variations, including 174 269 SNPs (one genetic variation in every 800 bp). In addition, we established a high-throughput SNP genotyping system by a combination of multiplex PCRs (simultaneous amplification of 96 genomic segments) with Invader assay.<sup>16</sup> Using this platform we constructed the JSNP database, including the information of nearly 190 000 gene-based SNP loci as well as the allelic frequencies of 768 Japanese individuals at nearly 79 000 SNP loci in 2002. This SNP database had the largest allelic-frequency information based on a large number of individuals ([http://snp.ims.u-tokyo.ac.jp/index\\_ja.html](http://snp.ims.u-tokyo.ac.jp/index_ja.html)) before the International HapMap database was constructed.

Since multiple large-scale SNP genotyping platforms were developed around 2003, construction of an international consortium for making a SNP database for three major populations had been discussed and the International HapMap project including six countries (Canada, China, Japan, Nigeria, UK and USA) was organized in 2003. The aims of the International HapMap Project were: (1) determination of the common patterns of one million or more DNA sequence variations in the human genome using DNA samples from populations with ancestry from parts of Africa, Asia and Europe, (2) construction of the LD map for all chromosomes and (3) making such information freely available to the scientific community. The HapMap results are expected to allow the discovery of sequence variants that affect common diseases, facilitate development of diagnostic tools and enhance our ability to choose targets for therapeutic intervention (The International HapMap Consortium 2003<sup>17</sup>).

After the very extensive efforts of the participating groups, the consortium constructed a database consisting of more than one million SNPs in 2005 (The International HapMap Consortium<sup>18</sup>) and subsequently reported an extended database in 2007 (The International HapMap Consortium 2007<sup>19</sup>). In the phase 1 project, we (the SNP Research Center, RIKEN) produced nearly 25% of the SNP typing data. 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 the human earwax type,<sup>20</sup> suggesting that, if they have effects on the quality or quantity of the gene product, SNPs uniquely observed in a certain population might be the genetic determinants characterizing them.

There were many criticisms and skepticisms about the ‘common variation–common diseases’ approach for identifying genes susceptible to common diseases at the beginning of the International HapMap project, but many published papers have shown the usefulness of the genome-wide association study (GWAS) to uncover various genetic factors associated with various diseases (not necessarily common diseases). Although not well recognized, in 2001 we began genome-wide association studies using nearly 90 000 gene-based SNPs that well represented covered regions containing genes. Through this approach, we identified genes susceptible to myocardial infarction.<sup>21</sup> In addition, we have reported many candidate genes possibly associated with susceptibility to diseases as summarized in Table 2.

Internationally, systematic GWAS was started in 2006 based on the accumulation of a large set of SNP information through the International HapMap project, as well as the development of cheap, commercially available and accurate high-throughput SNP analysis platforms. I have summarized representative examples of the GWAS approach for various common diseases in Table 3.

### COPY-NUMBER VARIATIONS (CNVs) AND DISEASES

Copy-number variation is defined as a form of genomic structural variation and refers to differences in the number of copies of a particular genomic region. We normally have two copies of each autosomal region, one per chromosome, but because of deletion or duplication (or multiplication) of particular genomic regions we observed a decrease or increase in the copy number. The involvement of copy-number alterations in human phenotypes was first reported as ‘genomic disorders’, which are often caused by *de novo* structural alterations,<sup>98</sup> and so far dozens of genomic disorders have been reported,<sup>99</sup> including inherited forms of CNVs that underlie Mendelian diseases such as autosomal-dominant leukodystrophy<sup>100</sup> and

**Table 2** Genes susceptible to various diseases reported from our centers (Center for Genomic Medicine, RIKEN and Human Genome Center, Institute of Medical Science, The University of Tokyo)

Disease	Gene(s)	References
Adult asthma	Fn-III-D	Matsuda <i>et al.</i> <sup>22</sup>
Aspirin-induced asthma	TBX21	Akahoshi <i>et al.</i> <sup>23</sup>
Cerebral infarction	AGTRL1	Hata <i>et al.</i> <sup>24</sup>
Cerebral infarction	PRKCH	Kubo <i>et al.</i> <sup>25</sup>
Childhood atopic asthma	IL-12B	Hirota <i>et al.</i> <sup>26</sup>
Childhood atopic asthma	RIP2	Nakashima <i>et al.</i> <sup>27</sup>
Crohn's disease	TNFSF15	Yamazaki <i>et al.</i> <sup>28</sup>
Diabetic nephropathy	ELMO1	Shimazaki <i>et al.</i> <sup>29</sup>
Diabetic nephropathy	NCALD	Kamiyama <i>et al.</i> <sup>30</sup>
Idiopathic pulmonary fibrosis	TERT	Mushiroda <i>et al.</i> <sup>31</sup>
IgA nephropathy	Selectin $\epsilon$	Takei <i>et al.</i> <sup>32</sup>
IgA nephropathy	HLA	Akiyama <i>et al.</i> <sup>33</sup>
IgA nephropathy	PIGR	Obara <i>et al.</i> <sup>34</sup>
IgA nephropathy	IGHMBP2	Ohtsubo <i>et al.</i> <sup>35</sup>
Kawasaki disease	ITPKC	Onouchi <i>et al.</i> <sup>15</sup>
Lumbar disc disease	CILP	Seki <i>et al.</i> <sup>36</sup>
Myocardial infarction	LTA	Ozaki <i>et al.</i> <sup>21</sup>
Myocardial infarction	LGALS2	Ozaki <i>et al.</i> <sup>37</sup>
Myocardial infarction	PSMA6	Ozaki <i>et al.</i> <sup>38</sup>
Myocardial infarction	MIAT	Ishii <i>et al.</i> <sup>39</sup>
Myocardial infarction	ITIH3	Ebana <i>et al.</i> <sup>40</sup>
Osteoarthritis	ASPN	Kizawa <i>et al.</i> <sup>41</sup>
Osteoarthritis	GDF5	Miyamoto <i>et al.</i> <sup>42</sup>
Osteoarthritis	EDG2	Mototani <i>et al.</i> <sup>43</sup>
Osteoarthritis	DWVA	Miyamoto <i>et al.</i> <sup>44</sup>
Rheumatoid arthritis	PADI4	Suzuki <i>et al.</i> <sup>45</sup>
Rheumatoid arthritis	SLC22A4	Tokuhiro <i>et al.</i> <sup>46</sup>
Rheumatoid arthritis	FCRL3	Kochi <i>et al.</i> <sup>47</sup>
Rheumatoid arthritis	CD244	Suzuki <i>et al.</i> <sup>48</sup>
Systemic lupus erythematosus	TNXB	Kamatani <i>et al.</i> <sup>49</sup>
Systemic lupus erythematosus	ITPR3	Oishi <i>et al.</i> <sup>50</sup>
Type 2 diabetes	WNT5B	Kanazawa <i>et al.</i> <sup>51</sup>
Type 2 diabetes	TFAP2B	Maeda <i>et al.</i> <sup>52</sup>
Type 2 diabetes	KLF7	Kanazawa <i>et al.</i> <sup>53</sup>
Type 2 diabetes	KCNQ1	Unoki <i>et al.</i> <sup>54</sup>
Association of VKORC1 and CYP2C9 polymorphisms with warfarin dose requirements in Japanese patients.		Mushiroda <i>et al.</i> <sup>55</sup>
Nevirapine-induced skin adverse drug reactions	HLA-B*3505	Chantarangsu <i>et al.</i> <sup>56</sup>

