

Cancer genomes: where do we go from here?

Jorge Meléndez-Zajgla^{1*} and Vilma Maldonado Lagunas²

¹Subdirección de Basic Research, Instituto Nacional de Medicina Genómica, México, D.F., ²Subdirección de Medical Research, Instituto Nacional de Medicina Genómica, México, D.F.

Abstract

In recent years, an exponential growth of knowledge on the molecular bases of cancer has occurred. Particularly, the creation of important initiatives to dilucidate the genomes of several types of cancer has allowed, for the first time, to have complete catalogues of most mutational events, which opens important possibilities for oncology and public health. The present opinion offers a perspective on the advances and future direction in Mexico. (Gac Med Mex. 2014;150:553-8)

Corresponding author: Jorge Meléndez-Zajgla, jmelendez@inmegen.gob.mx

KEY WORDS: Genomic medicine. Cancer. New generation sequencing. Cancer genomics.

Introduction

Cancer comprises a group of diseases that share a molecular pathogenic process. In this process, the acquisition of genetic, genomic and epigenomic alterations, together with interaction with tissue microenvironment, lead to the malignant phenotype. In turn, this phenotype depends not only on the tissue of origin and biologic particularities of the patient, but on specific alterations of the genome, resulting of a Darwinian evolution process in cancer precursor cells. Since this process is, in general terms, random, most part of the genomic changes has no relevance and is not perpetuated in the genome during the carcinogenic process and, therefore they are known as “passenger” modifications. Only a minority of the so-called driver modifications are selected and preserved during carcinogenesis. For these reasons, there is not only great heterogeneity among tumors of different patients, but also among a tumor’s own cancerous cells. This genomic heterogeneity reflects in phenotypic diversity that gives opportunity for acquired chemotherapy-resistance and the motor that drives the invasion and metastasis processes to occur.

From the hope, a few years ago, of finding a series of alterations common to all tumors, currently we have focused on the reality of an existing, much more varied genomic landscape, with a few high mountains (e.g., *p53* and *Ras*), several small mountains (e.g., *NF1* and *2*, *PDGFRA*) and many hills (e.g., *Caspase-8*, *ACVR18*) in a valley of irrelevant and varied mutations in the different tumor types. However, even these mountains and hills allow for the idea of the existence of a tumor genomic signature to be continued, which allows for diagnostic and prognostic tests and, most importantly, for targeted therapies to be developed.

The idea of targeted therapy is not new, having started with the search for antagonists to hormone receptors and, in particular, to active kinase inhibitors (ABL), resulting of the fusion present in the Philadelphia chromosome. However, the methods to detect genes responsible of cancer and the time required to have proofs of principle were based on complicated, long and individual molecular assays. With the advent of high-volume technology for screening and sequencing, these times have been dramatically reduced (Fig. 1).

Importance of the catalogue of alterations in cancer

Based on the notion that a specific and reduced group of genetic, genomic and epigenomic alterations lead to the malignant phenotype, over the past few years, many investigators have devoted their time on

Correspondence:

*Jorge Meléndez-Zajgla

Subdirección de Investigación Básica

Instituto Nacional de Medicina Genómica

Periférico Sur, 4809

Col. Arenal de Tepepan, Del. Tlalpan, C.P. 14610, México, D.F.

