Briefings in Functional Genomics Advance Access published June 28, 2013

doi:10.1093/bfgp/elt024

Genomics and epigenomics: new promises of personalized medicine for cancer patients

Michal-Ruth Schweiger, Christian Barmeyer and Bernd Timmermann

Abstract

Recent years have brought about a marked extension of our understanding of the somatic basis of cancer. Parallel to the large-scale investigation of diverse tumor genomes the knowledge arose that cancer pathologies are most often not restricted to single genomic events. In contrast, a large number of different alterations in the genomes and epigenomes come together and promote the malignant transformation. The combination of mutations, structural variations and epigenetic alterations differs between each tumor, making individual diagnosis and treatment strategies necessary. This view is summarized in the new discipline of personalized medicine. To satisfy the ideas of this approach each tumor needs to be fully characterized and individual diagnostic and therapeutic strategies designed. Here, we will discuss the power of high-throughput sequencing technologies for genomic and epigenomic analyses. We will provide insight into the current status and how these technologies can be transferred to routine clinical usage.

Keywords: *personalized medicine; individualized therapy; high-throughput sequencing; genomics; epigenomics*

PERSONALIZED MEDICINE

Recently the term 'personalized medicine' arose with the goal to establish patient-specific diagnosis and treatment strategies. Hand in hand with an enormous progress in technologies, new molecularly targeted agents have been developed. The combination of both, the fast deciphering of genomic alterations and the emergence of side-specific therapeutic agents, paved the way for a new concept in oncology: Treatment of tumors based on their molecular profile nearly irrespective of their localization and histology.

The advancement in high-throughput technologies covering DNA, RNA, proteins or metabolites makes it possible to characterize the disease process of an individual patient. Already first genome-wide sequences of individual patients revealed the complexity of cancer genomes: Contrary to the previous view most tumors not only harbor one mutation, but a multitude of different genetic and epigenetic events which, in their combination, provide tumor-specific patterns which need to be taken into consideration for optimal therapeutic concepts [1–3]. Nevertheless, in some instances tumor genomes seem to be quite stable and depend on single tumorigenic driver events [4]. This might be restricted to specific translocation-prone tumor subgroups such as chronic myelogenous leukemia with *BCR–ABL* fusion genes [5,6]. The development of targeted therapies against the *ABL* kinase domain, Imatinib, resulted in high response rates with an estimated 93% of patients which remained free from disease progression further

Corresponding author. Michal R. Schweiger, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany. Tel.: +49 30 84131339; Fax: +49 30 84131380; E-mail: mschweig@molgen.mpg.de

Michal-Ruth Schweiger is a biochemist (PhD) and medical geneticist (MD) with interest in systems biology approaches for genomics and epigenomics in oncology. She is head of the cancer genomics group at the Max Planck Institute for Molecular Genetics, Berlin.
Christian Barmeyer is a physician (MD) with interest in molecular physiology and its implementation in personalized medicine. As postdoc he worked at the Yale University department of internal medicine, section of digestive diseases. Currently, he is a resident in internal medicine, gastroenterology at the Charité Universitätsmedizin Berlin, Germany.
Bernd Timmermann (PhD) as head of the sequencing core facility of the Max Planck Society in Berlin is involved in many large-scale sequencing projects (e.g. 1000 Genomes, OncoTrack). His research interest is focused on human genetic variation and the translation of new sequencing technologies into diagnostic settings.

© The Author 2013. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

underlining the tumor's dependency on the oncogenic fusion gene [7].

The knowledge of the complexity of tumor genomes by far surpasses our knowledge about human biology of a few years ago. Although it took 10 years and ~1 billion dollars to complete the sequence of the human genome just 10 years ago, we are now able to generate that much information on one next-generation sequencing (NGS) instrument in 10 days (Illumina, Solid), and we expect an even further enhancement with whole-genome sequencing results within less than a few days at costs of below 1000$ (e.g. PacBio announcement). This progress makes the transfer of high-throughput sequencing (HTS) technologies to routine clinical diagnostics likely.

First chemotherapies developed, alkylating agents and antimetabolites, aimed on a general disruption of cell homeostasis. Only since a few years, targeted therapies by selective small molecule inhibitors are in use. These drugs aim at an inhibition or activation of specific cellular target proteins, most frequently protein kinases. Parallel to this development the recognition emerged that only a subset of patients significantly benefit from these chemotherapies. The others are unsuccessfully treated, suffering from side-effects without benefiting from the treatment. Thus, the driving force underlying personalized therapy is the enormous heterogeneity among tumors, even among tumors of the same class. Using newest high-throughput technologies knowledge arises that a wide spectrum of different genetic alterations from mutations over copy number to structural variations are found in each tumor. This complexity is even aggravated by epigenetic variations seen in many tumor entities.

HIGH-THROUGHPUT SEQUENCING

The worldwide personal genome project (PGP) and the 1000 genomes project aim at sequencing thousands of individual genomes to gain insight into genomic variabilities [8,9]. This aim seemed to be an illusion just a few years ago, but has now become reality. In this direction, the development of HTS technologies (454 (Roche), Illumina, SOLiD and Ion Torrent (Life Technologies)) has initiated a real revolution in genomics analyses. With these technologies, an enormous parallel analysis of genomic DNA has become possible in a time-frame of a few days. Key features of these technologies are the spatial immobilization of millions of short DNA fragments followed by a massively parallel sequencing process (Figure 1). Fluorescence markers incorporated into the DNA fragments either by ligation (SOLiD) or by polymerase activity (Illumina) during the sequencing process, or a light signal emitted from luciferase activity coupled to the incorporation of nucleotides (454FLX) are detected by high-resolution cameras [10,11]. The Ion Torrent technology is, even though it follows the same process of spatial localization of DNA fragments and a cycled wash mode, located between second and third generation sequencing technologies. A semiconductor technology is used to create micro-wells that carry out sequencing steps and sense the incorporation of nucleotides by the release of hydrogen ions. This eliminates the need of scanning cameras and accelerates the sequencing process. However, amplification and termination steps are still required which set an upper limit for the speed of the process.