hereditary pancreatitis (triplication of the trypsinogen locus).<sup>101</sup> Through genome-wide screening of CNVs, a large body of CNV information has been accumulated.<sup>102–106</sup> CNVs are considered to influence the quantity of the gene products because of the different copy number of certain genes if a part of a significant segment of the gene is deleted, or if a functional unit (including the regulatory region) of a particular gene is duplicated or multiplied. In fact, systematic analysis indicated an association between CNVs (and also SNPs) and gene expression variations, which might be a model of complex phenotypes.<sup>107</sup> The associations of some CNVs with common diseases have also been reported, including CNV of CCL3L1 and HIV-1/AIDS susceptibility,<sup>108</sup> CNV of *Fcgr3* and glomerulonephritis,<sup>109</sup> CNVs of complement component C4 and systemic lupus erythematosus (SLE),<sup>110</sup> and CNV of *FCGR3B* and systemic

**Table 3 Representative results of GWAS analysis of various diseases**

Disease	References
Amyotrophic lateral sclerosis	Van Es <i>et al.</i> <sup>57</sup>
Amyotrophic lateral sclerosis	Stefansson <i>et al.</i> <sup>58</sup>
Asthma	Ober <i>et al.</i> <sup>59</sup>
Asthma (childhood)	Moffatt <i>et al.</i> <sup>60</sup>
Atrial fibrillation	Gudbjartsson <i>et al.</i> <sup>61</sup>
Basal cell carcinoma (skin)	Stacey <i>et al.</i> <sup>62,63</sup>
Bipolar disorder	Ferreira <i>et al.</i> <sup>64</sup>
Bipolar disorder, coronary artery disease, Crohn's disease, rheumatoid arthritis, Type 1 diabetes, Type 2 diabetes	Wellcome Trust Case-Control Consortium <sup>65</sup>
Breast cancer	Easton <i>et al.</i> <sup>66</sup>
Breast cancer	Stacey <i>et al.</i> <sup>62,63</sup>
Breast cancer (postmenopausal)	Hunter <i>et al.</i> <sup>67</sup>
Chronic lymphocytic leukemia	Di Bernardo <i>et al.</i> <sup>68</sup>
Colorectal cancer	Tomlinson <i>et al.</i> <sup>69</sup>
Colorectal cancer	Zanke <i>et al.</i> <sup>70</sup>
Colorectal cancer	Broderick <i>et al.</i> <sup>71</sup>
Coronary artery disease	Samani <i>et al.</i> , 2007
Coronary artery disease	McPherson <i>et al.</i> <sup>72</sup>
Gallstone	Buch <i>et al.</i> <sup>73</sup>
Gastric cancer (diffuse-type)	Study Group of Millennium Genome Project for Cancer (Japan) <sup>74</sup>
Glaucoma	Thorleifsson <i>et al.</i> <sup>75</sup>
Gout	Dehghan <i>et al.</i> <sup>76</sup>
Lung cancer	Hung <i>et al.</i> <sup>77</sup>
Lung cancer	Amos <i>et al.</i> <sup>78</sup>
Lung cancer	Thorleifsson <i>et al.</i> <sup>79</sup>
Multiple sclerosis	International Multiple Sclerosis Genetics Consortium <sup>80</sup>
Myocardial infarction	Helgadóttir <i>et al.</i> <sup>81</sup>
Neuroblastoma	Maris <i>et al.</i> <sup>82</sup>
Osteoporosis	Richards <i>et al.</i> <sup>83</sup>
Prostate cancer	Gudmundsson <i>et al.</i> <sup>84,85</sup>
Restless legs syndrome	Winklemann <i>et al.</i> <sup>86</sup>
Sarcoidosis	Hofmann <i>et al.</i> <sup>87</sup>
Schizophrenia	O'Donovan <i>et al.</i> <sup>88</sup>
Systemic lupus erythematosus	International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) <sup>89</sup>
Systemic lupus erythematosus	Kozyrev <i>et al.</i> <sup>90</sup>
Systemic lupus erythematosus	Hom <i>et al.</i> <sup>91</sup>
Type 1 diabetes	Hakonarson <i>et al.</i> <sup>92</sup>
Type 2 diabetes	Sladek <i>et al.</i> <sup>93</sup>
Type 2 diabetes	Steinthorsdóttir <i>et al.</i> <sup>94</sup>
Type 2 diabetes	Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research <sup>95</sup>
Type 2 diabetes	Yasuda <i>et al.</i> <sup>96</sup>
Urinary bladder cancer	Kiemeny <i>et al.</i> <sup>97</sup>

autoimmunity.<sup>111</sup> However, we suspect that the association of CNVs of complement component C4 with SLE might be false-positively observed because of the linkage disequilibrium (LD) between the C4 CNVs and the SNP in *TNXB* for which we found a much stronger association with SLE.<sup>49</sup> As there are considerable ethnic differences in the genes susceptible to various diseases, particularly in those related

to autoimmune disease, we must be very careful in interpretations of the data, but our results at least indicate that the association between CNVs and diseases might be trapped by the use of SNPs for which very high-throughput and accurate analysis systems are widely available. It is certainly better to measure CNVs directly, but as the methods to accurately measure the copy numbers in an entire genome are not yet established, in particular in cases containing more than two copies or in cases of gene families with very high homology, the information of LD between SNPs and CNVs is essentially important.

