

# Genetic Variation and its Role in Malignancy

Bente A. Talseth-Palmer<sup>1, 2</sup>, Rodney J. Scott<sup>1, 2, 3</sup>

<sup>1</sup>*School of Biomedical Sciences and Pharmacy, University of Newcastle, Australia;*

<sup>2</sup>*Hunter Medical Research Institute, John Hunter Hospital, Newcastle, Australia;*

<sup>3</sup>*Hunter Area Pathology Service, Hunter New England Area Health, Newcastle, Australia*

## ABSTRACT

Genetic variation has long been thought associated with common complex disease and has therefore been widely studied. Genetic variation in the human genome is present in many forms and have been summarised in this review. The potential role of DNA damage, DNA repair and environmental influence on genetic variation in the development of cancer will be discussed, before significant genome projects are reviewed. All the various forms of genetic variation have been associated with malignancies and have been extensively studied and this is a review of the state of the field. We also discuss the road ahead in fulfilling the ultimate goal in all cancer genetic studies, which is decreasing deaths caused by cancer. (*Int J Biomed Sci* 2011; 7 (3): 158-171)

**Keywords:** DNA damage and repair; environmental influence; genetic variation; malignancy

## BACKGROUND

Most diseases are multifactorial and are a result of an interaction between genetic and environmental factors that play a role in disease development throughout life. There is accumulating evidence indicating that genetic variation accounts for a proportion of susceptibility to common diseases such as diabetes, cardiovascular disease and cancer (1-3). The identification of genetic variation associated with common complex disorders is therefore a priority in elucidating the pathophysiological processes that underlie common human afflictions. In the past decades there has been an increasing interest in the possible impact of common, functional germline polymorphisms on clinical outcomes among patients with cancer.

Genetic variation refers to the genomic differences seen in a population or species (4). Because of the great diversity in the human genome, genetic variation is regarded as a parameter, which controls an individual's phenotype (5). Genetic diversity refers to variation at the level of individual genes and provides a mechanism for population survival by their potential to adapt to an ever-changing environment. Within and between populations genetic variation has long been thought to be the key to the biology of human disease (6-8). Even though all humans are members of the same species no two individuals are exactly alike and even identical twins have slight differences in their DNA. Between any two humans, the amount of genetic variation is about 0.1 percent (9-11).

Genetic variation in the human genome is present in many forms and occurs at different frequencies throughout the genome. The different forms of genetic variation includes tandemly repeated DNA, single nucleotide polymorphisms (SNPs), small insertions/deletions, large-scale mutations, transposable elements, fragile sites and null alleles.

Tandemly repeated (satellite) DNA appears to be the main type of repeated DNA that accounts for the enormous

---

**Corresponding author:** Bente A. Talseth-Palmer, HMRI, Room 3642, Level 3, John Hunter Hospital, Lookout Rd, New Lambton Heights, NSW 2305, Australia. Tel: +61 (0)2 4985 5663; Fax: +61 (0)2 4985 5895; E-mail: Bente.Talseth-Palmer@newcastle.edu.au.

**Received** January 24, 2011; **Accepted** February 16, 2011

**Copyright:** © 2011 Bente A. Talseth-Palmer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.5/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

variability observed in genome size among eukaryotes (12, 13). Tandemly repeated DNA can be divided into satellite DNA (highly repetitive DNA with repeat lengths of one thousand to several thousand base pairs (bp)), minisatellite DNA (moderately repetitive, 9-100 bp) and microsatellites (di-, tri- and tetra nucleotide repeats). Minisatellites are extremely polymorphic, which have made them excellent markers for DNA finger printing and linkage analysis (14-16). Microsatellites are very useful genetic markers (linkage analysis) as they are highly polymorphic, co-dominant and are spread approximately every 50 kb across the entire human genome (17, 18). Microsatellites therefore quickly replaced minisatellites and represented the markers of choice for gene mapping until SNPs were discovered.

SNPs represent the major source of genetic variation in the human genome and account for approximately 90 percent of all human genetic variation occurring approximately every 100 to 300 bases. For a variation to be considered a SNP, the least frequent allele has to have an abundance of 1 percent or greater in a given population (19, 20). According to the neutral theory of molecular evolution, most SNPs are maintained in natural populations due to their location in non-coding regions and as such their distribution is not under selective pressure (21). When occurring in a gene, SNPs can be a causative genetic variant that changes protein expression, directly contributing to a disease phenotype (22). Together with SNPs, small insertions and deletions account for the vast majority of genetic variation observed in DNA. Small insertions and deletions (1-30 bp) in the coding regions of genes can, but not always, lead to frame-shift mutations causing a severely altered and potentially non-functional protein.

Large-scale mutations in chromosomal structure include amplification and deletion of large chromosome regions known as copy number variation (CNV), translocations and loss of heterozygosity (LOH). From about the beginning of the last decade researchers highlighted DNA CNV as a large under-explored source of human genetic variation that may be a factor underlying a range of genetic diseases (23-26). Copy number refers to the amount of copies of a particular gene and CNV represents a copy number change involving a DNA fragment of 1 kilobase or larger and contributes to genomic diversity observed between humans (10). CNVs can include genes and their regulatory regions (27-29), suggesting that these variants may be important in our understanding of disease or phenotypic variation. The copy number of a particular gene can be related to gene expression levels and consequently the protein that it encodes (30).

Less frequent genetic variation includes transposable elements which can be divided into two main classes; DNA transposons and retrotransposons. DNA transposons in humans make up approximately 3 percent of the genome and appear to have become entirely inactive (31). Retrotransposons can be separated into long interspersed nuclear elements (LINEs) - AT rich region of 6-8 kb with internal promoters for RNA polymerase III, short interspersed nuclear elements (SINEs) - GC rich dimeric sequences of ~300 bp in size located in untranslated intronic regions and long terminal repeat (LTR)-retrotransposons. They make up about 21 percent, 13.5 percent and 8.5 percent of the genome, respectively (32, 33). Another form of genetic variation resides in rare fragile sites, which occur in less than 5 percent of the human genome and are comprised of di- and tri nucleotide repeats that may cause spontaneous breaks during DNA replication, frequently affecting neighbouring genes (34). Rare fragile sites are associated with the expansion of unstable repeat sequences (35, 36) and segregate in specific families. Common fragile sites in the human genome are part of the normal chromosomal structure; they are large regions of genomic instability and targets for chromosomal rearrangements and deletions (34). Null alleles, also a form of genetic variation, are the outcome of a mutant copy of a gene that lacks normal function due to the absence of a gene product, or the expression of a non-functional gene product (37).

DNA sequence variation can also be classified according to frequency. Mutations are rare sequence variants that are found in less than 1 percent of the population and typically have a major influence on disease development. Deleterious mutations can be subdivided into inborn error (embryonic lethal), germline mutations (hereditary disease) and somatic mutations, which are spontaneous mutations in cells that occurs throughout the body that could result in the malfunctioning of genes and consequently lead to common disease development (38). Polymorphisms are common DNA sequence variants that are classified as being present in over 1 percent of the population, and they are thought to have either a weak or no influence on disease development (39). However, the presence of multiple variations could potentially influence individual disease risk. The consequence of DNA sequence variation is genetic variability. Interpreting the functional significance of genetic variation in a population poses a major challenge since genetic variation between individuals is required for the long-term adaptation to environmental challenges (40).

Epigenetic variation, non-sequence based alterations (DNA methylation, histone modification and chromatin

remodelling), that are inherited through cell division and variance in the reading frames of microRNAs (miRNAs) may also disrupt gene function (41-43). miRNA's are small noncoding RNAs that regulate gene expression post-transcriptionally by binding of target mRNAs to regulate their stability or translation (44).

Individual susceptibility to cancer is influenced by the ability to repair DNA damage, sensitivity to DNA damaging agents and defects in DNA repair proteins, which have been associated with several human hereditary cancer syndromes (45-47). Genetic variation in DNA repair genes as a risk factor for disease is very apparent in rare cancers such as Ataxia telangiectasia, Fanconi's anaemia, Nijmegen breakage syndrome, familial breast cancer, hereditary nonpolyposis colorectal cancer (HNPCC) and xeroderma pigmentosum (XP) (48-51). Genetic variation is now also being used as a tool to screen patients for targeted treatment. For example, the detection of somatically acquired mutations in the *KRAS* gene has been shown to interfere with the efficacy of new targeted treatments (EGFR-targeted mAbs) for bowel, head and neck, and lung cancers such that the treatment is only recommended to individuals harbouring the wild-type *KRAS* (52). The impact of genomic studies has been the ability to investigate the biology behind disease and cancer development in a comprehensive, unbiased, hypothesis-free manner (33, 53).