All these sequencing technologies provide digital information on DNA sequences which are assembled and aligned to reference genomes using bioinformatics tools. Digital information is the basis for re-sequencing approaches as well as quantification modules for gene expression analyses or chromatin immunoprecipitation experiments. The parallel sequencing of millions of DNA molecules is especially useful for sequencing heterogeneous material, as is the case with cancer tissues.

Sequencing of entire genomes is an important application of HTS. Although all types of genetic polymorphisms can be identified using whole-genome re-sequencing approaches, this method is still too cost-intensive to be conducted routinely. Instead, many research and diagnostic goals might be achieved by sequencing only a fraction of the genome.

For many diseases, specific sets of genes involved in the pathomechanism or implicated by whole-genome association studies are of major interest. In addition, there might be limitations on sequencing capacity to only re-sequence all protein-coding regions ('exome') which encompass ~1% of the genome. Several targeted sequence enrichment techniques to reduce DNA sequence complexities have been established [12–16]. In particular, microarray-based genomic selection (MGS), multiplex exon capture or bead-based enrichment methods are already commercially available and used for targeted

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

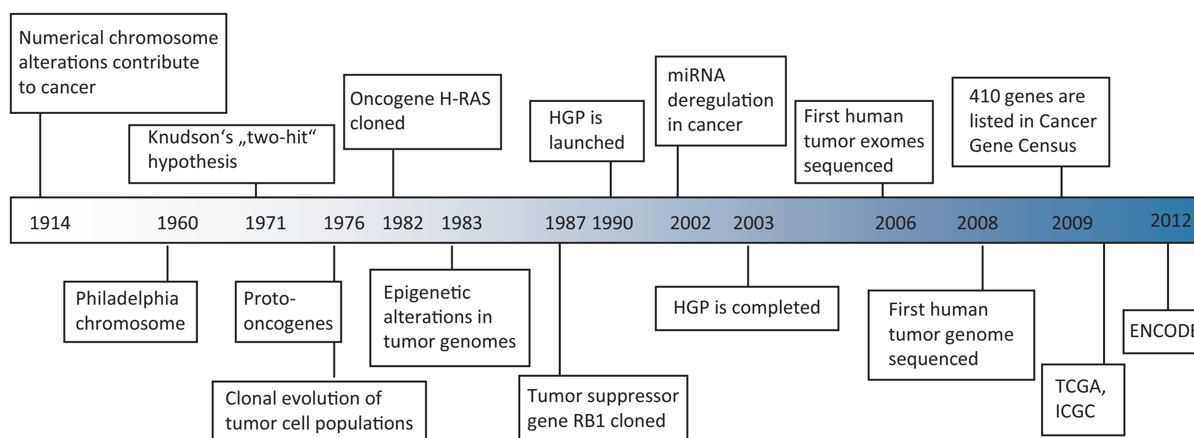

Figure 1: Time line of cancer research. The first human genome was sequenced in a world-wide effort, the HGP (Human Genome Project). The development of new technologies with high-throughput enabled, besides the fast sequencing of individual genomes, the parallel detection of mutations, structural variations and epigenetic alterations. Large-scale projects like the TCGA (The Cancer Genome Atlas) and the ICGC (International Cancer Genome Consortium) aim at sequencing more than 500 cases for each cancer entity.

sequencing approaches. Main differences are the amount of input DNA, the ease of performance and if they are hybridization- or synthesis based. In addition, as sequencing capacities per run continue to increase and are already at throughputs of up to 3 billion reads per run, the ability to easily multiplex multiple samples will get increasingly important.

Besides the detection of polymorphisms and mutations structural variations can be detected with HTS technologies. As such, genomic rearrangements resulting in aberrant transcriptional events are common features in human cancer. Thus, besides point mutations extended genome rearrangements are implicated in tumorigenesis such as translocations, inversions, small insertions/deletions (InDels) and copy number variations (CNVs). InDels are most often defined as deletions or insertions below 1 kb of DNA, whereas CNVs comprise alterations larger than 1 kb of DNA. Recent analyses by genome-wide approaches have uncovered the importance of structural genomic variations in health and disease. Furthermore, genetic association studies have implicated CNVs in cancer. The connection of changes in CNVs to several diseases has boosted the development of new technologies to investigate these rearrangements. First methods involve microscopic examination of chromosome bandings, PCR, fluorescence in situ hybridizations (FISH) and microarrays. More recently, HTS technologies have been utilized for these analyses. These approaches offer important advantages over conventional methods such as microarrays or array comparative genomic hybridization. In particular, in addition to quantitative information they provide data about qualitative mechanisms, e.g. balanced rearrangements such as reciprocal translocations and inversions which would have been otherwise overseen. Moreover, since the sequencing is based on digital modes, they are able to detect variants that are present in a subpopulation of cells. Given the short read-lengths of most HTS technologies, paired-end sequencing approaches have been developed. Here, two short DNA segments separated by a spacer of chosen length (typically 200 bp–2000 bp) are sequenced together. A comparison of the actual distance of the mapped segments on the reference genome with the chosen spacer length is able to identify insertions, deletions and intra-/inter-chromosomal rearrangements. Paired-end sequencing is in particular important for the detection of translocations and, if applied to RNA, for the detection of splice variants.