### GENES ASSOCIATED WITH RARE GENETIC DISEASES AND COMMON DISEASES

It should be noted that common genetic variations detected in genes responsible for rare genetic diseases (often severe and early-onset) were later found to be associated with common diseases with similar phenotypes. One typical example is shown in the studies of diabetes mellitus. Maturity-onset diabetes of the young (MODY, OMIM 606391) is an autosomal dominant form of diabetes mellitus, which is characterized by early onset of hyperglycemia and dysfunction of  $\beta$ -cells in the pancreas and is phenotypically very similar to type 2 diabetes. Six genes have so far been identified as the causes of MODY, including hepatic nuclear factor 4a (*HNF4A*; *MODY1*), glucokinase (*GCK*; *MODY2*), transcription factor 1 (*TCF1/HNF1A*; *MODY3*), insulin promoter factor 1 (*IPF1/PDX1*; *MODY4*), transcription factor 2 (*TCF2/HNF1B*; *MODY5*) and neurogenic differentiation 1 (*NEUROD1*; *MODY6*).<sup>112</sup> The common variants within these MODY genes have been extensively investigated for their association with type 2 diabetes. The *HNF4A* gene has two tissue-specific transcription initiation sites in the liver and pancreas. An SNP rs1884613 and several other variations in the pancreas-specific promoter region were shown to be associated with the susceptibility to type 2 diabetes.<sup>113–115</sup> Similarly, genetic variations within the other MODY genes were also shown to have modest effects on conferring susceptibility to type 2 diabetes; an SNP in *TCF2/HNF1B* was identified as a strong candidate for type 2 diabetes through GWASs using populations of European descent.<sup>84,85</sup> The results of multiple European GWAS analyses for type 2 diabetes also indicated that previously reported SNPs in *KCNJ11* (rs5219; E23K<sup>116</sup>) and *PPARG* (rs1801282; P12A<sup>117</sup>) were possible candidates susceptible to type 2 diabetes;<sup>118–120</sup> these genes were known to be causative genes for monogenic forms of diabetes mellitus. Another example is *WFS1*, a gene responsible for Wolfram syndrome, which encodes Wolfram, a membrane glycoprotein important for maintenance of calcium homeostasis in the endoplasmic reticulum.<sup>121</sup> Wolfram syndrome is characterized by diabetes insipidus, juvenile-onset non-autoimmune diabetes mellitus, optic atrophy and deafness (DIDMOAD, OMIM 222300). An association study for type 2 diabetes found two SNPs in *WFS1* (rs10010131, rs6446482) to be associated with a risk for type 2 diabetes.<sup>122</sup>

Recently we also found a similar and interesting result showing the link between rare genetic forms and a relatively common form of IPF (idiopathic pulmonary fibrosis), in which mutations and common variations in the *TERT* gene encoding a reverse transcriptase, a component of a telomerase, were involved. Two groups independently demonstrated that mutations in *TERT* were responsible for familial IPF in a Caucasian population. Mutations in the *TERT* gene were reported to be responsible for familial IPF in a Caucasian population on the basis of the fact that many affected individuals in the DKC (dyskeratosis congenital) families also had IPF.<sup>123</sup> A study by Tsakiri *et al.*<sup>124</sup> independently revealed that missense and frameshift mutations in *TERT* co-segregated with IPF in the two families. Through a genome-wide association study using patients with sporadic IPF, we

found a significant association of an SNP in intron 2 of the *TERT* gene (rs2736100) with IPF. These results indicated that the significance of the effect of genetic variations on the quality and/or quantity of the gene product influences the level of dysfunction; if the effect is severe it becomes a cause of a rare form of genetic disease and if the effect is modest it becomes a risk factor for a common disease.

## GENETIC VARIATIONS IN PHARMACOGENETICS AND PHARMACOGENOMICS

Although definitions of pharmacogenetics and pharmacogenomics are still unclear, it is certain that a subset of genetic variations in genes encoding drug-metabolizing enzymes, drug transporters, drug receptors, drug-target molecules and downstream molecules involved in their signaling pathways influence the effectiveness of and adverse reactions to drugs.

Adverse drug reactions (ADRs) can be classified into two groups: one can be explained by the mode of therapeutic drug actions; there is little or no information about the underlying mechanisms of the other. Typical examples for the former cases are leucocytopenia caused by cytotoxic anti-cancer drugs, and brain or intestinal bleeding caused by warfarin, an oral anticoagulant. Representative examples for the latter cases are toxic epidermal necrolysis and drug-induced liver injury caused by various drugs. Millions of patients are suspected to suffer from severe ADRs worldwide every year, and a subset of patients even lose their lives. Although many factors are involved in the etiology of ADRs, genetic factors (genetic variations) are implicated to be one of the critical ones. Hence, identification of genetic factors that increase the risk of ADRs is expected to improve the medical management of patients with high risk for ADRs, and might also contribute to significant reduction of unnecessary medical costs (see an example for warfarin at <http://www.aei-brookings.org/publications/abstract.php?pid=1127>).

In the past decade, extensive research has been performed to uncover the genetic factors underlying ADRs and has successfully identified the genetic variations that increase the risks of ADRs. Table 4 summarizes the genetic variations that have been validated and/or are listed in the 'Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels' in the US Food and Drug Administration homepage ([http://www.fda.gov/cder/genomics/genomic\\_biomarkers\\_table.htm](http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm)). The genes listed in Table 4 are mainly categorized into three groups: group 1—drug-metabolizing enzymes and drug transporters related to pharmacokinetics; group 2—proteins related to pharmacodynamics and group 3—human leukocyte antigens (HLAs). It is obvious that poor clearance of drugs from our body increases drug concentration up to a toxic level and results in the appearance of ADRs (group 1); a typical example is the case of anti-cancer drugs: genetic variants in the UDP glucuronosyltransferase 1A1 (*UGT1A1*) gene increase the risk of myelotoxicity caused by irinotecan. An example of group 2 is genetic variations in *VKORC1*, which is involved in vitamin K recycling and influences the bleeding risk of warfarin. HLAs have indicated its involvement in many types of ADRs, and the HLA-B\*1502 allele was shown to significantly increase serious dermatologic reactions to carbamazepine such as toxic epidermal necrosis (TEN) and Stevens–Johnson syndrome (SJS).<sup>129</sup> Similarly, we found a strong association of the HLA-B\*3505 allele with nevirapine-induced adverse drug reactions in the skin of Thai patients.<sup>56</sup> Although the molecular mechanisms of HLA's involvement in dermatologic ADRs are unclear, an interaction between a certain HLA molecule and a drug (or its metabolite) may trigger a cellular immune response.

