MINI-REVIEW

# Omics of Cancer

**Aniruddha Bhati, H Garg, A Gupta, H Chhabra, A Kumari, T Patel***

## Abstract

**With the advances in modern diagnostic expertise for cancer, certain approaches allowing scanning of the complete genome and the proteome are becoming very useful for researchers. These high throughput techniques have already proven power, over traditional detection methods, in differentiating disease sub-types and identifying specific genetic events during progression of cancer. This paper introduces major branches of omics-technology and their applications in the field of cancer. It also addresses current road blocks that need to be overcome and future possibilities of these methods in oncogenic detection.**

**Keywords:** Genomics - proteomics - epigenomics - transcriptomics

*Asian Pacific J Cancer Prev,* **13 (9)**, 4229-4233

## The Term '-omics'

In the new era of high end technology and ever growing field of research the word "omics" has evolved as a cynosure of many scientists across the world. It refers to "totality in terms of bio systems" bifurcated mainly as Genome and Proteome. In other words ''omics'' involves the measurement of large number of parameters, typically genes (genomics), proteins (proteomics), lipids (lipidomics) or metabolites (metabolomics) so on and so forth (Lay et al., 2006). Omics has enabled us to increase our ability to handle and measure large data sets. Omics techniques like mass spectrometry and arrays with the help of their specificity and selectivity has helped us in identifying, measuring and quantifying DNA, messenger RNAs and proteins derived from body (Morel et al., 2004).

## Genomics

In the modern epoch the genome represents a complete complement of genes in an organism's hereditary material. The term "genome" was adapted by Hans (1920), Professor of Botany at the University of Hamburg, Germany. In Greek, the word genome stands for "I become; I am born, to come into being". The Oxford English Dictionary suggests the name to be a blend of words gene and chromosome (Lederberg and McCray, 2001). Figure 1 puts through a newer assortment of the Gen-Omics

The Human Genome Project was initialized with the intent to map and decode the complete human DNA. Fiers et al. (1976) at the University of Ghent (Belgium) pioneered to establish the complete nucleotide sequence of viral RNA-genome bacteriophage MS2. Sanger et al. (1977) sequenced the first genome of phage Φ-X174 which was about the size of 5386 nucleotide base pairs. The inaugural bacterial genome to be sequenced was that of *Haemophilus influenzae* by the team at The Institute for Genomic Research (Rockville, Maryland, United States.) in 1995. Following that, the first eukaryotic genome completed was that of the budding yeast *Saccharomyces cerevisiae*. In early 1970s the DNA sequences obtained by the research groups used methods based on two-dimensional chromatography. Subsequent to the development of dye-based sequencing methods automated sequencers were invented (Olsvik et al., 1993)

Recent development of novel technologies has greatly decreased the cost and the complexity of sequencing and as a result of this the number of completed genome sequences is raising rapidly. As a consequence of high throughput technologies, the data being released is fairly huge and to cope up with this, a large number of databases are being constructed at high rate. Some prokaryotic and eukaryotic genome resources that are in progress or completed can be accessed through the internet (www.ncbi.nlm.nih.gov/genome). The DNA sequencing has become comparatively easier and orders of magnitude faster with invent of next generation sequencing technologies by different

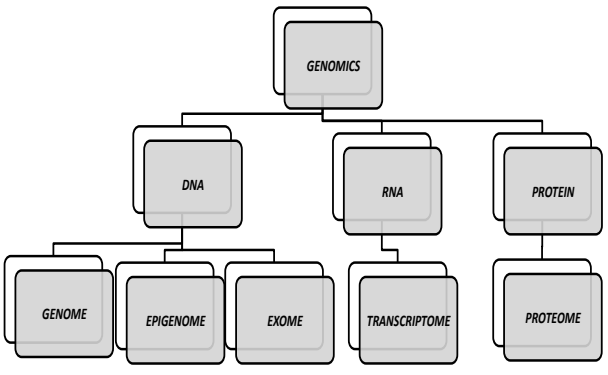


**Figure 1. Assortment of Gen-Omics**

*Division of Medical Biotechnology, School of BioSciences and Technology, VIT, Vellore, India *For correspondence: Dr.TNPatel@gmail.com*

companies worldwide (Pettersson, 2009; Schuster, 2008). A few of these sequencing assemblies are - 1) GS-FLX (Roche), 2) SOLiD™ (Applied Biosystems), 3) Ion Torrent™ (Applied Biosystems), 4) Illumina Solexa Sequencer

*Genomics of Cancer*

One of the most significant achievements in scientific history is the completion of human genome draft sequence (Lander et al., 2001; Venter et al., 2001). Analysis of these elucidated sequences represents a benchmark, transforming our knowledge of fundamental processes which govern human biology and pathology. Despite an enormous progress made in the field of genomics, the sequence based structural analysis of tumour genomes has been a very laborious and complex process until recent times (Raphael et al., 2003; Volik et al., 2003).