In comparison, 'third generation sequencing' approaches, relying on detecting the binding of the nucleotidetriphosphate to the polymerase in real time (Pacific Biosciences), nanopore (e.g. Oxford Nanopore) and scanning probe sequencing approaches [17–19], are directed toward sequencing of single DNA molecules without any prior amplification or labeling [20]. PacBio use optical techniques to monitor single polymerases in real time: In the procedure developed by Pacific Biosciences, multiple 'zero mode waveguide' (ZMW) structures on a chip define minute volumes containing single polymerase molecules and restrict the area of

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

detecting fluorescence signals to the bottom 30 nm of the ZMW. DNA sequences are read out by the series of desoxynuclotidetriphosphates that are labeled with different fluorescent dyes and illuminate during the incorporation step. Key to the detection process is that the incorporation process takes milliseconds, which is approximately three orders of magnitude longer than simple diffusion. This difference results in a higher signal intensity for incorporated versus unincorporated nucleotides. During incorporation of the triphosphate analog, the fluorescent label is cleaved off together with the pyrophosphate group, allowing the next incorporation step. Thus, the combination of the ZMW technology with the immediate release of the fluorescent dye during the incorporation enables a real-time sequencing detection. In addition to determining the sequence, this procedure has been shown to also be able to detect base modifications in the DNA, due to their influence on the kinetics of incorporation [21]. The basic principle of nanopore sequencing is such that a DNA strand or a cleaved nucleotide is passed through a nanopore and induces changes in the current applied [17]. The use of electrical currents for nucleotide identification promises the discrimination of all four nucleotides and, in addition, the identification of methylated cytosines. That would implicate that during one sequencing process all 'five' nucleotides (A, T, C, G and 5mC) could be distinguished (in addition to further nucleotide modifications) and no additional manipulation of DNA would be required for the construction of DNA methylation patterns. In another approach under development, called scanning probe sequencing, the DNA molecule is immobilized and the scanning instrument records the nucleotides [19].

TUMOR GENOMES

The nucleotide sequence is the primary level of genetic information and the basic principle of genetic inheritance. A first connection between alterations in the genome and cancer was discovered in 1960 by Nowell and Hungerford who found a consistent chromosomal abnormality in leukemic cells (Figure 2) [5]. This abnormality was later called 'Philadelphia' chromosome, a translocation between the long arm of chromosome 9 and 22, leading to the oncogenic activation of ABL ('Abelson Murine Leukemia Viral Oncogene Homolog 1') [22]. Further studies revealed copy number alterations (gain and loss) and mutations in single genes leading to an activation of oncogenes or inactivation of tumor suppressor genes. Prototype oncogenes are MYC (v-myc myelocytomatosis viral oncogene homolog) and the tumor suppressor gene RB1 (retinoblastoma 1). For a long time such genomic events were thought to be single events sufficient for the tumor progress. However, through large-scale sequencing studies it turned out that this is by far not the case and that a multitude of genomic alterations mark cancer genomes. The combination of these events make up a complex pattern underlying each individual cancer genome and the sum of these events may drive tumorigenesis. Thus, the understanding of oncogenic processes has turned from single-gene alterations to pathways which are frequently disrupted. This view fits very well to the notion that no single gene has been identified which is altered in all patients. As a consequence of this changed view it becomes difficult to identify genomic biomarkers with high specificity and sensitivity. Nevertheless, genomic information may be used for personalized medicine approaches where each cancer patient is treated with drugs tailored for his particular tumor. But even here, due to the broad array of different mutations found in each cancer, it becomes difficult to find clinical trials where special patients might be integrated. Either we need new forms of clinical trials in terms of personalized medicine studies or we need a broader array of drug testing studies [23].

TUMOR EPIGENOMES

Besides genomic information, another level of tumor complexity arises from epigenetic variations of DNA segments which are also underlying the inheritance of phenotypes from generation to generation as well as from cell to cell during cell division [24]. Genome-wide studies on epigenetic changes are now termed 'epigenomics'. Epigenetic variations can be grouped into covalent DNA modifications, in particular methylation of nucleotides, or post-transcriptional modifications of histones (e.g. acetylation, ubiquitinylation or methylation) and—on a higher order—chromatin remodeling processes. Here, chromatin remodeling means the dynamic compaction of the genome including the activity of ATP-dependent chromatin remodeling complexes and non-coding RNAs. Less compacted structures can be visualized as light-colored bands

