The Cancer Genome: What's New?

Based on materials from the 4th Annual Conference of Participants in the International Cancer Genome Consortium: December 1–3, 2010, Brisbane, Australia

E. B. Prokhorchuk
Centre "Bioengineering", Russian Academy of Sciences
E-mail: prokhortchouk@biengi.ac.ru

It's been eight months since the announcement of the establishment of the International Cancer Genome Consortium [1]. The conference was held with the intention of showing that the participants had started successfully. The Consortium was established in 2008 by predominantly European researchers, with the purpose of performing a consolidated and coordinated investigation of the genome of cancer cells. It should be noted that, in the United States, a similar project called TCGA – The Center Genome Atlas (http://tcga.cancer.gov–had been launched several years earlier. An agreement was made that TCGA would become part of ICGC. It is unquestionable that this decision benefited European researchers more, since it provided access to the numerous American databases that had accumulated. The establishment of the ICGC was for the most part stimulated by a technological breakthrough in determining the nucleotide sequence of DNA, enabling it to migrate from the analogue signal on micromatrices to a digital signal on NGS sequencers. At the time this article was being written, twelve countries were participating in the consortium: Italy, Spain, France, Germany, Great Britain, the United States, Canada, India, China, Australia, and Japan. Mexico is also reported to be on the verge of joining. Information on which country is "responsible" for a particular cancer type can be found on the Consortium website http://www.icgc.org. The total budget of the ICGC was announced at the Brisbane Conference. The sum is rather impressive: $500 million. This is the sum of contracts signed between the governments of participating countries and their national research centers, institutes, and universities, ensuring that at least 12,100 cancer genomes will be investigated. So, what is the status of the individual projects today, and what are the participants most concerned with?

One participant in each project (a mini-consortium devoted to a certain cancer type) is tasked with outlining the general state of the specific practice. The structure of almost all investigations appears to be the same. The first stage comprises a collection of clinical material. During this stage, the central role is most frequently played by hospitals, surgeons, and oncologists. They are responsible for collecting and storing material, recording medical history, and selecting the treatment regimen for a patient. A detailed histological description of the clinical material then has to be compiled. For example, three independent oncologists need to anonymously assess the sections of formalin fixed tumors in the Consortium devoted to kidney cancer CA-GEKID (Cancer Genomics of Kidney), supported by the Seventh Framework Program. Only if there is unanimity concerning the homogeneity of the material (at least 90% of cancer cells on the section) and the stage of tumor progression (which can be expressed quantitatively with no more than 5% inaccuracy) can the refrigerated adjacent section be used for the extraction of DNA and RNA. The problem of the heterogeneity of tumor material was mentioned by all speakers. The contamination of samples with stromal cells and the ingress of several tumor foci into the surgical material are the primary reasons why no more than 10% of several hundreds of tumor/control samples live up to the molecular-genetic investigations. Certain laboratories (John McPherson, Canada) have attempted to enrich the cell material using flow cytofluorimetric sorting (but the attempts have not been sufficiently successful) or used xenotransplantants in immune-deficient mice (in this case, increasingly considerable enrichment can be performed). All mini consortia use almost the same methods for analyzing nucleic acids at subsequent stages. The following are the most frequently sequenced by ICGC researchers in a tumor/control pair: 1) genome

