



# New tools for functional genomic analysis

Xin Chen<sup>1</sup>, Eric Jorgenson<sup>1,2,3</sup> and Siu Tim Cheung<sup>4</sup>

<sup>1</sup> Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, United States

<sup>2</sup> Department of Neurology, University of California, San Francisco, CA, United States

<sup>3</sup> Gallo Clinic and Research Center, Emeryville, CA, United States

<sup>4</sup> Department of Surgery, The University of Hong Kong, Hong Kong, China

For the past decade, the development of genomic technology has revolutionized modern biological research and drug discovery. Functional genomic analyses enable biologists to perform analysis of genetic events on a global scale and they have been widely used in gene discovery, biomarker determination, disease classification, and drug target identification. In this article, we provide an overview of the current and emerging tools involved in genomic studies, including expression arrays, microRNA arrays, array CGH, ChIP-on-chip, methylation arrays, mutation analysis, genome-wide association studies, proteomic analysis, integrated functional genomic analysis and related bioinformatic and biostatistical analyses. Using human liver cancer as an example, we provide further information of how these genomic approaches can be applied in cancer research.

Genomic analyses include a variety of tools that address the global changes of specific biological parameters. Genomic analyses that examine DNA, RNA, or protein levels provide powerful tools to characterize gene function and regulation, facilitate disease classification, biomarker identification, risk factor stratification and drug discovery (Figure 1).

## Global gene expression profile

Genome-wide expression studies empowered by microarray analysis enable the systematic analysis of complex biological systems and because expression microarrays were developed in the 1990s they represent the beginning of the genomic era. The principle of microarray expression profiling study is based on the hybridization of a single strand nucleic acid fragment to its complementary single strand with high specificity [1]. The target cDNA is first labeled using fluorescent dyes and then hybridized to the array surface. The retained labeled target will then be subjected to

stringent washes, capture and quantification of the fluorescent signals, followed by the data analysis process. Two platforms that are commonly in use are cDNA microarrays and oligonucleotide microarrays [1]. The cDNA microarrays contain a collection of probes generated by PCR amplification from cDNA libraries, expressed sequence tag clones or long genome cloned fragments. Oligonucleotide microarrays commonly contain short oligonucleotides (25–30 nt) or long oligonucleotides (50–80 nt). In both platforms, the short DNA segments are printed onto solid supports, usually microscopic slides, by direct contact mode (mechanical robotic spotting) or non-contact mode (ink jet technique). Oligonucleotide microarrays manufactured by Affymetrix, the GeneChip, employ a different technology platform, designated the photolithographic method, to synthesize the short oligonucleotides *in situ* on the chip. Different platforms each have their limitations and advantages, but in common, all have been shown reliably to capture gene expression signature on a genomic scale [1].

Global expression profiles enable a better understanding of the molecular signature of human diseases, including liver cancers [2,3] (Figure 2). For instance, we and others have reported genome-wide expression profiles of liver cancer and their clinico-pathological implications [4–7]. We have observed specific

☆ This article provides an overview of the current and emerging tools involved in genomic studies and analyses.

Corresponding author: Cheung, Chen, X. (xin.chen@ucsf.edu), S.T. (stcheung@hkucc.hku.hk)