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

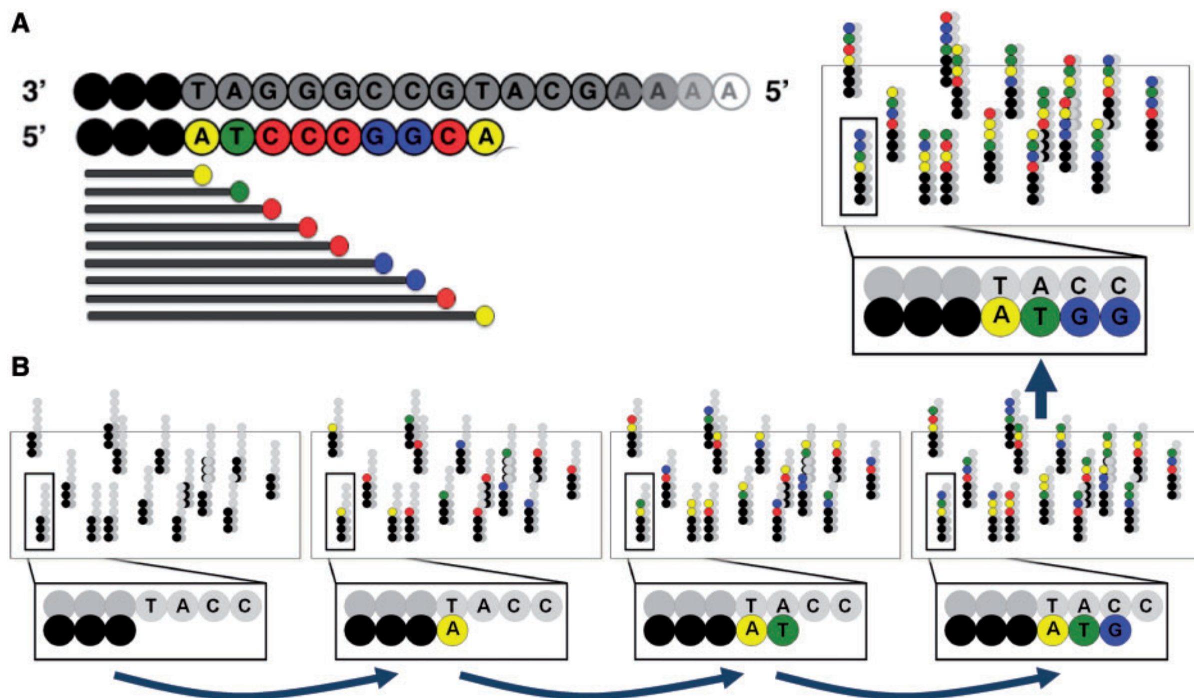

Figure 2: Schematics of the sequencing process. (**A**) Modern Sanger sequencing is based on the labeling of nucleotides with different fluorescence dyes and a termination chemistry. Size separation of the fragments resolves the sequence. (**B**) NGS technologies provide sequence information on millions of DNA fragments in parallel. The resolution of single fragments is achieved by an immobilization of the DNA fragments—as here illustrated by the Illumina process—and the scanning of fluorescence intensities after each nucleotide incorporation cycle. New nucleotides are incorporated stepwise after periodic scan-and-wash cycles.

when stained and observed under an optical microscope, whereas heterochromatin, tightly packed chromatin, stains darkly and is mainly associated with transcription-silent regions.

Epigenetic modifications play critical roles in all DNA-based processes including transcription, DNA repair and replication [25]. The development of HTS technologies has dramatically advanced our view of genome-wide epigenetic events. With this it became possible to provide comprehensive maps of nucleosome positioning and chromatin conformation [26,27]. In addition, transcription factor binding sites as well as post-transcriptional modifications of histones are localized with chromatin immunoprecipitation techniques followed by HTSs (ChIP-Seq) [28,29]. Here, antibodies against the protein or modification of interest are used to pull out the antigen/protein from cellular extracts. Through preceding cross-linking approaches proteins which have been pulled out carry their bound DNA as 'backpack' with them. HTS then provides specific site information where the protein of interest or the histone with the modification of interest binds. A prerequisite for this technology are antibodies with high specificities and high overall qualities. At the moment these antibodies are main limitations for high-throughput analyses and careful quality controls need to be performed for each antibody.

In humans, cytosine methylation was the first mark discovered. In the current paradigm it is required for the regulation of gene expression as well as for silencing transposons and other repetitive sequences [30]. The chemical modification occurs predominantly via a covalent attachment of a methyl group to the C5 position of the cytosine ring (5mC) in CpG dinucleotides. Thereby, the structure of cytosine is altered without changing its base-pairing properties. Altered DNA methylation patterns have been reported in a diverse array of complex human diseases such as cancer, systemic autoimmune and psychiatric diseases as well as in monogenic epigenetic diseases [31]. In this regard, the first molecular epigenetic change, a global reduction of DNA methylation in cancer cells, has been described by Feinberg and Vogelstein [32] and in the same year by Gama-Sosa *et al.* [33]. These changes

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

were found in both preinvasive and invasive cancers and implicate that alteration in the cytosine methylation patterns are among the earliest events in tumorigenesis. In addition, it has been shown that specific alterations in the cytosine methylation patterns of CpGs in promoter regions are associated with certain tumor entities or stages. Overall, the extend of epigenetic modifications in cancer tissues is extraordinary different from their normal counterparts (Figure 3). For DNA methylation more than 150 000 significantly differentially methylated regions can be identified [34]. This is, in particular, interesting for the identification of biomarkers, because it is more likely to identify reliable markers for any given clinical question when they can be selected out of a large set of alterations than if they are based on single and often infrequent events such as mutations. Consequently, the first biomarkers have been developed on the basis of these modifications [35].

Third generation sequencing instruments contain the promise to directly identify modified nucleotides. Otherwise, modified nucleotides need to be either enriched by, e.g. chromatin immunoprecipitations, or chemically treated to distinguish them from the 'background'. Over the past years, several epigenetic technologies have been developed either for profiling methylated genomic regions (indirect methods) or for typing the methylated base (direct methods). These approaches differ concerning the obtainable resolution with direct methods resulting in single-nucleotide patterns of methylated cytosines within genomes, whereas indirect methods measure average methylation levels across many molecules [36–39].

