EDITORIAL

Clinical and ethical considerations of massively parallel sequencing in transplantation science

Andreas Scherer

Andreas Scherer, Spheromics, 81100 Kontiolahti, Finland
Author contributions: Scherer A solely contributed to this paper.
Correspondence to: Andreas Scherer, PhD, APM, CEO, Spheromics, Lehmustie 16, 81100 Kontiolahti, Finland. andreas.scherer@spheromics.com
Telephone: +358-44-3629559 Fax: +358-44-3629559

Abstract

Massively parallel sequencing (MPS), alias next-generation sequencing, is making its way from research laboratories into applied sciences and clinics. MPS is a framework of experimental procedures which offer possibilities for genome research and genetics which could only be dreamed of until around 2005 when these technologies became available. Sequencing of a transcriptome, exome, even entire genomes is now possible within a time frame and precision that we could only hope for 10 years ago. Linking other experimental procedures with MPS enables researchers to study secondary DNA modifications across the entire genome, and protein binding sites, to name a few applications. How the advancements of sequencing technologies can contribute to transplantation science is subject of this discussion: immediate applications are in graft matching *via* human leukocyte antigen sequencing, as part of systems biology approaches which shed light on gene expression processes during immune response, as biomarkers of graft rejection, and to explore changes of microbiomes as a result of transplantation. Of considerable importance is the socio-ethical aspect of data ownership, privacy, informed consent, and result report to the study participant. While the technology is advancing rapidly, legislation is lagging behind due to the globalisation of data requisition, banking and sharing.

Key words: Sequencing; Diagnosis; Ethics; Consent; Data management

Core tip: Despite the great excitement about the opportunities that massively parallel sequencing (MPS) bears for the promotion of, *e.g.*, transplant science, personalized medicine faces the challenge to guarantee privacy of data and findings. Here, some applications of MPS in transplant science are mentioned, and concerns and challenges in data analysis and management are discussed.

Scherer A. Clinical and ethical considerations of massively parallel sequencing in transplantation science. *World J Transplant* 2013; 3(4): 62-67 Available from: URL: http://www.wjgnet.com/2220-3230/full/v3/i4/62.htm DOI: http://dx.doi.org/10.5500/wjt.v3.i4.62

MASSIVELY PARALLEL SEQUENCING

Massively parallel sequencing (MPS) is an alias of the probably more popular term next-generation sequencing (NGS). In this article the term next-generation sequencing will be avoided for the simple reason that "next" is a relative term in matter of time. If Sanger's and Maxam and Gilbert's ground breaking inventions of DNA sequencing in 1974 are counted as the first generation, then the automation of Sanger's method could be considered the next or second generation. Further, the next step which in 2005 lead to the development of machines which were able to sequence millions of fragments of DNA simultaneously would certainly have to be called the next-next-generation or third generation[1]. And the next generation of technological improvement is on its way: the 4th generation of sequencing methodology will utilize entire strands of DNA without the need of fragmentation, and will become cheaper, and more precise, and simpler to handle bioinformatically. In order

to avoid counting generations of technological advancement, the term MPS, or MPS for short, seems more applicable and will be used herein.

A variety of technical approaches to MPS exist, all of them have been reviewed in detail (for an excellent overview please see, *e.g.*[2]. Briefly, the general principle relies on (1) the fragmentation of DNA/RNA, optionally followed by fragment size selection; (2) amplification of the fragments; and (3) sequencing of the fragments. Currently, the length of those sequences can be up to 800 bp, depending on the vendor. As the fragmentation step generates random breakpoints in the DNA backbone, so will the sequenced fragments in Step 3 be at a random position of the DNA or RNA. This is where the individual small pieces of sequence information will have to be bioinformatically stitched together, "assembled", being a challenge to which there are plenty of approaches with slightly different quality, depending on the analysis pipeline of choice.

The applications of MPS are overwhelming and offer never seen before opportunities to study genomes, exomes, transcriptomes, and chromosomal rearrangements and secondary modifications like methylation of DNA and alkylation of RNA. Through its unbiased template-free approach, it is now also possible to sequence DNA and RNA of novel species in *de novo* assembly analyses and thus accelerate discovery of, *e.g.*, ontological relationships[3], and even discover novel RNA species[4]. Input amounts in the low ng-range for some MPS applications make it possible to study biological samples in a detail which could not have been envisioned before[5]. MPS has found its way to the analysis of single eukaryotic cells or even cell-free DNA in blood samples, *e.g.* for non-invasive prenatal diagnosis[6].

PERSONALIZED TREATMENT

Following transplantation, drug treatment must be carefully adjusted to prevent rejection. Drug metabolism is influenced by a large variety of factors such as age, gender, disease, dose, drug-drug interaction, and metabolic competence. Differences in the genotype (polymorphisms) can be linked with altered drug metabolism in transplant patients. Tacrolimus for example is primarily metabolized *via* the CYP450 enzymes. Non-expressors of CYP3A5 metabolize the drug slower than others, hence requiring lower doses than normal expressors[7]. Similarly, in conjunction with age as variant, polymorphisms in the transporter ABCB1 can determine the bioavailability of cyclosporine and mycophenolate mofetil[8]. Another example which illustrates the importance of studying genetic polymorphisms to optimize personal treatment is the occurrence of hypertension after transplantation. In genome-wide association studies polymorphisms have been identified in a number of genes affecting hypertension (*e.g.*[9]). For overviews of the field please see the recent reviews of D'Alessandro *et al*[10] and of Kurzawski *et al*[11]. As these examples and other studies, which cannot be discussed here for space limitation, show, the individual landscapes of polymorphisms in patients need to be assessed to optimize treatment efficacy. Sequencing of genes with standard methods is time-consuming and can deliver ambiguous. MPS technology can be used to study exomes of patients through targeted sequencing of candidate genes and determine polymorphisms which may affect treatment. However, MPS does not always deliver unambiguous results either due to sequence coverage differences and DNA sequence specifics such as guanine-cytosine (GC) content or homopolymers which cannot always be resolved by current MPS technologies alone. At times, one may need to verify the results by alternative technologies to obtain further sequence information.