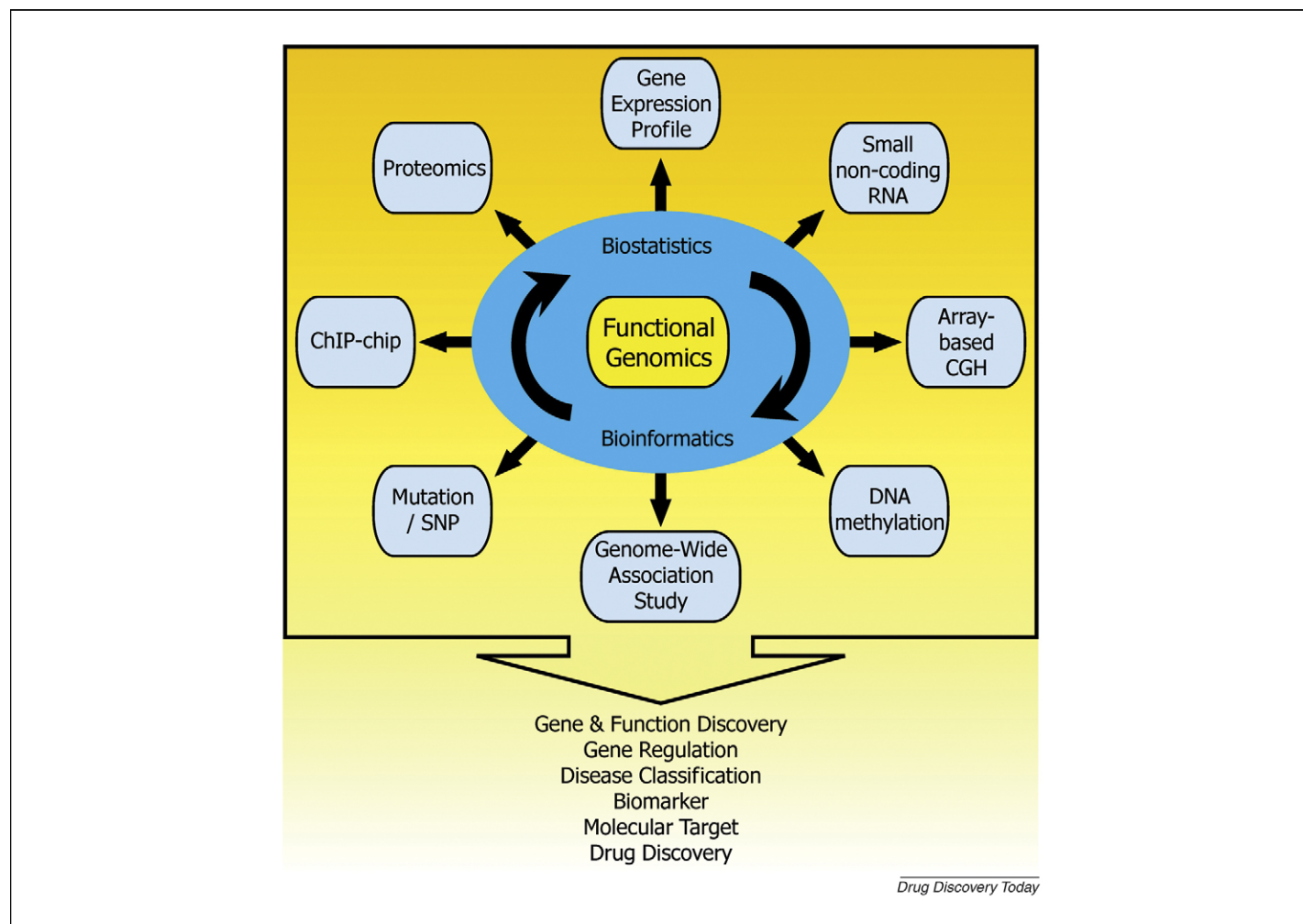


FIGURE 1

Schematic illustration of new tools for functional genomic analysis.

gene expression patterns between tumor and non-tumor [4], in association with p53 status [4], and clonality delineation for multinodular tumor [8]. There are reports on gene expression profiling in association with tumor metastasis [6], and patient outcome [5,7]. Differentially expressed genes demonstrated the potential to serve as prognostic biomarkers [9] and therapeutic targets [10].

### Genomic analysis small non-coding RNAs

An emerging number of studies have shown the regulation of gene expression by small non-coding RNAs in animals and plants [11,12]. MicroRNAs (miRNA) are an abundant class of small non-protein-coding RNAs that are 19–25 nucleotides in length that have been implicated in many human diseases including tumor initiation and progression. miRNA function as negative gene regulators, which can control hundreds of gene targets, and may function as either tumor suppressors or oncogenes [13]. Differential expression of miRNAs has revealed diagnostic, prognostic and therapeutic implications [14]. There are several approaches for genome scale miRNA expression profiling [15,16]. One of the most commonly used platforms is miRNA oligo arrays [17]. The overall experimental design of these miRNA arrays is very similar to regular expression arrays. Other miRNA profiling meth-