E-mail: jmelendez@inmegen.gob.mx

Modified version reception: 16-10-2013

Date of acceptance: 26-03-2014

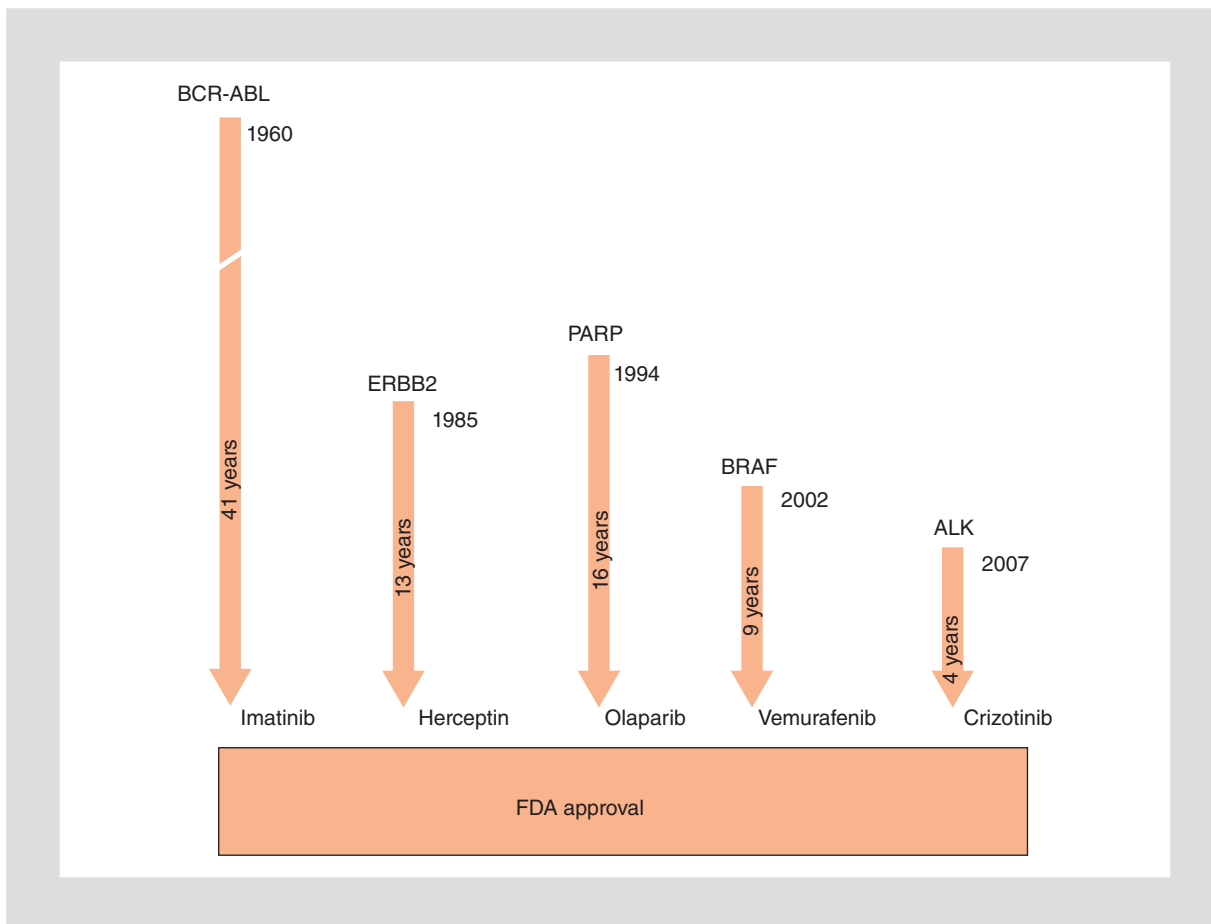


Figure 1. Historic graph of the time elapsed since the identification of genomic alterations of genes responsible of cancer and the approval of a drug targeted against them by the U.S. Food and Drug Administration.

trying to create a catalogue of these. In addition to basic interest on knowing the carcinogenic process, identifying these alterations has allowed and continues to allow for diagnostic and prognostic tests to be created and, most important, identification of new therapeutic targets for targeted therapy (Fig. 2). Since the beginning, this work was arduous, given that the molecular tools that were employed were low-volume and limited to one or a few genes at one time. With the human genome project and the drive of the creation of new high-volume tools, the strategy was deeply transformed. Now it is possible to conduct genomic studies where all the structural alterations and genic mutations present in a particular tumor are analyzed. In view of this, dozens of high-impact articles have been published globally describing most of these anomalies in small groups of patients, including some Mexicans¹⁻⁸. However, the picture is still incomplete, given that, for statistical reasons, a much larger group of tumor samples has to be sequenced. Furthermore, infrequent

tumors are still missing, non-coding regions and epigenomes have not been adequately covered, there is not much information on non-caucasian populations, etc. Therefore, it is required to keep generating information of this nature, through close collaboration between clinical and basic investigators; in addition, joining efforts and establishing co-operation between national and foreign institutions is necessary for the benefit of our patients.

The first benefit of these catalogues is the possibility of profoundly improving the design of clinical trials for new oncologic drugs. Knowing the molecular subgroups and biological idiosyncrasy of each patient will allow for statistical approaches to determine the effectiveness of each drug to be subdivided and improved. This should increase the number and effectiveness of therapies available for our patients and decrease times and costs associated with the therapeutic process.

