Economical, legal and ethical considerations on reevaluation and retesting in molecular diagnostics

Lucio Pastore[1,3], Barbara Lombardo[1,3], Maria Vitale[1,3], Lorella Tripodi[1,3], Felice Amato[1,3], Mario Cosenza[2], Gianluca Giannini[2]

[1]Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Napoli, Italia
[2]Dipartimento degli Studi Umanistici, Università di Napoli Federico II, Napoli, Italia
[3]CEINGE-Biotecnologie Avanzate, Napoli, Italia

ABSTRACT

In the *-omics* era, analysis of genome and transcriptome have become extremely relevant for the elucidation of the genetic cause of a number of diseases, previously undiagnosed. In addition, microbiome analysis is becoming relevant in many pathological conditions.Identification of genetic variants is very efficient with techniques such as comparative genome hybridization (CGH)-array and with whole genome sequence (WGS) or whole exome sequence (WES) performed with next-generation sequence (NGS) methodologies. Most importantly, correct classification of variants and elucidation of their clinical significance are tasks of extreme relevance for the correct diagnosis and, often, also to indicate the most efficient therapeutic choices. However, over the years our understanding of significance of genetic variants has dramatically improved, therefore many cases would require reevaluation and, on occasions, retesting. In this article, we reviewed the major advances in the genomic diagnostics field focusing, in particular, at addressing the relevance of periodic reevaluation of results and retesting patients when significantly novel technologies are developed, focusing also on economical, legal and ethical points.

Key words: *ethics, molecular diagnostic, retesting*

THE -OMICS ERA

Analysis of genome and transcriptome

The history of genome analysis begins in 1975 when Sanger introduced the concept of DNA sequencing that became extremely popular and allowed, ultimately, the completion of human genome first draft. Today, as well as during the second half of 20th century, DNA sequence analysis has a main role in the knowledge of genome structure, function and evolution. In time, new techniques allowed the collection of an increasing range of high-quality DNA sequence information and brought down the cost for generating genome-scale data. The introduction of these new methodologies, collectively defined as next-generation sequencing (NGS), have changed human and animal genome research allowing genotyping, identification of genome wide structural variations, *de novo* assembling and re-assembling of genome, detection of variants in mendelian and polygenic human diseases, mainly, introducing individual genome sequencing in the clinical practice (1).

NGS also allows sequencing of cellular RNAs, even at single cell level, introducing "transcriptomic analysis" (i.e., sequencing of all RNA transcripts, coding and non-coding, in an individual, a population of cells or a single cell). Several pathological conditions drive extensive change in transcriptome (2); therefore, changes in abundance of the transcripts can be used as diagnostic or prognostic markers and can provide information relevant for possible therapeutic choices.

In this opinion paper, we summarize some of the novel

Corresponding author: Lucio Pastore, Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università di Napoli Federico II, Napoli, Italia. Email: lucio.pastore@unina.it

Received: 04.04.2022 Revised: 19.04.2022 Accepted: 17.05.2022

Published on-line: 13.06.2022 DOI: 10.19186/BC_2022.038

improvement in genomic analysis methodologies and evaluate the significance of periodic case reevaluation and, eventually retesting with some economical, legal and ethical perspective.

Analysis of the "other genome": the microbiome

The community of commensal, symbiotic and pathogen microorganisms, including bacteria, archaea, viruses, and yeast, living in a specific environment is the perfect definition of a microbiota. On the other hand, a microbiome is the entire collection of all the genomic elements of a specific microbiota. Metagenomics is the field of molecular research that studies the complexity of microbiomes. Gut microbiome, which hosts over 1 000 bacterial species that encode about 5 million genes, performs many of the functions required for host physiology and survival. Human gut microbiome is not a static system and changes with host development. The dynamic and complexity nature of this system allows variations in the density and composition of bacteria the gut microbiome along longitudinal and transverse gradients (3). Human gut microbiota is composed primarily of Firmicutes and Bacteroidetes that represent 90% of gut microbiota (4). Gut microbiota has many functions and is responsible for metabolizing nutrients into bioactive food components: bacteria metabolize indigestible carbohydrates like cellulose, hemicelluloses, oligosaccharides, pectin and lignin into short chain fatty acids (SCFAs) such as butyric, propionic and acetic. These fatty acids escape from digestion in the upper gastrointestinal tract and enter the colon (5). Alterations of microbiota can result in dysfunction of the biosynthesis of SCFAs, associated to a number of pathological conditions (6). Gut microbiota exerts several other functions in human body such as modulating the immune system (7), affecting the neurological functions of the host through brain-gut communication (8). In addition, there is growing awareness that the microbiome influences tumor progression, in part through inflammatory and immune circuits (9,10).

