High-throughput sequencing in clinical oncology: from past to present

Mery Benoîte¹, Vallard Alexis², Rowinski Elise³, Magne Nicolas³

¹ Department of Medical Oncology, Lucien Neuwirth Cancer Institute, Saint Priest en Jarez, France
² Radiotherapy Department, Lucien Neuwirth Cancer Institute, Saint Priest en Jarez, France

Summary
The war on cancer remains a major challenge. One of the obstacles to additional progress is the complexity of the mechanisms underlying the disease. Cutting-edge technologies and computing tools have led to whole genome sequencing and an integrated and inclusive omic approach to cancers, from accurate molecular signatures of tumours to impressive progress in the field of next-generation sequencing (NGS). Genomic data may foster strategies for new drug development in addition to a better understanding of cancer genesis, opening a new era in oncological clinical practice. This review discusses the development of genomics approaches in cancer research and the potential of genomics for precision medicine, as well as clinical implications and remaining challenges.

Keywords: genomics, cancer, next-generation sequencing, omics technologies

Introduction
Despite the huge progress and efforts in cancer research to date, the war on cancer remains a major challenge and a never-ending battle. The global burden of cancer will increase to 22 million new cases each year by 2030 [1, 2]. One of the major obstacles to additional progress in the struggle against cancer is the complexity of the mechanisms underlying the disease. Unravelling cancer’s complexity by harnessing the power of cutting-edge technologies and computing tools has become a major objective of cancer research [3]. To this end, whole-genome sequencing and new measurement technologies let us pinpoint new therapeutic attack strategies [4]. The detection of alterations in gene-based structure and function could have major consequences for the diagnosis, prognosis and treatment of cancer [5]. Genome-analysis techniques provide data about genetic abnormalities, epigenetic changes and transcriptome modifications. Such work has led to a more integrated and inclusive approach to cancers and the accurate molecular characterisation of tumours [6]. Indeed, while cancers have long been classified according to clinical and histological signs, gene expression profiling leads to a new, genetic taxonomy of cancers according to their different molecular signatures. Such analyses are supposed to untangle the complex nature of cancer biology and to overcome the wide diversity of tumours through the identification of genes and pathways that could be deregulated to better detect and manage the disease [7]. Much effort is currently focused on the identification of molecular subtypes. These allow the best treatment option to be chosen, the risk of relapse to be predicted and therapeutic response to be assessed through a new, molecular taxonomy of cancer [8].

Major improvements in biotechnology and bioinformatics have been achieved over the past two decades, allowing the development of omic signatures of several molecular subtypes of cancer, each associated with different outcomes [9]. For instance, in breast cancer, which remains a devastating disease, with 5-year survival rates of metastatic breast cancer below 25%, several gene signatures have been devised which predict the risk of relapse, the response to specific treatments, and the benefit from adjuvant chemotherapy [10] [11]. As well as shedding light on cancer heterogeneity and enabling adequate treatment, genomics profiling may lead to a better understanding of the metastatic process and foster other strategies for new drug development [12]. Therefore, several programs combining informatics, biomedical and biotechnology data aspire to create new ways to diagnose and treat cancers. The assessment of novel bioinformatic methodologies in support of genomics approaches may be the beginning of a new era in oncological clinical practice [13].

This review discusses the development of genomics approaches in cancer research and the potential of genomics for precision medicine, as well as clinical implications and remaining challenges.