This review will discuss the potential role of DNA damage, DNA repair and environmental influence on genetic variation in the development of cancer, with a particular focus on the associations involving genetic variation and malignancy. As human genetic variation is an important topic in a rapidly changing genetics field this article is meant to be a review of the state of the field.

## DNA DAMAGE AND REPAIR

DNA is constantly exposed to external and internal mutagenic agents (i.e. free radicals, ionising agents, ultraviolet light, plant toxins and various chemical agents) that could potentially influence the integrity of the genome. If DNA damage is not repaired it can result in disruption of genomic integrity and alter the risk of malignancy. Many of the DNA lesions caused by mutagenic exposure are associated with structural damage, which can impact on the cells ability to function appropriately. DNA repair is not perfect and as such reflects a balance between maintaining genomic integrity and allowing sufficient differences to be transmitted from generation to generation thereby maintaining the potential of evolutionary change (54).

Failure to repair the effects of DNA damage can result in dysregulated cell growth which can ultimately result in cancer. It is therefore extremely important that the DNA repair system is constitutively active so it can respond to the induction of DNA damage. A wide range of intracellular products and bi-products can result in DNA damage which are considered to be major factors in mutagenesis, carcinogenesis and ageing (55, 56).

The DNA damage response is a series of molecular events that recognise, respond and result in DNA repair. The repair processes include removal of DNA damage, restoration of the integrity of the DNA helix, activation of DNA damage checkpoints, changes in the transcriptional profile that might be beneficial to cells and apoptosis (57, 58). Analysis of mutations in cancer cells has identified the importance of DNA damage response in preventing tumorigenesis (59, 60). The response to DNA damage implicates cell cycle checkpoint responses, preventing the expansion of cells that have sustained damage by initiating DNA repair or by activating the induction of apoptosis if there is too much damage (61).

In an individual cell, thousands of DNA lesions are repaired every day by different DNA repair systems that have overlapping repair specificities. Approximately 150 DNA repair gene products have been identified (62, 63) and there appears to be 6 major repair pathways: 1) DNA repair by direct reversal, 2) base excision repair, 3) mismatch repair, 4) nucleotide excision repair, 5) homologous recombination and 6) non-homologous end joining (58, 64-73). The six repair mechanisms function together to protect DNA from bi-products of cellular metabolism and environmental insult to maintain genome integrity.

The main function of chromatin is to package DNA into smaller volume and efficient DNA repair is complicated by the chromatin structure, a highly condensed structure that hinders DNA accessibility and its subsequent repair (64, 74). Despite this, DNA repair systems are extremely efficient. In normal cells, 50 percent of single- and double strand breaks induced by environmental ionising radiation are repaired within 15 minutes and complete repair occurs within 1-2 hours (75). Cancer patients have a significantly higher level of basal DNA damage than healthy controls (76). It has been shown that young female breast cancer patients have a much lower DNA repair capacity than their healthy female sibling, and that obesity might be a factor that is involved in DNA repair capacity (77).

Genetic variation, such as polymorphisms, present in DNA repair genes may have modifying effects on cancer risk influenced by the efficiency of DNA repair creating

genomic instability. According to Madhusudan *et al.* (78) studies suggest that variation in DNA repair capacity in the normal population may influence cancer susceptibility and outcome, and deficiencies in DNA repair capacity have been associated with an increased risk of breast cancer (79). DNA repair polymorphisms have also been shown to be associated with individual levels of DNA damage, thereby modulating cancer risk (80). The modifying effects of genetic variation will only explain a small percentage of the differences in mutagen sensitivity observed in healthy individuals, which together with genetic variation in DNA repair genes, highlights the complexity of incorporating genetic differences into quantitative estimates of risk associated with environmentally relevant exposures (81).

## ENVIRONMENTAL INFLUENCE ON GENETIC VARIATION

The sources of exogenous damaging agents are numerous (i.e. ultraviolet light, x-rays, thermal disruption and man-made mutagens), causing genetic lesions (somatic mutations) as a result of insufficient DNA repair or excessive exposure. Repetitive elements (as described above) show a unique capacity to respond to environmental signals and tend to cluster in genes associated with externally triggered processes. Mutations in repetitive elements are associated with adaptive changes of phenotype in natural populations and certain types of mutations convey adaptive benefits (82). It is thought that environmental exposures early in development influences epigenetic changes, which may play a role in susceptibility to diseases later in life and disease phenotypes through modification of the epigenome (83, 84).

The development of cancer is an interplay between the accumulation of unrepaired DNA damage, epigenetic variation, environmental factors and DNA repair capacity, and an intricate balance between them is necessary to maintain genome integrity. For example, individuals exposed to high levels of benzene may have an increased risk for genotoxicity influencing cancer risk, due to decreased DNA repair capacity as a result of polymorphisms in susceptibility genes involved in xenobiotic clearance (85). Since both genetic and environmental factors influence the levels of enzymes that metabolically activate and detoxify chemicals (86), they are also considered to influence cancer risk. The relationship between genetic variation and environmental influence on disease risk has also been demonstrated between two non-synonymous *xe-*

*roderma pigmentosum D (XPD)* polymorphisms and sun exposure on skin cancer (87), and between a *XPD* polymorphism and the risk of breast cancer, especially from polycyclic aromatic hydrocarbon (PAH)-DNA adducts found in cigarette smoke (88). Furthermore, it has been shown that diets consisting of protective micronutrients as well as carcinogens and mutagens may alter the risk of malignancy, particularly in individuals who are genetically susceptible as a result of genetic variation (89). In addition, the intestinal microenvironment has been shown to influence tumour formation in an animal model of the inherited predisposition to colorectal cancer called familial adenomatous polyposis (90).

At a population level, confirmation of the interaction between environment and genetic susceptibility is observed in populations that have migrated from a region of low cancer incidence to one where the incidence is much greater. Neuhausen *et al.* (91) suggested that ethnic variation in cancer risk is most likely a result of both genetic and epidemiological factors based on the observation that Japanese men living in Japan was shown to have the lowest incidence of prostate cancer but with migration to the United States (US) the rate of prostate cancer increased.

## SIGNIFICANT GENOME PROJECTS

The Human Genome Project (HGP) began in 1990 with the intention of determining the DNA sequence and identifying all the genes in the human genome. The first part of the HGP was completed in 2003, which resulted in the identification of all the genes in the human genome (92). The study of genetic variation in other species increases the understanding of our own and consequently a series of genome have now been sequenced that include, bacteria (*Escherichia Coli*), the fruit fly (*Drosophila melanogaster*), rice (*Oryza sativa*), yeast (*Saccharomyces cerevisiae*) and the laboratory mouse (*Mus musculus*) (93-100) to name but a few. The first phase of the HGP has been completed and currently extensive re-sequencing is being undertaken, which is revealing considerable genetic variation both between and within species. The HGP has revealed that the spectrum of protein-coding genes is smaller than originally assumed and that differences between species are more likely to be a result of variation in the control of gene expression via regions of the genome that had hitherto not been assigned a function (101). The beneficial outcomes of the HGP and genome projects studying other species are many fold and include; earlier detection of genetic predispositions to disease, better understanding of the mecha-

nisms modifying disease, energy and environmental applications (i.e. use microbial genomics to create new energy sources), risk assessment of disease, better understanding of human evolution and the common biology with other organisms, improved DNA forensics, and detailed knowledge of plant and other animal genomes, providing us the potential to develop stronger, more disease-resistant plants and animals (89, 93, 94, 102-106). The HGP genome technologies and bioinformatics tools have improved rapidly and the study of entire genomes, sets of expressed RNAs or proteins, gene families, variation among individuals, and the classes of gene regulatory elements are now being identified and their functional consequences assessed.

Only a fraction of the genome is comprised of genes encoding proteins, with the biological information contained in the genes nucleotide sequence. Coding regions in genes are thought to cover approximately 5 percent of the human genome. This is now being challenged by the results published by the ENCODE project (107), as the simple view of the genome as having a defined set of isolated loci transcribed independently does not seem to hold true. The human genome contains a considerable amount of information as it codes for not only the functioning of each and every one of us, but it is also a record for of an individual's ancestry and origin. The ENCODE project has enriched the annotation of the human DNA sequence by describing the functional elements encoded therein (108, 109).

The international HapMap project developed a haplotype map of the human genome (110, 111), which describes the common patterns of genetic variation. The information from these projects has been made available to the public to increase the identification rate of genes associated with disease, individual drug responses and response to other environmental factors. Because of the great heterogeneity across the genome in terms of patterns of genetic variation, HapMap is one of the main online databases providing information on human genetic diversity (112).