With these catalogues, now it is possible to design lower-cost projects intended to determine the usefulness

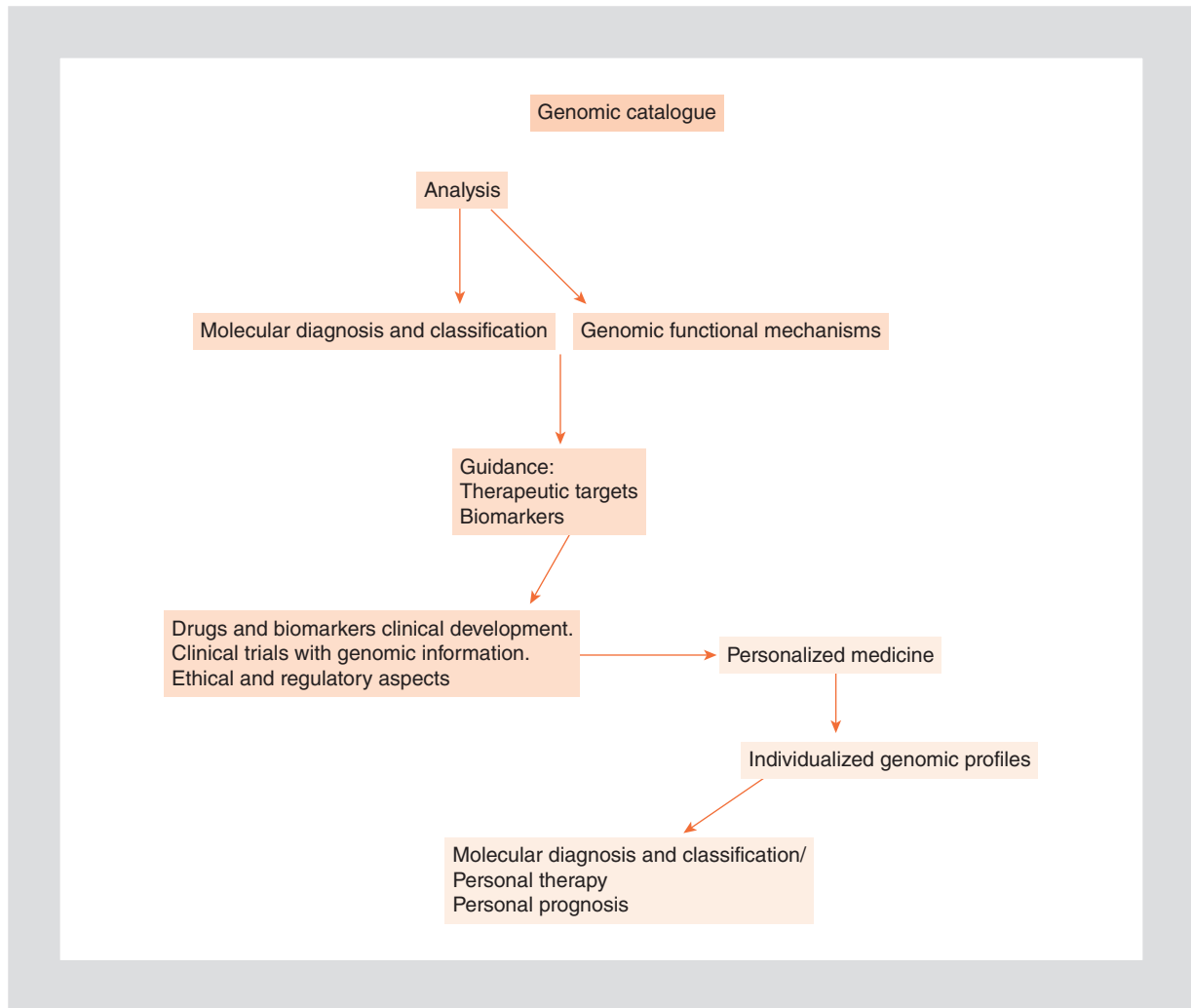


Figure 2. Phases of genomic discovery. Based on the catalogue of genomic alterations it is possible to perform assays for molecular diagnosis and classification of each neoplasm, as well as to perform functional research on affected mechanisms. This information is useful for the search of therapeutic targets and biomarkers that enable oriented clinical trials that will lead to personalized or precision medicine. In this last phase, personalized genomic profiles, both of the tumor and the patient, will not only allow for more accurate molecular diagnosis and classification, but also for much more personalized prognoses and therapies.

of molecular markers for diagnosis, prognosis and therapeutic prediction, which can be accomplished in a directed and complete manner.