In addition, some cases that influence the effectiveness of drugs have also been reported. A good example is the relation between

**Table 4 Valid genes associated with drug effectiveness or ADRs**

Gene	Drug	References
<i>(1) Recommended in the drug use</i>		
CYP2C9	Warfarin	Anderson <i>et al.</i> <sup>125</sup>
VKORC1	Warfarin	D'Andrea <i>et al.</i> <sup>126</sup>
TPMT	Azathioprine	Snow and Gibson <sup>127</sup>
UGT1A1	Irinotecan	Innocenti <i>et al.</i> <sup>128</sup>
HLA-B*1502	Carbamazepine	Chung <i>et al.</i> <sup>129</sup>
HLA-B*5701	Abacavir	Mallal <i>et al.</i> <sup>130</sup>
<i>(2) Information only</i>		
CYP2C19	Proton pump inhibitors	Furuta <i>et al.</i> <sup>131</sup>
CYP2D6	Atomoxetine	Sauer <i>et al.</i> <sup>132</sup>
	Fluoxetine	Sallee <i>et al.</i> <sup>133</sup>
	Codeine	Gasche <i>et al.</i> <sup>134</sup>
	Tamoxifen	Goetz <i>et al.</i> <sup>135</sup>
		Kiyotani <i>et al.</i> <sup>136</sup>
NAT2	Isoniazid	Ohno <i>et al.</i> <sup>137</sup>
	Sulfasalazine	Taniguchi <i>et al.</i> <sup>138</sup>
<i>(3) Future biomarker candidates</i>		
CYP2B6	Efavirenz	Gatanaga <i>et al.</i> <sup>139</sup>
CYP3A5	Tacrolimus	Haufroid <i>et al.</i> <sup>140</sup>
ABCB1	Paroxetine	Kato <i>et al.</i> <sup>141</sup>
SLCO1B1	Simvastatin	Link <i>et al.</i> <sup>142</sup>
SLC22A1	Metformin	Shu <i>et al.</i> <sup>143</sup>

genetic variations of CYP2D6 and the clinical outcome of breast cancer patients subjected to tamoxifen treatment. The clinical outcome of tamoxifen treatment was suggested to be influenced by the activity (genetic variations) of cytochrome P450 2D6 (CYP2D6) enzyme because tamoxifen is metabolized by CYP2D6 to its active forms of anti-estrogenic metabolite, 4-hydroxytamoxifen and endoxifen in our body. Higher incidences of recurrence after surgical treatment in breast cancer patients with null- or low-activity genotypes were shown in Caucasian and Japanese populations.<sup>135,136</sup> As patients need to take this drug over a long duration, the genetic diagnosis for tamoxifen treatment may contribute to the improvement of patients' outcomes and the reduction of unnecessary medical cost. Pharmacogenetics information should be useful for predicting the effectiveness, risk of adverse reactions and appropriate dose of drugs for individual patients, and may contribute to the establishment of 'personalized medicine', the concept of 'an appropriate dose of a right drug to a right patient'.

It is notable that although a large amount of information is being accumulated in pharmacogenetics and pharmacogenomics, the incidence of severe ADRs is very low. Hence, it is often difficult to verify the genetic variations that can predict the risk of deleterious ADRs. As collection of severe ADR cases is time-consuming, the global effort to collect cases and analyze genetic factors is eagerly awaited to reduce ADRs that lead to poor quality of life and unnecessary medical costs.

## CONCLUSION

In this review, I briefly describe the history of human genetic variations, as well as their significance in life science, particularly in medical genetics and pharmacogenetics. One of the major goals of life science is to improve our quality of lives. Through a better understanding of the molecular mechanisms causing diseases, we can achieve our goals of preventing diseases, avoiding their progression

and curing them. Genetic variations have played and will play very important roles for better medical and health care. However, we need to make extensive efforts to ensure that genetic variations are not used for genetic discrimination. Although we are all different, we should have equal rights and should respect each other's differences.

## ACKNOWLEDGEMENTS

I thank Drs Tatsuhiko Tsunoda, Shiro Maeda and Taisei Mushiroda for their kind assistance and useful advices regarding this paper.

- Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
- Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M. *et al.* Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616–1622 (1987).
- Jeffreys, A. J., Wilson, V. & Thein, S. L. Hypervariable 'minisatellite' regions in human DNA. *Nature (London)* **314**, 67–73 (1985).
- Odelberg, S. J., Plaetke, R., Eldridge, J. R., Ballard, L., O'Connell, P., Nakamura, Y. *et al.* Characterization of eight VNTR loci by agarose gel electrophoresis: Implications for parentage testing and forensic individualization. *Genomics* **5**, 915–924 (1989).
- Gatti, R., Nakamura, Y., Nussmeier, M., Susi, E., Shan, W. & Grody, W. Informativeness of VNTR genetic markers for detecting chimerism after bone marrow transplantation. *Disease Markers* **7**, 105–112 (1989).
- Vogelstein, B., Fearon, E. R., Kern, S. E., Hamilton, S. R., Preisinger, A. C., Nakamura, Y. *et al.* Allelotype of colorectal carcinomas. *Science* **244**, 207–211 (1989).
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S., Presinger, A. C., Leppert, M. *et al.* Genetic alterations during colorectal tumor development. *New Eng. J. Med.* **319**, 525–532 (1988).
- Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M. *et al.* Chromosome 17 deletions and p53 mutations in colorectal carcinomas. *Science* **244**, 217–221 (1989).
- Nakamura, Y., Lathrop, M., Bragg, T., Jones, C., O'Connell, P., Leppert, M. *et al.* An extended linkage map for human chromosome 10. *Genomics* **3**, 389–392 (1988).
- Gusella, J. F., Wexler, N. S., Connelly, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–238 (1983).
- Weber, J. L. & May, P. E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396 (1989).
- Lander, E. S. & Botstein, D. Homozygosity mapping—a way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567–1570 (1987).
- Kobayashi, K., Nakahori, Y., Mizuno, K., Miyake, M., Nakamura, Y., Tokunaga, K. *et al.* Founder-haplotype analysis in Fukuyama-type congenital muscular dystrophy (FCMD). *Hum. Genet.* **103**, 323–327 (1998).
- Tsujikawa, M., Kurahashi, H., Tanaka, T., Nishida, K., Shimomura, Y., Tano, Y. *et al.* Identification of the gene responsible for gelatinous drop-like corneal dystrophy. *Nat. Genet.* **21**, 420–423 (1999).
- Onouchi, Y., Gunji, T., Burns, J. C., Shimizu, C., Newburger, J. W., Yashiro, M. *et al.* ITPKC functional polymorphism associated with Kawasaki disease susceptibility and formation of coronary artery aneurysms. *Nat. Genet.* **40**, 35–42 (2008).
- Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. & Nakamura, Y. A high-throughput SNP typing system for genome-wide association studies. *J Hum. Genet.* **46**, 471–477 (2001).
- The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
- International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
- Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T. *et al.* A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat. Genet.* **38**, 324–330 (2006).
- Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T. *et al.* Functional SNPs in the lymphotoxin- $\alpha$  gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
- Matsuda, A., Hirota, T., Akahoshi, M., Shimizu, M., Tamari, M., Miyatake, A. *et al.* Coding SNP in tenascin-C Fn-III-D domain associates with adult asthma. *Hum. Mol. Genet.* **14**, 2779–2786 (2005).
- Akahoshi, M., Obara, K., Hirota, T., Matsuda, A., Hasegawa, K., Takahashi, N. *et al.* Functional promoter polymorphism in the TBX21 gene associated with aspirin-induced asthma. *Hum. Genet.* **117**, 16–26 (2005).
- Hata, J., Matsuda, K., Ninomiya, T., Yonemoto, K., Matsushita, T., Ohnishi, Y. *et al.* Functional SNP in a Sp1-binding site of AGR1L1 gene is associated with susceptibility to brain infarction. *Hum. Mol. Genet.* **16**, 630–639 (2007).
- Kubo, M., Hata, J., Ninomiya, T., Matsuda, K., Yonemoto, K., Nakano, T. *et al.* A nonsynonymous SNP in PRKCH increases the risk of cerebral infarction. *Nat. Genet.* **39**, 212–217 (2007).
- Hirota, T., Suzuki, Y., Hasegawa, K., Obara, K., Matsuda, A., Akahoshi, M. *et al.* Functional haplotypes of IL-12B are associated with childhood atopic asthma. *J Allergy Clin. Immunol.* **116**, 789–795 (2005).
- Nakashima, K., Hirota, T., Suzuki, Y., Matsuda, A., Akahoshi, M., Shimizu, M. *et al.* Association of the RIP2 gene with childhood atopic asthma. *Allergol. Int.* **55**, 77–83 (2006).
- Yamazaki, K., McGovern, D., Ragoussis, J., Paolucci, M., Butler, H., Jewell, D. *et al.* Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum. Mol. Genet.* **14**, 3499–3506 (2005).