Omics approaches in the cancer landscape: historical background
The Cancer Genome Atlas (TCGA) project was launched in 2005 as an effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing. The National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) jointly designed this major undertaking to improve the diagnosis, treatment and prevention of cancer.
through a complete identification of the molecular aberrations associated with cancer. The advent of polymerase chain reaction (PCR) and next-generation sequencing (NGS) have expanded the possibilities for genome analysis, methylation analysis and transcriptional profiling. Each biomolecular class, including RNA, DNA, proteins and lipids, has become the cornerstone of its own study, defining different omics approaches [14] [5]. The term genomics refers to DNA studies conducted through single-gene testing or multi-gene testing. The identification of all common cancer-related genomic aberrations would allow the assessment of tumours through their specific signatures. Genome sequencing data from TCGA project are now accessible for several tumours, including breast cancer, and should provide tools to assign the most appropriate therapeutic strategies in clinical practice [15]. Indeed, more than 11,000 patient samples and 33 tumour types are included in this database, with valuable data on key genomic changes providing a major contribution to the understanding of cancer genomics [16, 17]. Epigenetic features of cancer are the focus of another major project, the Roadmap Epigenomics Mapping Consortium, whose main objective is to generate a public human epigenomic database thanks to NGS technologies. The study of DNA methylation, histone modifications, RNA transcripts and chromatin accessibility aims to provide a reference of normal epigenomes that could help with the identification of relevant differences in comparison with epigenetics in cancer cells [18]. This work has led to the identification of the major genes and pathways involved in cancer genesis, as well as molecular events conferring oncogenic properties. It has also unveiled the full complexity of tumour heterogeneity between even histologically similar tumours [19]. Several other programs have since been formed, and the International Cancer Genome Consortium (ICGC) was created to coordinate a large number of research projects that have the common aim of comprehensively elucidating the genomic changes present in various forms of cancer. The main goal is to generate data for the entire research community on genomic abnormalities, including somatic mutations, abnormal expression of genes and epigenetic modifications, in tumours from different types of cancer. The ICGC data portal currently contains more than 24 cancer projects and copes with high computational volumes, high complexity and high heterogeneity of data [20]. The conceptual evolution of oncological clinical practice through high-throughput sequencing is described in figure 1.

High-throughput technologies: pushing genomics knowledge

Improvements in high-throughput DNA sequencing, along with computational and algorithmic advances, have led to tremendous progress in accessing omics data. NGS technologies provide accurate analyses of tumours’ pan-genomic profiles at genomic and transcriptomic levels, enabling the possibility of designing the right cancer drug for the right patient [21]. DNA sequencing technologies for human genomic medicine emerged more than 30 years ago and have undergone major technical improvements. Whereas the first automated genome sequencing machine (AB370), launched in 1987, was able to detect 96 bases simultaneously and 500,000 bases a day, the current AB3730 can detect up to 2.88 million bases per day [22, 23]. The first high-throughput NG sequencer, known as Illumina’s MiSeq Dx, was launched in 2013 and paved the way for the development of new genome-based tests. Several NGS sequencing techniques, including whole genome sequencing, whole exome sequencing, transcriptome sequencing and targeted panel sequencing, have been designed since. NGS has become increasingly reliable, cheaper and faster, and allows the identification of crucial somatic mutations thanks to advances in nanotechnology and continued developments in bioinformatics [24]. Whole-genome sequencing sequences the complete genome of a sample, whereas whole-exome sequencing sequences the protein-coding genes. This approach aims to assess the full sequence of cancer-related gene panels. In another approach, certain regions of selected genes can be sequenced, focusing on cancer gene “hotspot” regions with recurrent mutations. The common goal remains the ability to perform almost any type of analysis in order to identify potential therapeutic targets at genomic and transcriptomic levels which can be used to classify tumours and predict outcomes [25]. Indeed, tumours are biologically diverse and contain complex genetic alterations. Well over 100 genes have been found which are frequently mutated in one kind of cancer or another, and the sheer number of cancer genes has frustrated attempts to deduce which ones are necessary and sufficient to cause the disease. Can-
Cancer genomes are highly rearranged, with several large-scale mutations including translocations, inversions, fusions and copy number changes [26]. In this respect, NGS could partly solve the partial map problem. Specific tools have been designed to process NGS data. Notably, several mapping software programs allow the analysis of tumour and normal genome pairs simultaneously, so that germline and somatic mutations can be distinguished [27] [28]. NGS data create huge bioinformatics challenges including storage, transmission, manipulation and analysis, making the downstream processing of these data a daunting task. DNA sequencing strategies are detailed in table 1.