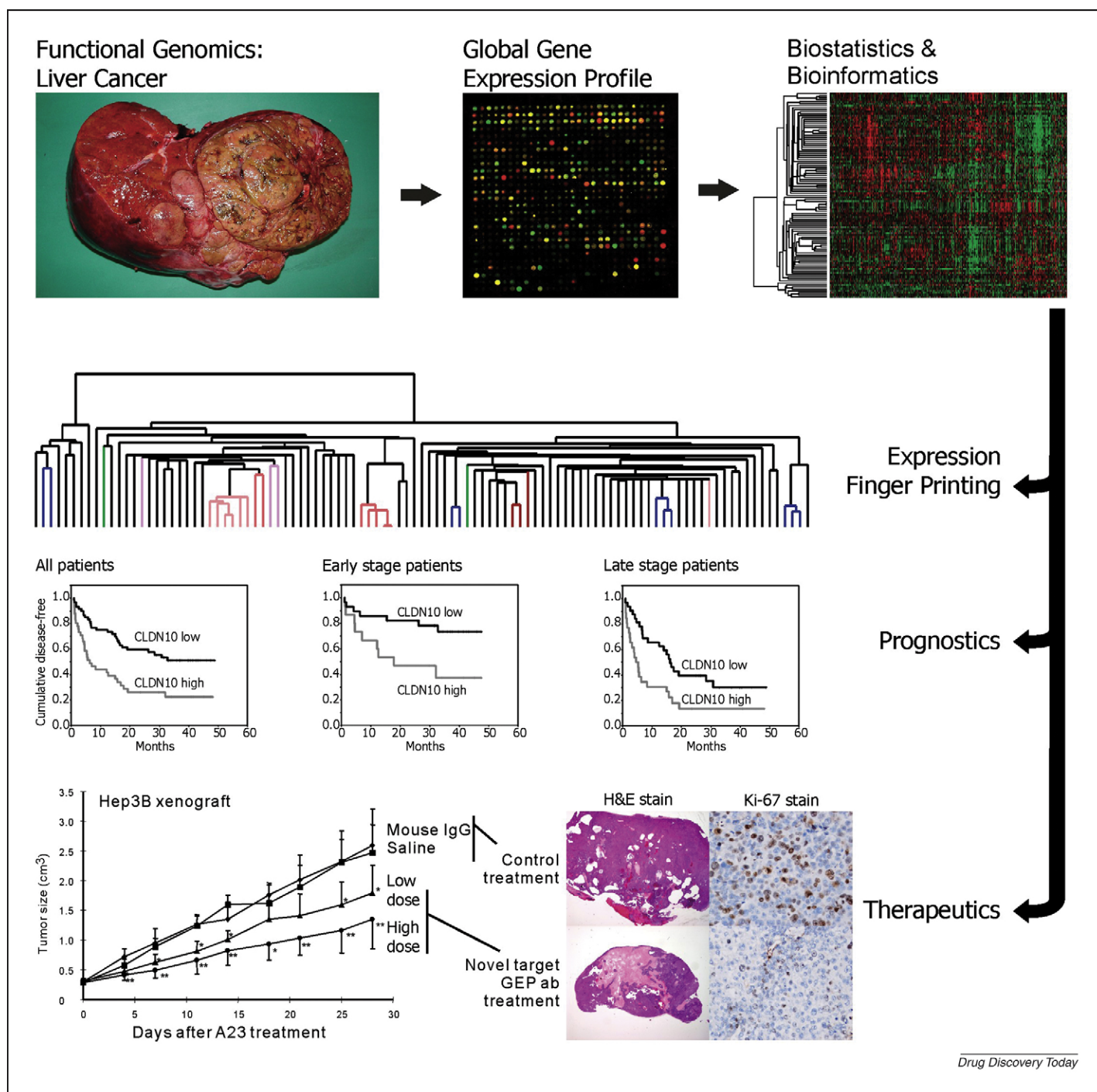
ods included multiplexed q-RT-PCR assays and bead-based methods [15,16].

### Array-based CGH for genome-wide analysis of DNA copy number variation

Chromosomal imbalances, including deletions and amplifications, are common in human tumors. Comparative genome hybridization (CGH) has been widely used to examine for the global analysis of DNA copy number since its first report in the early 1990s. The resolution of conventional CGH is limited by the length of the metaphase chromosomes, which is approximately 10 megabases and could contain hundreds of genes. Microarray-based CGH has been developed, which combines microarray technology with the CGH approach [18,19]. Defined DNA fragments (BACs, cDNAs or oligo) have been used to replace metaphase chromosomes and results in higher resolution. Microarray-based CGH allows a precise mapping for the regions of genetic aberrations [20,21], including human liver cancers [22].