- Shimazaki, S., Kawamura, Y., Kanazawa, A., Sekine, A., Saito, S., Tsunoda, T. *et al.* Genetic variations in the gene encoding ELMO1 are associated with susceptibility to diabetic nephropathy. *Diabetes* **54**, 1171–1178 (2005).
- Kamiyama, M., Kobayashi, M., Araki, S., Iida, A., Tsunoda, T., Kawai, K. *et al.* Polymorphisms in the 3'UTR in the neurocalcin  $\delta$  gene affect mRNA stability, and confer susceptibility to diabetic nephropathy. *Hum. Genet.* **122**, 397–407 (2007).
- Mushiroda, T., Wattanapokayakit, S., Takahashi, A., Nukiwa, T., Kudoh, S., Ogura, T., *et al.*, Clinical Study Group. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. *J Med. Genet.* **45**, 654–656 (2008).
- Takei, T., Iida, A., Nitta, K., Tanaka, T., Ohnishi, Y., Yamada, R. *et al.* Association between single-nucleotide polymorphisms in selectin genes and IgA nephropathy **70**, 781–786 (2002).
- Akiyama, F., Tanaka, T., Yamada, R., Ohnishi, Y., Tsunoda, T., Maeda, S. *et al.* Single-nucleotide polymorphisms in the class II region of the major histocompatibility complex in Japanese patients with immunoglobulin A nephropathy. *J Hum. Genet.* **47**, 532–538 (2002).
- Obara, W., Iida, A., Suzuki, Y., Tanaka, T., Akiyama, F., Maeda, S. *et al.* Association of single-nucleotide polymorphisms in the polymeric immunoglobulin receptor gene with Immunoglobulin A nephropathy (IgAN) in Japanese patients. *J Hum. Genet.* **48**, 293–299 (2003).
- Ohtsubo, S., Iida, A., Nitta, K., Tanaka, T., Yamada, R., Ohnishi, Y. *et al.* Association of a single-nucleotide polymorphism in the immunoglobulin u-binding protein 2 gene with immunoglobulin A nephropathy. *J Hum. Genet.* **50**, 30–35 (2005).
- Seki, S., Kawaguchi, Y., Chiba, K., Mikami, Y., Kizawa, H., Oya, T. *et al.* A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. *Nat. Genet.* **37**, 607–612 (2005).
- Ozaki, K., Inoue, K., Sato, H., Iida, A., Ohnishi, Y., Sekine, A. *et al.* Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin- $\alpha$  secretion *in vitro*. *Nature* **429**, 72–75 (2004).
- Ozaki, K., Sato, H., Iida, A., Mizuno, H., Nakamura, T., Miyamoto, Y. *et al.* A functional SNP in PSMA6 confers risk of myocardial infarction in the Japanese population. *Nat. Genet.* **38**, 921–925 (2006).
- Ishii, N., Sato, H., Iida, A., Mizuno, H., Nakamura, T., Miyamoto, Y. *et al.* Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum. Genet.* **51**, 1087–1099 (2006).
- Ebana, Y., Ozaki, K., Inoue, K., Sato, H., Iida, A., Lwin, H. *et al.* A functional SNP in ITIH3 is associated with susceptibility to myocardial infarction. *J Hum. Genet.* **52**, 220–229 (2007).
- Kizawa, H., Kou, I., Iida, A., Sudo, A., Miyamoto, Y., Fukuda, A. *et al.* An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis. *Nat. Genet.* **37**, 138–144 (2005).
- Miyamoto, Y., Mabuchi, A., Shi, D., Kubo, T., Takatori, Y., Saito, S. *et al.* A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat. Genet.* **39**, 529–533 (2007).
- Mototani, H., Iida, A., Nakajima, M., Furuichi, T., Miyamoto, Y., Tsunoda, T. *et al.* A functional SNP in EDG2 increases susceptibility to knee osteoarthritis in Japanese. *Hum. Mol. Genet.* **17**, 1790–1797 (2008).
- Miyamoto, Y., Shi, D., Nakajima, M., Ozaki, K., Sudo, A., Kotani, A. *et al.* Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat. Genet.* **40**, 994–998 (2008).
- Suzuki, A., Yamada, R., Chang, X., Tokuhira, S., Sawada, T., Suzuki, M. *et al.* Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
- Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T. *et al.* An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat. Genet.* **35**, 341–348 (2003).
- Kochi, Y., Yamada, R., Suzuki, A., Harley, J. B., Shirasawa, S., Sawada, T. *et al.* A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–485 (2005).
- Suzuki, A., Yamada, R., Kochi, Y., Sawada, T., Okada, Y., Matsuda, K. *et al.* Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population. *Nat. Genet.* **40**, 1224–1229 (2008).
- Kamatani, Y., Matsuda, K., Ohishi, T., Ohtsubo, S., Yamazaki, K., Iida, A. *et al.* Identification of a significant association of a single nucleotide polymorphism in TNXB with systemic lupus erythematosus in a Japanese population. *J Hum. Genet.* **53**, 64–73 (2008).
- Oishi, T., Iida, A., Ohtsubo, S., Kamatani, Y., Usami, M., Takei, T. *et al.* A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to