Using HTS technologies for the interrogation of DNA methylation patterns, the classification into indirect and direct approaches can be maintained and extended: Indirect methods include affinity-enrichment sequencing (AE-Seq: MeDIP-Seq and MBP-Seq) methods and methods that use endonuclease digests followed by sequencing (Enzyme-Seq: Methyl-Seq, MCA-Seq, HELP-Seq, MSCC) [24]. Methylation profiles are then inferred by subsequent sequencing, read alignment and counting of reads per genomic interval. Direct methods—BS-Seq, BC-Seq, BSPP and RRBS—in contrast rely on bisulfite conversion of unmethylated cytosines and consecutive sequencing, which allows methylation profiling with a resolution on single base level.

Indirect approaches provide information as a methylation score for regions of ~100–200 bp length. All methods are based on the enrichment of methylated DNA. The fragments captured by any of those methods can then be identified by either hybridization to known sequences or by sequencing. The use of HTS instead of custom-designed hybridization-arrays to identify precipitated DNA fragments provides genome-wide information about methylated regions. This implies that all DNA fragments can be identified and not only pre-selected regions which are immobilized on an array. The completeness of the data is especially advantageous in generating methylation profiles outside of CpG-islands and promoter regions, for example in gene bodies where DNA methylation changes have recently been shown to occur [40,41].

MeDIP-Seq (methylation-dependent immunoprecipitation) and MBP-Seq (methyl-binding protein) rely on precipitations of DNA fragments containing methylated cytosines (5mC) and use an anti-5mC antibody or methyl-binding proteins (MBPs) [42–45]. Both methods belong to the class of affinity-enrichment sequencing approaches (AE-Seq).

The MeDIP-enrichment depends upon the 5mC content in a way that a threshold level of methylation, ~2–3%, is required for a successful enrichment [34]. Regions with high CpG content are, therefore, more likely to be enriched than regions with low CpG content. First MeDIP-seq experiments indicate that ~30–40 million reads are required for a human genome-wide analysis [36,46]. MeDIP-seq approaches have been performed so far using Illumina's Genome Analyzer technology [46] but we recently established several methylation analysis methods for SOLiD sequencers, because of improved throughput [34]. MBPs preferentially bind double-stranded DNA with symmetrically methylated CpG sequences and, in contrast to MeDIP-protocols where the DNA is denatured and single stranded, the adapter ligation step is less critical and can be performed after the affinity purification. A challenge of both AE-Seq methods is that 'no signal' can be explained either by very low methylation levels or experimental failure and hypomethylation patterns are, therefore, very difficult to assess.

Protocols that use endonucleases (Enzyme-Seq technologies) like Methyl-Seq [47], MCA-Seq [48], HELP-Seq [49] and MSCC [41,50] exploit the fact, that restriction enzymes exists which target sequences that comprise CpG sites in a methylation sensitive manner. Following DNA digestion all Enzyme-Seq methods encompass a size selection

