We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

The Investigation of Gene Regulation and Variation in Human Cancers and Other Diseases

Shihori Tanabe and Sun Ha Jee

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/51271

1. Introduction

Dynamic regulation of genes is an important part of the cell life cycle in health and disease. The regulation includes the variety and alteration of genome and gene expression, and the concept such as quality of genome will be useful to predict and assess the developmental stages of the cells, disease status and drug sensitivity. Recent technologies and worldwide sequencing projects have revealed 26,383 annotated genes in the 2.91-Gigabase human genome [1,2]. The main molecular functions of the annotated genes, as categorized by Gene Ontology (GO), are enzyme, signal transduction, nucleic acid binding, cell adhesion, chaperone, cytoskeletal structural protein, extracellular matrix, immunoglobulin, ion channel, motor, structural protein of muscle, protooncogene, select calcium binding protein, intracellular transporter, and transporter [1,3]. Despite a wealth of knowledge, the function of 42% of the annotated genes remains unknown [1]. When the human genome sequence was published in 2001 [1], there were a predicted 39,114 genes, of which 59% were of unknown function. According to the International Human Genome Sequencing Consortium, the number of identified genes is approximately 32,000, of which 51% show a match within InterPro, a database that integrates diverse information about protein families, domains, and functional sites [2-5]. In 2001, InterPro combined sequence and pattern information from four databases (PRINTS, PROSITE, Pfam, Prosite Profile); however, it now includes information from an additional eight databases (SMART, ProDom, PIRSF, SUPERFAMILY, PANTHER, CATH-Gene3D, TIGRFAM, and HAMAP) [2,4-16]. In [2], the InterPro entries are collapsed into 12 broad categories: cellular processes, metabolism, DNA replication/modification, transcription/translation, intracellular signaling, cell–cell communication, protein folding and degradation, transport, multifunctional proteins, cytoskeletal/structural, defense and immunity, and miscellaneous function. The

rate of single nucleotide polymorphism (SNP) variation has been reported as 1 in 1250 base pairs [1] and more than 1.4 million SNPs have been identified [2] (Table 1).

Size of the genome	2.91 Gbp	[1]
Number of annotated genes	26,383	[1]
Main molecular functions of annotated genes	enzyme, signal transduction, nucleic acid binding, cell adhesion, chaperone, cytoskeletal structural protein, extracellular matrix, immunoglobulin, ion channel, motor, structural protein of muscle, protooncogene, select calcium binding protein, intracellular transporter, transporter	[1]
Percentage of annotated genes with unknown function	42%	[1]
Number of hypothetical and annotated genes	39,114	[1]
Percentage of hypothetical and annotated genes with unknown function	59%	[1]
Number of identified genes	approx. 32,000	[2]
Percentage of matches with InterPro	51%	[2]
Rate of SNP variation	1/1250 bp	[1]
SNPs identified	more than 1.4 million	[2]

Table 1. Genomic and gene characteristics revealed by the Human Genome Project.

Among the databases combined in InterPro (Table 2), PRINTS, PROSITE, and Pfam contain protein families in which the homology between each protein is predicted by the degree of sequence similarity [8]. The others—SMART, ProDom, PIRSF, SUPERFAMILY, PANTHER, CATH-Gene3D, TIGRFAM, and HAMAP [4-16]—have unique characteristics and URLs, and have been developed sharing information among each other and incorporating information from GO. In detail, PRINTS is a collection of diagnostic protein family "fingerprints", which are groups of conserved motifs, evident in multiple sequence alignments [6]; PROSITE is a protein domain database for functional characterization and annotation that consists of documentation entries describing protein domains, families, and functional sites as well as associated patterns and profiles to identify them [7]; Pfam contains collections of protein families, each represented by multiple sequence alignments and hidden Markov models, available via servers in the UK, the USA, and Sweden [8]; SMART (Simple Modular Architecture Research Tool) is an online resource for the identification and annotation of protein domains and the analysis of protein domain architectures [9]; ProDom is a comprehensive set of protein domain families generated automatically from the UniProt database [10]; PIRSF is a classification system that reflects evolutionary relationships among full-length proteins and domains [11]; SUPERFAMILY is a database of structural and functional annotation for all proteins and genomes [12]; PANTHER is a classification system that clas-

sifies genes by their functions using published experimental evidence and evolutionary relationships to predict function even in the absence of direct experimental evidence [13]; CATH-Gene3D is a comprehensive database of protein domain assignments for sequences from the major sequence databases [14]; TIGRFAM is a collection of protein family definitions built to aid high-throughput annotation of specific protein functions [15]; and HAMAP is composed of two databases: the proteome database and the family database, and of an automatic annotation pipeline mainly focused on microbial proteomes [16]. Hidden Markov models are usually used for the database algorithm.

Database Name	Context	URL	Reference
InterPro	integrative predictive models of protein families, domain and functional sites of multiple databases such as PRINTS, PROSITE, Pfam, SMART, ProDom, PIRSF, SUPERFAMILY, PANTHER, CATH-Gene3D, TIGRFAM, and HAMAP	http://www.ebi.ac.uk/interpro/	[4], [5]
PRINTS	a collection of diagnostic protein family "fingerprints" which are groups of conserved motifs, evident in multiple sequence alignments	http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/index.php	[6]
PROSITE	a protein domain database for functional characterization and annotation which consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them	http://prosite.expasy.org/	[7]
Pfam	a database of collection of protein families, each represented by multiple sequence alignments and hidden Markov models, available via servers in the UK, the USA and Sweden	http://pfam.sanger.ac.uk/http://pfam.janelia.org/http://pfam.sbc.su.se/	[8]
SMART	an online resource for the identification and annotation of protein domains and the analysis of protein domain architectures, of which abbreviation is Simple Modular Architecture Research Tool	http://smart.embl.de/	[9]
ProDom	a comprehensive set of protein domain families automatically generated from the uniProt knowledge Database	http://prodom.prabi.fr/prodom/current/html/home.php	[10]
PIRSF	the classification system which reflects evolutionary relationships of full-length proteins and domains	http://pir.georgetown.edu/pirsf/	[11]
SUPERFAMILY	a database of structural and functional annotation for all proteins and genomes	http://supfam.org/SUPERFAMILY/	[12]
PANTHER	the classification system which classifies genes by their functions using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct experimental evidence	http://www.pantherdb.org/	[13]
CATH-Gene3D	a comprehensive database of protein domain assignments for sequences from the major sequence databases	http://gene3d.biochem.ucl.ac.uk/	[14]

Database Name	Context	URL	Reference
TIGRFAM	a collection of protein family definitions built to aid in high-throughput annotation of specific protein functions	http://www.jcvi.org/cgi-bin/tigrfams/index.cgi	[15]
HAMAP	a system which composed of two databases, the proteome database and the family database, and of an automatic annotation pipeline	http://hamap.expasy.org/	[16]

Table 2. Database information.

2. Gene regulation

2.1. Gene markers for cancer and cancer stem cells

Several molecular markers of cancer have been identified [17]. Metastatic cancer cells can transfer into bodily fluids through the cellular epithelia, which enables the detection of cancer markers in bodily fluids such as blood plasma, urine, or saliva [17]. The different types of cancer markers include genomic DNA point mutations, microsatellite alterations, promoter hypermethylation, viral sequences, aberrant chromosomal copy number, chromosomal translocations, deletions, or loss of heterozygosity, telomere extension, alterations in RNA or protein expression, and mitochondrial DNA mutations [17].

Molecular markers of cancer include *TP53* (encoding p53), which has been shown to be mutated in head and neck, lung, colon, pancreatic, and bladder cancer [17,18]; colon, lung, esophagus, breast, liver, brain, reticuloendothelial tissue, and hematopoietic tissue cancers [19]; and bladder cancer [20]. Mutation of the epidermal growth factor receptor (*EGFR*) gene is an important predictive/prognostic factor for EGFR-tyrosine kinase inhibitor therapy in non-small cell lung cancer [21]. *RAS* oncogene mutations have been identified in colorectal tumors [22]. Microsatellites, which are tandem iterations of simple di-, tri-, or tetranucleotide repeats, have been reported to be unstable in some inherited diseases and in some types of cancer [23], including head and neck, lung, breast, and bladder cancer [17,23].

The expression levels of the cell cycle-related proteins p21 (*CDKN1A*), p53 (*TP53*), cyclin D1 (*CCND1*), and aurora kinase A (*AURKA*) may be used as prognostic markers to predict recurrence in stage II and stage III colon cancer [24]. In addition, markers of the epithelial–mesenchymal transition (EMT)–such as reduced expression of keratins, a switch from E-Cadherin to N-Cadherin, and enhanced migration in D492M cells—might be a useful marker in breast cancer [25]. Furthermore, expression of the stem cell markers cytokeratins 15 and 19 was altered in squamous cell carcinoma: cytokeratin 15 levels were decreased and the localization of cytokeratin 19 was altered [26]. *KLK3*, which encodes prostate-specific antigen, a member of the kallikrein family of serine proteases, is a biomarker for prostate cancer detection and disease monitoring [27,28]. Mitochondrial DNA mutations have been associated with bladder, head and neck, lung, colorectal, and pancreatic cancer [29-32] (Table 3).