### Genome-wide analysis of DNA methylation

DNA methylation is one of the most important covalent modifications of genomic DNA in eukaryotic cells. Human tumor samples frequently show abnormal patterns of DNA methylation, which

**FIGURE 2**

Functional genomic approach in liver cancer studies using gene expression profiling. Liver tissues and liver cancer tissues were collected with patient consent. The purified nucleic acid samples would be labeled with fluorescent dyes and hybridized to the arrays. The fluorescent signal intensities would be further analyzed by biostatistic and bioinformatic methods [4]. The clinical implication for functional genomics includes expression finger printing to delineate the clonality relationship of the multinodular tumors [8], identify novel biomarkers for prognostication [9] and therapeutic development [10].

may contribute to the aberrant patterns of gene expression and regulation. Traditional, methylation can only be determined on a gene-by-gene basis using methods that include bisulfite conversion followed by PCR or sequencing, methylation-sensitive restriction enzymes or affinity purification. Recent technology development has enabled the analysis of DNA methylation in a genome-wide scale [23,24]. Genome-wide methylation analyses

can be divided into two categories: array-based or non-array-based. Several companies provide commercially available arrays for methylation analysis [24]. The arrays are designed to analyze bisulfite-converted DNA (for example, bead arrays from Illumina), or use the restriction enzyme-based methylation analysis (for example, oligo arrays from NimbleGen or Agilent). Array hybridization and analysis is similar to what has been described in

expression or CGH arrays. Nonmicroarray-based experimental designs include Restriction Landmark Genome Scanning (RLGS), methylation specific digital karyotyping (MSDK), and high-throughput sequencing after bisulfate conversion [23,24].

#### *Chromatin immunoprecipitation followed by microarray analysis (ChIP-chip)*

'ChIP-chip' applies antibodies specific to a regulatory factor, in most cases, a transcriptional factor, for genome scale chromatin-immunoprecipitation combined with microarrays spotted with intergenic sequences to identify their bound targets [25,26]. Protein-bound DNA fragments retrieved by the antibody are hybridized to microarrays to identify the retrieved sequences. ChIP-chip analysis is now being widely used as a reliable tool to identify targets of crucial transcriptional regulators in a high-throughput manner.

#### **Genome-wide association studies of DNA sequence variation**

DNA sequence variation can influence disease risk and response to drug therapy by altering gene expression, RNA processing, or the amino acid sequence of proteins. Genome-wide association studies use DNA microarrays to investigate the effect of millions of common DNA sequence variants in the human genome, of which the most common type are single nucleotide polymorphisms (SNPs). The genome-wide association approach is statistically powerful [27] and has led to the discovery of many new genetic variants that underlie variation in human traits ([www.genome.gov/26525384](http://www.genome.gov/26525384)), including a number of cancers [28]. A number of these disease variants influence the risk of multiple diseases, including shared risk variants for several common cancers [29]. Thus, the genome-wide studies may have important implications in drug development by assisting to identify novel therapeutic targets and genetic biomarkers that for drug discovery [30].

Genome-wide association studies have become possible owing to several recent technological advances. Improvements in DNA microarray technology have rapidly reduced the cost of genotyping SNPs, allowing for the testing of up to one million SNPs using a single microarray. At the same time, the HapMap Project validated nearly four million SNPs in multiple diverse populations, and determined the extent of linkage disequilibrium (LD) between SNPs [31]. LD refers to the non-random association of SNPs, typically those that are closest together. The presence of LD allows for SNPs on genotyping platforms to serve as a proxy for other nearby SNPs [32]. As a result, current DNA microarrays can assay most common SNPs in the HapMap. In this way, LD reduces both the genotyping costs of genome-wide association studies and the multiple testing burden (see the 'Biostatistical Analysis' Section below).

It is important to note that genome-wide association studies are better suited to investigate the potential association of common variants, typically defined as those with a minor allele frequency of greater than 5%, with disease than rare variants [33]. Since strongly deleterious alleles are likely to face selection pressure, variants with large effects will be rare; common variants will have more modest effects on gene function. Most variants associated with disease in genome-wide association studies are common and have modest or small effects. Because of this, individual variants will not serve as

strong predictors of disease risk [34]. They may, however, explain a large amount of the risk of disease in the population, or population attributable fraction. For this reason, interventions that developed to counteract these risk variants could substantially reduce the incidence of disease in a population.