## GENETIC VARIATION AND MALIGNANCIES

As early as 1974 it was suggested that cancers must exhibit a mutator phenotype (a series of mutations) as a result of differences between the paucity of spontaneous mutations and the large number of mutations found in human tumours (113). This observation has also been confirmed in more recent studies (reviewed by Prindle *et al.* (114)). Accumulation of mutations during neoplasia represents an imbalance between DNA damage, the efficiency of DNA

damage repair and the response to un-repaired damage (115). Cancers arise as a result of an accumulative series of genetic and epigenetic changes that drive the progressive transformation of normal cells into highly malignant derivatives. An important factor for almost all cancer cells is genomic instability, ranging from the steady accumulation of mutations to gross chromosomal rearrangements and alterations in chromosome numbers (116-118).

Genetic variation does influencing the development of disease phenotype and cancer through different avenues, including genomic instability, chromatin structure and transcriptional activity (119-122). Most cancers are genetically unstable and most of the instability is observed at the chromosome level, with frequent gains and losses of large chromosome segments or entire chromosomes (123). A common pattern of unbalanced translocations, leading to loss of chromosomal material and gain of selected genes have been reported in the acute myeloid leukaemia complex karyotype (defined by the presence of abnormalities involving at least three chromosomes) (124). While losses and gains of chromosomal material have been observed in primary prostate tumours (125). Additionally, the Philadelphia chromosome, a specific chromosomal abnormality due to a reciprocal translocation involving chromosome 9 and 22, has been associated with chronic myelogenous leukaemia and is found in 95 percent of cases (126). This special translocation, also known as Bcr-Abl, has also been observed in acute lymphoblastic leukaemia and occasionally in acute myelogenous leukaemia (126).

The less frequently observed genetic variation (retrotransposons, null alleles and fragile sites) are also associated with different types of cancer. The distribution of retrotransposons has been implicated as a potential source of disease by insertional mutagenesis or their ability to influence transcription of neighbouring genes (32, 127). SINES have been associated with different types of cancers; leukaemia, ovarian carcinoma and breast cancer (128-131). Null alleles of the gene glutathione-S-transferase (GST) M1 and/or T1 have been implicated as a risk factor for lung cancer, ovarian cancer, breast cancer, bladder cancer and cancer occurrence in hereditary non-polyposis colorectal cancer (HNPCC) (132-136). There have been studies showing that fragile sites, and associated genes, are frequently deleted or rearranged in cancer cells and this has demonstrated their importance in genomic instability in tumorigenesis (137-146). FRA2B (3p14.2) and FRA16D (16q23.2) are the two most frequently expressed common fragile sites in the human genome (147, 148). *FHIT* and *WWOX* (the genes located within FRA3B and

FRA16D respectively) have both been shown to function as tumour suppressors genes, and their inactivation have been associated with a poor clinical prognosis in cancer (61, 149). Minisatellites have been associated with fragile sites and are proximal to a number of recurrent translocation breakpoints (150). A minisatellite downstream of the H-ras proto-oncogene has been associated with the risk of cancer (151). However, from 1985 and for the following 15 years, contradictory results were published on the association of this minisatellite near the H-ras gene and cancer risk. A more recent study using improved genotyping and analysis method has failed to reproduce the association (152) suggesting that variation at this site is more complex than originally thought.

Approximately 10 percent of all cancers are familial, which is defined as cancer that occurs within families at relatively high frequency and at a younger diagnostic age compared to the general population. Conversely, approximately 90 percent of cancer cases consist of non-familial, sporadic forms of cancer (153). Genetic predispositions to sporadic cancer are considered to be multifactorial, but a study on non-familial breast cancer indicated that predispositions to sporadic cancer are strongly influenced by genetic factors (154). Predisposition to disease is a combination of weak genetic variants that may be of much more significance to public health than the marked individual risk seen in the inherited cancer syndromes (155, 156). Nonsense mutations have frequently been identified in a number of inherited predispositions to cancer and include adenomatous polyposis (157), hereditary nonpolyposis colorectal cancer and sporadic colorectal cancer (158, 159), familial breast cancer (160) and multiple endocrine neoplasia type 1 (161). The identification of tumour susceptibility genes has significantly aided our understanding of the pathogenetic mechanisms underlying cancers that appear not to be associated with an inherited predisposition (162).

Over the past decade there has been an increasing interest in the possible impact of common, functional germline polymorphisms on clinical outcomes among patients with cancer (163-171). For example, a Caspase-8 (*CASP8*) polymorphism has been associated with reduced susceptibility to multiple cancers (172) while low-penetrance CRC susceptibility loci have been shown to increase the risk of developing colorectal cancer in Lynch syndrome patients (173, 174). When searching for SNPs associated with disease it is important to consider that the frequency of the variation of interest can vary significantly between populations. The allele frequency for any given SNP tends to be

population-specific (175-177), but it has also been shown that for most of the common disease associated SNPs, ethnicity is likely to be a poor predictor of an individual's genotype (178).

Identifying genes that contribute to complex disease has been and remains a major challenge. Substantial scientific debate has been generated regarding optimal strategies to localise and identify genes for complex human disorders (9, 179-184). Linkage analysis has been used successfully to map highly penetrant genes associated with monogenic disease (185), but has been less successful for the identification of low-penetrant susceptibility genes. Association studies utilising candidate gene polymorphisms that are likely to affect the tumour development were excellent for the purpose of identifying common genetic variation that confers modest disease risk. Both these approaches have now however been superseded by a genome wide approach (186-189). Both large population based studies, comparing thousand of subjects with equal number of controls, and inherited predisposition disorders are being utilised to examine genetic variation, and its association with disease (190-192).

Genome-wide association studies (GWAS's) based on the common disease - common variant hypothesis (193) have appeared as a relatively new approach for investigating the genetic basis of complex disease (194). A GWAS is designed to examine the entire genome using a large number of markers (some of which are linked to a disease allele) to discover gene loci that are different between individuals with disease compared to those without disease (195). These studies require thousands of samples (both cases and controls) to have sufficient power to detect susceptibility loci as they suffer from the problem of multiple testing that must be corrected for. GWAS's has served as an attractive approach to search for novel moderate to high-penetrance genes in high risk cancer families, as both common and rare variation may cause cancer susceptibility. In 2008 it was suggested that the variation identified by GWAS's only explain a small fraction of the overall disease risk in any given disease and from a population-wide perspective their impact seems limited (196, 197). By 2009, 600 human GWAS's examining 150 diseases were reported that found 800 SNP associations (198). Many of these studies have investigated a variety of cancers types, which started to appear in leading journals around 2007 (190, 192, 199). GWAS's have identified over 100 low-penetrant cancer susceptibility loci associated with modest disease risk ( $OR < 1.5$ ) (200, 201) and this has cast some doubts over the validity of the common disease - common variant hypoth-

esis (193). Rather than identifying single gene associations it is more likely that GWAS results will reveal molecular pathways associated with disease. The importance of replication studies for GWAS has been emphasised as only a limited number of observed variants are true risk alleles (202). Even though GWAS's have enhanced our ability to study genetic variation, the targets identified by these studies require validation through functional studies before the findings can be used in cancer prediction and prevention (200, 202). GWAS's will, however, continue to reveal new insights into tumour biology (203). Nevertheless, to fully understand the genetic basis of common malignancies a more integrated approach that includes a combination of SNP, CNV and whole-genome sequencing data will be required to provide clinically relevant information.

CNVs were discovered only after the results of the HGP became available. Early studies utilised bacterial artificial chromosomes (BAC) clones for the identification of CNVs (204-207), however, with the development of SNP array technology CNV detection has become much more straightforward (24, 208, 209). The first complete CNV map of the human genome was reported in 2006 (210) and many studies have since investigated CNV and cancer risk (reviewed by Kuiper *et al.*, 2010 (196)). CNV's have been shown to have the potential to influence cancer risk by varying the gene dosage of genes involved in tumour development and progression (211), with studies focusing on the comparison between tumour and normal cells from the same individual aimed at identifying "driver genes" for the purpose of predicting prognosis and treatment (212, 213). A CNV study of over 3000 cancer specimens identified 158 CNV regions altered across the genome and found that most of the somatic CNV's within any cancer type were common to other cancers, suggesting the existence of a combination of a limited number of functionally relevant events for cancer development (214).

Epigenetic changes are a common feature of all human cancers (215), i.e. many hyper-methylated genes have been associated with various human neoplasias (216), and can lead to genetic alteration as a result of a breakdown in key DNA repair processes like DNA mismatch repair, nucleotide excision repair and recombination repair (217). Accumulating evidence suggests that aberrant regulation or mutations of miRNAs may contribute to the pathogenesis of cancer and genomic regions containing miRNAs often are targeted for amplification, loss of heterozygosity and structural breakpoint's in tumours (218). Recently, the complete characterisation of the microRNAome in a patient with acute myeloid leukaemia was reported,

which identified novel miRNAs that were differentially expressed between the tumour and normal cells demonstrating that somatic mutations can affect gene expression (219). miRNA's have also been shown to play a role for invasion and metastasis during cancer progression (220). It has been suggested that SNPs in miRNA genes affect cancer susceptibility, response to treatment and prognosis (221). However, as miRNA SNPs are rare and minor allele frequencies low, large studies are required to confirm their relative significance.