Highly parallel identification of cancer-related genes using small hairpin RNA screening has revealed that the expression of known and putative oncogenes, such as *EGFR*, *KRAS*, *MYC*, *BCR-ABL*, *MYB*, *CRKL*, and *CDK4* that are essential for cancer proliferation, is altered in cancer cells [33]. Other genes such as *PTPN1*, *NF1*, *SMARCB1*, and *SMARCE1* have been identified as essential for the imatinib response of leukemia cells, and *TOPOIIA* expression is involved in resistance to etoposide, an anti-topoisomerase II agent, in small cell lung cancer [33-36].

Marker	Cancer Type	Reference
	head and neck cancer	[18]
TP53 mutation	bladder cancer	[20]
	lung cancer (small cell lung cancer and non-small cell lung cancer); breast, colon, esophagus, liver, bladder, ovary, and brain cancers; sarcomas, lymphomas, and leukemias	[19]
EGFR mutation	non-small cell lung cancer	[21]
RAS mutation	colorectal tumors	[22]
DNA microsatellite alterations	bladder cancer	[23]
alteration in cell cycle mRNA expression	colon cancer	[24]
alteration in cytokeratin mRNA expression	squamous cell carcinoma	[26]
alteration in kallikrein mRNA expression	prostate cancer	[27]
mitochondrial DNA mutations	bladder cancer, head and neck cancer, lung cancer	[29]
	colorectal tumors	[30], [32]
	pancreatic cancer	[31]

Table 3. Genomic markers of cancer.

2.2. Genes related to cell proliferation

Cyclins, which regulate the cell cycle, play important roles in cell proliferation and the uncontrolled cell proliferation that is the most important factor in tumorigenesis [37]. Tumor cells accumulate mutations that result in constitutive mitogenic signaling and defective responses to anti-mitogenic signals that contribute to unscheduled proliferation [38]. In cancer, unscheduled proliferation, genomic instability, and chromosomal instability are the three major factors in cell cycle dysregulation [38]. Regulation of the cell cycle is mainly conducted by complexes of cyclins and cyclin-dependent kinases [38]. Cyclin D1 in cell migration and proliferation is temporo-spatially separated by its biphasic expression induced by thrombin, a G protein-coupled receptor agonist, which is mediated by nuclear factor of activated T cells c1 (*NFATC1*) and signal transducer and activator of transcription 3 (*STAT3*) [39]. Cyclin D1 regulates kinase activity and the G_1–S phase tran-

sition in the cell cycle; deregulated cyclin D1 expression is well documented in breast, colon, and prostate cancers [39,40]. The expression of cyclin D1 is regulated by several factors including cytokines such as interleukin 3 and interleukin 6 *via* STAT3 and STAT5, or extracellular matrix factors such as collagen, fibronectin, and vitronectin, which activate focal adhesion kinase upon integrin clustering, and hepatocyte nuclear factor 6 [41]. Cyclin D1 is a crucial regulator of Wnt- and Notch-regulated development [41,42]. The binding of Wnt to its receptor, Frizzled, causes release of β-catenin to translocate from the cytoplasm to the nucleus, where it forms a complex with the ternary complex factor and/or the lymphoid enhancer-binding factor [41,43]. Cyclin D1 is induced by overexpression of β-catenin, which is a major component of adherens junctions that link the actin cytoskeleton to members of the cadherin family of transmembrane cell–cell adhesion receptors. It plays an important role in linking the cytoplasmic side of cadherin-mediated cell–cell contacts to the actin cytoskeleton [43]. Beta-catenin is upregulated in colorectal cancer, which is considered to trigger cyclin D1 gene expression followed by uncontrolled progression of the cell cycle [43]. In addition, β-catenin plays another role in signaling that involves transactivation, in complex with transcription factors of the lymphoid enhancing factor family in the nucleus [43]. The pathway involving β-catenin/LEF1 and elevation of cyclin D1 might be crucial for tumorigenesis [43]. Inhibiting EglN2, a member of the EglN (also called PHD or HPH) family of prolyl hydroxylases that regulates the heterodimeric transcription factor hypoxia-inducible factor (HIF), causes a decrease in the expression of its interaction partner cyclin D1 in cancer cells and impairs the cells' ability to proliferate *in vivo* [44].

Progression of the eukaryotic cell cycle is driven by cyclin-dependent protein kinases (CDKs), which are binding partner of cyclins. The CDK oscillator acts as the primary organizer of the cell cycle [45]. Phosphorylation of cyclin-Cdk complexes is one of the primary mechanisms of cell cycle regulation [46]. Cyclins are degraded by ubiquitin-mediated proteolysis [46]. The ubiquitylation and degradation of cyclin 1 and cyclin 2 are mediated by the SCF complex, a multi-subunit ubiquitin ligase that contains Skp1, a member of the cullin family (Cdc53) and an F-box protein, as well as a RING-finger-containing protein [46]. CDKs including CDK1, CDK2, CDK4, CDK6, and CDK11 have various functions that have been investigated using loss-of-function, target validation, and gain-of-function mouse models [38]. CDK1 is a mitotic CDK, also known as cell division control protein 2 (CDC2). It is one of the master regulators of mitosis as it controls the centrosome cycle as well as mitotic onset; deficiency in CDK1 results in embryonic lethality in the first cell divisions [38,47]. CDK2, CDK4, and CDK6 are interphase CDKs that are not essential for the mammalian cell cycle; they are, however, required for the proliferation of specific cell types [38]. Deficiency in CDK2, CDK4, and CDK6 caused mid-gestation embryonic lethality because of hematopoietic defects [38,47].

2.3. Genes related to cell differentiation

Inhibitor of differentiation 1 (Id1) is associated with the induction of cell proliferation and invasion [48], as well as the invasive features of cancer and the EMT [48]. The *HOX* genes

encode homeodomain-containing transcription factors involved in the regulation of cellular proliferation and differentiation during embryogenesis [49]. The expression of HOXA1, which plays an important role in proliferation, apoptosis, adhesion, invasion, the EMT, and anchorage-independent growth, was significantly increased in oral squamous cell carcinoma compared with in healthy oral mucosa [49], and it might be a useful prognostic marker for patients with this disease [49].

Wnt/β-catenin signaling controls skeletal development and differentiation [50]. The initiating step of skeletal development is mesenchymal condensation, during which mesenchymal progenitor cells are at least bipotentiate [50]. Osteochondral progenitor cells differentiate into osteoblasts instead of chondrocytes when Wnt/β-catenin signaling is activated [50]. *In vitro* models using human pluripotent stem cell-derived neural progenitor cells have been used to examine whether G11778A-mutated mitochondrial DNA, which is associated with Leber's hereditary optic neuropathy, might be involved in the differentiation of neural progenitor cells into neurons, oligodendrocytes, and astrocytes [51]. The differentiation of neural progenitor cells can be visualized by staining for the neuronal marker class III beta-tubulin [51]. Alternative splicing of exons play an important role in cellular differentiation and pathogenesis [52]. Alternative splicing in colorectal cancer and renal cell cancer samples has been analyzed by the Bioinformatics Exon Array Tool (BEAT, http://beat.ba.itb.cnr.it/) using an Affymetrix GeneChip Exon Array [52]. When the dataset was analyzed using GO terms, the cell differentiation (GO:0030154)-related gene delta-like 1 (Drosophila) (*DLL1*) was found to be involved in colorectal cancer [52].

2.4. Genes related to apoptosis

Cell proliferation and death are regulated by various molecules. Recently, microRNAs have been revealed to play important roles during death receptor-mediated apoptosis (programmed cell death) [53]. Transfection with miR-133b caused a proapoptotic effect on tumor necrosis factor alpha (TNFα)-stimulated HeLa cells [53]: the expression of apoptosis regulatory proteins such as transgelin 2 (TAGLN2), myosin, heavy chain 9, non-muscle (MYH9), cytoskeleton-associated protein 4 (CKAP4), polypyrimidine tract binding protein 1 (PTBP1), glutathione-S-transferase pi 1 (GSTP1), and copine III (CPNE3) were down-regulated compared with in control cells [53]. The BCL protein family plays a major role in regulation of the apoptotic cascade [54]. BCL2-associated protein (BAX) promotes apoptosis and delays disease progression, and has been associated with longer disease-free survival in patients with a number of gastrointestinal cancers, such as esophageal, stomach, small intestine, and colon cancer; moreover, high BCL6 expression is correlated with worse prognosis in patients with other gastrointestinal tumors, such as esophageal adenocarcinoma [54]. There are two major cell death pathways that transduce the effects of various death inducers: the extrinsic death pathway that is mediated through cell death receptors of the TNF receptor family, such as the Fas receptor; and the intrinsic death pathway that proceeds through mitochondria [55]. The expression of apoptosis signal-regulating kinase (ASK1), which plays an important role as a mitogen-activated protein kinase kinase kinase in apoptosis signaling, is in-

creased in gastric cancer [56]. Furthermore, the levels of cyclin D1 and phosphorylated JNK were higher in gastric cancer than in non-tumor epithelium [56]. ASK1 may play a role in the development of gastric cancer [56].