International cancer genome consortium is the teamed effort of various international laboratories which coordinates large-scale genome studies in 500 tumours from each of 50 different cancer types and subtypes in both adults and children, adding up to a total of 25,000 cancer genomes. This effort is expected to reveal the major repertoire of oncogenic mutations, thus allowing the definition of clinically relevant subtypes and the development of newer cancer diagnostics and therapeutics (Hudson et al., 2010). Massive parallel sequencing is a single most common method that can reveal important information related to genes and their functional changes due to variations in the genome of cancer (Bentley, 2006). GWAS (Genome-wide association studies) has been performed in nearly all common malignancies and scientist have identified more than 100 common genetic risk variants that confer a modest increased risk to cancer (Stadler et al., 2010). The accumulation of such information from the various cancers can help in revolutionizing the basic and clinical cancer research (Bardelli et al., 2003). Among the multiple strategies defined for the systematic identification of mutational profile of tumour genomes, one of them (Benvenuti et al., 2005) is represented in Figure 2.

Complete genome studies of various cancers have given striking mutational statistics of these cancers.

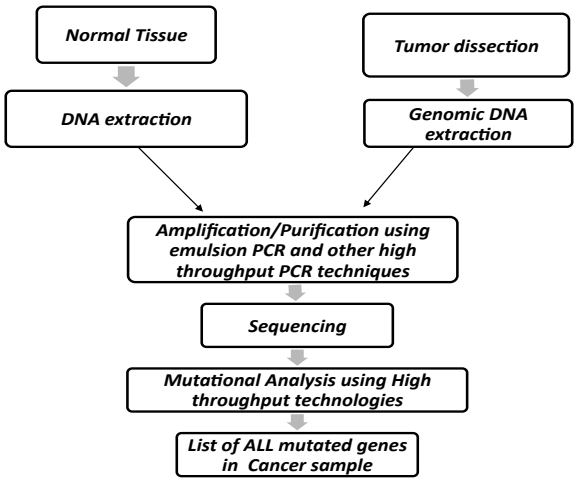


**Figure 2. Strategy for Systematic Identification of Mutational Profile of Tumor Genome.**

Sjoblom et al. (2006) studied the breast cancer genome and revealed that out of 18191 genes analysed, around 2380 genes showed different mutations. Weir et al. (2007) carried out a similar experiment in colorectral cancer and showed a similar statistics of non synonymous mutations. Genomic studies on acute myeloid leukemia (Lay et al., 2006) filtered 10 genes that were frequently mutated and nearly 500-1,000 having non silent mutation. A large number of cancer genomes were studied using high throughput techniques (Stephens et al., 2007; Ding et al., 2008; Jones et al., 2008) and all of these studies have given useful statistics in terms of mutations that occur in various cancer and certain signatures that can be common to all of the cancers.

Recent major achievement that we perceive in the cancer research is sequencing of complete cancer genomes using high throughput techniques and identification of complete sets of mutation in a given subtype of cancer. Lately sequencing of an entire breast cancer tumour has been reported (Shah et al., 2009).The International Cancer Genome Consortium (ICGC) announced launch of a $20 million Canadian research project on February 15, 2011. This project is intended to map the genetic structure of prostate cancer and provide critical information that could greatly improve the diagnosis and treatment.

## Epigenomics

Epigenetics study aims to understand the heritable gene regulations that are not directly encoded in the DNA sequence. These changes may remain through cell divisions for the remaining of the cell's life or may also last for multiple generations (Bird, 2007). Non-genetic factors cause the organism's genes to express themselves differently leading to significant changes in genome which are manifested at the phenotypic level. Epigenetics is also defined as "the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms (Holliday, 1990).