## THE ROAD AHEAD

Next-generation sequencing, also known as massively parallel sequencing, is introducing a new era in which the poorly explored regions of the genome and their association with disease susceptibility may be revealed. In an attempt to identify the genomic landscape of cancer; targeted sequencing (for the detection of somatic mutations in cancer genomes), whole genome sequencing (matched tumour and normal genomes of a single patient) or whole transcriptome sequencing (how the somatic mutations are manifest in the genes expressed) using next-generation sequencing platforms are being utilized (reviewed by Mardis *et al.*, 2009 (189)). A major challenge for all using this new technology is how to develop appropriate bioinformatics-based approaches for the data-analysis. Next-generation sequencing has already entered the clinical research arena, with targeted sequencing of 21 genes in women with severe family histories of breast or ovarian cancer proving that widespread genetic testing for personalised risk assessment can be reliably undertaken with this technology (222). New and unexpected oncogenic mechanisms have been suggested by patterns of somatic mutations discovered after sequencing 38 multiple myeloma genomes and matched normal DNA, including genes involved in protein translation, histone methylation and blood coagulation (53). Chromatin remodelling has been shown to contribute to the pathogenesis of ovarian clear cell carcinoma through exome sequencing of 8 tumours and normal cells, which identified four genes that were mutated in at least two tumours (223).

Rare genetic variants in common cancer might explain the "missing heritability" of cancer (224), however, next-generation sequencing could prove invaluable in uncovering the roles of rare variants of major effects in common disease (225). This technology will not be able to fully address the role of rare variants (including non-coding and structural variants) in cancer until it becomes a rapid cost-

effective approach to analysing the entire genome (193). Nevertheless, if the cost and assay time decreases in near future, next-generation sequencing will probably be used as a general-purpose tool to characterise cancer genome for more accurate prognosis and tailored treatment of cancer patients (226)

Integrated oncology research, where a combination of genetics, epigenetics and epidemiology is used has been suggested as the key for future discoveries (227). Even sporadic tumour formation is a combination of genomic, genetic and epigenetic events and the combination of changes in gene-dosage, methylation-based silencing and polymorphisms causing reduced gene function greatly complicates the search for cancer genes (228).

## CONCLUSION

The genetic variation studies reviewed here only begin to describe the complex networks of change that seem to be involved in the development of malignancy. The pathways involved in the control of the genome are complex and poorly understood, underlining the difficulties in disease loci identification. The study of genetic variation that can contribute to complex disease is therefore a major challenge. Two main approaches exist, the candidate gene approach and the genome-wide approach and both methods have merit. The candidate gene approach can be used when possible targets exist, whereas genome-wide approach is more applicable when target of interest are yet to be identified or to identify new targets for disease development.

Linkage analysis has been used successfully in mapping genes associated with monogenic diseases (185, 229, 230), but is influenced by genetic and environmental heterogeneity and is not appropriate when identifying low penetrance tumour susceptible genes. Association studies with polymorphisms in candidate genes that are likely to affect tumour development and progression are excellent for the purpose of identifying common genetic variants that confer modest disease risk (187). But it has become very obvious to use well defined study populations when searching for genes with the candidate gene approach as it is more likely to reveal true associations between the genetic variant and disease (231). For example, a study using the candidate gene approach has identified common genetic variants in *vascular endothelial growth factor (VEGF)*, a gene critical for angiogenesis that might influence bladder cancer risk (186). Genome-wide association studies have been highly effective in exploring the role of genetic

variation underlying common familial diseases (232), and have identified several susceptibility loci in common malignancies (190, 192, 199, 233).

Genetic association studies may not result in a clear understanding of the causative role of any associated genes, and are not always replicated in other studies (234). This is most likely due to the fact that the majority of the genetic variation linked to complex disease has only a modest affect on disease development, and does not adequately take into account the contribution of environmental factors to disease risk. For example, a breast cancer study examining SNPs involved in the metabolism of tamoxifen found no association with any single SNP, but when a combined SNP analysis was performed, harbouring two variant alleles in the genes *sulfotransferase family cytosolic 1A phenol-preferring member 1 (SULT1A1)* and *UPD glucuronosyl transferase 2 family polypeptide B15 (UGT2B15)* revealed an increased risk of recurrence and reduced survival (235). This demonstrates that it is important to consider linked polymorphisms that tend to travel through a population together, creating haplotype blocks, when searching for disease susceptibility genes (236-238).

Future cancer sequencing projects will discover the importance of mutations in cancer development and the list of cancer genes will continue to grow. It is important to incorporate epidemiological knowledge together with genetic, epigenetic and environmental studies to increase our understanding of cancer development. The appropriate application of new technologies with what we already know about cancer will lead to new screening tests and early-detection programs for high-risk relatives as well as effective population screening for common malignancies (239). There are enormous expectations about the power of next-generation sequencing and it is to be expected that significant improvements in patient outcomes will be forthcoming.

## FUNDING

This work was supported by grants from the Hunter Medical Research Institute, Cancer Institute NSW, Gladys M. Brawn Memorial Fund through the University of Newcastle and National Health & Medical Research Council, Australia.

## REFERENCES

- Schmith VD, Campbell DA, Sehgal S, *et al.* Pharmacogenetics and disease genetics of complex diseases. *Cell Mol. Life Sci.* 2003; 60 (8): 1636-1646.

2. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100 (1): 57-70.
3. Eccles D, Tapper W. The influence of common polymorphisms on breast cancer. *Cancer Treat Res*. 2010; 155: 15-32.
4. Nevo E. Genetic variation in natural populations: patterns and theory. *Theor. Popul. Biol.* 1978; 13 (1): 121-177.
5. Kaneko K, Furusawa C. An evolutionary relationship between genetic variation and phenotypic fluctuation. *J. Theor. Biol.* 2005.
6. McKeigue PM. Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am. J. Hum. Genet.* 1997; 60 (1): 188-196.
7. Shriver MD. Ethnic variation as a key to the biology of human disease. *Ann. Intern. Med.* 1997; 127 (5): 401-403.
8. Shriver MD, Mei R, Parra EJ, *et al.* Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. *Hum. Genomics*. 2005; 2 (2): 81-89.
9. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet.* 2001; 17 (9): 502-510.
10. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat. Rev. Genet.* 2006; 7 (2): 85-97.
11. Przeworski M, Hudson RR and Di Rienzo A: Adjusting the focus on human variation. *Trends Genet.* 2000; 16 (7): 296-302.
12. Gregory TR, Nicol JA, Tamm H, *et al.* Eukaryotic genome size databases. *Nucleic. Acids Res.* 2007; 35 (Database issue): D332-338.
13. Vinogradov AE. Evolution of genome size: multilevel selection, mutation bias or dynamical chaos? *Curr. Opin. Genet. Dev.* 2004; 14 (6): 620-626.
14. Das B, Seshadri M. Genetic variability at the DIS80 minisatellite: predominance of allele 18 among some indian populations. *Ann. Hum. Biol.* 2004; 31 (5): 541-553.
15. Supply P, Warren RM, Banuls AL, *et al.* Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol. Microbiol.* 2003; 47 (2): 529-538.
16. Tamaki K, Jeffreys AJ. Human tandem repeat sequences in forensic DNA typing. *Leg. Med. (Tokyo)*. 2005; 7 (4): 244-250.
17. Bennett P. Demystified ... microsatellites. *Mol. Pathol.* 2000; 53 (4): 177-183.
18. Hearne CM, Ghosh S, Todd JA. Microsatellites for linkage analysis of genetic traits. *Trends Genet.* 1992; 8 (8): 288-294.
19. Brookes AJ. The essence of SNPs. *Gene*. 1999; 234 (2): 177-186.
20. Riva A, Kohane IS. SNPper: retrieval and analysis of human SNPs. *Bioinformatics*. 2002; 18 (12): 1681-1685.
21. Shastry BS. SNP alleles in human disease and evolution. *J. Hum. Genet.* 2002; 47 (11): 561-566.
22. Sunyaev S, Ramensky V, Koch I, *et al.* Prediction of deleterious human alleles. *Hum. Mol. Genet.* 2001; 10 (6): 591-597.
23. Dhami P, Coffey AJ, Abbs S, *et al.* Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am. J. Hum. Genet.* 2005; 76 (5): 750-762.
24. Iafrate AJ, Feuk L, Rivera MN, *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* 2004; 36 (9): 949-951.
25. Stankiewicz P, Lupski JR. Molecular-evolutionary mechanisms for genomic disorders. *Curr. Opin. Genet. Dev.* 2002; 12 (3): 312-319.
26. Wilson GM, Flibotte S, Chopra V, *et al.* DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. *Hum. Mol. Genet.* 2006; 15 (5): 743-749.
27. Sharp AJ, Locke DP, McGrath SD, *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* 2005; 77 (1): 78-88.
28. Stefansson H, Helgason A, Thorleifsson G, *et al.* A common inversion under selection in Europeans. *Nat. Genet.* 2005; 37 (2): 129-137.
29. Tuzun E, Sharp AJ, Bailey JA, *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* 2005; 37 (7): 727-732.
30. Freeman JL, Perry GH, Feuk L, *et al.* Copy number variation: new insights in genome diversity. *Genome Res.* 2006; 16(8):949-61.
31. Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* 1999; 9 (6): 657-663.
32. Druker R and Whitelaw E. Retrotransposon-derived elements in the mammalian genome: a potential source of disease. *J. Inherit. Metab. Dis.* 2004; 27 (3): 319-330.
33. Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature*. 2001; 409 (6822): 860-921.
34. Arlt MF, Durkin SG, Ragland RL, Glover TW. Common fragile sites as targets for chromosome rearrangements. *DNA Repair. (Amst)*. 2006; 5 (9-10): 1126-1135.
35. Richards RI. Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Hum. Mol. Genet.* 2001; 10 (20): 2187-2194.
36. Sutherland GR, Baker E. The clinical significance of fragile sites on human chromosomes. *Clin. Genet.* 2000; 58 (3): 157-161.
37. Sprenger R, Schlagenhauser R, Kerb R, *et al.* Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation. *Pharmacogenetics*. 2000; 10 (6): 557-565.
38. Maki H. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annu. Rev. Genet.* 2002; 36: 279-303.
39. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 1998; 8 (12): 1229-1231.
40. Hamilton BA. Variations in abundance: genome-wide responses to genetic variation and vice versa. *Genome Biol.* 2002; 3 (10): reviews 1029.
41. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 2006; 7 (1): 21-33.
42. Rakyan VK, Beck S. Epigenetic variation and inheritance in mammals. *Curr. Opin. Genet. Dev.* 2006; 16 (6): 573-577.
43. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev. Cell.* 2010; 19 (5): 698-711.
44. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116 (2): 281-297.
45. Markham AF, Coletta PL, Robinson PA, *et al.* Screening for cancer predisposition. *Eur. J. Cancer.* 1994; 30A (13): 2015-2029.
46. Tamura K, Utsunomiya J, Iwama T, *et al.* Mechanism of carcinogenesis in familial tumors. *Int. J. Clin. Oncol.* 2004; 9 (4): 232-245.
47. Vineis P. Individual susceptibility to carcinogens. *Oncogene*. 2004; 23 (38): 6477-6483.
48. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature*. 2001; 411 (6835): 366-374.
49. Hsieh P. Molecular mechanisms of DNA mismatch repair. *Mutat. Res.* 2001; 486 (2): 71-87.
50. Thompson LH, Schild D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. *Mutat. Res.* 2001; 477 (1-2): 131-153.
51. Welsh PL, Owens KN, King MC. Insights into the functions of BRCA1 and BRCA2. *Trends Genet.* 2000; 16 (2): 69-74.
52. Markman B, Javier Ramos F, Capdevila J, Taberner J. EGFR and KRAS in colorectal cancer. *Adv. Clin. Chem.* 2010; 51: 71-119.
53. Chapman MA, Lawrence MS, Keats JJ, *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011; 471 (7339): 467-472.

54. Maresca B, Schwartz JH. Sudden origins: a general mechanism of evolution based on stress protein concentration and rapid environmental change. *Anat. Rec. B New Anat.* 2006; 289 (1): 38-46.
55. Lindahl T. Instability and decay of the primary structure of DNA. *Nature.* 1993; 362 (6422): 709-715.
56. Loeb LA. Endogenous carcinogenesis: molecular oncology into the twenty-first century--presidential address. *Cancer Res.* 1989; 49 (20): 5489-5496.
57. McGowan CH, Russell P. The DNA damage response: sensing and signaling. *Curr. Opin. Cell Biol.* 2004; 16 (6): 629-633.
58. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* 2004; 73: 39-85.
59. Attardi LD. The role of p53-mediated apoptosis as a crucial anti-tumor response to genomic instability: lessons from mouse models. *Mutat. Res.* 2005; 569 (1-2): 145-157.
60. Ishikawa T, Zhang SS, Qin X, et al. DNA repair and cancer: lessons from mutant mouse models. *Cancer Sci.* 2004; 95 (2): 112-117.
61. Schmitt E, Paquet C, Beauchemin M, Bertrand R. DNA-damage response network at the crossroads of cell-cycle checkpoints, cellular senescence and apoptosis. *J. Zhejiang Univ. Sci. B.* 2007; 8 (6): 377-397.
62. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes, 2005. *Mutat. Res.* 2005; 577 (1-2): 275-283.
63. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. *Science.* 2001; 291 (5507): 1284-1289.
64. Ataian Y, Krebs JE. Five repair pathways in one context: chromatin modification during DNA repair. *Biochem. Cell Biol.* 2006; 84 (4): 490-504.
65. Bassing CH, Alt FW. The cellular response to general and programmed DNA double strand breaks. *DNA Repair. (Amst).* 2004; 3 (8-9): 781-796.
66. Costa RM, Chigancas V, Galhardo Rda S, et al. The eukaryotic nucleotide excision repair pathway. *Biochimie.* 2003; 85 (11): 1083-1099.
67. Franco S, Alt FW, Manis JP. Pathways that suppress programmed DNA breaks from progressing to chromosomal breaks and translocations. *DNA Repair. (Amst).* 2006; 5 (9-10): 1030-1041.
68. Gerson SL. MGMT: its role in cancer aetiology and cancer therapeutics. *Nat. Rev. Cancer.* 2004; 4 (4): 296-307.
69. Lees-Miller SP, Meek K. Repair of DNA double strand breaks by non-homologous end joining. *Biochimie.* 2003; 85 (11): 1161-1173.
70. Mills KD, Ferguson DO, Alt FW. The role of DNA breaks in genomic instability and tumorigenesis. *Immunol. Rev.* 2003; 194: 77-95.
71. Nordstrand LM, Ringvoll J, Larsen E, Klungland A. Genome instability and DNA damage accumulation in gene-targeted mice. *Neuroscience.* 2007.
72. Schofield MJ, Hsieh P. DNA mismatch repair: molecular mechanisms and biological function. *Annu. Rev. Microbiol.* 2003; 57: 579-608.
73. Wyman C, Ristic D, Kanaar R. Homologous recombination-mediated double-strand break repair. *DNA Repair. (Amst).* 2004; 3 (8-9): 827-833.
74. Peterson CL, Cote J. Cellular machineries for chromosomal DNA repair. *Genes Dev.* 2004; 18 (6): 602-616.
75. Rojas E, Lopez MC, Valverde M. Single cell gel electrophoresis assay: methodology and applications. *J. Chromatogr. B Biomed. Sci. Appl.* 1999; 722 (1-2): 225-254.
76. Nadin SB, Vargas-Roig LM, Drago G, et al. DNA damage and repair in peripheral blood lymphocytes from healthy individuals and cancer patients: a pilot study on the implications in the clinical response to chemotherapy. *Cancer Lett.* 2006; 239 (1): 84-97.
77. Berwick M, Vineis P. Measuring DNA repair capacity: small steps. *J. Natl. Cancer Inst.* 2005; 97 (2): 84-85.
78. Madhusudan S, Middleton MR. The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer. *Cancer Treat. Rev.* 2005; 31 (8): 603-617.
79. Kennedy DO, Agrawal M, Shen J, et al. DNA repair capacity of lymphoblastoid cell lines from sisters discordant for breast cancer. *J. Natl. Cancer Inst.* 2005; 97 (2): 127-132.
80. Zijno A, Verdina A, Galati R, et al. Influence of DNA repair polymorphisms on biomarkers of genotoxic damage in peripheral lymphocytes of healthy subjects. *Mutat. Res.* 2006; 600 (1-2): 184-192.
81. Mohrenweiser HW. Genetic variation and exposure related risk estimation: will toxicology enter a new era? DNA repair and cancer as a paradigm. *Toxicol. Pathol.* 2004; 32 Suppl 1: 136-145.
82. Schmidt AL, Anderson LM. Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biol. Rev. Camb. Philos. Soc.* 2006; 81 (4): 531-543.
83. Cutfield WS, Hofman PL, Mitchell M, Morison IM. Could Epigenetics Play a Role in the Developmental Origins of Health and Disease? *Pediatr. Res.* 2007.
84. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* 2007; 8 (4): 253-262.
85. Chanvaivit S, Navasumrit P, Hunsonti P, et al. Exposure assessment of benzene in Thai workers, DNA-repair capacity and influence of genetic polymorphisms. *Mutat. Res.* 2007; 626 (1-2): 79-87.
86. Wogan GN, Hecht SS, Felton JS, et al. Environmental and chemical carcinogenesis. *Semin. Cancer Biol.* 2004; 14 (6): 473-486.
87. Han J, Colditz GA, Liu JS, Hunter DJ. Genetic variation in XPD, sun exposure, and risk of skin cancer. *Cancer Epidemiol. Biomarkers Prev.* 2005; 14 (6): 1539-1544.
88. Terry MB, Gammon MD, Zhang FF, et al. Polymorphism in the DNA repair gene XPD, polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 2004; 13 (12): 2053-2058.
89. Reszka E, Wasowicz W, Gromadzinska J. Genetic polymorphism of xenobiotic metabolising enzymes, diet and cancer susceptibility. *Br. J. Nutr.* 2006; 96 (4): 609-619.
90. Colnot S, Niwa-Kawakita M, Hamard G, et al. Colorectal cancers in a new mouse model of familial adenomatous polyposis: influence of genetic and environmental modifiers. *Lab. Invest.* 2004; 84 (12): 1619-1630.
91. Neuhausen SL. Ethnic differences in cancer risk resulting from genetic variation. *Cancer.* 1999; 86 (11 Suppl): 2575-2582.
92. Human\_Genome\_Project: Human Genome Program, U.S. Department of Energy, Genomics and Its Impact on Science and Society: A 2003 Primer. 2003.
93. Klose J. Genotypes and phenotypes. *Electrophoresis.* 1999; 20 (4-5): 643-652.
94. Bennett MD, Leitch IJ. Plant genome size research: a field in focus. *Ann. Bot. (Lond).* 2005; 95 (1): 1-6.
95. Griffing B, Scholl RL. Qualitative and quantitative genetic studies of Arabidopsis thaliana. *Genetics.* 1991; 129 (3): 605-609.
96. Latchman DS. Heat shock proteins and human disease. *J. R. Coll. Physicians Lond.* 1991; 25 (4): 295-299.
97. Lonai P, Orr-Urtreger A. Homeogenes in mammalian development and the evolution of the cranium and central nervous system. *Faseb. J.* 1990; 4 (5): 1436-1443.
98. Nagy A, Perrimon N, Sandmeyer S, Plasterk R. Tailoring the genome: the power of genetic approaches. *Nat. Genet.* 2003; 33 Suppl: 276-284.
99. Ross-Macdonald P. Functional analysis of the yeast genome. *Funct. Integr. Genomics.* 2000; 1 (2): 99-113.
100. Schaechter M. Escherichia coli and Salmonella 2000: the view from