A wide diversity of microbiome-suitable molecular analyses can be performed on biological samples, each with strengths and weaknesses. The correct type of analyses for an experiment is dependent on the scientific or diagnostic question. Amplicon analysis is the most popular characterization of gut microbiome: it consists in the amplification of 16S rRNA for bacteria and archaea and the internal transcribed spacer for fungi, highly conserved regions. Bacterial 16S rRNA genes contain 9 hypervariable regions (V1–V9) that show sequence diversity and therefore often are used as a barcode-like method to differentiate many bacterial *taxa*, sometimes at species level. Sequences are then placed into a phylogenetic tree or matched to a database (11) obtaining information relevant for a number of pathological conditions. Analysis of gut microbiome has already a recognized clinical significance in diseases such as Inflammatory Bowel Disease (IBD), (12) and is getting a relevant role as biomarker prior to cancer immunotherapy (13,14).

Genomic analyses and identification and evaluation of variants

In the past two decades, the development of comparative genomic hybridization (CGH)-array has allowed the identification of structural variants at genomic level. Among these, copy-number variants (CNVs) are genetic variations larger than 50 bp in size (usually several kb in length) that involve gain or loss of DNA segments that can include an entire gene or part of it; most often CNVs include larger genomic region encompassing multiple genes (15). CNVs have been associated with the development of several genetic diseases, including autism spectrum disorders, neurodevelopmental disorders, multiple congenital anomalies and autoimmune diseases (16). At present, genome-wide assessment of CNVs is recommended as a first level test in patients with intellectual disability, developmental delay, autism spectrum disorder, and congenital anomalies.

While many recurrent CNVs (such as those flanked by segmental duplications) have been well characterized, most CNVs are unique, requiring further investigation to determine their clinical significance. Accurate interpretation of the clinical significance of CNVs requires consistent methods for the evaluation of gene content and correlation of the patient clinical findings with those observed in patients with similar variants, with the ultimate goal of producing consistent, evidence-based clinical classification across laboratories (17). Inconsistency among laboratories can create confusion for clinicians and their patients, leaving them unable to confidently use genetic information to manage health-care decisions (18). In order to reduce discordance between CNVs classifications, a good contribution is provided by the new guidelines which take into account the clinical characteristics of a wide range of CNVs and allow a complete analysis and an accurate classification of the variants (19). However, implementation of these guidelines on a large scale is challenging, as each CNV requires considerable time to obtain a definitive classification (20).

Each CNV is classified, according to the American College of Medical Genetics (ACMG) (19), into one of the following categories: benign, likely benign, variant of uncertain significance (VOUS), likely pathogenic, or pathogenic. Benign CNVs are found with a frequency >1% in the population. These CNVs should be reported in at least 3 individuals (preferably in different datasets). Likely benign CNVs in large case-control studies show no significant disease association. VOUS are CNVs for which there is not yet sufficient knowledge to understand if they are benign or potentially associated with diseases because of conflicting or insufficient evidence. Likely pathogenic CNVs are reported in one or few cases with similar or partially overlapping phenotypes in affected individuals in which the causative gene has not yet been identified or reported and include genes whose functions can likely cause the observed clinical phenotype. Pathogenic CNVs are well documented in the literature and reported in databases in patients with similar phenotypes. Furthermore, population studies suggest

that over 99% of all benign CNVs are inherited; therefore, inherited CNVs are less likely to be pathogenic than *de novo* ones. However, presence of a CNV in one of the parents does not necessarily exclude pathogenicity (21).