- Systemic Lupus Erythematosus in Japanese population. *J Hum. Genet.* **53**, 151–162 (2008).
- 51 Kanazawa, A., Tsukada, S., Sekine, A., Tsunoda, T., Takahashi, A., Kashiwagi, A. *et al.* Association of the gene encoding wingless-type mammary tumor virus integration-site family member 5B (WNT5B) with type 2 diabetes. *Am. J. Hum. Genet.* **75**, 832–843 (2004).
- 52 Maeda, S., Tsukada, S., Kanazawa, A., Sekine, A., Tsunoda, T., Koya, D. *et al.* Genetic variations in the gene encoding TFAP2B are associated with type 2 diabetes mellitus. *J Hum. Genet.* **50**, 283–292 (2005).
- 53 Kanazawa, A., Kawamura, Y., Sekine, A., Iida, A., Tsunoda, T., Kashiwagi, A. *et al.* Single nucleotide polymorphisms in the gene encoding Kruppel-like factor 7 are associated with type 2 diabetes. *Diabetologia* **48**, 1315–1322 (2005).
- 54 Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G. *et al.* SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **40**, 1098–1102 (2008).
- 55 Mushiroya, T., Ohnishi, Y., Saito, S., Takahashi, A., Kikuchi, Y., Saito, S. *et al.* Association of VKORC1 and CYP2C9 polymorphisms with warfarin dose requirements in Japanese patients. *J Hum. Genet.* **51**, 249–253 (2006).
- 56 Chantarangsu, S., Mushiroya, T., Mahasirimongkol, S., Kiertiburanakul, S., Sungkanuparph, S., Manosuthi, W. *et al.* HLA-B\*3505 allele is a strong predictor for nevirapine-induced skin adverse drug reactions in Thai HIV-infected patients. *Pharmacogenetics Genomics* (2008) (in press).
- 57 van Es, M. A., van Vught, P. W., Blauw, H. M., Franke, L., Saris, C. G., Van den Bosch, L. *et al.* Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.* **40**, 29–31 (2008).
- 58 Stefansson, H., Rye, D. B., Hicks, A., Petursson, H., Ingason, A., Thorgeirsson, T. E. *et al.* A genetic risk factor for periodic limb movements in sleep. *N. Engl. J. Med.* **357**, 639–647 (2007).
- 59 Ober, C., Tan, Z., Sun, Y., Possick, J. D., Pan, L., Nicolae, R. *et al.* Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. *N. Engl. J. Med.* **358**, 1682–1691 (2008).
- 60 Moffatt, M. F., Kabesch, M., Liang, L., Dixon, A. L., Strachan, D., Heath, S. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007).
- 61 Gudbjartsson, D. F., Arnar, D. O., Helgadóttir, A., Gretarsdóttir, S., Holm, H., Sigurdsson, A. *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357 (2007).
- 62 Stacey, S. N., Gudbjartsson, D. F., Sulem, P., Bergthorsson, J. T., Kumar, R., Thorleifsson, G. *et al.* Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat. Genet.* **40**, 1313–1318 (2008).
- 63 Stacey, S. N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S. A., Jonsson, G. F. *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **40**, 703–706 (2008).
- 64 Ferreira, M. A., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.* **40**, 1056–1058 (2008).
- 65 Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* **447**, 661–678 (2007).
- 66 Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., *et al.*, SEARCH collaborators; kConFab; AOCs Management Group. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
- 67 Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**, 870–874 (2007).
- 68 Di Bernardo, M. C., Crowther-Swanepoel, D., Broderick, P., Webb, E., Sellick, G., Wild, R. *et al.* A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat. Genet.* **40**, 1204–1210 (2008).
- 69 Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., *et al.*, CORGI Consortium. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39**, 984–988 (2007).
- 70 Zanke, B. W., Greenwood, C. M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S. M. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **39**, 989–994 (2007).
- 71 Broderick, P., Carvajal-Carmona, L., Pittman, A. M., Webb, E., Howarth, K., Rowan, A., *et al.*, CORGI Consortium. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.* **39**, 1315–1317 (2007).
- 72 McPherson, R., Pertsemidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
- 73 Buch, S., Schafmayer, C., Völzke, H., Becker, C., Franke, A., von Eller-Eberstein, H. *et al.* A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat. Genet.* **39**, 995–999 (2007).
- 74 Study Group of Millennium Genome Project for Cancer, Sakamoto, H., Sakamoto, H., Yoshimura, K., Saeki, N., Katai, H., Shimoda, T., Matsuno, Y. *et al.* Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.* **40**, 730–740 (2008).
- 75 Thorleifsson, G., Magnusson, K. P., Sulem, P., Walters, G. B., Gudbjartsson, D. F., Stefansson, H. *et al.* Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* **317**, 1397–1400 (2007).
- 76 Dehghan, A., Köttgen, A., Yang, Q., Hwang, S. J., Kao, W. L., Rivadeneira, F. *et al.* Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet* (2008). e-pub ahead of print 30 Sept 2008.
- 77 Hung, R. J., McKay, J. D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D. *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**, 633–637 (2008).
- 78 Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T. *et al.* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* **40**, 616–622 (2008).
- 79 Thorgeirsson, T. E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K. P. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–642 (2008).
- 80 International Multiple Sclerosis Genetics Consortium, Hafler, D. A., Compston, A., Sawcer, S., Lander, E. S., Daly, M. J., De Jager, P. L. *et al.* Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* **357**, 851–862 (2007).
- 81 Helgadóttir, A., Thorleifsson, G., Manolescu, A., Gretarsdóttir, S., Blondal, T., Jonasdóttir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
- 82 Maris, J. M., Mosse, Y. P., Bradfield, J. P., Hou, C., Monni, S., Scott, R. H. *et al.* Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N. Engl. J. Med.* **358**, 2585–2593 (2008).
- 83 Richards, J. B., Yuan, X., Geller, F., Waterworth, D., Bataille, V., Glass, D. *et al.* Male-pattern baldness susceptibility locus at 20p11. *Nat. Genet.* **40**, 1282–1284 (2008).
- 84 Gudmundsson, J., Sulem, P., Manolescu, A., Amundadóttir, L. T., Gudbjartsson, D., Helgason, A. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* **39**, 631–637 (2007).
- 85 Gudmundsson, J., Sulem, P., Steinthorsdóttir, V., Bergthorsson, J. T., Thorleifsson, G., Manolescu, A. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
- 86 Winkelmann, J., Schormair, B., Lichtner, P., Ripke, S., Xiong, L., Jalilzadeh, S. *et al.* Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat. Genet.* **39**, 1000–1006 (2007).
- 87 Hofmann, S., Franke, A., Fischer, A., Jacobs, G., Nothnagel, M., Gaede, K. I. *et al.* Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat. Genet.* **40**, 1103–1106 (2008).
- 88 O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskva, V., *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40**, 1053–1055 (2008).
- 89 International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley, J. B., Alarcón-Riquelme, M. E., Criswell, L. A., Jacob, C. O., Kimberly, R. P., Moser, K. L. *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat. Genet.* **40**, 204–210 (2008).
- 90 Kozryev, S. V., Abelson, A. K., Wojcik, J., Zaghlool, A., Linga Reddy, M. V., Sanchez, E. *et al.* Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 211–216 (2008).
- 91 Hom, G., Graham, R. R., Modrek, B., Taylor, K. E., Ortmann, W., Garnier, S. *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.* **358**, 900–909 (2008).
- 92 Hakonarson, H., Grant, S. F., Bradfield, J. P., Marchand, L., Kim, C. E., Glessner, J. T. *et al.* A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* **448**, 591–594 (2007).
- 93 Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
- 94 Steinthorsdóttir, V., Thorleifsson, G., Reynisdóttir, I., Benediktsson, R., Jonsdóttir, T., Walters, G. B. *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770–775 (2007).
- 95 Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena, R., Voight, B. F., Lyssenko, V., Burt, N. P., de Bakker, P. I., Chen, H. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
- 96 Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H. *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **40**, 1092–1097 (2008).
- 97 Kiemeny, L. A., Thorlacius, S., Sulem, P., Geller, F., Aben, K. K., Stacey, S. N. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* **40**, 1307–1312 (2008).
- 98 Inoue, K., Lupski, J. R. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* **3**, 199–242 (2002).
- 99 Lupski, J. R. Genomic rearrangements and sporadic disease. *Nat. Genet.* **39**, S43–S47 (2007).
- 100 Padiath, Q. S., Saigoh, K., Schiffmann, R., Asahara, H., Yamada, T., Koepfen, A. *et al.* Lamin B1 duplications cause autosomal dominant leukodystrophy. *Nat. Genet.* **38**, 1114–1123 (2006).
- 101 Le Marechal, C., Masson, E., Chen, J. M., Morel, F., Ruzsniwski, P. *et al.* Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat. Genet.* **38**, 1372–1374 (2006).