*Epigenomics in Cancer*

An 'epigenome' is a representation of all epigenetic phenomena which mainly include DNA methylation and histone modification across the genome. The study of the total methylation of the cell is called as methylome (Bernstein et al., 2007). (Out of the two epigenetic changes i.e. histone modification and DNA methylation, the latter one is the most common and extensively studied marker) There are compelling evidences to support the importance of DNA methylation alterations in cancer development much earlier to the manifestation of the disease. Both losses and gains of gene expression have been observed due to methylation contributing to pathophysiological changes in the tumour. The resultant effect is inactivation of tumour suppressor genes and ectopical expression of certain genes inducing chromosomal instability (Estécio and Issa, 2011). The DNA methylation pattern in normal and cancerous cells is diagrammatically represented in Figure 3.

New technologies for genome wide DNAm (methylome) analysis, such as MeDIP-seq, have been

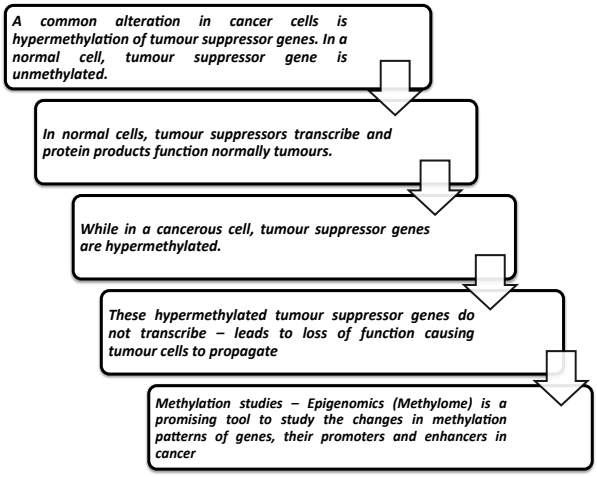


**Figure 3. DNA Methylation Patterns in Normal and Cancerous Cells.**

developed; enabling the unbiased analysis of cancer methylome. Using MeDIP-seq, sequencing-based comparative methylome analysis of malignant peripheral nerve sheath tumours (MPNST), benign neurofibromas, and normal Schwann cells was done by Feber (2010). Two hundred candidate genes that cluster throughout the genome of which 25 were previously reported as harbouring cancer-specific promoter methylation were studied on 20 cancer cell lines of about 5 major cancer types using bisulfite sequencing or methylation- specific PCR (MSP). It was found that cancer-specific methylation of at least one gene with high frequency in all cancer types. Identification of a large number of genes with cancer-specific methylation leads to new targets for diagnostic and therapeutic intervention and thus opens fertile avenues for therapeutics in tumor biology (Hoque et al., 2008).

## Exomics

Exome is a part of the genome formed by all the exon -the coding portions of genes in the genome. It provides the genetic blueprint used in the synthesis of proteins and other functional gene products. Exome is functionally most relevant part of the genome with respect to coding for proteins and thus it is most likely to govern the phenotype of an organism. Sequencing of the 'exome' (all the exons in the genome) involves chopping the genome into millions of pieces and capturing and sequencing only selected exomic regions. It differs from transcriptome sequencing by focusing on DNA rather than the expressed RNA in a given cell and it promises to be vastly cheaper than whole-genome sequencing. The complete exome of human genome is estimated to comprise about 1.5 % of the total genome (Ng et al., 2008). Exome capture allows a comprehensive analysis of the complete protein-coding regions in the genome. Researchers can use exome capture to focus on critical part of the human genome. This method in fact gives way to larger numbers of samples studies with low economical inputs and more defined data output (Estécio and Issa, 2011). Next generation high throughput sequencing was utilized for sequencing of 12 human exomes demonstrating the sensitive and specific identification of rare and common variants in over 300 mega bases of coding sequence (Sarah et al., 2009).

*Exomics in Cancer*

It has been a significant focus of International Cancer Genome Consortium to carry out "exome sequencing" for multiple different cancer types. The most recent studies published in April 2011, in Nature Genetics identified GRIN2A as frequently mutated in melanoma using Exome sequencing (Wei et al., 2011). Similar studies by another group has shown frequent mutations of the SWI/SNF complex gene *PBRM1* in renal carcinoma (Varela et al., 2011). Future studies on exome sequencing integrated with methylome studies could give an idea about early changes in the methylation patterns of the exons in a genome and could help elucidate newer diagnostic and therapeutic targets. The authors are presently working towards the similar aim.