Next Generation Sequencing analysis applied to inherited diseases

In the recent years, the development of NGS-based technologies has introduced genome-level sequencing into the clinical practice. A number of genetic disorders have overlapping clinical manifestations; therefore, clinical observation may not be sufficient to identify the gene that needs to be analyzed. Differential diagnosis has been tremendously improved by NGS analysis of panels of genes that cause diseases with partially overlapping clinical conditions; this has also led to the identification of patients with multiple variants contributing to the observed phenotype. However, in some conditions the number of possible causative genes can be extremely large: therefore, other strategies may be used. Whole-genome sequencing (WGS), consisting in the sequencing of the entire genome, and whole-exome sequencing (WES), consisting in the sequencing of all the coding regions, have become less expensive and technically feasible and represent the test of choice for these conditions (22). WES allows sequencing of exons, that include 85% of all the disease-causing mutations; on the other hand, WGS allows the identification of variants in regulatory regions; however, evaluation of their clinical significance and reporting is more difficult and time-consuming. For clinical conditions caused by a large number of genes, WES has replaced gene panels analysis; in fact, after WES, variant analysis can be restricted to the desired panels of genes for detection of mutations in established disease-genes. In patients that lack causative pathogenic variants, analysis can subsequently be carried out on the entire exome. As described for CNVs, variants identified by NGS can also be classified as benign, likely benign, variant of uncertain significance (VOUS), likely pathogenic, or pathogenic.

NGS-based analyses require specific set-up in the diagnostic laboratory: data storage has to be carefully planned and access to genomic data needs to be restricted to specific operators. In addition to the technical laboratory procedures, data analysis and evaluation are time-consuming and require experienced operators. Recently, software certified for diagnostic purpose is available for data analysis and mutation detection: however, experienced operators are still required for careful data evaluation.

REEVALUATION OF MOLECULAR DIAGNOSTICS DATA

Clinical value

Data obtained with genomic screening methodologies, such as CGH-array and NGS, are subject to change in results interpretation over time for the accumulation of new knowledge and the identification of new disease/gene relations (Figure 1). In fact, a VOUS can subsequently be reclassified as pathogenic or benign after additional studies. In addition, even the significance of specific

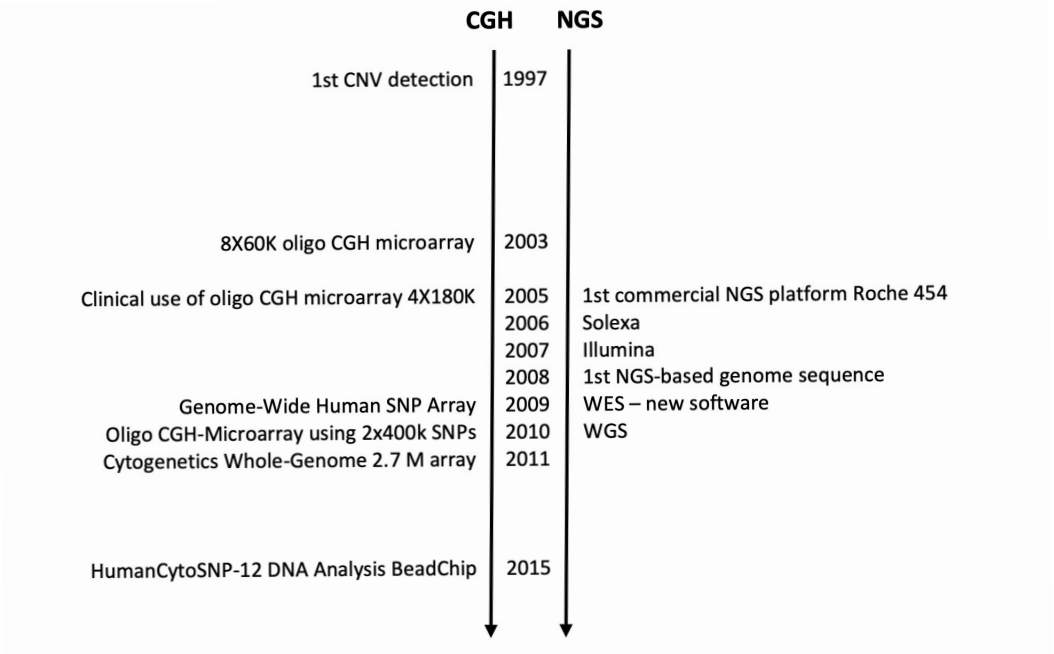

Figure 1
Major advances in the development of NGS and CGH methodologies and analyses of data generated with these technologies. After 2010, technological improvements and new and certified analysis pipelines have significantly modified clinical application of these methodologies.
CGH, comparative genome hybridization; NGS, next generation sequencing