2.5. Detection of cell proliferation or apoptosis

Several methods have been suggested for the diagnosis of cancer [57]. Protein markers for cancer include prostate-specific antigen for prostate cancer, CA125 for ovarian cancer, carcinoembryonic antigen for colon cancer, human chorionic gonadotropin for trophoblastic cancer, and a-fetoprotein for hepatocellular carcinoma and germ cell tumors [57]. Assays to detect telomerase activity in clinical samples include the TRAP (telomere repeat amplification protocol) assay, which involves protein extraction and subsequent primer-directed PCR amplification of telomere extensions [57].

Assays for the detection of kinases that regulate cell growth, proliferation, differentiation, and metabolism have been developed [58]. The assay technology includes fluorescence polarization to detect protein phosphorylation, scintillation proximity to detect protein dephosphorylation by phosphatases, fluorescence resonance energy transfer to detect protein cleavage or modification, immunosorbent assays to detect phosphorylation state, luciferase-based ATP detection to detect the kinase-dependent depletion of ATP, luminescent oxygen channeling to detect phosphorylation, time-resolved fluorescence resonance energy transfer to detect phosphopeptide formation, and enzyme fragment complementation to detect molecular interactions with kinases [58,59]. Cell proliferation can also be determined by the tetrazolium hydroxide (XTT) cell proliferation assay, in which absorbance is measured by an ELISA reader under 490-nm-wavelength light (Biological Industries) [60].

Cell proliferation assays and apoptosis assays have been used to examine the effects of inhibitors on cancer cells [61]. The cell proliferation of Neuro-2A cells, neuroblastoma cells, can be determined using the CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay reagent (Promega) [61]. A colony formation assay using Neuro-2A cells was used to determine the effect of an inhibitor of GSK-3β [61]. In this experiment, colonies were allowed to form for 10 days, after which the cells were fixed with 70% ethanol and stained with 1% methylene blue. Apoptosis was then measured by flow cytometry using an Annexin V-allophycocyanin (APC) /propidium iodide (PI) detection kit (BD PharMingen) [61]. Apoptosis was also determined using 4'6-diamidino-2-phenylindole (DAPI) staining, observing apoptotic nuclear morphology, and immunoblotting with antibodies to β-catenin, X-linked inhibitor of apoptosis, and BCL2 [61]. Cell cycle analysis using PI to quantify the proportions of cells in the G_1/G_0 or G_2–M phases was used to examine cell cycle status [61].

Viable cells can be determined using MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) colorimetric assays [62]. Absorbance at 570 nm is used to detect the incorporation of MTT. Apoptosis can also be determined by caspase activation using an anti-poly ADP-ribose polymerase (PARP) antibody [62]. Viable cells can also be determined using a 3-(4,5-dimethyl-thiazol-2yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) kit (Promega) [63]. The terminal transferase dUTP nick end labeling (TUNEL) assay is commonly used to detect apoptosis [63]. Harvested cells are resuspended in DNA labeling solution

consisting of TdT reaction buffer, TdT enzyme, and BrdUTP, then stained with PI to detect a fluorescein isothiocyanate-labeled anti-BrdU antibody [63]. Cell viability and proliferation assays were used to validate internal tandem duplication mutations in *FLT3* as a therapeutic target for human acute myeloid leukemia [64]. Cell viability and proliferation can be determined using a Vi-cell XR automated cell viability analyzer (Beckman Coulter) [64].

3. Genomic variation in disease

3.1. Genome-wide association studies in cancer

Despite extensive research efforts for several decades, the genetic basis of common human diseases such as cancers remains largely unknown [65]. Genome-wide association studies (GWAS) have emerged as an important tool for the discovery of genomic regions that harbor genetic variants conferring risk for various cancers [66,67]. Family-based linkage studies and studies comprising tens of thousands of gene-based SNPs can also assay genetic variation across the genome [68], but the National Institutes of Health guidelines for GWAS require a sufficient density of genetic markers to capture a large proportion of the common variants in the study population, measured in enough individuals to provide sufficient power to detect variants of modest effect [67]. The recent success of GWAS can be attributed to the convergence of new technologies that can genotype hundreds of thousands of SNPs in hundreds or thousands of samples [66,69].

GWAS have been conducted in the five of the most common cancer types: breast, prostate, colorectal, lung, and melanoma (Table 4) and have identified more than 20 novel disease loci, confirming that susceptibility to these diseases is polygenic [70]. For many years, human genetics has been used to map rare mutations with large effect sizes in families or genetically homogeneous populations, such as *BRCA1/BRCA2* mutations in Ashkenazi women with breast cancer and ovarian cancer [71]. A number of SNPs have now been associated with breast cancer; for example, a SNP in intron 2 of the *FGFR2* gene, which encodes a receptor tyrosine kinase that is amplified and overexpressed in 5–10% of breast tumors [72,73], and SNPs on chromosomes 16q and 5q. The locus on 16q contains a gene *TNRC9* and a hypothetical gene *LOC643714*. The function of TNRC9 is unknown but the presence of an HMG box motif suggests that it might act as a transcription factor. The 5q locus includes *MAP3K1*, which encodes a protein involved in signal transduction (but not previously known to be involved in cancer) and two other genes: *MGC33648* and *MIER3*. In addition, several of the breast cancer loci appear to be associated with specific subtypes of the disease. In particular, the *FGFR2* association is strongly associated with estrogen receptor-positive breast cancer, while the *TNRC9* SNP is associated with both estrogen receptor-positive and -negative breast cancer [74,75]. It is surprising that none of the strongest associations map to regions harboring estrogen/progesterone genes in women of European background, particularly because a GWAS in Asian women reported a convincing association with markers near the estrogen receptor alpha (*ESR1*) gene [76]. In prostate cancer, the first and most important region to emerge was 8q24. This region was first associated with prostate cancer through

linkage studies by the deCode group, was followed up by association analyses [77], and has been confirmed in subsequent GWAS [78-81]. Another signal, on chromosome 10q13, points to a variant in the promoter of the *MSMB* gene, which encodes the PSP94 protein; this is now under intense investigation as a biomarker for prostate cancer [80,81].

In general, the susceptibility alleles discovered thus far are common—that is, with a frequency in one or more population of >10%, and each allele confers a small contribution to the overall risk of the disease. For nearly all regions conclusively identified by GWAS, the effect sizes per allele are estimated at <1.3. It was not anticipated that GWAS in certain cancers would yield many novel regions when other cancers strongly associated with particular environmental exposures have yielded so few regions. For example, prostate cancer, breast cancer, and colon cancer have been associated with 29, 13, and 10 regions of the genome, respectively, while there are only three associated regions for lung cancer in smokers, and three for bladder cancer despite analysis of sufficiently large data sets [67]. Several GWAS for lung cancer have identified the same locus on 15q25, suggesting that this is an important susceptibility locus for this disease [82-87]. This locus contains the nicotinic acetylcholine receptor subunit genes *CHRNA3* and *CHRNA5*, suggesting that susceptibility may be mediated through smoking behavior [86,87].

GWAS represent an important advance in discovering genetic variants influencing disease but have important limitations. There is a high potential for false-positive results, they do not yield information on gene function, they are insensitive to rare and structural variants, they require large sample sizes, and incur possible biases because of case and control selection and genotyping errors [88]. Clinicians and scientists must understand the unique aspects of these studies and be able to assess and interpret GWAS results for themselves and their patients. However, at present these studies mainly represent a valuable discovery tool for examining genomic function and clarifying pathophysiological mechanisms. However, through GWAS, the identification of variants, genes, and pathways involved in multiple cancers offers a potential route to new therapies, improved diagnosis, and better disease prevention [65].

Cancer type	Reference	Year	Platform [SNP passing QC]	Ethnic group	Initial sample size		Replication sample size		
					Cases	Controls	Ethnic groups	Cases	Controls
Breast cancer	[89]	2012	Affymetrix [555,525]	Korean	2,273	2,052	Korean	4,049	3,845
	[90]	2012	Affymetrix [690,947]	Chinese	2,918	2,324	Chinese	6,838	6,888
							Han Chinese	1,297	1,585
							Taiwan Chinese	1,066	1,065
							Korean	5,038	6,869
							Japanese	1,934	1,875