## Proteomics

Word "proteome" is a combination of "**prote**in" and "gen**ome**". Proteome is a complete complement of proteins, including the changes occurring in a particular set of proteins, which deviate with time and distinct requirements, or stresses, that a cell or organism undergoes. The idea of observing protein expression of genomes in a holistic manner rather than one protein at a time arose with the advent of 2-dimensional gel electrophoresis (2-DE). The term "proteome" was coined by Marc Wilkins in the year 1994 (Wilkins, 1994). Proteomics is generally divided in three main areas: 1) Micro-characterization of proteins for large-scale identification, 2) Comparison of protein expression and abundance at two different time period in a same set of cells, 3) Studies of protein-protein interactions and post-translational changes (Murray and Linardou, 2004).

A comprehensive proteomic profile of a cell can be obtained by analyzing various sub cellular compartments like, membrane, cytoplasmic and nuclear proteins and in cell culture - the secretome obtained from cell-conditioned medium after cells and cell debris have been removed. For proteome study sub cellular protein fractions are extracted using various techniques or commercially available kits, subjected to two dimensional gel electrophoresis followed by mass spectrometry coupled with chromatography. Subsequently the proteins are identified using appropriate tools and databases which can yield potential biomarkers (Hanash et al., 2008) for further use. Proteome sequencing is currently also used to study differential splicing, post-translational modifications and data integration.

*Proteomics in Cancer*

The egression of novel technologies in field of proteomics have exerted dramatic elevation in pace of cancer research. It has helped us understand the use of proteomics for detection of cancer biomarkers that are normally not recognizable by traditional techniques, identification of specific genetic events that occurs during progression of cancer, detection and differentiation of malignant and benign tumor cells. Recent study using in depth quantitative protein profiling revealed protein

signature for Titf1/Nkx2-1, a known lineage-survival oncogene in lung cancer, in plasma of mouse models of lung adenocarcinoma. Different EGFR (Epidermal growth factor receptor) signatures were also revealed in the same study of mouse model which were in concordance with human lung cancer (Taguchi et al., 2011).

A study was conducted using CA 125 (Cancer Antigen 125) immunoassay and surface-enhanced laser desorption and ionization time-of-flight mass spectrometry (SELDI-TOF) to measure 7 proteins (apolipoprotein A1, truncated transthyretin, transferrin, hepcidin, β-2 microglobulin, connective tissue activating protein III and interalpha-trypsin inhibitor heavy-chain) known to be expressed in early stages of ovarian cancer in sera samples (Moore et al., 2012). This study showed that proteomics has been an important tool to identify serum biomarkers that are used to differentiate early stage ovarian cancer from otherwise healthy individuals. Proteomics has evolved as one of the promising tool for detecting early stage malignancy in Prostate, Lung, Colorectal, and Ovarian (PLCO) by CA 125 and TVS (Trans-vaginal sonography) screening. A similar screening of CA125 and 7 other biomarkers panel has help to differentiate malignant and benign pelvic masses (Moore et al., 2012).

## Transcriptomics

Transcriptome is a complete set of all the RNA species, including mRNA, rRNA, tRNA, and other non-coding RNA generated by the genome at any given time in a cell/group of cells. Unlike the genome, the transcriptome is extremely dynamic and has varied patterns and abundance of expression. Genome on the other hand is similar in all the cells regardless of cell type, environmental condition or stage of development. Transcriptomics is considered more robust, high-throughput, cost-effective technology which is capable of simultaneously quantifying tens of thousands of defined mRNA species in a miniaturized automated formats and comparatively more manageable than proteomics. The transcriptomics study uses two main methods for detection of RNA expression: 1) High Density DNA microarrays- that includes used of oligonucleotide probes on a gene chip and hybridizing m-RNA strands onto it followed by staining and scanning. Different computational tools like Expressionist R (Genedata solutions) are used for further analysis of principle component, hierarchical clustering and data filtering. Real time PCR is one of the methods used for confirmation and validation of such data. It is more global analysis, 2) Low Density DNA microarrays use colorimetric detection, Quantitative real time-PCR (qRT-PCR), Molecular beacons for gene expression studies. It is more focused analysis.

A completely different transcriptomics approach has lately been used is coupling conventional method of SAGE (serial analysis of gene expression) with highly parallel multiplex new-generation sequencing technology from Solexa (Illumina) and 454 (Roche). Microarrays quantify gene expression levels indirectly using fluorescence intensity whereas SAGE determines absolute mRNA levels without using pre existing clone knowledge. Massive parallel sequencing, a next generation sequencing (NGS) encompasses several high-throughput approaches to DNA sequencing utilizing miniaturized and parallelized platforms for sequencing of 1–100 million of short reads (50-400 bases). The various platforms used include Roche 454, Life Technology Solid 4, HiSeq 2000a and many more. These techniques were used for surveying the polyA+ transcripts from four stages of the nematode *Caenorhabditis elegans* along with novel statistical analysis that evaluated the coverage of annotated features of the genome and of candidate processed transcripts, including splice junctions, trans-spliced leader sequences, and polyadenylation tracts (La Deana et al., 2009).