- here. *Microbiol. Mol. Biol. Rev.* 2001; 65 (1): 119-130.
101. Lander ES. Genome-sequencing anniversary. The accelerator. *Science.* 2011; 331 (6020): 1024.
  102. Fecher LA, Cummings SD, Keefe MJ, Alani RM. Toward a molecular classification of melanoma. *J. Clin. Oncol.* 2007; 25 (12): 1606-1620.
  103. Maillet P, Chappuis PO, Khoshbeen-Boudal M, et al. Twenty-three novel BRCA1 and BRCA2 sequence variations identified in a cohort of Swiss breast and ovarian cancer families. *Cancer Genet. Cytogenet.* 2006; 169 (1): 62-68.
  104. Pillay VV, Menezes RG, Krishnaprasad R, et al. Biotechnology in forensic science: the revolution continues. *Nepal. Med. Coll. J.* 2007; 9 (1): 57-62.
  105. Van Lanen SG, Shen B. Microbial genomics for the improvement of natural product discovery. *Curr. Opin. Microbiol.* 2006; 9 (3): 252-260.
  106. Verschure PJ, Visser AE, Rots MG. Step out of the groove: epigenetic gene control systems and engineered transcription factors. *Adv. Genet.* 2006; 56: 163-204.
  107. The ENCODE Project Consortium: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007; 447 (7146): 799-816.
  108. Weinstock GM. ENCODE: more genomic empowerment. *Genome. research.* 2007; 17 (6): 667-668.
  109. Gerstein MB, Bruce C, Rozowsky JS, et al. What is a gene, post-ENCODE? History and updated definition. *Genome. research.* 2007; 17 (6): 669-681.
  110. Int\_HapMap Consortium: The International HapMap Project. *Nature.* 2003; 426 (6968): 789-796.
  111. Int\_HapMap Consortium: A haplotype map of the human genome. *Nature.* 2005; 437 (7063): 1299-1320.
  112. McVean G, Spencer CC, Chaix R. Perspectives on human genetic variation from the HapMap Project. *PLOS Genet.* 2005; 1 (4): e54.
  113. Loeb LA, Springgate CF, Battula N. Errors in DNA replication as a basis of malignant changes. *Cancer Res.* 1974; 34 (9): 2311-2321.
  114. Prindle MJ, Fox EJ, Loeb LA. The mutator phenotype in cancer: molecular mechanisms and targeting strategies. *Curr. Drug Targets.* 2010; 11 (10): 1296-1303.
  115. Coates PJ, Lorimore SA, Wright EG. Cell and tissue responses to genotoxic stress. *J. Pathol.* 2005; 205 (2): 221-235.
  116. Kolodner RD, Putnam CD, Myung K. Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science.* 2002; 297 (5581): 552-557.
  117. Maser RS, DePinho RA. Connecting chromosomes, crisis, and cancer. *Science.* 2002; 297 (5581): 565-569.
  118. Rouse J, Jackson SP. Interfaces between the detection, signaling, and repair of DNA damage. *Science.* 2002; 297 (5581): 547-551.
  119. Hasler J, Strub K. Alu elements as regulators of gene expression. *Nucleic. Acids Res.* 2006; 34 (19): 5491-5497.
  120. Matsutani S. Links between repeated sequences. *J. Biomed. Biotechnol.* 2006; 2006 (1): 13569.
  121. Zuckerkandl E, Cavalli G. Combinatorial epigenetics, "junk DNA", and the evolution of complex organisms. *Gene.* 2007; 390 (1-2): 232-242.
  122. Armour JA. Tandemly repeated DNA: why should anyone care? *Mutat. Res.* 2006; 598 (1-2): 6-14.
  123. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature.* 1998; 396 (6712): 643-649.
  124. Alvarez S, Cigudosa JC. Gains, losses and complex karyotypes in myeloid disorders: a light at the end of the tunnel. *Hematol. Oncol.* 2005; 23 (1): 18-25.
  125. Pettus JA, Cowley BC, Maxwell T, et al. Multiple abnormalities detected by dye reversal genomic microarrays in prostate cancer: a much greater sensitivity than conventional cytogenetics. *Cancer Genet. Cytogenet.* 2004; 154 (2): 110-118.
  126. Kurzrock R, Kantarjian HM, Druker BJ, Talpaz M. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann. Intern. Med.* 2003; 138 (10): 819-830.
  127. Deininger PL, Batzer MA. Alu repeats and human disease. *Mol. Genet. Metab.* 1999; 67 (3): 183-193.
  128. Economou-Pachnis A, Tsiachlis PN. Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus. *Nucleic. Acids Res.* 1985; 13 (23): 8379-8387.
  129. Miki Y, Katagiri T, Kasumi F, et al. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.* 1996; 13 (2): 245-247.
  130. Rowe SM, Coughlan SJ, McKenna NJ, et al. Ovarian carcinoma-associated TaqI restriction fragment length polymorphism in intron G of the progesterone receptor gene is due to an Alu sequence insertion. *Cancer Res.* 1995; 55 (13): 2743-2745.
  131. Varon R, Gooding R, Steglich C, et al. Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. *Nat. Genet.* 2003; 35 (2): 185-189.
  132. Hung RJ, Boffetta P, Brennan P, et al. GST, NAT, SULT1A1, CYP1B1 genetic polymorphisms, interactions with environmental exposures and bladder cancer risk in a high-risk population. *Int. J. Cancer.* 2004; 110 (4): 598-604.
  133. Lee KM, Park SK, Kim SU, et al. N-acetyltransferase (NAT1, NAT2) and glutathione S-transferase (GSTM1, GSTT1) polymorphisms in breast cancer. *Cancer Lett.* 2003; 196 (2): 179-186.
  134. Shin JH, Ku JL, Shin KH, et al. Glutathione S-transferase M1 associated with cancer occurrence in Korean HNPCC families carrying the hMLH1/hMSH2 mutation. *Oncol. Rep.* 2003; 10 (2): 483-486.
  135. Sorensen M, Autrup H, Tjonneland A, et al. Glutathione S-transferase T1 null-genotype is associated with an increased risk of lung cancer. *Int. J. Cancer.* 2004; 110 (2): 219-224.
  136. Spurdle AB, Webb PM, Purdie DM, et al. Polymorphisms at the glutathione S-transferase GSTM1, GSTT1 and GSTP1 loci: risk of ovarian cancer by histological subtype. *Carcinogenesis.* 2001; 22 (1): 67-72.
  137. Arlt MF, Casper AM, Glover TW. Common fragile sites. *Cytogenet. Genome. Res.* 2003; 100 (1-4): 92-100.
  138. Balsara BR, Pei J, De Rienzo A, et al. Human hepatocellular carcinoma is characterized by a highly consistent pattern of genomic imbalances, including frequent loss of 16q23.1-24.1. *Genes Chromosomes Cancer.* 2001; 30 (3): 245-253.
  139. Hansen LL, Yilmaz M, Overgaard J, et al. Allelic loss of 16q23.2-24.2 is an independent marker of good prognosis in primary breast cancer. *Cancer Res.* 1998; 58 (10): 2166-2169.
  140. Kovacs G, Erlandsson R, Boldog F, et al. Consistent chromosome 3p deletion and loss of heterozygosity in renal cell carcinoma. *Proc. Natl. Acad. Sci. USA.* 1988; 85 (5): 1571-1575.
  141. Maitra A, Wistuba II, Gibbons D, et al. Allelic losses at chromosome 3p are seen in human papilloma virus 16 associated transitional cell carcinoma of the cervix. *Gynecol. Oncol.* 1999; 74 (3): 361-368.
  142. Naylor SL, Johnson BE, Minna JD, Sakaguchi AY. Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer. *Nature.* 1987; 329 (6138): 451-454.
  143. Paris PL, Witte JS, Kupelian PA, et al. Identification and fine mapping of a region showing a high frequency of allelic imbalance on chromosome 16q23.2 that corresponds to a prostate cancer susceptibility locus. *Cancer Res.* 2000; 60 (13): 3645-3649.
  144. Rader JS, Kamarasova T, Huettner PC, et al. Allelotyping of all chromosomal arms in invasive cervical cancer. *Oncogene.* 1996; 13 (12): 2737-2741.
  145. Wang N, Perkins KL. Involvement of band 3p14 in t(3;8) hereditary