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

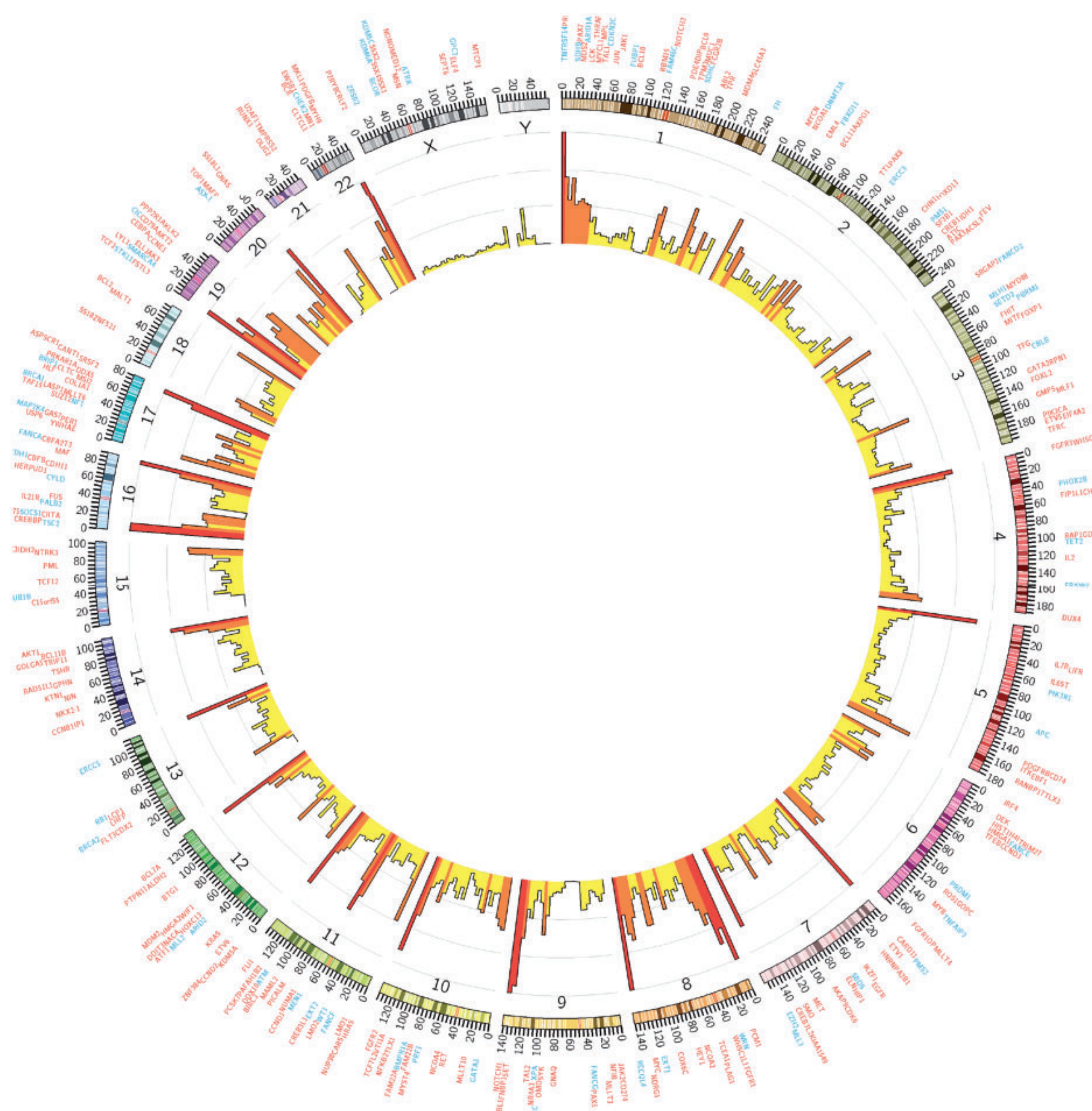

Figure 3: Circos plot for the distribution of differentially methylated regions in cancer. All chromosomes are arranged in a circle (middle). The inner circle shows the number of significantly differentially methylated regions. Peak heights and color represents the number of alterations—with dark high peaks indicating a high number of differential methylation. The outer circle indicates locations of oncogenes and tumor suppressor genes.

step to ensure that the fragments selected for sequencing are close to the CpG site [47]. Analysis is done by counting the reads per genomic region and combined evaluation of treatment and control samples. If no control samples exist methylation-sensitive (e.g. HpaII, SmaI) and methylation-insensitive (e.g. MspI, XMaI) preparations can be compared, a step which is also advisable if copy number variants are expected to be present [49]. A drawback of the Enzyme-Seq methods is that any region showing at least one read in the methylation-sensitive digest is currently called 'unmethylated'. Thereby the quantitative methylation state of the individual region is lost, and partial methylation remains unidentified [41,47].

Direct assessment techniques like BS-Seq [51–53], BC-Seq [54], BSPP [41,50,55,56] or RRBS [57,58] determine methylation profiles directly from the sequence enabling base pair resolution. Methylated DNA is marked through a 'bisulfite (BS) conversion' reaction for which genomic DNA is treated with

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

sodium bisulfite under denaturing conditions. Cytosine residues get deaminated and converted to uracil leaving methylated cytosine moieties unaffected [59]. Identification of the resulting DNA sequence leads to a detection of converted and unconverted cytosine residues and subsequent identification of the prior methylation status of the nucleotide. The analysis deduces that cytosine residues within a CpG context were methylated if they have not been converted by bisulfite. Common to all direct investigation techniques are pitfalls leading to false positive methylation calls due to incomplete conversion reactions, degraded DNA caused by harsh conversion conditions and methylation in pseudogenes [60,61]. With the aid of HTS technologies do whole human genome m5C patterns become feasible [62]. Major challenges for whole-genome BS sequencing are the sequencing capacities and costs required, which are still relatively high. Thus, it is more practical to investigate only parts of the genome if one wants to gain insight into methylation patterns of mammals, especially if large numbers of samples need to be analysed.

DNA methylation analysis methods cannot easily be compared as many approaches have competing strengths and weaknesses. The number of samples, which can be analysed in parallel, the quantity of DNA and the desired resolution are the central decision points.

Taken together, the number of different HTS technologies is large, and each has its own advantages and disadvantages. The selection of the right technology for the research question investigated is crucial in making the most out of the enormous power HTS has for basic and clinical directions of research.

CLINICAL APPLICATIONS OF HIGH-THROUGHPUT DATA

A number of tumor biomarkers have been developed based on aberrant genomic and epigenomic (in particular DNA methylation) profiles and are used as diagnostic, predictive and prognostic tools. As predictive biomarkers they enable the stratification of patients in subgroups which should either receive a specific treatment or not [63,64]. Well-established examples include the sequencing of *KRAS* before Cetuximab treatment or the determination of the estrogen receptor status for Tamoxifen [63]. Only colorectal cancer patients with wild-type *KRAS* benefit from Cetuximab, an antibody directed against *EGFR*. Tamoxifen is an antagonist of the estrogen receptor, thus only hormone receptor-positive breast cancer patients will respond to the antagonist. An example for an epigenetic biomarker, hypermethylation of *MLH1*, a protein of the DNA mismatch repair, predicts sensitivity to cisplatin in colorectal cancer [64]. There are many more examples available for specific biomarkers either already in clinical use or in clinical testing. However, one commonality is that they are used in a hypothesis-driven manner: Patients with colorectal cancer are exclusively screened for *KRAS* mutations; or breast cancer patients are examined for the expression of the estrogen receptor. For these biomarker—directed decisions HTSs are not required. In contrast to these specific analyses new concepts of personalized oncology are emerging: Here the treatment of tumors is directed by their genome-wide molecular profile. This means that whole genome, whole exome and transcriptome profiles of a tumor patient will be generated within days. Afterwards a tumor board consisting of oncologists, radiologists, molecular geneticists and pathologists, and other related disciplines, will then determine the appropriate treatment of the patient based on his genetic information [65]. First proof-of-concept studies illustrate the effectiveness of this approach [23,66]. However, much more experience need to be acquired before a conclusion can be drawn. At the moment different strategies are explored how to streamline the analysis process and how quality controls can be set up. The transfer of the large amount of data into clinical usable information is far from easy and routinely preformed. We are right now at the boarder to clinical applications of HTS and it will be a question of time until large-scale genomic and epigenomic technologies are used as routine tools for the finding of diagnoses and for making therapy decisions.