*Transcriptomics in Cancer*

Tripathi and co-researchers carried out a complete transcriptomic analysis in horn cancer of Indian Jebu Cattle using Roche 454 (Tripathi et al., 2011). Scientists employed next generation sequencing platform to sequence *Bos indicus* cancerous and normal horn tissue transcripts. This study showed that the transcript abundance increased and decreased for certain genes - aberrant gene expression, presence of non coding RNA (nc-RNA), pseudogenes and unusual pathways. It also demonstrated that transcriptomics study of HC (high cytoplasmic staining group) and HN (high cytoplasmic, low nuclear staining group) tissues could result in discovery of various cancer transcripts which are previously uncharacterized.

Another study by research team of China, focused on 4 kind of tumours - nasopharyngeal carcinoma (NPC), breast cancer, colorectal cancer, and glioma to carry out expression and regulation of tumor transcriptomics; identification of tumor suppressor/susceptible genes; mechanism of tumor epigenetics including miRNAs and comparative study of specific gene/protein cluster of tumor transcriptomics and proteomics. The study revealed a common list of transcriptional genes viz., SPLUNC1, LTF, BRD7, NOR1, BRCA1/2, PALB2, AF1Q, SOX17, NGX6, SOX7, and LRRC4 during initial and invasive stages of these four tumours (Li et al., 2011).

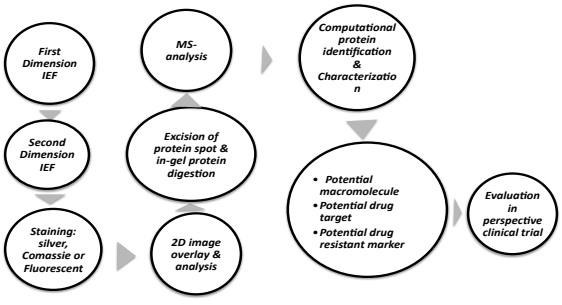


**Figure 4. Age-specific Colorectal Cancer Incidence Rates per 100 000**

## Conclusion

Laboratory based high throughput omics studies are giving promising result and translational application of these could be a boon to cancer patients helping in early diagnosis and appropriate therapeutics. However, the cost per sample study for commercial use is still not in the

economy range. Nonetheless, Next Generation Sequencing individually and coupled with other techniques have been greatly useful in defining the signatures in specific cancers and discovering certain potential biomarkers which could be used for early cancer diagnosis and target based therapies.

## References

Bardelli A, Parsons DW, Silliman N, et al (2003). Mutational analysis of the tyrosine kinome in colorectal cancers. *Science*, **300**, 949.

Bentley DR (2006). Whole-genome re-sequencing. *Curr Opin Genet Dev*, **16**, 545-52.

Benvenuti S, Arena S, Bardelli A (2005). Identification of cancer genes by mutational profiling of tumor genomes. *FEBS Lett*, **579**, 1884-90.

Bernstein BE, Meissner A, Lander ES (2007). The mammalian epigenome. *Cell*, **2**, 669-81.

Bird A (2007). Perceptions of epigenetics. *Nature*, **447**, 396-8.

Ding L, Getz G, Wheeler DA, et al (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, **455**, 1069-75.

Estécio MR, Issa JP (2011). Dissecting DNA hypermethylation in cancer. *FEBS Letters*, **585**, 2078-86.

Feber A (2010). Global profiling of methylation in the cancer genome. *Cancer Genet Cytogenet*, **203**, 44-65.

Fiers W, Contreras R, Duerinck F (1976). Complete nucleotide-sequence of bacteriophage MS2-RNA - primary and secondary structure of replicase gene. *Nature*, **260**, 500-7.

Hanash SM, Pitteri SJ, Faca VM (2008). Mining the plasma proteome for cancer biomarkers. *Nature*, **452**, 571-9.

Holliday R (1990). Mechanisms for the control of gene activity during development. *Biol Rev Cambr Philos Soc*, **65**, 431-71.