pathogenetic mutations can vary with the development of novel therapeutic strategies that may lead to possible medical actions. In fact, development of novel genetic and pharmacological therapies has called for medical actions on patients diagnosed many years earlier with "classical" technique. One paramount example is the case of mutations in cystic fibrosis patients. To date, about 2 000 genetic variants in the *CFTR* gene are known, of which only about 400 have been shown to be pathogenetic. Since 2012, new drugs have been approved to correct and improve the function of the CFTR protein in a mutation-dependent manner. This makes it necessary to reevaluate a huge number of variants to clearly establish their pathological effect and to establish their response to treatment with new CFTR protein modulators. A great contribution to solving these issues comes from the development of protocols for functional characterization and assessment of responsiveness to drugs. If this has been possible in cystic fibrosis, thanks to the availability of *ex vivo* cellular models that allow this type of analysis (23,24), much still needs to be done for other pathologies where it is not always easy to obtain from patients' samples that can be used for drug testing and variant characterization (e.g., neuronal disease, polycistic kidney and so on).

As mentioned, CGH-array analysis can identify CNVs whose clinical significance may change over time: identification of a variant in a significant number of patients and, conversely, its absence in the unaffected population may indicate causality or strong contribution to the development of a particular disease. Similarly, variants identified by NGS are, at first, often classified as VOUS until further evidence of pathogenicity *in silico*, *in vitro*, in animal models or in patients are accumulated allowing a correct classification.

Change of classification of variants identified by NGS or CGH-array can have a relevant clinical impact for patients. Identification of the pathogenicity of a variant can be, in fact, extremely relevant for family planning of future pregnancies even in absence of specific therapies, opening the way to questions that include ethical issues especially when penetrance and expressivity of the molecular alteration is uncertain. On occasions, however, a variant reclassified as pathogenic may be medically actionable, therefore changing the clinical history of the patient. Therefore, periodic reevaluation of the significance of variants can have an impact on a number of patients. Although, at first, periodic reassessment of the significance of variants can probably have an impact on a small number of patients, nevertheless, the clinical significance in these few cases has the potential to be extremely relevant, especially since as statistics and studies grow, so does the relationship to possible choices in individual cases.

Economical, legal and ethical issues in re-evalution and retesting in molecular diagnostics

We can state, at present, that periodic reevaluation of variants is a service that a diagnostic laboratory can and should provide, and it is clear that, given the "public" impact of such health services, decisions cannot be postponed for long. Patients' data can be reevaluated simply updating the clinical classification of variants previously identified as VOUS: this type of data analysis is well defined and confined to a small number of variants per patients, therefore requires a reasonable effort for the diagnostic laboratory, obviously depending on the number of the diagnoses released. Case reevaluations are definitively more expensive and time-consuming, but they can indeed provide clinical information of high significance.

However, a number of issues should be clarified and possibly defined with legislative interventions to regulate laboratory policies concerning data reevaluations. Clinical significance of variants is likely to be updated for the next decade or more: therefore, periodic data reevaluation should be an integral part of the tests and clearly specified in the informed consent at sample collection. Reevaluation of variants or cases should be added as part of the cost of a test: at present, at least in the Italian healthcare system, these costs are not included and therefore, not reimbursed to the diagnostic laboratory; discussion at governmental level, whether national or regional, is therefore required. In addition, reevaluation is not usually included in the informed consent and, therefore, a laboratory should receive a specific authorization prior to communicate new evidence; for the same reason, a mechanism for the patients to opt out and not receive further information needs to be devised.

The diagnostic laboratory could add policies for periodical variant reclassification or case reanalysis and can therefore become the actual promoter. Reevaluation can be periodical (every 2-5 years) or, as the American College of Medical Genetics suggests, when a new resource or database becomes available or new methodologies for data analysis are devised or, finally, when new disease/gene associations are discovered (25). Results of case reevaluations should be sent then to the referring physicians or directly to the patients.

Reevaluation could be requested also by the referring physicians, especially when new clinical manifestations appear in the patient; in this case, specific reasons for granting such requests should be defined as well as the applicable costs. In addition, patient retesting with a novel methodology should be considered if the laboratory test had been performed many years earlier; in fact, both CGH-array and NGS have significantly progressed not only in data analysis pipelines but also in the actual testing methodologies. However, cost/benefit ratio should be evaluated considering the patient disease and the possible medical actionability.