Cancer type	Reference	Year	Platform [SNP passing QC]	Ethnic group	Initial sample size		Replication sample size		
					Cases	Controls	Ethnic groups	Cases	Controls
	[91]	2012	Affymetrix [613,031]	Chinese	1,950	-	Chinese	4,160	-
	[92]	2012	Illumina [470,796]	Japanese	240	-	Japanese	222	-
	[93]	2011	Affymetrix [684,457]	East Asian	2,062	2,066	East Asians	15,091	14,877
	[94]	2011	Affymetrix [782,838]	European	302	321	European	1,153	1,215
	[95]	2011	Illumina [~296,114]	British	1,694	2,365	British	7,317	8,124
							European	1,145	1,142
	[96]	2010	Illumina [285,984]	Swedish & Finnish	617	4,583	European	1,001	7,604
	[97]	2010	Affymetrix [592,163]	European	899	804	European	1,264	1,222
	[98]	2010	Illumina [285,984]	European	2,702	5,726	European	7,386	7,576
	[99]	2010	Illumina [582,886]	UK	3,659	4,897	European	12,576	12,223
	[100]	2010	Illumina [528,252]	British	1,145	-	British	4,335	-
	[101]	2009	Illumina [528,173]		1,145	1,142		8,625	9,657
	[102]	2009	Affymetrix [up to 607,728]	Chinese	1,505	1,522	Chinese	1,554	1,576
	[103]	2008	Affymetrix [200,220]		30	30	-	-	-
	[104]	2008	Affymetrix [492,900]	Ashkenazi Jewish	249	299	Ashkenazi Jewish	1,193	1,166
	[105]	2007	Affymetrix [70,897]	Framing-ham	1,345	-	-	-	-
	[106]	2007	Perlegen [205,586]		390	634		26,646	24,889
	[107]	2007	Illumina [528,173]		1,145	1,142		1,176	2,072

Cancer type	Reference	Year	Platform [SNP passing QC]	Ethnic group	Initial sample size		Replication sample size		
					Cases	Controls	Ethnic groups	Cases	Controls
Prostate cancer	[108]	2012	Illumina [509,916]	European	1,176	1,101	European	1,964	3,172
	[109]	2012	Affymetrix & Illumina [1,117,531] (imputed)		4,723	4,792	-	-	-
	[110]	2011	NR [2.6 million] (imputed)	European	6,621	6,939	European	22,957	23,234
							Japanese	285	298
							Chinese	135	135
							African American	112	298
								7,140	5,455
	[111]	2011	Illumina [571,243]	European	2,782	4,458	European	8,217	6,732
	[112]	2011	Illumina [1,047,198]	African American	3,425	3,290	African American	1,275	1,695
							Senegalese	86	414
							Ghanaian	271	968
							Barbadian	246	253
	[113]	2011	Affymetrix [387,384]	European	202	100	European	1,122	1,167
	[114]	2010	Affymetrix [419,613]	Caucasian	222	415	Caucasian	500	155
	[115]	2010	Illumina [510,687]	Japanese	1,583	3,386	Japanese	3,001	5,415
	[116]	2009	Illumina [541,129]	European	1,854	1,894	European, Chinese, Japanese, African American, Latino, and Hawaiian	19,879	18,761
	[117]	2009	Illumina [310,520]	Icelandic	1,968	35,382	European	11,806	12,387
	[118]	2008	Illumina [541,129]	European	1,854	1,894		3,268	3,366

Cancer type	Reference	Year	Platform [SNP passing QC]	Ethnic group	Initial sample size		Replication sample size		
					Cases	Controls	Ethnic groups	Cases	Controls
	[119]	2008	Illumina [527,869]		1,172	1,157		3,941	3,964
	[120]	2007	Affymetrix & Illumina [60,275]		1,235	1,599		1,242	917
	[121]	2007	Affymetrix [70,897]	Framingham	1,345		-	-	-
	[122]	2007	Illumina [316,515]		1,453	3,064	East Asia	1,210	2,445
	[123]	2007	Illumina [538,548]		1,172	1,157		3,124	3,142
Colorectal cancer	[124]	2011	Illumina [378,739]	European	2,906	3,416	European	8,161	9,101
	[125]	2010	Illumina [~550,000] (imputed)	European	3,334	4,628	European	14,851	15,569
	[126]	2010	Affymetrix [460,945]	German	371	1,263	German Czech	4,121 794	7,344 815
	[127]	2008	Illumina [~548,586]		1,902	1,929		4,878	4,914
	[128]	2008	Illumina [541,628]		981	1,002		16,476	15,351
	[129]	2008	Illumina [547,647]		922	927		17,872	17,526
	[130]	2007	Illumina [547,647]		940	965		7,473	5,984
	[131]	2007	Illumina [547,647]		930	960		7,334	5,246
	[132]	2007	Affymetrix & Illumina [99,632]		1,257	1,336		6,223	6,443
Lung cancer	[133]	2011	Affymetrix [906,703]	Chinese	2,331	3,077	Chinese	6,313	6,409
	[134]	2011	Illumina [307,260]	White	327		European	587	

Cancer type	Reference	Year	Platform [SNP passing QC]	Ethnic group	Initial sample size		Replication sample size		
					Cases	Controls	Ethnic groups	Cases	Controls
	[135]	2011	Illumina [620,901] (pooled)	Italian	600	-	Italian	317	-
	[136]	2010	Affymetrix [265,996]	Han Chinese	245	-	Han Chinese	305	-
	[137]	2010	Affymetrix [246,758]	Korean	621	1,541	Korean	804	1,470
	[138]	2010	Illumina [542,050]		377	377		511	1,007
	[139]	2009	Illumina [515,922]	European	5,739	5,848	European	7,561	13,818
	[140]	2009	Illumina [511,919]	European	1,952	1,438	European	5,608	6,767
	[141]	2008	Illumina [317,498]		1,154	1,138		2,724	3,694
	[142]	2008	Illumina [306,207]		10,995 smokers			4,848 smokers	
	[143]	2008	Illumina [310,023]		1,926	2,522		2,513	4,752
	[144]	2007	Affymetrix [~116,204] (pooled)	Italian	338	335	Norwegian	265	356
Melanoma	[145]	2011	Illumina [594,997]	European	2,804	7,618	European	5,551	7,449
	[146]	2011	Illumina [5,480,804 (imputed)]	European	2,168	4,387	European	5,193	15,144
	[147]	2011	Illumina [818,977]	European	1,804	1,026	European	6,483	23,324
	[148]	2011	Illumina [491,227]	European	156	2,150	NR	-	-
	[149]	2009	Illumina [~317,000]	European	1,539	3,917	European	2,312	1,867

Table 4. Summary of GWAS for the five of the most common types of cancer.

3.2. Genetic risk score in cancer and diabetes

Type 2 diabetes mellitus and cancers are major health problems worldwide [150,151]. The recent increase in the prevalence of these diseases is largely attributable to environmental factors. However, convincing evidence shows that genetic factors may play an important role in these diseases [152,153]. Recent GWAS have led to the identification of a series of SNPs that are robustly associated with either the risk of diabetes or cancers [151,154-159]. For type 2 diabetes mellitus, common SNPs have been identified in the *PPARG*, *KCNJ11*, and *TCF7L2* genes, and have been widely replicated in populations of various ethnicities [160-162]. Other potential new loci include *HHEX*, *CDKAL1*, *CDKN2A/B*, *IGF2BP2*, *SLC30A8*, and *WFS1* [65,155-159,163,164]. A number of SNPs have been identified as associated with breast cancer risk, including *FGFR2*, *CASP8*, *ERBB4*, *TAB2*, *BARX2*, *TMEM45B*, *ESR1*, *FGFR2*, *TNRC9*, *MAP3K1*, *MGC33648*, *MIER3*, and *RAD51L1* [74,75,151] (Table 5).

Combining multiple loci with modest effects into a global genetic risk score (GRS) might improve the identification of those at risk for common complex diseases such as type 2 diabetes and cancers [165-167]. Several studies have developed methods to predict the risk of certain diseases, such as coronary heart disease, type 2 diabetes, and breast cancer, aggregating information from multiple SNPs into a single GRS [151,168,169]. For example, in the Atherosclerosis Risk in Communities study, the aggregation of multiple SNPs into a single GRS was responsible for improving the prediction of coronary heart disease incidence [168]. In a study that used a GRS to determine the risk of type 2 diabetes in US men and women, individuals in the highest quintile of GRS had a significantly increased risk of type 2 diabetes compared with those in the lowest quintile; however, the addition of a GRS to the conventional model consisting of lifestyle risk factors only increased the area under the curve by only 1% (AUC=0.78). In this instance, the GRS was determined to be useful only when combined with the body mass index or a family history of diabetes [169]. For breast cancer, a GRS was created using 14 SNPs previously associated with breast cancer, and was substantially more predictive of estrogen receptor-positive breast cancer than of estrogen receptor-negative breast cancer, particularly for absolute risk [151]. Further studies are needed to confirm whether a GRS improves disease risk prediction.

The GRS is calculated on the basis of reproducible tagging of SNP-associated loci reaching genome-wide levels of significance. The GRS can be created by two methods: a simple count method (count GRS) and a weighted method (weighted GRS) [169,170]. Both methods anticipate each SNP to be independently associated with risk. An additive genetic model is used for each SNP, applying a linear weighting of 0, 1, or 2 to genotypes containing 0, 1, or 2 risk alleles, respectively. This model is known to perform well even when the true genetic model is unknown or wrongly specified [171]. The count model assumes that each SNP in the panel contributes equally to the disease risk and is calculated by summing the values for each of the SNPs. The weighted GRS is calculated by multiplying each B-coefficient, the estimates resulting from an analysis carried out on variables that have been standardized, by the number of corresponding risk alleles (0, 1, or 2).