Although data generation has increased exponentially, we are faced with new challenges to transform these data into useful models that help predicting the outcome of genomic aberrations (e.g. in the cancer field) and to develop novel diagnostic and therapeutic strategies. The challenges are enormous and require completely new types of infrastructure, for example data storage that dynamically adapts to data volumes not seen before. International quality and data format standardization are essential to compare data and utilize them in different modeling approaches. Another challenge will be to translate

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

the gained knowledge and models into day-to-day medical applications to finally benefit the patient with a personalized systems medicine. Systems biology technologies and modeling will contribute on the level of disease target detection, diagnostic and therapeutic approaches.

FUTURE PERSPECTIVES

The complexity of individual genomes is too large to be comprehended just by looking at the list of changes. There are two possibilities how to treat this large amount of data: Either by extraction of a subset of data, e.g. by looking at all drug targets and known biomarkers (following a conventional way of data analysis), or by application of computational modeling approaches. Different kinds of mathematical and computational modeling approaches exist [67]. Spatial models take the spatial organization of individual cells into account. Non-spatial stochastic models are adapted to specific purposes such as the role of chromosomal instabilities in tumor initiation or the role of Wnt signaling. Compartmental models investigate the transformation of cell types without taking their position within the colonic crypt into account.

Difficulties exist with the comparability of the models in regard to their functions included, e.g. investigation of cell differentiation, growth, turnover, etc. Even more difficult is their evaluation in regard to their capability of describing biological processes. Not only qualitative measurements, but also quantitative ratings would be desirable in order to improve the models and to assess their ability to predict disease stages or therapeutic strategies. First attempt are undertaken to integrate different subcellular levels (genomic, proteomic) combined with cellular (cell–cell communication) and tissue (movement and migration) levels within a unifying model. It is also desirable to add kinetic information. One way may be to integrate missing parameters, which are sampled from appropriate random distributions, e.g. by applying a Monte Carlo approach [68]. Further work on these models will refine relationships and will give predictive output with an improved degree of certainty. Difficulties arise because valid and rapid *in vivo* test systems for the output information are missing. In oncology it is important that an optimal treatment can be assigned without delay and that it is adapted to the individual's specific properties. Thus, even though the computational modeling approaches are just beginning to emerge, the hope is that the computer models will help with in *silico* predictions to optimize treatment strategies, which might further on have a significant impact on the outcome of the disease.

Key Points

- Most often tumor patients are treated regardless of alterations within their genomes and epigenomes leading to a large number of patients treated without benefit.
- Recent progress in HTS technologies has paved the way for personalized medicine: Tumors are sequenced and depending on their mutations, structural variations and epigenetic modification individual drug combinations are proposed. First clinical studies encompassing this concept are under way.
- The development of HTS technologies has initiated a large number of different applications with several pros and cons. Some of them are on the border to routine clinical usage; others are destined for basic research. Knowledge of the different technologies is key for understanding the concept of personalized medicine.

ACKNOWLEDGEMENTS

We would like to thank Axel Fischer and Dr Martin Kerick for bioinformatics support.

FUNDING

This work was supported by the Federal Ministry of Education and Research [grant numbers BMBF HNPCCSys 0316065E, EPITREAT 0316190A-E].

References

1. Timmermann B, Kerick M, Roehr C, *et al*. Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLoS One* 2010;**5**:e15661.
2. Schweiger MR, Hussong M, Rohr C, *et al*. Genomics and epigenomics of colorectal cancer. *Wiley Interdiscip Rev Syst Biol Med* 2013;**5**:205–19.
3. Mardis ER, Ding L, Dooling DJ, *et al*. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 2009;**361**:1058–66.
4. Macconaill LE, Garraway LA. Clinical implications of the cancer genome. *J Clin Oncol Off J Am Soc Clin Oncol* 2010;**28**: 5219–28.
5. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst* 1960;**25**:85–109.
6. Heisterkamp N, Stam K, Groffen J, *et al*. Structural organization of the BCR gene and its role in the Ph' translocation. *Nature* 1985;**315**:758–61.
7. Druker BJ. Translation of the Philadelphia chromosome into therapy for CML. *Blood* 2008;**112**:4808–17.