Variant reclassifications and case reevaluations present an additional hidden cost: data storage. While storage of CGH-array data is usually well within the capability of a diagnostic laboratory, NGS data require a specific set-up. Primary NGS sequencing files (fastq) and alignments result files (.bam) need gygabites of space; on the other hand, variant lists are relatively small. Primary data should be maintained for legal reason for an appropriate amount of time (5 years at least); variant

reclassification and case reanalysis can be performed on the files containing data about the variants or, eventually, keeping the significant alignment files. However, data storage should be considered when defining costs of the analyses in both private and public settings, since the growing societal tendency to organize itself is based on big data. In this sense, the problem addressed here is a practical correlate of a historical trend with a definitive impact on biological sciences. A considerate possibility could be the set-up of regional or national databases where data could be uploaded and safely stored: informatic security can, in fact, be a major problem for many laboratories and can definitively addressed more efficiently with centralized storage. Development and running of such databases should be under governmental control and should also benefit from the expertise of the major research and diagnostic institutes involved in NGS. Legislative decisions would be required to define proper access to the resource and use of the data, perhaps even at international level. Actors requiring data reevaluation (patients, tutor or referring physician) should also be clearly addressed by the legislator; types of reevaluations, costs and communication between diagnostic laboratory and counterpart should be defined as well.

CONCLUSIONS

The -omics era is characterized by collection of an enormous amount of informatic data and periodic reassessment of the significance of laboratory findings. Therefore, the present considerations will soon apply to other types of analysis, such as untargeted metabolomics, in terms of data storage and reevaluation; addressing in a timely manner all these issues on economic, legal and ethical point of view will definitively provide benefit for patients.

From a more strictly ethical point of view, the issues offered for debate cannot be considered irrelevant. There is a general consent on the communication of medically actionable data to the patients: implications of the data for the patient and for the health care personnel must be clarified. To this end, communication of molecular diagnostics data based on clarity is necessary, also at the aim of motivating and improving therapeutic adherence. These issues are of paramount relevance to improve the patient/physician relationship and leave no room for lack of understanding.

Relationship between big data and their impact on individuals focuses on philosophical categories such as free-will and control, opening new scenarios in the relationship between patients and disease. In addition, NGS data storage introduce a new meaning of bio-banking. The first biobanks were born with the aim of collecting biological samples for specific research projects; however, in recent years, these have changed into real “deposits”, economically supported by governments and institutions. Individual genomic data introduces a new meaning of biobanks, that has gone hand in hand with the progress of the scientific world.

Many ethical and legislative aspects must revolve around the theme of the new concept of biobanks of the future. Nevertheless, the moral imperative to be placed at the basis of the very activities of study, knowledge and care must remain strong: man always as an end and never as a means. In order to avoid dangerous drifts, it will be necessary to have the human being as an ever-present perspective, never forget the fundamental motive of research in biological, human and applied sciences (26,27).

Ethical issues will be increasingly present in management of genomic data. Genomic or biological sample collection will always deal with human samples, invading individual autonomy or limiting self-control and, consequently, raising a number of ethical questions (28). For these reasons, it will be very important to provide researchers with an up-to-date review of the literature on bio-banking ethics in a systematic way, to document the latest consensus on ethical issues in bio-banking and to highlight emerging issues. All of this should stimulate policymakers and legislators to create an appropriate legal framework for genomic research (29).

More generally, the role of public organization become extremely relevant for a transparent control of genomic data. As it happens in other areas, in modern society ownership or simple knowledge of large amount of data is a source of wealth and social control (direct and indirect); therefore, actions are necessary to prevent or limit the danger of monopolies of genomic data, since the jurisprudence is not always able to keep up with technological changes. In addition, it would also be relevant to determine the level of control of genomic data (local, national or at European Union level) and right of access to it. These decisions will eventually impact in a major way academic and industrial research, as well as patients’ privacy and rights towards generated data.

CONFLICT OF INTEREST

None.