Hoque MO, Kim MS, Ostrow KL, et al (2008). Genome-wide promoter analysis uncovers portions of the cancer methylome. *Cancer Res*, **68**, 2661-70.

http://www.ncbi.nlm.nih.gov/genome.

Hudson TJ, Anderson W, Artez A, et al (2010). International Cancer Genome Consortium. International network of cancer genome projects. *Nature*, **464**, 993-8.

Jones S, Hruban RH, Kamiyama M, et al (2009). Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science*, **324**, 217.

La Deana WH, Valerie R, Philip G, et al (2009). Massively parallel sequencing of the polyadenylated transcriptome of C. elegans. *Genome Res*, **19**, 657-66.

Lander ES, Linton LM, Birren B, et al (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

Lay JO, Borgmann S, Liyanage R, Wilkins C L (2006). Problems with the ''omics''. *Trends in Analytical Chemistry*, **25**, 11.

Lederberg J, McCray AT (2001). 'Ome Sweet 'Omics -- A genealogical treasury of words. *The Scientist*, **15**, 8.

Li X, Shen S, Wu M, et al (2011). Transcriptomic regulation and molecular mechanism of polygenic tumor at different stages. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*, **36**, 585-91.

Moore LE, Pfeiffer RM, Zhang Z, et al (2012). Proteomic biomarkers in combination with CA125 for detection of epithelial ovarian cancer using pre-diagnostic serum samples from the prostate lung colon and ovary (PLCO) cancer screening trial. *Cancer*, **118**, 91-100.

Morel NM, Holland JM, van der Greef J, et al (2004). Primer on medical genomics part XIV: Introduction to systems biology-A new approach to understanding disease and treatment. *Mayo Clin Proc*, **79**, 651-8.

Murray S, Linardou H (2004). Proteomics and Cancer. STEP ΚΛΙΝΙΚΗΣ ΟΓΚΟΛΟΓΙΑΣ, **3**, 49-56.

Ng PC, Levy S, Huang J, et al (2008). Genetic variation in an individual human exome. *PLoS Genetics*, **4**, 1000160.

Olsvik O, Wahlberg J, Petterson B, et al (1993).- Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in Vibrio cholerae O1 strains. *J Clin Microbiol*, **31**, 22-5.

Pettersson E, Lundeberg J, Ahmadian A (2009). Generations of sequencing technologies. *Genomics*, **93**, 105-11.

Raphael BJ, Volik S, Collins C, Pevzner PA (2003). Reconstructing tumor genome architectures. *Bioinformatics*, **19**, 162-71.

Sanger F, Air GM, Barrell BG, et al (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, **265**, 687-95.

Sarah BN, Emily HT, Peggy DR, et al (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, **461**, 272-76.

Schuster SC (2008). Next-generation sequencing transforms today's biology. *Nat Methods*, **5**, 16-8.

Shah SP, Morin RD, Khattra J, et al (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, **461**, 809-13.

Sjöblom T, Jones S, Wood LD, et al (2006). The consensus coding sequences of human breast and colorectal cancers. *Science*, **314**, 268-74.

Stadler ZK, Vijai J, Thom P, et al (2010). Genome-wide association studies of cancer predisposition. *Hematol Oncol Clin North Am*, **24**, 973-96.

Stephens PJ, McBride DJ, Lin ML, et al (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature*, **462**, 1005-10.

Taguchi A, Politi K, Pitteri, et al (2011). Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. *Cancer Cell*, **20**, 289-99.

Tripathi AK, Koringa PG, Jakhesara SJ, et al (2012). A preliminary sketch of horn cancer transcriptome in Indian zebu cattle. *Gene*, **1**, 124-31.

Varela I, Tarpey P, Raine K, et al (2011). Exome sequencing identifies frequentmutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature*, **469**, 539-42.

Venter CJ, Adams DM, Myers WE, et al (2001). The sequence of the human genome. *Science*, **291**, 1304-51.

Volik S, Zhao S, Chin K, et al (2003). End-sequence profiling: Sequence-based analysis of aberrant genomes. *Proc Natl Acad Sci*, **100**, 7696-701.

Wei X, Walia V, Lin JC, et al (2011). Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature Genetics*, **43**, 442-6.

Weir BA, Michele S. Woo MS, Getz G, et al (2007). Characterizing the cancer genome in lung adenocarcinoma. *Nature*, **450**, 893-8.

Wilkins MR, Appel R, Jennifer EVE, et al (1994), Guidelines for next 10 years of proteins. *Proteomics*, **6**, 4-8.