Disease	Reference	Year	Ethnic group	Participants	No. of SNPs	Genes found from GWAS
Type 2 diabetes	[165]	2008	Framingham	2,377 diabetic patients	18	NOTCH2 (rs10923931), BCL11A (rs10490072), THADA (rs7578597), IGF2BP2 (rs1470579), PPARg (rs1801282), ADAMTS9 (rs4607103), CDKAL (rs7754840), VEGFA (rs9472138), JAZF1 (rs86475), SLC30A8 (rs13266634), CDKNA/2B (rs10811661), HHEX (rs1111875), CDC123, CAMK1D (rs12779790), TCF7L2 (rs7903146), KCNJ11 (rs5219), INS (rs689), DCD (rs1153188), TSPAN8, LGR5 (rs7961581)
	[169]	2009	European	2,809 diabetic patients & 3,501 health controls	10	WFS1 (rs10010131), HHEX (rs1111875), CDKAL1 (rs7756992), IGF2BP2 (rs4402960), SLC30A8 (rs13266634), CDKN2A/B (rs10811661), TCF7L2 (rs12255372), PPARG (rs1801282), KCNJ11 (rs5219)
Breast cancer	[151]	2010	UK	10,306 breast cancer patients & 10,393 controls	14	FGFR2 (rs2981582), TNRC9 (rs3803662), 2q35 (rs13387042), MAP3K1 (rs889312), 8q24 (rs13281615), 2p (rs4666451), 5pas (rs981782), CASP8 (rs104548), LSP1 (rs3817198), 5p (rs30099), TGFB1 (rs198/2073), ATM (rs1800054), TNRC9 (rs8051542), TNRC9 (rs12443621)

Table 5. Studies using a genetic risk score for cancers and diabetes, comprising SNPs identified in GWAS.

3.3. Cancer Cell Line Encyclopedia

The Cancer Cell Line Encyclopedia (CCLE) has made predictive modeling of anticancer drug sensitivity a realistic proposition, by determining genomic markers of drug sensitivity in cancer cells [172,173]. The CCLE contains information from 947 human cancer cell lines including data on gene expression, chromosomal copy number, and massively parallel sequencing data. It has been used to identify genetic, lineage-specific, and gene expression-based predictors of drug sensitivity [172]. This has revealed, for example, that the plasma cell lineage is correlated with sensitivity to IGF1 receptor inhibitors, aryl hydrocarbon receptor (*AHR*) expression is associated with MEK inhibitor efficacy in *NRAS*-mutant lines, and *SLFN11* expression is associated with sensitivity to topoisomerase inhibitors [172]. Genomic markers of drug sensitivity in cancer cells have also been systematically identified using the Genomics of Drug Sensitivity in Cancer database (http://www.cancerRxgene.org) [173]. These databases will enable to overview genome quality.

4. Conclusion

There are dramatic changes in the genomes of cancer cells, which vary according to cancer subtype. Integrative and wide investigations of cancer cell genomes have revealed mutations and alterations in gene expression that are associated with the disease. Databases that include abundant data related to gene and protein conformation, gene expression, and genomic mutations enable the construction of dynamic cellular simulations and disease models. New sequencing tools such as next-generation sequencing will reveal new horizons in the prediction of disease and drug sensitivity, which play an important role in personalized medicine. Appropriate translation of the abundance of information to clinical practice is one of most important future challenges for medicine. The quality of genome would be one of the important factors for detecting the development of the disease.

Acknowledgements

Author details

Shihori Tanabe[1*] and Sun Ha Jee[2]

*Address all correspondence to: stanabe@nihs.go.jp

1 National Institute of Health Sciences, Tokyo, Japan

2 Institute for Health Promotion, Yonsei University, Seoul, Korea

References

[1] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science*, 291(5507), 1304-1351.

[2] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. Paper presented at International Human Genome Sequencing Consortium. *Nature*, 409(6822), 860-921.

[3] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, 25(1), 25-29.

[4] Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., et al. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res.*, 37, D211-D215.

[5] Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.*, 40, D306-D312.

[6] Attwood, T. K., Coletta, A., Muirhead, G., Pavlopoulou, A., Philippou, P. B., Popov, I., et al. (2012). The PRINTS database: a fine-grained protein sequence annotation and analysis resource--its status in 2012. *Database*, Oxford, bas019.

[7] Sigrist, C. J. A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A., et al. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.*, 38, D161-D166.

[8] Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.*, 40, D290-D301.

[9] Letunic, I., Doerks, T., & Bork, P. (2012). SMART7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.*, 40, D302-D305.

[10] Bru, C., Courcelle, E., Carrère, S., Beausse, Y., Dalmar, S., & Kahn, D. (2005). The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Res.*, 33, D212-D215.

[11] Nikolskaya, A. N., Arighi, C. N., Huang, H., Barker, W. C., & Wu, C. H. (2007). PIRSF family classification system for protein functional and evolutionary analysis. *Evol. Bioinform*, Online, 2, 197-209.

[12] de Lima, Morais. D. A., Fang, H., Rackham, O. J. L., Wilson, D., Pethica, R., Chothia, C., et al. (2011). SUPERFAMILY 1.75 including a domain-centric gene ontology method. *Nucleic Acids Res.*, 39, D427-D434.

[13] Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., & Thomas, P. D. (2010). PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.*, 38, D204-D210.

[14] Lees, J., Yeats, C., Perkins, J., Sillitoe, I., Rentzsch, R., Dessailly, B. H., et al. (2012). Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. *Nucleic Acids Res.*, 40, D465-D471.

[15] Selengut, J. D., Haft, D. H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W. C., et al. (2007). TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res.*, 35, D260-D264.

[16] Lima, T., Auchincloss, A. H., Coudert, E., Keller, G., Michoud, K., Rivoire, C., et al. (2009). HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/ Swiss-Prot. *Nucleic Acids Res.*, 37, D471-D478.

[17] Sidransky, D. (2002). Emerging molecular markers of cancer. *Nat. Rev. Cancer*, 2(3), 210-219.

[18] Brennan, J. A., Mao, L., Hruban, R. H., Boyle, J. O., Eby, Y. J., Koch, W. M., et al. (1995). Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.*, 332(7), 429-435.

[19] Hollstein, M., Sidransky, D., Vogelstein, M., & Harris, C. C. (1991). p53 mutations in human cancers. *Science*, 253(5015), 49-53.

[20] Sidransky, D., Von Eschenbach, A., Tsai, Y. C., Jones, P., Summerhayes, I., Marshall, F., et al. (1991). Identification of p53 gene mutations in bladder cancers and urine samples. *Science*, 252(5006), 706-709.

[21] Rossi, A., & Galetta, D. (2012). Biomarkers for the targeted therapies of non-small cell lung cancer. *Current Biomarker Findings*, 2, 7-17.

[22] Sidransky, D., Tokino, T., Hamilton, S. R., Kinzler, K. W., Levin, B., Frost, P., et al. (1992). Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science*, 256(5053), 102-105.

[23] Gonzalez-Zulueta, M., Ruppert, J. M., Tokino, K., Tsai, Y. C., Spruck, I. I. I. C. H., Miyano, N., et al. (1993). Microsatellite instability in bladder cancer. *Cancer Res.*, 53(23), 5620-5623.

[24] Belt, E. J. T., Brosens, R. P. M., Delis-van Diemen, P. M., Bril, H., Tijssen, M., van Essen, D. F., et al. (2012). Cell cycle proteins predict recurrence in stage II and III colon cancer. *Ann. Surg. Oncol.*, DOI: 10.1245/ s10434-012-2216-7, Published online: 04 Feb 2012.

[25] Sigurdsson, V., Hilmarsdottir, B., Sigmundsdottir, H., Fridriksdottir, A. J. R., Ringnér, M., Villadsen, R., et al. (2011). Endothelial induced EMT in breast epithelial cells with stem cell properties. *PLoS One*, 6(9), e23833.

[26] Abbas, O., Richards, J. E., Yaar, R., & Mahalingam, M. (2011). Stem cell markers (cytokeratin 15, cytokeratin 19 and p63) in in situ and invasive cutaneous epithelial lesions. *Mod. Pathol.*, 24(1), 90-97.

[27] Kote-Jarai, Z., Amin, Al., Olama, A., Leongamornlert, D., Tymrakiewicz, M., Saunders, E., et al. (2011). Identification of a novel prostate cancer suscepti-bility variant in the KLK3 gene transcript. *Human Genet.*, 129(6), 687-694.

[28] Schröder, F., Hugosson, J., Roobol, M. J., Tammela, T. L. J., Ciatto, S., Nelen, V., et al. (2012). ERSPC Investigators. Prostate-cancer mortality at 11 years of follow-up. *N. Engl. J. Med.*, 366(11), 981-990.

[29] Fliss, M. S., Usadel, H., Caballero, O. L., Wu, L., Buta, M. R., Eleff, S. M., et al. (2000). Facile detection of mitochondrial DNA mutations in tumors and bodily fluid. *Science*, 287(5460), 2017-2019.

[30] Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J. K. V., Markowitz, S. D., et al. (1998). Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat. Genet.*, 20(3), 291-293.

[31] Lam, E. T., Bracci, P. M., Holly, E. A., Chu, C., Poon, A., Wan, E., et al. (2012). Mitochondrial DNA sequence variation and risk of pancreatic cancer. *Cancer Res.*, 72(3), 686-695.