Some achievements that are starting to emerge from this type of high-volume strategy are the development of diagnostic tests, such as Mammaprint® and Oncotype® for breast cancer, the establishment of therapeutic targets with the use of biomarkers, such as *KIT* and *PDGFRA* mutations for the use of imatinib or lapatinib in gastrointestinal stromal tumors; mutations in *EGFR* for the use of gefitinib and erlotinib in the case of lung cancer; mutations in *BRCA1* to indicate the use of olaparib in breast cancer; *BRAF* mutations for the use of vemurafenib in the case of melanoma, *ALK* fusions for crizotinib, etc. There is plenty of potential

for personalized and combined use of these drugs in the adjuvant setting.

However, it is necessary for these achievements to be considered in the light of the challenges that have to be contemplated for their application. In the first place, there is the cost of these drugs, which cannot be absorbed by the national health system at this moment. However, with the entry into force of the Mexican Official Standard on pharmacogenomics and the Official Standard 177 amendments by the Federal Commission for Protection Against Sanitary Risks (COFEPRIS – *Comisión Federal para la Protección contra Riesgos Sanitarios*), the possibility is opened to the production of interchangeable generic drugs in the case of drugs with expired patent, which in this sense is a

start to have more affordable drugs for everyone. The second challenge is to establish enough oncology centers qualified to perform the biomarker analyses required for these tests. There are different models in first-world countries, which range from centralized, federally supported models, as in the case of France, to disseminated, commercial models, as in the case of the U.S.A.

Is a change of paradigm towards precision medicine possible in oncology?

In order to capitalize on the use of new technologies, we have to operate a change in the paradigm of oncology. We need to pass from medicine based exclusively on pathological diagnoses to those deeply supported on mechanisms; from grouping by organ to molecular and biological subclassification; from standardized to personalized treatment and, with all this, to early detection and intervention, using methods to determine the relative risk⁹.

As we mentioned previously, the first step for this is to start with the creation of informed catalogues of alterations in cancer. This involves high-volume tools, particularly the use of new generation sequencing. In this type of assays, libraries are produced of informational molecules to be studied, either total DNA, total RNA or fractions enriched in functional elements such as exons, methylated regions or regions bound to specific transcription factors, small RNAs, non-coding RNAs, panels of genes of interest, etc. There are diverse enrichment strategies, either using polymerase chain reaction, specific probes to purify regions, fractioning by size or presence of specific genomic marks (e.g., polyadenine tails). Based on these libraries, equipments with different platforms are used (Illumina, SOLiD, 454, etc.), which, with few exceptions, produce short readings of sequences read in an unordered manner. This approach (shot-gun) makes sequencing easier, but complicates subsequent arrangement. This can be relatively trivial when it comes to sequencing of gene panels, but it becomes a limiting step in the case of complete genomes, which require using high-performance computing equipment and experience on bioinformatics. In view of these considerations, in addition to the required storing space and associated costs, most part of efforts to sequence cancer genomes has taken place within large consortiums and alliances between investigation groups, among which two deserve to be mentioned: The Cancer Genome Atlas (TCGA)¹⁰ of American origin and the International

Cancer Genome Consortium (ICGC)¹¹, a global initiative. From the efforts by all these international groups, we have found many new data of interest, such as a large number of genes responsible for cancer, new mutational mechanisms, cell processes not previously known to intervene in carcinogenesis and new therapeutic targets. These studies also reinforced the idea of the high heterogeneity present in this group of diseases, not only between neoplasms originated in different tissues, but also between patients and even cancerous cells populations.

In the ICGC context, Mexico has participated in four projects: breast cancer⁴, head and neck³, non-Hodgkin lymphoma⁸ and cervical cancer. In the case of breast cancer, we found new genes involved with this group of neoplasms such as *CBFB* and *RUNX1*, whose alterations had only been described in lymphohematopoietic tumors. The head and neck cancer project demonstrated the existence of new genes with conducting mutations, involved with epithelial differentiation, such as *NOTCH1*, *IRF6* and *TP63*. Finally, non-Hodgkin lymphomas analyzed in the initiative showed novel conducting mutations in *MEF2B*, *MLL2*, *BTG1*, *GNA13*, *ACTB*, *P2RY8*, *PCLO* and *TNFRSF14* genes, in addition to a new and interesting possible mechanism for *Bcl-2* gene alteration, mediated by an hypersomatic mutation in the immunoglobulin H locus, probably due to the action of cytidine deaminase. In the case of cervical cancer, several non-mutated genes not previously described were also found¹². It should be mentioned that only in the case of breast and cervical cancer, samples of Mexican patients were analyzed. In the remaining projects, the participation of national groups was focused on collaborative analysis work. Based on these projects and the generated experience, several international initiatives, where Mexico takes part, have been created¹³, in addition to national initiatives, including projects in Mexico City and Monterrey that are analyzing different neoplasms such as soft tissue sarcomas, pediatric tumors, pulmonary, pancreatic, cervical and testicular carcinomas, etc^{14,25}.