REFERENCES

1. Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev 2006;16:545-52.
2. Chambers DC, Carew AM, Lukowski SW, et al. Transcriptomics and single-cell RNA-sequencing. Respirology 2019;24:29-36.
3. Tropini C, Earle KA, Huang KC, et al. The gut microbiome: connecting spatial organization to function. Cell host microbe 2017;21:433-42.
4. Arumugam A, Agullo P, Boopalan T, et al. Neem leaf extract inhibits mammary carcinogenesis by altering cell proliferation, apoptosis, and angiogenesis. Cancer Biol Ther 2014;15:26-34.
5. Thursby E, Juge N. Introduction to the human gut microbiota. Biochem J 2017;474:1823-36.
6. Perry RJ, Peng L, Barry NA, et al. Acetate mediates a microbiome–brain–β-cell axis to promote metabolic syndrome. Nature (Internet) 2016;534:213-7. https://www.nature.com/articles/nature18309
7. Mills S, Stanton C, Lane JA, et al. Precision nutrition and the microbiome, part i: current state of the science. Nutrients 2019;11:923.

8. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Research 2020;30:1-15.
9. Zitvogel L, Ayyoub M, Routy B, et al. Microbiome and anticancer immunosurveillance. Cell 2016;165:276-87.
10. Tripodi L, Passariello M, D'Argenio V, et al. Evaluation of the antiproliferative effect of Bifido bacterium longum BB-536 in solid tumor cell lines, co-cultured with murine splenocytes. Biochim Clin 2021;45:242-7.
11. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. Isme J 2012;6:1621-4.
12. Lee M, Chang EB. Inflammatory bowel diseases (ibd) and the microbiome-searching the crime scene for clues. Gastroenterology 2021;160:524-37.
13. Yi M, Yu S, Qin S, et al. Gut microbiome modulates efficacy of immune checkpoint inhibitors. J Hematol Oncol 2018;11:47. https://jhoonline.biomedcentral.com/articles/10.1186/s13045-018-0592-6
14. Georgiou K, Marinov B, Farooqi AA, et al. Gut microbiota in lung cancer: where do we stand? Int J Mol Sci 2021;22:10429.
15. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet 2011;12:363-76.
16. Shaikh TH. Copy number variation disorders. Curr Genetic Medicine Reports 2017;5:183-90.
17. Brothman AR, Dolan MM, Goodman BK, et al. College of American Pathologists/American College of Medical Genetics proficiency testing for constitutional cytogenomic microarray analysis. Genet Med 2011;13:765-9.
18. Tsuchiya KD, Shaffer LG, Aradhya S, et al. Variability in interpreting and reporting copy number changes detected by array-based technology in clinical laboratories. Genet Med 2009;11:866-73.
19. Riggs ER, Andersen EF, Cherry AM, et al. Technical standards for the interpretation and reporting of constitutional copy number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genetics Medicine Official J Am Coll Medical Genetics 2020;22:245-57.
20. Gurbich TA, Ilinsky VV. ClassifyCNV: a tool for clinical annotation of copy-number variants. Sci Rep-uk. 2020;10(1):20375.
21. Tizaoui K. De novo vs. inherited copy number variations in multiple sclerosis susceptibility. Cell Mol Immunol 2018;15:812-4.
22. Sun H, Shen XR, Fang ZB, et al. Next-Generation Sequencing Technologies and Neurogenetic Diseases. Life 2021;11:361.
23. Terlizzi V, Amato F, Castellani C, et al. Ex vivo model predicted in vivo efficacy of CFTR modulator therapy in a child with rare genotype. Mol Genetics Genom Medicine 2021;9:e1656.
24. Amato F, Scudieri P, Musante I, et al. Two CFTR mutations within codon 970 differently impact on the chloride channel functionality. Hum Mutat 2019;40:742-8.
25. Deignan JL, Chung WK, Kearney HM, et al. Points to consider in the reevaluation and reanalysis of genomic test results: a statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med 2019;21:1267-70.
26. Salvaterra E, Lecchi L, Giovanelli S, et al. Banking together. Embo Rep 2008;9:307-13.
27. Favalli C, Galletti M. Biobanche e informazioni genetiche. Problemi etici e giuridici. Aprilia (LT): Aracne 2011.
28. Cambon-Thomsen A, Rial-Sebbag E, Knoppers BM. Trends in ethical and legal frameworks for the use of human biobanks. Eur Respir J 2007;30:373-82.
29. Caulfield T, McGuire AL, Cho M, et al. Research ethics recommendations for whole-genome research: consensus statement. Plos Biol 2008;6:e73.