#### **Genome-wide somatic mutation analysis**

Human cancer is widely considered to be induced by somatic alterations within the cancer genome, leading to mutations of oncogenes or tumor suppressors [35]. Sequencing of tumor cell genomic DNA, also known as 'deep sequencing', has been applied to identify 'driver mutations', which will clearly have an important impact on our understanding of carcinogenesis. The development of next-generation sequencing technologies has enabled us to perform genome-wide mutation analysis on cancer cells [36–38]. In most cases, these deep sequencing analyses involve amplifying all individual exons for most known genes, followed by large scale sequencing to examine the amplicon for possible mutations [36,38]. For instance, the recent studies have discovered somatic mutations that affect key signaling pathways in acute myeloid leukemia and lung cancers [39,40].

#### **Proteomics analysis**

Expression profiling (mRNA-based for expression level changes) and genomic profiling (DNA-based for copy number or sequence variation) cannot provide a complete picture on the heterogeneity of complex diseased tissue. The level of mRNA or DNA changes do not always correspond to protein level changes, or to the post-translational modifications, for example, phosphorylation, which are crucial in regulating protein activity. Indeed, a number of the targeted therapeutic agents are designed to inhibit the activity of a protein, for example, tyrosine kinases. Therefore, protein profiling is essential in providing the protein molecular signatures for bioassay and therapeutic development. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) approach is commonly employed to study protein profiles [41,42]. The proteins can be separated according to their size (molecular mass) and charge (isoelectric point) properties and their abundance then determined accordingly. The difficulty in elucidating the identity of the protein spots remains, however, the major obstacle in clinical application. Since the late 1980s, matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) has been advanced to allow the rapid measurement for the molecular weights of different proteins with a time-of-flight (TOF) MS. MALDI-TOF MS has limitations on mass resolution and accuracy, however, to identify peptides with high confidence. Alternative approaches have been used to provide a more reliable determination of peptide sequence, including collision-induced dissociation (CID) with a tandem MS, electron capture dissociation (ECD), infrared multiphoton dissociation (IRMPD), and electron transfer dissociation (ETD) [43]. Furthermore, protein arrays, also known as antibody arrays, are an emerging technology that provides parallel analysis of multiple proteins [44]. In addition, protein arrays can be applied to profile specific protein post-translational modifications, such as phosphorylation or neddylation, and to measure enzyme activities and protein cell-surface expression. Proteomic approaches have been widely used for biomarker discovery in human tumors [45]. Protein profiling of blood samples has been the focus in recent



years, because it allows repeated measurements (especially important in monitoring treatment response) and without the need to obtain tumor tissues. Blood samples, prepared as serum or plasma fractions, have been used for biomarker discovery [46–48].

### Biostatistical analysis of genomic studies

Most genomic experiments produce large amounts of data. For example, in a gene expression microarray study, 22 000 genes  $\times$  100 samples will generate 2.2 million data points. In addition, genomic experiments are often noisy and not normally distributed, and usually contain missing values in the expression matrix. Robust biostatistical analyses are required to obtain biological relevant interpretations of the genomic data [49,50].

Specific statistical tools need to be applied to specific genomic studies. Therefore, people have to choose biostatistical software that is best suited for their specific experiments and those questions that they are trying to address. In general, statistical analyses of genomic data can be divided into two major categories: supervised and unsupervised methods [49,50]. Supervised approaches try to identify the genetic events that fit a predetermined pattern. For example, supervised analysis is used to identify genes that are differentially expressed between groups of samples, as well as to find genes that can be used to accurately predict the characteristics of groups. In contrast to the supervised method, the unsupervised approaches characterize genomic data without prior input or knowledge of predetermined pattern. Unsupervised analysis is used to identify internal structure in the genomic data set. The most commonly used unsupervised analysis tool is Hierarchical clustering and Principal Components Analysis (PCA).

Because genomic studies examine thousands or millions of data points, stringent significance criteria are applied to the association results. One method is to undertake a Bonferroni correction by dividing the significance criteria by the number of tests being conducted. For example, correcting for a million tested common SNPs in a genome-wide association study, an association would need a  $p$ -value of  $5 \times 10^{-8}$  (0.05 divided by 1 000 000) to be considered 'genome-wide significant.' It is for this reason that genome-wide association studies typically involve thousands of subjects to achieve sufficient statistical power. Other multiple testing adjustments that are less conservative than a Bonferroni correction, such as permutation derived  $p$ -values and false discovery rates, are often employed to maintain statistical power and to clarify the strength of a reported finding in light of the genomic scale of the experiment [51,52].