The data generated by genomic analyses, deposited both in the TCGA and the ICGC, can now be used to lead to the development of diagnostic, prognostic and predictive biomarkers for targeted therapy, to ideally arrive to a personalized or more accurate medicine. In this type of medicine, tumor and germinal genomic profiles that allow for personalized diagnosis, prognosis and therapy have to be obtained. In several countries, the first initiatives to create specialized laboratories for this purpose are already underway¹⁶⁻¹⁹.

The most important challenges to achieve this advance are several and include:

- Availability of the sample. For many tumors it is clearly difficult to obtain a sample of sufficient size. Due to inaccessibility, use of neoadjuvancy, ethical reasons to perform incisional biopsies, etc., in the initial studies, where no direct benefit is derived for the patients, sufficient tumor material is difficult to obtain. An alternative option is the use of archived material (paraffin-embedded specimens). Although ideal, there are still several methodological problems that make its routine use difficult and expensive²⁰. In particular, problems are centered on the recovery of good quality nucleic acids and in sufficient quantity for analyses^{21,22}. In addition to investigative efforts in this regard, the development of pathologic samples collection standard procedures will help to solve the problems.
- Available platforms. Type and implementation. There are several new- (second) generation sequencing platforms, with those provided by Illumina (GAllx, Hi-Seq, MiSeq), Life Technologies (SOLiD, Ion PGM, Ion Proton), 454 (GS FLX, GS FLX junior), etc. standing out²³. The decision on the use of each one of these is based on work volume, experience on the analysis and costs. Ideally, these platforms should be concentrated in a small group of tertiary care hospitals, since this way costs could be largely reduced, until the advance on technology enables to have small equipments that perform the analysis with a better cost-benefit profile.
- Analysis and interpretation. This is one of the most complicated points. Depending on the platform and the extent of the analysis, specialized computing equipment and trained personnel are required. Requirements range from the analysis of gene panels, which practically do not require experience or sophisticated equipment, to the analysis of complete genomes, which require using super-computers and highly experienced personnel. Although there are several commercial programs, there is no general consensus on their use. The open access software, developed by the investigators themselves, is still the most widely used for this type of analysis^{24,25}. Furthermore, an important challenge is the ability to confer clinical value to found mutations, either by the existence of targeted therapy, prognostic usefulness or possible therapeutic synergism.

- Costs. Costs have decreased importantly over the past few years. The cost of the human genome was about 3 billion dollars, whereas, currently, it is possible to sequence a genome for 3,000-5,000 dollars²⁶. Prices keep falling with platform efficiency improvement. The possibility of multiplexing samples and to enrich specific regions or genes may reduce costs enough to generate sufficient clinically useful information, depending on the equipment and the volume of samples.

An important question that we should consider is the scope of the tests. This must consider both measured clinical usefulness and impact on the course of the disease, including the cost of the tests themselves. It is important to consider that an additional variable, inherent to this type of tests, is cost decrease when high-volume assays are performed. The capacity of the platforms and the possibility of marking each sample with a bar code enable the processing of many samples simultaneously, which considerably reduces costs. The alternative is establishing low-volume tests that, although more expensive per gen, offer advantages by making interpretation and assembly easier.

Other point to consider is the priorities on which the medical community must focus. Both the impact on healthcare costs and morbidity and mortality of the neoplasm must be weighted, without forgetting to anticipate a space for vulnerable groups, such as pediatric populations. Therefore, it is important to reach a consensus on the subject.