- renal carcinoma. *Cancer Genet Cytogenet.* 1984; 11 (4): 479-481.
146. Yokoyama S, Yamakawa K, Tsuchiya E, *et al.* Deletion mapping on the short arm of chromosome 3 in squamous cell carcinoma and adenocarcinoma of the lung. *Cancer Res.* 1992; 52 (4): 873-877.
  147. Smeets DF, Scheres JM, Hustinx TW. The most common fragile site in man is 3p14. *Hum. Genet.* 1986; 72 (3): 215-220.
  148. Sutherland GR, Richards RI. The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet Dev.* 1995; 5 (3): 323-327.
  149. Ramos D, Aldaz CM. WWOX, a chromosomal fragile site gene and its role in cancer. *Adv. Exp. Med. Biol.* 2006; 587: 149-159.
  150. Barros P, Blanco MG, Boan F, Gomez-Marquez J. Heteroduplex analysis of minisatellite variability. *Electrophoresis.* 2005; 26 (22): 4304-4309.
  151. Krontiris TG, DiMartino NA, Colb M, Parkinson DR. Unique allelic restriction fragments of the human Ha-ras locus in leukocyte and tumour DNAs of cancer patients. *Nature.* 1985; 313 (6001): 369-374.
  152. Langdon JA, Armour JA. Evolution and population genetics of the H-ras minisatellite and cancer predisposition. *Hum. Mol. Genet.* 2003; 12 (8): 891-900.
  153. Ponder BA. Cancer genetics. *Nature.* 2001; 411 (6835): 336-341.
  154. Pharoah PD, Antoniou A, Bobrow M, *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.* 2002; 31 (1): 33-36.
  155. Bell J. The new genetics in clinical practice. *BMJ.* 1998; 316 (7131): 618-620.
  156. Holtzman NA, Marteau TM. Will genetics revolutionize medicine? *N. Engl. J. Med.* 2000; 343 (2): 141-144.
  157. Beroud C, Soussi T. APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic. Acids Res.* 1996; 24 (1): 121-124.
  158. Kim CJ, Lee JH, Song JW, *et al.* Chk1 frameshift mutation in sporadic and hereditary non-polyposis colorectal cancers with microsatellite instability. *Eur. J. Surg. Oncol.* 2007; 33 (5): 580-585.
  159. Peltomaki P, Vasen HF. Mutations predisposing to hereditary non-polyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Non-polyposis Colorectal Cancer. *Gastroenterology.* 1997; 113 (4): 1146-1158.
  160. Wooster R, Bignell G, Lancaster J, *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature.* 1995; 378 (6559): 789-792.
  161. Nuzzo V, Tauchmanova L, Falchetti A, *et al.* MEN1 family with a novel frameshift mutation. *J. Endocrinol. Invest.* 2006; 29 (5): 450-456.
  162. Fijneman RJ. Genetic predisposition to sporadic cancer: how to handle major effects of minor genes? *Cell Oncol.* 2005; 27 (5-6): 281-292.
  163. Bond GL, Hu W, Bond EE, *et al.* A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell.* 2004; 119 (5): 591-602.
  164. Braude I, Vukovic B, Prasad M, *et al.* Large scale copy number variation (CNV) at 14q12 is associated with the presence of genomic abnormalities in neoplasia. *BMC Genomics.* 2006; 7: 138.
  165. Davies SM, Bhatia S, Ross JA, *et al.* Glutathione S-transferase genotypes, genetic susceptibility, and outcome of therapy in childhood acute lymphoblastic leukemia. *Blood.* 2002; 100 (1): 67-71.
  166. DeMichele A, Martin AM, Mick R, *et al.* Interleukin-6 -174G->C polymorphism is associated with improved outcome in high-risk breast cancer. *Cancer Res.* 2003; 63 (22): 8051-8056.
  167. Frank B, Bermejo JL, Hemminki K, *et al.* Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. *Carcinogenesis.* 2007.
  168. Heinimann K, Scott RJ, Chappuis P, *et al.* N-acetyltransferase 2 influences cancer prevalence in hMLH1/hMSH2 mutation carriers. *Cancer Res.* 1999; 59 (13): 3038-3040.
  169. Jones JS, Chi X, Gu X, *et al.* p53 polymorphism and age of onset of hereditary nonpolyposis colorectal cancer in a Caucasian population. *Clin. Cancer Res.* 2004; 10 (17): 5845-5849.
  170. Krajcinovic M, Lemieux-Blanchard E, Chiasson S, *et al.* Role of polymorphisms in MTHFR and MTHFD1 genes in the outcome of childhood acute lymphoblastic leukemia. *Pharmacogenomics J.* 2004; 4 (1): 66-72.
  171. Sotamaa K, Liyanarachchi S, Mecklin JP, *et al.* p53 codon 72 and MDM2 SNP309 polymorphisms and age of colorectal cancer onset in Lynch syndrome. *Clin. Cancer Res.* 2005; 11 (19 Pt 1): 6840-6844.
  172. Sun T, Gao Y, Tan W, *et al.* A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. *Nat. Genet.* 2007; 39 (5): 605-613.
  173. Talseth-Palmer BA, Brenne IS, Ashton KA, *et al.* Colorectal cancer susceptibility loci on chromosome 8q23.3 and 11q23.1 as modifiers for disease expression in lynch syndrome. *J. Med. Genet.* 2011; 48 (4): 279-284.
  174. Wijnen JT, Brohet RM, van Eijk R, *et al.* Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. *Gastroenterology.* 2009; 136 (1): 131-137.
  175. Govan VA, Carrara HR, Sachs JA, *et al.* Ethnic differences in allelic distribution of IFN-gamma in South African women but no link with cervical cancer. *J. Carcinog.* 2003; 2 (1): 3.
  176. Hoffmann SC, Stanley EM, Cox ED, *et al.* Ethnicity greatly influences cytokine gene polymorphism distribution. *Am. J. Transplant.* 2002; 2 (6): 560-567.
  177. Inserra P, Abrahamson M, Papenfuss M, Giuliano AR. Ethnic variation of the P53 codon 72 polymorphism, HPV persistence, and cervical cancer risk. *Int. J. STD AIDS.* 2003; 14 (12): 800-804.
  178. Lohmueller KE, Mauney MM, Reich D, Braverman JM. Variants associated with common disease are not unusually differentiated in frequency across populations. *Am. J. Hum. Genet.* 2006; 78 (1): 130-136.
  179. Martin ER, Lai EH, Gilbert JR, *et al.* SNPping away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. *Am. J. Hum. Genet.* 2000; 67 (2): 383-394.
  180. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum. Mol. Genet.* 2002; 11 (20): 2417-2423.
  181. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996; 273 (5281): 1516-1517.
  182. Suh Y, Vijg J. SNP discovery in associating genetic variation with human disease phenotypes. *Mutat. Res.* 2005; 573 (1-2): 41-53.
  183. Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr. Opin. Biotechnol.* 1998; 9 (6): 578-594.
  184. Wright AF, Hastie ND. Complex genetic diseases: controversy over the Croesus code. *Genome. Biol.* 2001; 2 (8): COMMENT 2007.
  185. Kristinsson SY, Thorolfsson ET, Talseth B, *et al.* MODY in Iceland is associated with mutations in HNF-1alpha and a novel mutation in NeuroD1. *Diabetologia.* 2001; 44 (11): 2098-2103.
  186. Garcia-Closas M, Malats N, Real FX, *et al.* Large-scale evaluation of candidate genes identifies associations between VEGF polymorphisms and bladder cancer risk. *PLOS Genet.* 2007; 3 (2): e29.
  187. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. *Nat. Rev. Cancer.* 2004; 4 (11): 850-860.
  188. Pharoah PD, Tyrer J, Dunning AM, *et al.* Association between Common Variation in 120 Candidate Genes and Breast Cancer Risk.