Despite all these advantages of biostatistical analysis, there is no standard or one-size-fits-all solution for statistical approach or even a single way to pick the significant  $p$ -value to balance type I and type II errors in statistical readouts. Each biologist has to approach this question on the basis of his/her own biological questions in each specific setting.

### Bioinformatics analysis

Genomic studies generally generate large amounts of data. Even after statistical analysis, one may identify large number of deregulated genes, for example, genes that are methylated or mutated in tumor samples. Bioinformatics analysis tools have been developed to assist scientists to extract meaningful data and interpret the genomic data in a functional manner.

One of the most commonly used methods to annotate the gene function is through Gene Ontology (GO, <http://www.geneontology.org/>) [53,54]. GO classifies gene function according to three organizing principles: molecular function, biological process and cellular component. When certain GO terms are statistically enriched in a cluster, it may suggest possible functional significance of the cluster of genes.

Another commonly used bioinformatic analysis tool is Gene Module Analysis [55,56]. Just as GO term analysis builds on pre-existing knowledge for the interpretation of microarray data, one can interrogate the global gene expression profile with respect to known sets of genes by gene module analysis. In brief, gene module analysis asks whether the genes whose level of expression changes in an experiment are similar to those that have been observed in another setting. The gene modules may be defined by function (e.g. GO terms or other annotations), the presence of specific *cis*-regulatory motifs or *trans*-regulatory motifs for transcription factor or miRNA binding, or known responsiveness to specific signaling pathways or drugs. Gene Set Enrichment Analysis (GSEA) is the most popular modular analysis method that is publicly available (<http://www.broad.mit.edu/gsea/>) [57]. Ingenuity Pathway analysis software (<http://www.ingenuity.com/>) is a popular commercially available modular analysis tool.

### Integrated functional genomic analysis

Integrated functional genomics is an approach that combines results from multiple genomic analyses or genomic analysis and functional experiments to identify important genetic signals underlying biological processes. Integrated functional genomics is crucial for cancer research. As one can imagine, genome-wide studies are likely to produce large numbers of genes with expression alternations, mutations, abnormal methylation or DNA copy number variations in cancer cells. Only a small number of these genetic alternations, however, have functional roles during tumorigenesis. The majority of the genes are likely to be passenger genes that are either the products of genomic instability or secondary changes during carcinogenesis, but have no direct contribution to the malignant transformation. By combining multiple genomic analyses, one can significantly narrow down the list of genes that may contain functionally significant oncogenes or tumor suppressor genes. For example, DNA copy number gains and losses contribute to cancer development by increased and decreased expression of oncogenes and tumor suppressor genes, respectively. In our recent studies, we combined expression arrays and array CGH studies of human HCC samples, and were able to identify a relatively small set of genes as candidate oncogenes or tumor suppressors whose expression levels are associated with DNA copy number changes [58]. In another example, ChIP-chip experiments can be integrated with gene expression analysis to delineate how specific transcriptional factors regulate global gene expression. In a recent study, Acevedo *et al.* performed ChIP-chip experiments to assay binding of RNA polymerase II, H3me3K27, and H3me3K9 and DNA methylation in 25 000 promoter regions in normal liver and liver tumor samples [59]. The experiments successfully identified changes in active and silenced regions of the genome in liver tumor cells, and in so doing identified novel molecular mechanisms that mediate tumor specific changes in gene expression in the liver. In addition, by combining genomic

analysis and functional screenings, such as siRNA mediated gene silencing, we can rapidly identify potential driver genetic events. For example, in a recent study, Zender *et al.* identified small regions of recurrent deletions in human liver cancer by genomic analyses [60]. Using microRNA-based short-hairpin RNA libraries, targeting genes within these deleted regions, the group conducted *in vivo* RNAi screening to identify genes that, when silenced, cooperate with Myc to promote liver cancer development. The study successfully identified and validated 13 tumor suppressor genes for liver cancer [60].

## Conclusion

In summary, in the next decade, with these tools for genomic analyses being widely used in biomedical research, one can foresee the emerging of large amounts of data tangling different biological questions. Functional genomic analyses will probably have multiple implications for drug discovery and development. For example, integrated functional genomic studies will probably

identify driver mutations or genes that tumor cells depend on for growth and metastasis. These genes can be used as targets for drug development, and it will lead to drugs for specific genetic events that are likely to be more efficient and less toxic for cancer treatment. Genomic analyses will also identify genetic signatures, such as gene expression profiles or specific mutation status, which can be used to predict drug responsiveness. These biomarkers will clearly increase the power and efficiency of clinical trials by selecting the appropriate patient populations and may lead to successful clinical drug development. Altogether, the application of genomic analysis to drug development will facilitate drug discovery and the development process in a more efficient manner.