**Classification and new taxonomy of tumours**

While NGS allows lower cost, higher-throughput genome sequencing and opens up exciting new approaches in terms of personalised cancer therapy, the huge number and variety of genetic aberrations found in cancer creates substantial analytical complexity. However, while the abundance of information generated may complicate decisions regarding cancer therapy, genome sequencing has clearly led to the development of novel therapeutic targets through the identification of driver mutations [29]. The relevance and therapeutic consequences of specific mutations are assessed through biomarker-driven research that recruits selected patients into clinical trials to assess the efficacy of targeted therapies [30, 31]. Chen et al. have given an overview of the NGS work flow, from DNA isolation to the sequencing and data analysis which determine therapy. The ultimate goal is the identification of alterations to actionable genes which are associated with their functional impact and have therapeutic implications [32]. Glioblastoma, for instance, was the first cancer type to be sequenced and deposited into TCGA. The analysis of its transcriptome and genome signatures allowed the classification of these tumours into proneural, neural, classical and mesenchymal subtypes [33]. NGS clearly improves our ability to refine the nosology of tumours. Furthermore, mutations in chromatin-modifier genes, as well as in genes for which targeted therapies aimed at other diseases have been developed, including BRAF, FGFR1, FGFR2 and FGFR3, have been found in most glioblastoma tumours, demonstrating the potential clinical impact of such NGS data [34]. Potential targets for therapy were identified among a panel of 130 genes in more than 70% of glioblastomas [35]. Additionally, new drugs such as crizotinib have advanced to late-phase clinical trials for their anticancer effects on non-small cell lung cancers carrying EML4-ALK translocations [36]. Other therapies that target recurrent alterations, including EGFR mutations, MET amplification and ROS1 fusions, have also been developed in lung adenocarcinoma [37]. In fact, molecular profiling of lung cancer has become crucial for predicting the response to targeted therapies [38]. A landscape of driver mutations in melanoma has also been established, and BRAF V600 mutations, which are present in 50% of melanomas, predict clinical efficacy of RAF inhibitors such as vemurafenib, a tyrosine kinase inhibitor [39]. Furthermore, the ability of NGS to identify MEK mutations associated with resistance to vemurafenib may help to guide patient treatment. Similarly, the discovery of mutations in other driver oncogenes, such as NRAS, GNAQ and GNA11, could help to decide whether to use immunotherapy with CTLA4 or PD1 inhibitors [38]. Analyses to identify drug targets and drug-resistance mutations are well documented for numerous cancer types. In a prospective study, Malapelle et al. recently identified mutations in KRAS, NRAS and BRAF which are associated with resistance to anti-EGFR therapy [40]. In head and neck cancer, distinct mutation profiles and targetable mutations could allow the identification of subgroups of patients with poor outcomes after adjuvant chemo radiation [41]. In triple negative breast cancer, integrated analysis of differentially expressed genes and pathways is ongoing. This aims to identify novel target molecules for therapy, as it remains a heterogeneous disease characterised by an aggressive phenotype and reduced survival [42]. A recent study using several NGS analyses on 439 patients with various types of cancer showed that 20% of patients had an actionable mutation targeted by on-label drugs, whereas 50% of patients had actionable mutations targeted by an off-label but approved drug [43]. Another prospective study of 800 patients demonstrated that more than 60% of patients had clinically relevant mutations, with 26% displaying a mutation with therapeutic implications [44]. Finally, a much more consequent study of 2,221 patients also concluded that relevant mutations were identified in 76% of cases [45–48]. These initial studies demonstrate the role of NGS in identifying a huge range of somatic alterations in cancer, leading to major therapeutic features.

**Future omics directions: challenges and perspectives**

Several clinical trials to assess the mutational profiles of cancer patients, especially metastatic ones, have been launched recently, leading to relevant results. In a Korean prospective trial which enrolled 407 patients between 2013 and 2014, 84% of patients had at least one aberration detected. When patients were matched to molecularly targeted agents, the response rate was significantly higher in the mutation-matched than in the non-matched treatment group [49]. The currently ongoing National Cancer Institute-Molecular Analysis for Therapy Choice (NCI-MATCH) trial is a national, signal-finding, precision medicine study that relies on genomic assays to screen and