- 102 Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- 103 Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
- 104 Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
- 105 Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M. *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **37**, 727–732 (2005).
- 106 Conrad, D. F., Andrews, T. D., Carter, N. P., Hurler, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* **38**, 75–81 (2006).
- 107 Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
- 108 Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G. *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
- 109 Aitman, T. J., Dong, R., Vyse, T. J., Norsworthy, P. J., Johnson, M. D., Smith, J. *et al.* Copy number polymorphism in Fcγ3 predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
- 110 Yang, Y., Chung, E. K., Wu, Y. L., Savelli, S. L., Nagaraja, H. N., Zhou, B. *et al.* Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am. J. Hum. Genet.* **80**, 1037–1054 (2007).
- 111 Fanciulli, M., Norsworthy, P. J., Petretto, E., Dong, R., Harper, L., Kamesh, L. *et al.* FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat. Genet.* **39**, 721–723 (2007).
- 112 Fajans, S. S., Bell, G. I. & Polonsky, K. S. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N. Engl. J. Med.* **345**, 971–980 (2001).
- 113 Winckler, W., Weedon, M. N., Graham, R. R., McCarroll, S. A., Purcell, S., Almgren, P. *et al.* Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* **56**, 685–693 (2007).
- 114 Johansson, S., Raeder, H., Eide, S. A., Midtjell, K., Hveem, K., Søvik, O. *et al.* Studies in 3523 Norwegians and meta-analysis in 11,571 subjects indicate that variants in the hepatocyte nuclear factor 4 alpha (HNF4A) P2 region are associated with type 2 diabetes in Scandinavians. *Diabetes* **56**, 3112–3117 (2007).
- 115 Holmkvist, J., Almgren, P., Lyssenko, V., Lindgren, C. M., Eriksson, K. F., Isomaa, B. *et al.* Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes. *Diabetes* **57**, 1738–1744 (2008).
- 116 Gloy, A. L., Weedon, M. N., Owen, K. R., Turner, M. J., Knight, B. A., Hitman, G. *et al.* Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* **52**, 568–572 (2003).
- 117 Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J. *et al.* The common PPAR gamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
- 118 Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
- 119 Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
- 120 Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., *et al.*, Wellcome Trust Case Control Consortium (WTCCC). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
- 121 Inoue, H., Tanizawa, Y., Wasson, J., Behn, P., Kalidas, K., Bernal-Mizrachi, E. *et al.* A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat. Genet.* **20**, 143–148 (1998).
- 122 Sandhu, M. S., Weedon, M. N., Fawcett, K. A., Wasson, J., Debenham, S. L., Daly, A. *et al.* Common variants in WFS1 confer risk of type 2 diabetes. *Nat. Genet.* **39**, 951–953 (2007).
- 123 Armanios, M. Y., Chen, J. J., Cogan, J. D., Alder, J. K., Ingersoll, R. G., Markin, C. *et al.* Telomerase mutations in families with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* **356**, 1317–1326 (2007).
- 124 Tsakiri, K. D., Cronkhite, J. T., Kuan, P. J., Xing, C., Raghu, G., Weissler, J. C. *et al.* Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc. Natl. Acad. Sci. USA* **104**, 7552–7557 (2007).
- 125 Anderson, J. L., Horne, B. D., Stevens, S. M., Grove, A. S., Barton, S., Nicholas, Z. P. *et al.* Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. *Circulation* **116**, 2563–2570 (2007).
- 126 D'Andrea, G., D'Ambrosio, R. L., Di Perna, P., Chetta, M., Santacrose, R., Brancaccio, V. *et al.* A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin. *Blood* **105**, 645–649 (2005).
- 127 Snow, J. L. & Gibson, L. E. The role of genetic variation in thiopurine methyltransferase activity and the efficacy and/or side effects of azathioprine therapy in dermatologic patients. *Arch. Dermatol.* **131**, 193–197 (1995).
- 128 Innocenti, F., Undevia, S. D., Iyer, L., Chen, P. X., Das, S., Kocherginsky, M. *et al.* Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J. Clin. Oncol.* **22**, 1382–1388 (2004).
- 129 Chung, W. H., Hung, S. I., Hong, H. S., Hsieh, M. S., Yang, L. C., Ho, H. C. *et al.* Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* **428**, 486 (2004).
- 130 Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A. M., Moore, C. *et al.* Association between presence of HLA-B\*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* **359**, 727–732 (2002).
- 131 Furuta, T., Shirai, N., Kodaira, M., Sugimoto, M., Nogaki, A., Kuriyama, S. *et al.* Pharmacogenomics-based tailored versus standard therapeutic regimen for eradication of *H. pylori*. *Clin. Pharmacol. Ther.* **81**, 521–528 (2007).
- 132 Sauer, J. M., Ponsler, G. D., Mattiuz, E. L., Long, A. J., Witcher, J. W., Thomasson, H. R. *et al.* Disposition and metabolic fate of atomoxetine hydrochloride: the role of CYP2D6 in human disposition and metabolism. *Drug Metab. Dispos.* **31**, 98–107 (2003).
- 133 Sallee, F. R., DeVane, C. L. & Ferrell, R. E. Fluoxetine-related death in a child with cytochrome P-450 2D6 genetic deficiency. *J. Child. Adolesc. Psychopharmacol.* **10**, 27–34 (2000).
- 134 Gasche, Y., Daali, Y., Fathi, M., Chiappe, A., Cottini, S., Dayer, P. *et al.* Codeine intoxication associated with ultrarapid CYP2D6 metabolism. *N. Engl. J. Med.* **351**, 2827–2831 (2004).
- 135 Goetz, M. P., Knox, S. K., Suman, V. J., Rae, J. M., Safgren, S. L., Ames, M. M. *et al.* The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Treat.* **101**, 113–121 (2007).
- 136 Kiyotani, K., Mushiroma, T., Sasa, M., Bando, Y., Sumitomo, I., Hosono, N. *et al.* Impact of CYP2D6\*10 on recurrence-free survival in breast cancer patients receiving adjuvant tamoxifen therapy. *Cancer Science* **99**, 995–999 (2008).
- 137 Ohno, M., Yamaguchi, I., Yamamoto, I., Fukuda, T., Yokota, S., Maekura, R. *et al.* Slow N-acetyltransferase 2 genotype affects the incidence of isoniazid and rifampicin-induced hepatotoxicity. *Int. J. Tuberc. Lung Dis.* **4**, 256–261 (2000).
- 138 Taniguchi, A., Urano, W., Tanaka, E., Furihata, S., Kamitsuji, S., Inoue, E. *et al.* Validation of the associations between single nucleotide polymorphisms or haplotypes and responses to disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a proposal for prospective pharmacogenomic study in clinical practice. *Pharmacogenet. Genomics* **17**, 383–390 (2007).
- 139 Gatanaga, H., Hayashida, T., Tsuchiya, K., Yoshino, M., Kuwahara, T., Tsukada, H. *et al.* Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2B6 \*6 and \*26. *Clin. Infect. Dis.* **45**, 1230–1237 (2007).
- 140 Haufroid, V., Mourad, M., Van Kerckhove, V., Wawrzyniak, J., De Meyer, M., Eddour, D. C. *et al.* The effect of CYP3A5 and MDR1 (ABCB1) polymorphisms on cyclosporine and tacrolimus dose requirements and trough blood levels in stable renal transplant patients. *Pharmacogenetics* **14**, 147–154 (2004).
- 141 Kato, M., Fukuda, T., Serretti, A., Wakeno, M., Okugawa, G., Ikenaga, Y. *et al.* ABCB1 (MDR1) gene polymorphisms are associated with the clinical response to paroxetine in patients with major depressive disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **32**, 398–404 (2008).
- 142 Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F. *et al.* SLC01B1 variants and statin-induced myopathy—a genomewide study. *N. Engl. J. Med.* **359**, 789–799 (2008).
- 143 Shu, Y., Sheardown, S. A., Brown, C., Owen, R. P., Zhang, S., Castro, R. A. *et al.* Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J. Clin. Invest.* **117**, 1422–1431 (2007).