## Acknowledgements

This work is supported by NIH K01CA096774 and R21CA131625 to X.C. as well as RGC and NSFC (HKU 7560/06 M and N\_HKU 709/07) to S.T.C.

## References

- 1 Kimmel, A. and Oliver, B. (2006) *DNA Microarrays*. Academic Press
- 2 Chung, C.H. *et al.* (2002) Molecular portraits and the family tree of cancer. *Nat. Genet.* 32 (Suppl.), 533–540
- 3 Liu, E.T. (2008) Functional genomics of cancer. *Curr. Opin. Genet. Dev.* 18, 251–256
- 4 Chen, X. *et al.* (2002) Gene expression patterns in human liver cancers. *Mol. Biol. Cell.* 13, 1929–1939
- 5 Iizuka, N. *et al.* (2003) Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *Lancet* 361, 923–929
- 6 Ye, Q.H. *et al.* (2003) Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat. Med.* 9, 416–423
- 7 Lee, J.S. *et al.* (2006) A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat. Med.* 12, 410–416
- 8 Cheung, S.T. *et al.* (2002) Identify metastasis-associated genes in hepatocellular carcinoma through clonality delineation for multinodular tumor. *Cancer Res.* 62, 4711–4721
- 9 Cheung, S.T. *et al.* (2005) Claudin-10 expression level is associated with recurrence of primary hepatocellular carcinoma. *Clin. Cancer Res.* 11 (2 Pt 1), 551–556
- 10 Ho, J.C. *et al.* (2008) Granulin-epithelin precursor as a therapeutic target for hepatocellular carcinoma. *Hepatology* 47, 1524–1532
- 11 Ruvkun, G. (2008) The perfect storm of tiny RNAs. *Nat. Med.* 14, 1041–1045
- 12 Stefani, G. and Slack, F.J. (2008) Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell. Biol.* 9, 219–230
- 13 Esquela-Kerscher, A. and Slack, F.J. (2006) Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* 6, 259–269
- 14 Fabbri, M. *et al.* (2008) *MicroRNAs Cancer J.* 14, 1–6
- 15 Blenkiron, C. and Miska, E.A. (2007) miRNAs in cancer: approaches, aetiology, diagnostics and therapy. *Hum. Mol. Genet.* 16, R106–113
- 16 Kong, W. *et al.* (2009) Strategies for profiling microRNA expression. *J. Cell. Physiol.* 218, 22–25
- 17 Yin, J.Q. *et al.* (2008) Profiling microRNA expression with microarrays. *Trends Biotechnol.* 26, 70–76
- 18 Kallioniemi, A. (2008) CGH microarrays and cancer. *Curr. Opin. Biotechnol.* 19, 36–40
- 19 Shinawi, M. and Cheung, S.W. (2008) The array CGH and its clinical applications. *Drug Discov. Today* 13, 760–770
- 20 Carter, N.P. (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.* 39 (7 Suppl.), S16–21
- 21 McCarroll, S.A. and Altshuler, D.M. (2007) Copy-number variation and association studies of human disease. *Nat. Genet.* 39 (7 Suppl.), S37–42
- 22 Patil, M.A. *et al.* (2005) Array-based comparative genomic hybridization reveals recurrent chromosomal aberrations and Jab1 as a potential target for 8q gain in hepatocellular carcinoma. *Carcinogenesis* 26, 2050–2057
- 23 Shen, L. and Waterland, R.A. (2007) Methods of DNA methylation analysis. *Curr. Opin. Clin. Nutr. Metab. Care* 10, 576–581
- 24 Zilberman, D. and Henikoff, S. (2007) Genome-wide analysis of DNA methylation patterns. *Development* 134, 3959–3965
- 25 Wu, J. *et al.* (2006) ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res.* 66, 6899–6902
- 26 Buck, M.J. and Lieb, J.D. (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 83, 349–360
- 27 Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* 273, 1516–1517
- 28 Jorgenson, E. and Witte, J.S. (2007) Genome-wide association studies of cancer. *Future Oncol.* 3, 419–427
- 29 Witte, J.S. (2009) Prostate cancer genomics: towards a new understanding. *Nat. Rev. Genet.* 10, 77–82
- 30 Kingsmore, S.F. *et al.* (2008) Genome-wide association studies: progress and potential for drug discovery and development. *Nat. Rev. Drug Discov.* 7, 221–230
- 31 Frazer, K.A. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861
- 32 Jorgenson, E. and Witte, J.S. (2006) Coverage and power in genomewide association studies. *Am. J. Hum. Genet.* 78, 884–888
- 33 Jorgenson, E. and Witte, J.S. (2006) A gene-centric approach to genome-wide association studies. *Nat. Rev. Genet.* 7, 885–891
- 34 Hardy, J. and Singleton, A. (2009) Genomewide association studies and human disease. *N. Engl. J. Med.* 360, 1759–1768
- 35 Weir, B. *et al.* (2004) Somatic alterations in the human cancer genome. *Cancer Cell.* 6, 433–438
- 36 Parmigiani, G. *et al.* (2009) Design and analysis issues in genome-wide somatic mutation studies of cancer. *Genomics* 93, 17–21
- 37 Sjoblom, T. (2008) Systematic analyses of the cancer genome: lessons learned from sequencing most of the annotated human protein-coding genes. *Curr. Opin. Oncol.* 20, 66–71
- 38 Simpson, A.J. (2009) Sequence-based advances in the definition of cancer-associated gene mutations. *Curr. Opin. Oncol.* 21, 47–52
- 39 Ley, T.J. *et al.* (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72
- 40 Ding, L. *et al.* (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455, 1069–1075
- 41 Issaq, H. and Veenstra, T. (2008) Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *Biotechniques* 44, 697–698
- 42 Carrette, O. *et al.* (2006) State-of-the-art two-dimensional gel electrophoresis: a key tool of proteomics research. *Nat. Protoc.* 1, 812–823
- 43 Chen, C.H. (2008) Review of a current role of mass spectrometry for proteome research. *Anal. Chim. Acta* 624, 16–36
- 44 Haab, B.B. (2006) Applications of antibody array platforms. *Curr. Opin. Biotechnol.* 17, 415–421