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

8. Ball MP, Thakuria JV, Zaranek AW, *et al*. A public resource facilitating clinical use of genomes. *Proc Natl Acad Sci USA* 2012;**109**:11920–27.
9. Abecasis GR, Auton A, Brooks LD, *et al*. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491**:56–65.
10. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genom Human Genet* 2008;**9**:387–402.
11. Margulies M, Egholm M, Altman WE, *et al*. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;**437**:376–80.
12. Albert TJ, Molla MN, Muzny DM, *et al*. Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 2007;**4**:903–5.
13. Gnirke A, Melnikov A, Maguire J, *et al*. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009;**27**:182–9.
14. Choi M, Scholl UI, Ji W, *et al*. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009;**106**:19096–101.
15. Hodges E, Xuan Z, Balija V, *et al*. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007;**39**: 1522–7.
16. Porreca GJ, Zhang K, Li JB, *et al*. Multiplex amplification of large sets of human exons. *Nat Methods* 2007;**4**:931–6.
17. Clarke J, Wu HC, Jayasinghe L, *et al*. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 2009;**4**:265–70.
18. Greenleaf WJ, Block SM. Single-molecule, motion-based DNA sequencing using RNA polymerase. *Science* 2006;**313**: 801.
19. Sugiyama S, Yoshino T, Tsukamoto K, *et al*. Application of scanning probe microscopy to genetic analysis. *Jpn J App Phys* 2006;**45**:2305–9.
20. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet* 2010;**19**: R227–40.
21. Murray IA, Clark TA, Morgan RD, *et al*. The methylomes of six bacteria. *Nucleic Acids Res* 2012;**40**:11450–62.
22. Rowley JD. Letter: a new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 1973; **243**:290–3.
23. Le Tourneau C, Kamal M, Tredan O, *et al*. Designs and challenges for personalized medicine studies in oncology: focus on the SHIVA trial. *Targeted Oncol* 2012;**7**: 253–65.
24. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 2010;**11**:191–203.
25. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell* 2012;**150**:12–27.
26. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 2012;**26**:11–24.
27. Segal E, Widom J. From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat Rev Genet* 2009;**10**: 443–56.
28. Farnham PJ. Insights from genomic profiling of transcription factors. *Nat Rev Genet* 2009;**10**:605–16.
29. Rando OJ, Chang HY. Genome-wide views of chromatin structure. *Annu Rev Biochem* 2009;**78**:245–71.
30. Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 1998;**20**:116–7.
31. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007;**447**:433–40.
32. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983;**301**:89–92.
33. Gama-Sosa MA, Slagel VA, Trewyn RW, *et al*. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res* 1983;**11**:6883–94.
34. Borno ST, Fischer A, Kerick M, *et al*. Genome-wide DNA methylation events in TMPRSS2:ERG fusion negative prostate cancers implicate an EZH2 dependent mechanism with miRNA-26a hypermethylation. *Cancer Discov* 2012; **2**(11):1024–35.
35. Banerjee HN, Verma M. Epigenetic mechanisms in cancer. *Biomark Med* 2009;**3**:397–410.
36. Beck S, Rakyan VK. The methylome: approaches for global DNA methylation profiling. *Trends Genet* 2008;**24**: 231–7.
37. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 2010;**11**:191–203.
38. Lister R, Ecker JR. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* 2009;**19**: 959–66.
39. Pomraning KR, Smith KM, Freitag M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* 2009;**47**:142–50.
40. Rakyan VK, Down TA, Thorne NP, *et al*. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res* 2008;**18**:1518–29.
41. Ball MP, Li JB, Gao Y, *et al*. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 2009;**27**:361–8.
42. Weber M, Davies JJ, Wittig D, *et al*. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005;**37**:853–62.
43. Keshet I, Schlesinger Y, Farkash S, *et al*. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* 2006;**38**:149–53.
44. Cross SH, Charlton JA, Nan X, *et al*. Purification of CpG islands using a methylated DNA binding column. *Nat Genet* 1994;**6**:236–44.
45. Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab Invest J Tech Methods Pathol* 2005;**85**:1172–80.
46. Down TA, Rakyan VK, Turner DJ, *et al*. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* 2008;**26**:779–85.
47. Brunner AL, Johnson DS, Kim SW, *et al*. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res* 2009;**19**:1044–56.
48. Toyota M, Ho C, Ahuja N, *et al*. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* 1999;**59**: 2307–12.

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013

49. Oda M, Glass JL, Thompson RF, *et al*. High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res* 2009;**37**:3829–39.
50. Berman BP, Weisenberger DJ, Laird PW. Locking in on the human methylome. *Nat Biotechnol* 2009;**27**:341–2.
51. Carr IM, Valleley EM, Cordery SF, *et al*. Sequence analysis and editing for bisulphite genomic sequencing projects. *Nucleic Acids Res* 2007;**35**:e79.
52. Korshunova Y, Maloney RK, Lakey N, *et al*. Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res* 2008;**18**:19–29.
53. Wang RY, Gehrke CW, Ehrlich M. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res* 1980;**8**:4777–90.
54. Hodges E, Smith AD, Kendall J, *et al*. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* 2009;**19**: 1593–1605.
55. Deng J, Shoemaker R, Xie B, *et al*. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol* 2009;**27**:353–60.
56. Li JB, Gao Y, Aach J, *et al*. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res* 2009;**19**:1606–15.
57. Meissner A, Gnirke A, Bell GW, *et al*. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 2005;**33**: 5868–77.
58. Meissner A, Mikkelsen TS, Gu H, *et al*. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 2008;**454**:766–70.
59. Frommer M, McDonald LE, Millar DS, *et al*. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA* 1992;**89**:1827–31.
60. Esteller M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* 2002;**21**:5427–40.
61. Warnecke PM, Stirzaker C, Melki JR, *et al*. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* 1997; **25**:4422–6.
62. Eckhardt F, Lewin J, Cortese R, *et al*. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006;**38**:1378–85.
63. Weng L, Zhang L, Peng Y, *et al*. Pharmacogenetics and pharmacogenomics: a bridge to individualized cancer therapy. *Pharmacogenomics* 2013;**14**:315–24.
64. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* 2012;**13**: 679–92.
65. Levy MA, Lovly CM, Pao W. Translating genomic information into clinical medicine: lung cancer as a paradigm. *Genome Res* 2012;**22**:2101–8.
66. Roychowdhury S, Iyer MK, Robinson DR, *et al*. Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med* 2011;**3**:111ra121.
67. De Matteis G, Graudenzi A, Antoniotti M. A review of spatial computational models for multi-cellular systems, with regard to intestinal crypts and colorectal cancer development. *J Math Biol* 2013;**66**:1409–62.
68. Wierling C, Kuhn A, Hache H, *et al*. Prediction in the face of uncertainty: a Monte Carlo-based approach for systems biology of cancer treatment. *Mutat Res* 2012;**746**: 163–70.

Downloaded from http://bfg.oxfordjournals.org/ at Universidade de SÃ£o Paulo on August 1, 2013