[32] Skonieczna, K., Malyarchuk, B. A., & Grzybowski, T. (2012). The landscape of mitochondrial DNA variation in human colorectal cancer on the back-ground of phylogenetic knowledge. *Biochim. Biophys. Acta*, 1825(2), 153-159.

[33] Luo, B., Cheung, H. W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., et al. (2008). Highly parallel identification of essential genes in cancer cells. *Proc. Natl. Acad. Sci.*, USA, 105(51), 20380-20385.

[34] Nitiss, J. L., Liu, Y., & Hsiung, Y. (1993). A temperature sensitive topoiso-merase II allele confers temperature dependent drug resistance on amsacrine and etoposide: a genetic system for determining the targets of topoisomerase II inhibitors. *Cancer Res.*, 53(1), 89-93.

[35] Osheroff, N. (1989). Effect of antineoplastic agents on the DNA cleavage/ religation reaction of eukaryotic topoisomerase II: inhibition of DNA religa-tion by etoposide. *Biochemistry*, 28(15), 6157-6160.

[36] Hande, K. R. (1998). Etoposide: four decades of development of a topoiso-merase II inhibitor. *European J. Cancer*, 34, 1514-1521.

[37] Sherr, C. J. (1996). Cancer cell cycles. *Science*, 274(5293), 1672-1677.

[38] Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer*, 9(3), 153-166.

[39] Kundumani-Sridharan, V., Quyen, D. V., Subramani, J., Singh, N. K., Chin, Y. E., & Rao, G. N. (2012). Novel interactions between NFATc1 (nuclear factor of activated T cells c1) and STAT-3 (signal transducer and activator of transcription-3) mediate G protein-coupled receptor agonist, thrombin-induced biphasic expression of cyclin D1, with first phase influencing cell migration and second phase directing cell proliferation. *J. Biol. Chem.*, 287(27), 22463-22482.

[40] Landis, M. W., Pawlyk, B. S., Li, T., Sicinski, P., & Hinds, P. W. (2006). Cyclin D1-dependent kinase activity in murine development and mammary tumorigenesis. *Cancer Cell*, 9(1), 13-22.

[41] Klein, E. A., & Assoian, R. K. (2008). Transcriptional regulation of the cyclin D1 gene at a glance. *J. Cell Sci.*, 121, Pt 23, 3853-3857.

[42] Hsu, W., Shakya, R., & Costantini, F. (2001). Impaired mammary gland and lymphoid development caused by inducible expression of Axin in transgenic mice. *J. Cell Biol.*, 155(6), 1055-1064.

[43] Shtutman, M., Zhurinsky, J., Simcha, I., Albanese, C., D'Amico, M., Pestell, R., et al. (1999). The cyclin D1 gene is a target of the beta-catenin/LEF-1 pathway. *Proc. Natl. Acad. Sci.*, USA, 96(10), 5522-5527.

[44] Zhang, Q., Gu, J., Li, L., Liu, J., Luo, B., Cheung, H., et al. (2009). Control of cyclin D1 and breast tumorigenesis by the EglN2 prolyl hydroxylase. *Cancer Cell*, 16(5), 413-424.

[45] Coudreuse, D., & Nurse, P. (2010). Driving the cell cycle with a minimal CDK control network. *Nature*, 468(7327), 1074-1079.

[46] Bloom, J., & Cross, F. R. (2007). Multiple levels of cyclin specificity in cell-cycle control. *Nat. Rev. Mol. Cell Biol.*, 8(2), 149-160.

[47] Santamaría, D., Barrière, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., et al. (2007). Cdk1 is sufficient to drive the mammalian cell cycle. *Nature*, 448(7155), 811-815.

[48] Tobin, N. P., Sims, A. H., Lundgren, K. L., Lehn, S., & Landberg, G. (2011). Cyclin D1, Id1 and EMT in breast cancer. *BMC Cancer*, 11, 417.

[49] Bitu, C., Destro, M., Carrera, M., Silva, S., Graner, E., Kowalski, L. P., Soares, F., et al. (2012). HOXA1 is overexpressed in oral squamous cell

carcinomas and its expression is correlated with poor prognosis. *BMC Cancer*, 12, 146.

[50] Yang, Y. (2012). Wnt signaling in development and disease. *Cell & Bioscience*, 2, 14.

[51] Iyer, S., Xiao, E., Alsayegh, K., Eroshenko, N., Riggs, M. J., Bennett, J. P., et al. (2012). Mitochondrial gene replacement in human pluripotent stem cell-derived neural progenitors. *Gene Ther.*, 19(5), 469-475.

[52] Consiglio, A., Carella, M., De Caro, G., Delle, Foglie. G., Giovannelli, C., Grillo, G., et al. (2012). BEAT: Bioinformatics Exon Array Tool to store, analyze and visualize Affymetrix GeneChip Human Exon Array data from disease experiments. *BMC Bioinformatics*, 13(Suppl 4), S21.

[53] Patron, J. P., Fendler, A., Bild, M., Jung, U., Müller, H., Arntzen, M., et al. (2012). MiR-133b targets antiapoptotic genes and enhances death receptor-induced apoptosis. *PLoS One*, 7(4), e35345.

[54] Alevizos, L., Gomatos, I. P., Smparounis, S., Konstadoulakis, M. M., & Zografos, G. (2012). Review of the molecular prolife and modern prognostic markers for gastric lymphoma: how do they affect clinical practice? *Can. J. Surg.*, 55(2), 117-124.

[55] Aoudjit, F., & Vuori, K. (2012). Integrin signaling in cancer cell survival and chemoresistance. *Chemother. Res. Pract.*, 2012, 283181.

[56] Hayakawa, Y., Hirata, Y., Nakagawa, H., Sakamoto, K., Hikiba, Y., Kinoshita, H., et al. (2011). Apoptosis signal-regulating kinase 1 and cyclin D1 compose a positive feedback loop contributing to tumor growth in gastric cancer. *Proc. Natl. Acad. Sci.*, USA, 108(2), 780-785.

[57] Cairns, P., & Sidransky, D. (1999). Molecular methods for the diagnosis of cancer. *Biochim. Biophys. Acta*, 1423(2), C11-C18.

[58] Glickman, J. F., Mc Gee, J., & Napper, A. (2004). *Assay development for protein kinase enzyme. Assay Guidance Manual [Internet]. Sittampalam GS, Weidner J, Auld D, et al., editors*, Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2012.

[59] Eglen, R. M., & Singh, R. (2003). Beta galactosidase enzyme fragment complementation as a novel technology for high throughput screening. *Comb. Chem. High Throughput Screen.*, 6(4), 381-387.

[60] Baran, Y., Zencir, S., Cakir, Z., Ozturk, E., & Topcu, Z. (2011). Imatinib-induced apoptosis: a possible link to topoisomerase enzyme inhibition. *J. Clin. Pharm. Ther.*, 36(6), 673-679.

[61] Dickey, A., Schleicher, S., Leahy, K., Hu, R., & Hallahan, D. (2011). Thotala DK GSK-3β inhibition promotes cell death, apoptosis, and in vivo tumor

growth delay in neuroblastoma Neuro-2A cell line. *J. Neurooncol.*, 104(1), 145-153.

[62] Wang, L., Hurley, D. G., Watkins, W., Araki, H., Tamada, Y., Muthukaruppan, A., et al. (2012). Cell cycle gene networks are associated with melanoma prognosis. *PLoS One*, 7(4), e34247.

[63] Kim, J., Chae, M., Kim, W. K., Kim, Y., Kang, H. S., Kim, H. S., et al. (2011). Salinomycin sensitizes cancer cells to the effects of doxorubicin and etoposide treatment by increasing DNA damage and reducing p21protein. *Br. J. Pharmacol.*, 162(3), 773-784.

[64] Smith, C. C., Wang, Q., Chin, C., Salerno, S., Damon, L. E., Levis, M. J., et al. (2012). Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. *Nature*, 485(7397), 260-263.

[65] Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661-668.

[66] National Institutes of Health. (2007). Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS). *Federal Regist.*, 72(166), 49290-49297, http://www.grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html, accessed 3 July 2012.

[67] Chung, C. C., Magalhaes, W. C., Gonzalez-Bosquet, J., & Chanock, S. J. (2010). Genome-wide association studies in cancer--current and future directions. *Carcinogenesis*, 31(1), 111-120.

[68] Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., et al. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.*, 39(2), 207-211.

[69] Manolio, T. A. (2010). Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.*, 363(2), 166-176.

[70] Easton, D. F., & Eeles, R. A. (2008). Genome-wide association studies in cancer. *Hum. Mol. Genet.*, 17(R2), R109-R115.

[71] Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266(5182), 66-71.

[72] Adnane, J., Gaudray, P., Dionne, C. A., Crumley, G., Jaye, M., Schlessinger, J., et al. (1991). BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene*, 6(4), 659-663.

[73] Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature*, 446(7132), 153-158.

[74] Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., Richesson, D., et al. (2008). Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PloS Genet.*, 4(4), e1000054.

[75] Antoniou, A. C., Spurdle, A. B., Sinilnikova, O. M., Healey, S., Pooley, K. A., Schmutzler, R. K., et al. (2008). CIMBA. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am. J. Hum. Genet.*, 82(4), 937-948.