- 45 Sun, S. *et al.* (2007) Oncoproteomics of hepatocellular carcinoma: from cancer markers' discovery to functional pathways. *Liver Int.* 27, 1021–1038
- 46 Steel, L.F. *et al.* (2003) A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. *Proteomics* 3, 601–609
- 47 Smalley, D.M. and Ley, K. (2008) Plasma-derived microparticles for biomarker discovery. *Clin. Lab.* 54, 67–79
- 48 Zolla, L. (2008) Proteomics studies reveal important information on small molecule therapeutics: a case study on plasma proteins. *Drug Discov. Today* 13, 1042–1051
- 49 Slonim, D.K. (2002) From patterns to pathways: gene expression data analysis comes of age. *Nat. Genet.* 32 (Suppl.), 502–508
- 50 Allison, D.B. *et al.* (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.* 7, 55–65
- 51 Browning, B.L. (2008) PRESTO: rapid calculation of order statistic distributions and multiple-testing adjusted *P*-values via permutation for one and two-stage genetic association studies. *BMC Bioinform.* 9, 309
- 52 Farcomeni, A. (2008) A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Stat. Methods Med. Res.* 17, 347–388
- 53 Rhee, S.Y. *et al.* (2008) Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* 9, 509–515
- 54 Harris, M.A. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–261
- 55 Segal, E. *et al.* (2005) From signatures to models: understanding cancer using microarrays. *Nat. Genet.* 37 (Suppl.), S38–45
- 56 Wang, X. *et al.* (2008) Gene module level analysis: identification to networks and dynamics. *Curr. Opin. Biotechnol.* 19, 482–491
- 57 Subramanian, A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550
- 58 Lee, S.A. *et al.* (2008) Integration of genomic analysis and *in vivo* transfection to identify sprouty 2 as a candidate tumor suppressor in liver cancer. *Hepatology* 47, 1200–1210
- 59 Acevedo, L.G. *et al.* (2008) Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors. *Cancer Res.* 68, 2641–2651
- 60 Zender, L. *et al.* (2008) An oncogenomics-based *in vivo* RNAi screen identifies tumor suppressors in liver cancer. *Cell* 135, 852–864