<table>
<thead>
<tr>
<th>Whole genome sequencing</th>
<th>Whole exome sequencing</th>
<th>Targeted panels sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most comprehensive and unbiased examination of the cancer genome</td>
<td>Detection of 85% of disease-causing mutations</td>
<td>Rapid and reliable identification of the most common molecular aberrations</td>
</tr>
<tr>
<td>Discovery of new mutations</td>
<td>Detection of unknown variants</td>
<td>Amplicon-based or hybridisation capture-based NGS</td>
</tr>
<tr>
<td>Most expensive</td>
<td>High bioinformatics level</td>
<td>NGS</td>
</tr>
</tbody>
</table>
enrol patients with relapsed or refractory cancer after standard treatments using validated NGS processes [50]. The MPACT trial (molecular profiling-based assignment of cancer therapy; NCT01827384) was one of the first randomised studies to assess the response of patients treated with a drug matched to the mutational profile obtained through NGS analysis [51]. These underlying premises are clearly opening up a new era of genomic medicine. Besides the main goal of novel therapeutic targets specifically adapted to the features of a tumour, another important objective of omics technologies is to re-examine adjuvant strategies, especially in breast cancer, by finding molecular signatures to identify patients who need close monitoring rather than aggressive chemotherapy. Indeed, some gene signatures which predict the risk of relapse, the risk of distant metastasis or the response to specific therapies have been identified. In patients with breast cancer and a lymph-node-negative status, the 21-gene Oncotype Dx signature, the 70-gene MammaPrint signature and the 76-gene Rotterdam signature currently identify patients at high risk of developing distant metastases within five years. They show comparable performance, despite few genes being shared between them. The oncotype recurrence score in particular provides data concerning the benefit of adjuvant chemotherapy [52–55]. Such screening strategies offer a more sensitive detection of metastasis-prone cancers than only the immunohistochemical tumour characteristics used to classify breast cancer subtypes until now [56]. Improvements in outcomes may come either from new treatment strategies or from the identification of aggressive tumours which require dense and sharp therapeutic treatment, before metastatic dissemination. Other additional applications of NGS are still under development and include the evaluation of circulating tumour cells or free-plasma DNA to detect early relapse or residual cancer [57]. Once genomic alterations have been identified, PRC assays could be used to detect circulating tumour cells or free plasma DNA harbouiring the same alterations, allowing the assessment of disease status, drug responsiveness and relapse. Accurate monitoring of the mutations harboured by a tumour may also allow the screening of new alterations that occur under treatment pressure, and thereby give insights into the mechanisms of acquired resistance [58]. Omics approaches are thus useful and promising tools to overcome the obstacle of tumour heterogeneity [59]. Another application of NGS is to improve cancer diagnosis in cases where histological identification is challenging through high sensitivity and specificity analysis, leading to the determination of the origin of the tumour [60]. Beyond genomics, the development of pharmacogenomics will also aid study of the association between genetic variation and anti-cancer drug response, as well as the development of predictive tools to anticipate tolerance and efficacy of treatment strategies, opening a totally new vision for therapeutics [61]. The prospects and challenges of NGS are summarised in figure 2.
Conclusion

Omics technologies have brought unprecedented advances in our understanding of the biology of cancer through the identification of relevant germline and somatic mutations. This has major clinical implications and paves the way to individualised medicine. Based on the massively parallel sequencing of DNA, with subsequent data processing and sequence alignment, NGS allows the simultaneous analysis of multiple genetic aberrations, including single nucleotide variants, small insertions/deletions, copy number variants and complex genomic rearrangements. The vast amounts of data generated by NGS have broadened our understanding of cancer. The concepts of omics, ranging from basic genomics to integrated systemomics, provide new insights into the genomics of tumour cells, which have subsequently led to crucial improvements in anti-cancer drugs [62]. In an article entitled “Cancer research: quo vadis – to war?”, Wheatley et al. emphasise that every cancer is as unique as the patient, and that each case must be seen as a challenge in itself. This is why there is such interest in omics technologies that open the way for a new era of genetic investigations which make cancer an idiosyncratic entity that should be treated as such [63].

The overall clinical potential of omics approaches in cancer research is already well known, but its utility probably goes beyond the hype, notably through the individualisation of treatment strategies. Long-term outcomes of better survival and clinically relevant benefits for patients are still being explored in clinical trials. Molecular databases are currently being updated to make it easier to understand the genetic profiles and to therefore provide specific and unique treatment recommendations. This novel paradigm will probably shape and enhance our understanding of tumour biology in the decades to come and be the cornerstone of the development of new anti-cancer drugs. The continued participation and collaboration of clinical oncologists, cancer researchers, computational biologists, bioinformaticians and, most importantly, patients remains the sine qua non condition for steady progress in genomic medicine. In default of the entire war, one part of the battle in the daunting fight against cancer may be about to be won [64].

Disclosure statement
No financial support and no other potential conflict of interest relevant to this article was reported.

References