A specific example is that of the National Institute of Cancer in France, which, in partnership with various institutions, has developed a genomic biomarkers panel for several tumors²⁷. This panel considers specifically “actionable” mutations, for which targeted therapy approved by their healthcare system exists. In 2001, more than 50,000 French patients received one of these tests to guide their therapy. This is an example of tests established in low-volume conditions, without involving analyses of complete or partial genomes. The analyses are carried out in a consortium of genetics hospital centers with global and specific guidelines²⁷. This approach achieves cost reduction and a more rational use of resources, although the entities themselves establish this series of tests as a transitional process towards more complete analyses.

Finally, we must consider who should be responsible for technological development. The scheme can be public, as in the case of France and the Great Britain; private as in the U.S.A., or a combination of both. The possibility of developing tests in research institutions

or tertiary care hospitals to subsequently disseminate them to other healthcare centers is an attractive idea, given the design of our healthcare system. Participation of private initiative in specific niches should be encouraged the same way.

Final considerations

In the dawn of the present century, elucidation of the human genome has become a reality. This has resulted in multiple larger magnitude projects focused on health problems worldwide. In particular, the cancer genome projects have matured and provided many new data that now we can use for the benefit of our patients. It is essential applying these data in Mexico and, through local and guided efforts, use them to create useful and low-cost clinical tools that have an important impact on health. There are several challenges to be met, but research in Mexico is solid enough to be able to overcome them.

References

1. Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature*. 2007;450(7171):893-8.
2. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470(7333):214-20.
3. Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science*. 2011;333(6046):1157-60.
4. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012;486(7403):405-9.
5. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. 2012;44(6):685-9.
6. Berger MF, Hodis E, Heffernan TP, et al. Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature*. 2012;485(7399):502-6.
7. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012;150(6):1107-20.
8. Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A*. 2012;109(10):3879-84.
9. Garraway LA, Verweij J, Ballman KV. Precision oncology: an overview. *J Clin Oncol*. 2013;31(15):1803-5.
10. The Cancer Genome Atlas. Disponible en: <http://cancergenome.nih.gov/>.
11. International Cancer Genome Consortium. Disponible en: <http://www.icgc.org>.
12. Ojesina AI. Landscape of genomic alterations in cervical cancer. En revisión.
13. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214-8.
14. Sistemas de Proyectos. Instituto Nacional de Medicina Genómica; 2013.
15. Barrera H. Comunicación personal. 2013.
16. Meric-Bernstam F, Farhangfar C, Mendelsohn J, Mills GB. Building a personalized medicine infrastructure at a major cancer center. *J Clin Oncol*. 2013;31(15):1849-57.
17. Tsimberidou AM, Iskander NG, Hong DS, et al. Personalized medicine in a phase I clinical trials program: the MD Anderson Cancer Center initiative. *Clin Cancer Res*. 2012;18(22):6373-83.
18. Tursz T, Andre F, Lazar V, Lacroix L, Soria JC. Implications of personalized medicine--perspective from a cancer center. *Nat Rev Clin Oncol*. 2011;8(3):177-83.
19. Fenstermacher DA, Wenham RM, Rollison DE, Dalton WS. Implementing personalized medicine in a cancer center. *Cancer J*. 2011;17(6):528-36.
20. MacConaill LE. Existing and emerging technologies for tumor genomic profiling. *J Clin Oncol*. 2013;31(15):1815-24.
21. Vui-Kee K, Mohd Dali AZ, Mohamed Rose I, Ghazali R, Jamal R, Mokhtar NM. Molecular markers associated with nonepithelial ovarian cancer in formalin-fixed, paraffin-embedded specimens by genome wide expression profiling. *Kaohsiung J Med Sci*. 2012;28(5):243-50.
22. Nadauld L, Regan JF, Miotke L, et al. Quantitative and Sensitive Detection of Cancer Genome Amplifications from Formalin Fixed Paraffin Embedded Tumors with Droplet Digital PCR. *Transl Med (Sunnyvale)*. 2012;2(2).
23. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)*. 2013;6:287-303.
24. Kim SY, Speed TP. Comparing somatic mutation-callers: beyond Venn diagrams. *BMC Bioinformatics*. 2013;14:189.
25. Tarczy-Hornoch P, Amendola L, Aronson SJ, et al. A survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record. *Genet Med*. 2013;15(10):824-32.
26. Lunshof JE, Bobe J, Aach J, et al. Personal genomes in progress: from the human genome project to the personal genome project. *Dialogues Clin Neurosci*. 2010;12(1):47-60.
27. Institut National du Cancer. Molecular genetic tests for access to targeted therapies in france in 2012. En: *Activity Reports and Assessment*. 2012.