[76] Zheng, W., Long, J., Gao, Y. T., Li, C., Zheng, Y., Xiang, Y. B., et al. (2009). Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.*, 41(3), 324-328.

[77] Amundadottir, L. T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B. A., et al. (2006). A common variant associated with prostate cancer in European and African populations. *Nat. Genet.*, 38(6), 652-658.

[78] Yeager, M., Orr, N., Hayes, R. B., Jacobs, K. B., Kraft, P., Wacholder, S., et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.*, 39(5), 645-649.

[79] Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.*, 39(5), 631-637.

[80] Eeles, R. A., Kote-Jarai, Z., Giles, G. G., Olama, A. A., Guy, M., Jugurnauth, S. K., et al. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.*, 40(30), 316-321.

[81] Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.*, 40(3), 310-315.

[82] Chanock, S. J., & Hunter, D. J. (2008). Genomics: when the smoke clears. *Nature*, 452(7187), 537-538.

[83] Hung, R. J., Mc Kay, J. D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., et al. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*, 452(7187), 633-637.

[84] Mc Kay, J. D., Hung, R. J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G., et al. (2008). Lung cancer susceptibility locus at 5p15. *Nat. Genet.*, 40(12), 1404-1406.

[85] Wang, Y., Broderick, P., Webb, E., Wu, X., Vijayakrishnan, J., Matakidou, A., et al. (2008). Common 5p15and 6p21.33 variants influence lung cancer risk. *Nat. Genet.*, 40(12), 1407-1409.

[86] Bierut, L. J., Madden, P. A., Breslau, N., Johnson, E. O., Hatsukami, D., Pomerleau, O. F., et al. (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum. Mol. Genet.*, 16(1), 24-35.

[87] Caporaso, N., Gu, F., Chatterjee, N., Sheng-Chih, J., Yu, K., Yeager, M., et al. (2009). Genome-wide and candidate gene association study of cigarette smoking behaviors. *PloS One*, 4(2), e4653.

[88] Pearson, T. A., & Manolio, T. A. (2008). How to interpret a genome-wide association study. *JAMA*, 299(11), 1335-1344.

[89] Kim, H. C., Lee, J. Y., Sung, H., Choi, J. Y., Park, S. K., Lee, K. M., et al. (2012). A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study. *Breast Cancer Res.*, 14, R56.

[90] Long, J., Cai, Q., Sung, H., Shi, J., Zhang, B., Choi, J. Y., et al. (2012). Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet.*, 8(2), e1002532.

[91] Shu, X. O., Long, J., Lu, W., Li, C., Chen, W. Y., Delahanty, R., et al. (2012). Novel genetic markers of breast cancer survival identified by a genome-wide association study. *Cancer Res.*, 72(5), 1182-1189.

[92] Kiyotani, K., Mushiroda, T., Tsunoda, T., Morizono, T., Hosono, N., Kubo, M., et al. (2012). A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. *Hum. mol. genet.*, 21(7), 1665-1672.

[93] Cai, Q., Long, J., Lu, W., Qu, S., Wen, W., Kang, D., et al. (2011). Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. *Hum. Mol. Genet.*, 20(24), 4991-4999.

[94] Sehrawat, B., Sridharan, M., Ghosh, S., Robson, P., Cass, C. E., Mackey, J. R., et al. (2011). Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Hum. Genet.*, 130(4), 529-537.

[95] Fletcher, O., Johnson, N., Orr, N., Hosking, F. J., Gibson, L. J., Walker, K., et al. (2011). Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J. Natl. Cancer Inst.*, 103(5), 425-435.

[96] Li, J., Humphreys, K., Darabi, H., Rosin, G., Hannelius, U., Heikkinen, T., et al. (2010). A genome-wide association scan on estrogen receptor-negative breast cancer. *Breast Cancer Res.*, 12(6), R93.

[97] Gaudet, M. M. , Kirchhoff, T., Green, T., Vijai, J., Korn, J. M., Guiducci, C., et al. (2010). Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. *PloS Genet.*, 6(10), e1001183.

[98] Li, J., Humphreys, K., Heikkinen, T., Aittomäki, K., Blomqvist, C., Pharoah, P. D., et al. (2011). A combined analysis of genome-wide association studies in breast cancer. *Breast Cancer Res Treat.*, 126(3), 717-727.

[99] Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., et al. (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.*, 42(6), 504-507.

[100] Azzato, E. M., Pharoah, P. D., Harrington, P., Easton, D. F., Greenberg, D., Caporaso, N. E., et al. (2010). A genome-wide association study of prognosis in breast cancer. *Cancer Epidemiol. Biomarkers Prev.*, 19(4), 1140-1143.

[101] Thomas, G., Jacobs, K. B., Kraft, P., Yeager, M., Wacholder, S., Cox, D. G., et al. (2009). A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11and 14q24.1 (RAD51L1). *Nat. Genet.*, 41(5), 579-584.

[102] Zheng, W., Long, J., Gao, Y. T., Li, C., Zheng, Y., Xiang, Y. B., et al. (2009). Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.*, 41(3), 324-328.

[103] Kibriya, M. G., Jasmine, F., Argos, M., Andrulis, I. L., John, E. M., Chang-Claude, J., et al. (2009). A pilot genome-wide association study of early-onset breast cancer. *Breast Cancer Res. Treat.*, 114(3), 463-477.

[104] Gold, B., Kirchhoff, T., Stefanov, S., Lautenberger, J., Viale, A., Garber, J., et al. (2008). Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc. Natl. Acad. Sci.*, USA, 105(11), 4340-4345.

[105] Murabito, J. M., Rosenberg, C. L., Finger, D., Kreger, B. E., Levy, D., Splansky, G. L., et al. (2007). A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med. Genet.*, 8(Suppl 1), S6.

[106] Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447(7148), 1087-1093.

[107] Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., et al. (2007). A genome-wide association study identifies alleles in FGFR2

associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, 39(7), 870-874.

[108] Tao, S., Feng, J., Webster, T., Jin, G., Hsu, F. C., Chen, S. H., et al. (2012). Genome-wide two-locus epistasis scans in prostate cancer using two European populations. *Hum. Genet.*, 131, 1225-1234.

[109] Tao, S., Wang, Z., Feng, J., Hsu, F. C., Jin, G., Kim, S. T., et al. (2012). A genome-wide search for loci interacting with known prostate cancer risk-associated genetic variants. *Carcinogenesis*, 33(3), 598-603.

[110] Kote-Jarai, Z., Olama, A. A., Giles, G. G., Severi, G., Schleutker, J., Weischer, M., et al. (2011). UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators, The Australian Prostate Cancer BioResource; PRACTICAL Consortium. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat. Genet.*, 43(8), 785-791.

[111] Schumacher, F. R., Berndt, S. I., Siddiq, A., Jacobs, K. B., Wang, Z., Lindstrom, S., et al. (2011). Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum. Mol. Genet.*, 20(19), 3867-3875.

[112] Haiman, C. A., Chen, G. K., Blot, W. J., Strom, S. S., Berndt, S. I., Kittles, R. A., et al. (2011). Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat. Genet.*, 43(6), 570-573.

[113] Fitz, Gerald L. M., Kwon, E. M., Conomos, M. P., Kolb, S., Holt, S. K., Levine, D., et al. (2011). Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer. *Cancer Epidemiol. Biomarkers Prev.*, 20(6), 1196-1203.

[114] Penney, K. L., Pyne, S., Schumacher, F. R., Sinnott, J. A., Mucci, L. A., Kraft, P. L., et al. (2010). Genome-wide association study of prostate cancer mortality. *Cancer Epidemiol. Biomarkers Prev.*, 19(11), 2869-2876.

[115] Takata, R., Akamatsu, S., Kubo, M., Takahashi, A., Hosono, N., Kawaguchi, T., et al. (2010). Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat. Genet.*, 42(9), 751-754.

[116] Eeles, R. A., Kote-Jarai, Z., Al Olama, A. A., Giles, G. G., Guy, M., Severi, G., et al. (2009). UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators; PRACTICAL Consortium, Easton DF Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat. Genet.*, 41(10), 1116-1121.

[117] Gudmundsson, J., Sulem, P., Gudbjartsson, D. F., Blondal, T., Gylfason, A., Agnarsson, B. A., et al. (2009). Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet.*, 41(10), 1122-1126.

[118] Eeles, R. A., Kote-Jarai, Z., Giles, G. G., Olama, Guy. M., Jugurnauth, S. K., et al. (2008). UK Genetic Prostate Cancer Study Collaborators; British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.*, 40(3), 316-321.

[119] Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.*, 40(3), 310-315.

[120] Duggan, D., Zheng, S. L., Knowlton, M., Benitez, D., Dimitrov, L., Wiklund, F., et al. (2007). Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J. Natl. Cancer Inst.*, 99(24), 1836-1844.

[121] Murabito, J. M., Rosenberg, C. L., Finger, D., Kreger, B. E., Levy, D., Splansky, G. L., et al. (2007). A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med. Genet.*, 8(Suppl 1), S6.

[122] Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.*, 39(5), 631-637.