with a coverage depth of ×30–×50; 2) transcriptome; 3) exome; 4) repertoire of micro-RNAs; and 5) highly methylated DNAs. On average, all mini consortia (probably, with the exception of the "advanced" TCGA and relatively young consortia, such as the German project "Genomics of prostate cancer") have reached appreciably the same level. In general, the genomes of no more than 10 tumor/control pairs and no more than several tens of transcriptomes or exomes had been successfully sequenced using NGS (Illumina or SOLiD) by December 2010. Neither consortium has provided data on DNA methylation or the repertoire of micro-RNAs. This is not surprising, since most teams only started receiving financial support in the beginning of, or middle of, 2010. However, ahead of the fifth meeting of the ICGC, to be held in June 2011 in Tokyo, almost all teams have promised to approach 30–50 genomes and several hundreds exomes/transcriptomes. Only American researchers of the Consortium have promised to approach 3,000 re-sequenced cancer genomes in two years. To be completely fair, it should be noted that the 3,000 genomes mark emerged in the American program as the joint effort of all institutes on all oncopathologies that they study. The policy of the consortium towards confirmation of the data obtained using NGS attracts significant interest. For example, in the Broad Institute, United States (Gad Getz), the following procedure is used: first, data on 30–40 genomes are accumulated, and mutations in cancer genomes are detected, and only then is the existence of mutations in amplicones attested using bar-coding and alternative methods of sequencing (Sanger sequencing or 454). Finally, the bioinformatics analysis caps these technological chains. This method allows to isolate somatic mutations which have occurred only in tumour cells, but not in the normal tissues adjacent to the tumour/or in blood cells. Databases of these mutations created at the Sanger Institute (Great Britain) – COSMIC (http://www.sanger.ac.uk/genetics/CGP/cosmic/) – have been widely mentioned, as well as the TCGA databases (http://www.broadinstitute.org/tcga/ password: tcga; login: tcga). None of the speakers mentioned a mutation in the intergenic regions or in promoters. On the contrary, more than sufficient data were given on coding regions and the exone-intron junction. On the basis of the preliminary data provided by NGS, it appears that 100 somatic mutations on average emerge in tumor cells (with the threshold value $p = 10^{-5}$); these values being very close in tumors of different etiologies. So, which genes are most frequently subjected to mutation? *TP53* is the absolute leader – from 85 to 96% depending on the tumor type. The following genes were also mentioned: *VHL*, in the case of kidney cancer; *MUC17*, upon gastric cancer; *CTNB1* (β-catenin) upon intestinal cancer, etc. This begs the very reasonable question as to whether it was necessary to spend US$500 million to determine that well-known genes –tumor growth suppressors – are mutated, while oncogenes are either amplified or characterized by an elevated level of transcription. The answer was given in the form of a lecture by Prof. Rob Sutherland, who was the first to propose anti-estrogen therapy for breast cancer. He explicitly stated that the scheme of therapy for each particular patient will be selected depending on the "genotype" of the mutations in his tumor. Herceptin, which is efficient only upon HER2+ malignant neoplasms of the breast, can serve as a striking example. The opposite is also true: mutations may disturb a certain methabolic pathway in tumor cells; therefore, a therapeutic drug, which would block this disturbance, could be found among the drugs that are very unlike those used in oncology. Hence, both concepts ("the right drug for a particular tumor" and "the right tumor for a certain drug") are valid. Such a change in the paradigm of pharmaceutics could considerably accelerate and improve the results in clinical trials of anti-tumor drugs. This means that in the future we are bound to witness a number of experiments devoted to finding the correlation between a whole-genome genotype of mutations in a tumor and the most efficient method for anti-tumor therapy. It is the underlying purpose of the ICGC consortium; namely, a transition to personified therapy for oncology patients. The issues of bioinformatic processing of genetic information were discussed in addition. With a considerable reduction in the cost of genome sequencing, the amount of data generated increases. Increasing computer resources are necessary in order to process, store, and provide access to the results obtained. The costs of computer resources compensate for the fall in prices and make the cost of the entire investigation even higher. Time will tell for how long the pursuit of cancer genomes will continue. Thus far, the amount of investigations is surely on the increase; the consortium is expected to have new participants with new projects. ●

REFERENCES

1. Hudson T.J., Anderson W., Artez A., Barker A.D., Bell C., Bernabé R.R., Bhan M.K., Calvo F., Eerola I., Gerhard D.S., et al. // Nature. 2010. V. 464(7291). P. 993–998.