- PLOS Genet.* 2007; 3 (3): e42.
189. Mardis ER, Wilson RK. Cancer genome sequencing: a review. *Hum. Mol. Genet.* 2009; 18 (R2): R163-168.
  190. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* 2007; 39 (8): 984-988.
  191. Bala S, Peltomaki P. CYCLIN D1 as a genetic modifier in hereditary nonpolyposis colorectal cancer. *Cancer Res.* 2001; 61 (16): 6042-6045.
  192. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* 2007; 39 (8): 989-994.
  193. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nature reviews. Genetics.* 2009; 10 (4): 241-251.
  194. Roberts R, Wells GA, Stewart AF, et al. The genome-wide association study--a new era for common polygenic disorders. *J. Cardiovasc. Transl. Res.* 2010; 3 (3): 173-182.
  195. Stadler ZK, Gallagher DJ, Thom P, Offit K. Genome-wide association studies of cancer: principles and potential utility. *Oncology. (Williston Park).* 2010; 24 (7): 629-637.
  196. Kuiper RP, Ligtenberg MJ, Hoogerbrugge N, Geurts van Kessel A. Germline copy number variation and cancer risk. *Curr. Opin. Genet. Dev.* 2010; 20 (3): 282-289.
  197. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science.* 2008; 322 (5903): 881-888.
  198. Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. *BMC Med. Genet.* 2009; 10: 6.
  199. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007.
  200. Stadler ZK, Vijai J, Thom P, et al. Genome-wide association studies of cancer predisposition. *Hematology/oncology clinics. of North America.* 2010; 24 (5): 973-996.
  201. Yoshida T, Ono H, Kuchiba A, et al. Genome-wide germline analyses on cancer susceptibility and GeMDBJ database: Gastric cancer as an example. *Cancer Sci.* 2010; 101 (7): 1582-1589.
  202. Huo D, Olopade OI. Interpretation of genome-wide association study results. *Oncology.* 2010; 24 (7): 643, 646.
  203. Speicher MR, Geigl JB, Tomlinson IP. Effect of genome-wide association studies, direct-to-consumer genetic testing, and high-speed sequencing technologies on predictive genetic counselling for cancer risk. *The lancet oncology.* 2010; 11 (9): 890-898.
  204. Cho EK, Tchinda J, Freeman JL, et al. Array-based comparative genomic hybridization and copy number variation in cancer research. *Cytogenet. Genome. Res.* 2006; 115 (3-4): 262-272.
  205. Mantripragada KK, Buckley PG, de Stahl TD, Dumanski JP. Genomic microarrays in the spotlight. *Trends Genet.* 2004; 20 (2): 87-94.
  206. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 1998; 20 (2): 207-211.
  207. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 1999; 23 (1): 41-46.
  208. Peiffer DA, Le JM, Steemers FJ, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome. Res.* 2006; 16 (9): 1136-1148.
  209. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature.* 2006; 444 (7118): 444-454.
  210. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature.* 2006; 444 (7118): 444-454.
  211. Shlien A, Tabori U, Marshall CR, et al. Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. *Proc. Natl. Acad. Sci. USA.* 2008; 105 (32): 11264-11269.
  212. Dear PH. Copy-number variation: the end of the human genome? *Trends Biotechnol.* 2009; 27 (8): 448-454.
  213. Chen Y, Chen C. DNA copy number variation and loss of heterozygosity in relation to recurrence of and survival from head and neck squamous cell carcinoma: a review. *Head Neck.* 2008; 30 (10): 1361-1383.
  214. Beroukhi R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature.* 2010; 463 (7283): 899-905.
  215. Paluszczak J, Baer-Dubowska W. Epigenetic diagnostics of cancer--the application of DNA methylation markers. *J. Appl. Genet.* 2006; 47 (4): 365-375.
  216. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* 2007; 8 (4): 286-298.
  217. Toyota M, Suzuki H. Epigenetic drivers of genetic alterations. *Adv. Genet.* 2010; 70: 309-323.
  218. Zhang L, Huang J, Yang N, et al. microRNAs exhibit high frequency genomic alterations in human cancer. *Proc. Natl. Acad. Sci. USA.* 2006; 103 (24): 9136-9141.
  219. Ramsingh G, Koboldt DC, Trissal M, et al. Complete characterization of the microRNAome in a patient with acute myeloid leukemia. *Blood.* 2010; 116 (24): 5316-5326.
  220. Le XF, Merchant O, Bast RC, Calin GA. The Roles of MicroRNAs in the Cancer Invasion-Metastasis Cascade. *Cancer Microenviron.* 2010; 3 (1): 137-147.
  221. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. *Nature reviews. Cancer.* 2010; 10 (6): 389-402.
  222. Walsh T, Lee MK, Casadei S, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc. Natl. Acad. Sci. USA.* 2010; 107 (28): 12629-12633.
  223. Jones S, Wang TL, Shih Ie M, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science.* 2010; 330 (6001): 228-231.
  224. Bodmer W, Tomlinson I. Rare genetic variants and the risk of cancer. *Curr. Opin. Genet. Dev.* 2010; 20 (3): 262-267.
  225. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* 2010; 11 (6): 415-425.
  226. Lander ES. Initial impact of the sequencing of the human genome. *Nature.* 2011; 470 (7333): 187-197.
  227. Risch A, Plass C. Lung cancer epigenetics and genetics. *International journal of cancer. Journal international du cancer.* 2008; 123 (1): 1-7.
  228. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nature genetics.* 2003; 33 Suppl: 238-244.
  229. Gusella JF, MacDonald ME. Huntington's disease. *Semin. Cell Biol.* 1995; 6 (1): 21-28.
  230. Tsui LC. The spectrum of cystic fibrosis mutations. *Trends Genet.* 1992; 8 (11): 392-398.
  231. Seddon JM, Francis PJ, George S, et al. Association of CFH Y402H and LOC387715 A69S with progression of age-related macular degeneration. *JAMA.* 2007; 297 (16): 1793-1800.
  232. The\_Welcome\_Trust\_Case\_Control\_Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447 (7145): 661-678.
  233. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* 2008; 40 (5): 623-630.

234. Rebbeck TR, Martinez ME, Sellers TA, *et al.* Genetic variation and cancer: improving the environment for publication of association studies. *Cancer Epidemiol. Biomarkers Prev.* 2004; 13 (12): 1985-1986.
235. Nowell SA, Ahn J, Rae JM, *et al.* Association of genetic variation in tamoxifen-metabolizing enzymes with overall survival and recurrence of disease in breast cancer patients. *Breast Cancer Res. Treat.* 2005; 91 (3): 249-258.
236. Blomhoff A, Olsson M, Johansson S, *et al.* Linkage disequilibrium and haplotype blocks in the MHC vary in an HLA haplotype specific manner assessed mainly by DRB1(\*)03 and DRB1(\*)04 haplotypes. *Genes Immun.* 2006; 7 (2): 130-140.
237. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in the human genome. *Science.* 2002; 296 (5576): 2225-2229.
238. Patil N, Berno AJ, Hinds DA, *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science.* 2001; 294 (5547): 1719-1723.
239. Voidonikolas G, Kreml SS, Chen C, *et al.* Basic principles and technologies for deciphering the genetic map of cancer. *World journal of surgery.* 2009; 33 (4): 615-629.