[123] Yeager, M., Orr, N., Hayes, R. B., Jacobs, K. B., Kraft, P., Wacholder, S., et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.*, 39(5), 645-649.

[124] Peters, U., Hutter, C. M., Hsu, L., Schumacher, F. R., Conti, D. V., Carlson, C. S., et al. (2012). Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum. Genet.*, 131(2), 217-234.

[125] Houlston, R. S., Cheadle, J., Dobbins, S. E., Tenesa, A., Jones, A. M., Howarth, K., et al. (2010). Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.*, 42(11), 973-977.

[126] Lascorz, J., Försti, A., Chen, B., Buch, S., Steinke, V., Rahner, N., et al. (2010). Genome-wide association study for colorectal cancer identifies risk polymorphisms in German familial cases and implicates MAPK signalling pathways in disease susceptibility. *Carcinogenesis*, 31(9), 1612-1619.

[127] Houlston, R. S., Webb, E., Broderick, P., Pittman, A. M., Di Bernardo, M. C., Lubbe, S., et al. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Paper presented at International Colorectal Cancer Genetic Association Consortium. *Nat. Genet.*, 40(12), 1426-1435.

[128] Tenesa, A., Farrington, S. M., Prendergast, J. G., Porteous, M. E., Walker, M., Haq, N., et al. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.*, 40(5), 631-637.

[129] Tomlinson, I. P., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A. M., et al. (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14and 8q23.3. *Nat. Genet.*, 40(5), 623, 630.

[130] Broderick, P., Carvajal-Carmona, L., Pittman, A. M., Webb, E., Howarth, K., Rowan, A., et al. (2007). CORGI Consortium. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.*, 39(11), 1315-1317.

[131] Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., et al. (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.2. *Nat. Genet.*, 39(8), 984-988.

[132] Zanke, B. W., Greenwood, C. M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S. M., et al. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.*, 39(8), 989-994.

[133] Hu, Z., Wu, C., Shi, Y., Guo, H., Zhao, X., Yin, Z., et al. (2011). A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.*, 43(8), 792-796.

[134] Wu, X., Ye, Y., Rosell, R., Amos, C. I., Stewart, D. J., Hildebrandt, et., & al, . (2011). Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy. *J. Natl. Cancer Inst.*, 103(10), 817-825.

[135] Frullanti, E., Galvan, A., Falvella, F. S., Manenti, G., Colombo, F., Vannelli, A., et al. (2011). Multiple genetic loci modulate lung adenocarcinoma clinical staging. *Clin. Cancer Res.*, 17(8), 2410-2416.

[136] Wu, C., Xu, B., Yuan, P., Miao, X., Liu, Y., Guan, Y., et al. (2010). Genome-wide interrogation identifies YAP1 variants associated with survival of small-cell lung cancer patients. *Cancer Res.*, 70(23), 9721-9729.

[137] Yoon, K. A., Park, J. H., Han, J., Park, S., Lee, G. K., Han, J. Y., et al. (2010). A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population. *Hum. Mol. Genet.*, 19(24), 4948-4954.

[138] Li, Y., Sheu, C. C., Ye, Y., de Andrade, M., Wang, L., Chang, S. C., et al. (2010). Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol.*, 11(4), 321-330.

[139] Landi, M. T., Chatterjee, N., Yu, K., Goldin, L. R., Goldstein, A. M., Rotunno, M., et al. (2009). A genome-wide association study of lung cancer identifies a region of chromosome 5p15associated with risk for adenocarcinoma. *Am. J. Hum. Genet.*, 85(5), 679-691.

[140] Broderick, P., Wang, Y., Vijayakrishnan, J., Matakidou, A., Spitz, M. R., Eisen, T., et al. (2009). Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res.*, 69(19), 6633-6641.

[141] Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T., et al. (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.*, 40(5), 616-622.

[142] Thorgeirsson, T. E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K. P., et al. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*, 452(7187), 638-642.

[143] Hung, R. J., Mc Kay, J. D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., et al. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*, 452(7187), 633-637.

[144] Spinola, M., Leoni, V. P., Galvan, A., Korsching, E., Conti, B., Pastorino, U., et al. (2007). Genome-wide single nucleotide polymorphism analysis of lung cancer risk detects the KLF6 gene. *Cancer Lett.*, 251(2), 311-316.

[145] Barrett, J. H., Iles, M. M., Harland, M., Taylor, J. C., Aitken, J. F., Andresen, P. A., et al. (2011). GenoMEL Consortium. Genome-wide association study identifies three new melanoma susceptibility loci. *Nat. Genet.*, 43(11), 1108-1113.

[146] Mac Gregor, S., Montgomery, G. W., Liu, J. Z., Zhao, Z. Z., Henders, A. K., Stark, M., et al. (2011). Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. *Nat. Genet.*, 43(11), 1114-1118.

[147] Amos, C. I., Wang, L. E., Lee, J. E., Gershenwald, J. E., Chen, W. V., Fang, S., et al. (2011). GenoMEL Investigators. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum. Mol. Genet.*, 20(24), 5012-5023.

[148] Teerlink, C., Farnham, J., Allen-Brady, K., Camp, N. J., Thomas, A., Leachman, S., et al. (2012). A unique genome-wide association analysis in extended Utah high-risk pedigrees identifies a novel melanoma risk variant on chromosome arm 10q. *Hum. Genet.*, 131(1), 77-85.

[149] Bishop, D. T., Demenais, F., Iles, M. M., Harland, M., Taylor, J. C., Corda, E., et al. (2009). Genome-wide association study identifies three loci associated with melanoma risk. *Nat. Genet.*, 41(8), 920-925.

[150] King, H., Aubert, R. E., & Herman, W. H. (1998). Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. *Diabetes Care*, 21(9), 1414-1431.

[151] Reeves, G. K., Travis, R. C., Green, J., Bull, D., Tipper, S., Baker, K., et al. (2010). Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *JAMA*, 304(4), 426-434.

[152] Kaprio, J., Tuomilehto, J., Koskenvuo, M., Romanov, K., Reunanen, A., Eriksson, J., et al. (1992). Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia*, 35(11), 1060-1067.

[153] Risch, N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.*, 46(2), 222-228.

[154] Grant, S. F., Thorleifsson, G., Reynisdottir, I., Manolescu, A., Sainz, J., Helgason, A., et al. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.*, 38(3), 320-323.

[155] Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445(7130), 881-885.

[156] Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H., et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316(5829), 1331-1336.

[157] Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, 316(5829), 1341-1345.

[158] Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., et al. (2007). Wellcome Trust Case Control Consortium (WTCCC). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316(5829), 1336-1341.

[159] Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., et al. (2008). Meta-analysis of genome-wide association data and large-scale

replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.*, 40(5), 638-645.

[160] van Dam, R. M., Hoebee, B., Seidell, J. C., Schaap de, Bruin. T. W., & Feskens, E. J. (2005). Common variants in the ATP-sensitive K+ channel genes KCNJ11 (Kir6.2) and ABCC8 (SUR1) in relation to glucose intolerance: population-based studies and meta-analyses. *Diabet. Med.*, 22(5), 590-598.

[161] Ludovico, O., Pellegrini, F., Di Paola, R., Minenna, A., Mastroianno, S., Cardellini, M., et al. (2007). Heterogeneous effect of peroxisome proliferator-activated receptor gamma2 Ala12 variant on type 2 diabetes risk. *Obesity (Silver Spring)*, 15(5), 1076-1081.

[162] Cauchi, S., El Achhab, Y., Choquet, H., Dina, C., Krempler, F., Weitgasser, R., et al. (2007). TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *J. Mol. Med.*, 85(7), 777-782.

[163] Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G. B., et al. (2007). A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.*, 39(6), 770-775.

[164] Sandhu, M. S., Weedon, M. N., Fawcett, K. A., Wasson, J., Debenham, S. L., Daly, A., et al. (2007). Common variants in WFS1 confer risk of type 2 diabetes. *Nat. Genet.*, 39(8), 951-953.

[165] Meigs, J. B., Shrader, P., Sullivan, L. M., Mc Ateer, J. B., Fox, C. S., Dupuis, J., et al. (2008). Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N. Engl. J. Med.*, 359(21), 2208-2219.

[166] Weedon, M. N., Mc Carthy, M. I., Hitman, G., Walker, M., Groves, C. J., Zeggini, E., et al. (2006). Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PloS. Med.*, 3(10), e374.

[167] Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.*, 17(10), 1520-1528.

[168] Morrison, A. C., Bare, L. A., Chambless, L. E., Ellis, S. G., Malloy, M., Kane, J. P., et al. (2007). Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am. J. Epidemiol.*, 166(1), 28-35.

[169] Cornelis, M. C., Qi, L., Zhang, C., Kraft, P., Manson, J., Cai, T., et al. (2009). Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann. Intern. Med.*, 150(8), 541-550.

[170] Ripatti, S., Tikkanen, E., Orho-Melander, M., Havulinna, Silander. K., Sharma, A., et al. (2010). A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet*, 376(9750), 1393-1400.

[171] Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.*, 7(10), 781-791.

[172] Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, Kim. S., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391), 603-607.

[173] Garnett, N. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391), 570-575.