HUMAN LEUKOCYTE ANTIGEN MATCHING

Alleles of the human leukocyte antigen (*HLA*) genes are commonly used for organ and bone marrow matching prior to transplantation. Humans vary widely in the composition of antigens arising from alleles of those six *HLA* genes (*A, B, C, DR, DQ, DP*). Detection of foreign HLA antigens by the host can lead to strong antibody mediated reactions, thus they can be considered important mediators of immune response. During the graft selection process, it is therefore essential to detect donor-host HLA mismatches, a process commonly performed by Sanger sequencing of the HLA locus. While Sanger sequencing certainly has its merits, technical limitations such as relatively high sequence inaccuracy resulting in sequence ambiguity due to highly polymorphic DNA regions, and limited sequence coverage in a single experiment (only a small number of exons is sequenced systematically, and some important polymorphisms may be located outside the sequenced regions) may make another round of experimental verification necessary in many cases. With ever decreasing costs, MPS has the potential to deliver high-quality sequence data which cover a large proportion of the entire HLA locus[12,13].

IMMUNE SYSTEM

MPS can be applied to many aspects of biological research in the transplantation arena. Exon arrays and RNA sequencing was applied to address the question whether alternative splicing takes place during immune response post-transplant. The group of Grigoryev *et al*[14] purified human CD2(+) T or CD19(+) B cells, activated them to model early post-transplant immune events and continued to sample from those cell pools over time. Indeed they were able to show that these two cell populations not only regulate gene expression following *in vitro* stimuli, but also regulate exon usage to generate alternative panels of transcripts which may contribute to the biological pattern of immune response. MPS now permits devising experiments which aim at studying the methylation status of DNA of T cells and B cells before, during and after immune response, *e.g.,* graft rejection. Methylation of promoter regions plays an important role in gene regula-

tion[15]. Changes of the methylation status of genes during immune response during and after treatment hence may give clues about how the expression of genes is regulated, for example in combination with DNA-protein motif discovery. For excellent overviews of MPS methods for the investigation of epigenetic modifications of DNA and RNA, please see[16,17].

METAGENOMICS

16S rRNA pyrosequencing is a variety of MPS used to selectively sequence the highly variable 16S rRNA regions of bacterial genomes, thus providing qualitative and quantitative genus and species information of bacteria present in a sample[18]. The group of Diaz et al[19] used 16S rRNA pyrosequencing to study the bacteriome of the human oral cavity after transplantation. They demonstrated a shift in the composition of the microbiome of the oral cavity during immunosuppression following transplantation[19]. The authors speculate that immunosuppression may create an environment in the oral cavity which could be more permissive for opportunistic pathogens.

A number of groups have focused on characterizing the microbiome of alveolar fluid in relation to lung transplantation. For instance, Borewizc et al[20] have applied 16S rRNA pyrosequencing to study the human lung microbiome after lung transplantation. The authors compiled sequencing data from 12 bronchoalveolar lavage fluid samples from four patients over three time points, two additional samples from healthy, non-transplanted individuals served as controls. Interestingly they found that the microbial diversity increased after transplantation, and that the dominating phyla after transplantation were different from those in healthy lungs. The authors suggest to follow those results under the aspect of the bronchiolitis obliterans syndrome, which is a marker of chronic lung transplant rejection[21].

DIAGNOSIS

In 2012, Wen et al[22] demonstrated that the number of circulating endothelial cells (CECs) increased in whole blood of renal transplant patients undergoing acute rejection, acute tubular necrosis, and chronic allograft nephropathy, when compared to control samples. CEC count decreased after immunosuppressive therapy. The authors attributed the increased CEC count to injury of vessel endothelium in conjunction with endarteritis, and conclude that monitoring CEC numbers can be used as minimally invasive tool to diagnose or prognose poor short-term outcome of renal allografts. Technically, it is not farfetched to design scenarios in which MPS technologies could be applied to monitor the number of CECs in whole blood samples. Whole genome sequencing would not be necessary; an exon-capture set specific for exons of endothelial genes would suffice for qualitative measurement of CECs. On the quantitative side read numbers would have to be normalized against a set of stably expressed genes, identification of which can be

challenging, as seen in the microarray arena.

Certainly, similar to other conventional approaches such as microarrays, MPS can be used to develop biomarkers of rejection or tolerance. Despite striving to identify the best matching grafts for hosts, the best matches are not always tolerated. The reason for tolerance, or lack thereof, may be found not within coding region of the HLA, but possibly in surrounding (introns, promoters) or even distant genomic areas. With ever decreasing costs of MPS it will soon be possible to sequence not only exons or exomes in a larger scale than possible or affordable today, but entire genomes. As is the case in other research disciplines it will be necessary to gather genomic sequence information from a sufficient number of individuals to draw significant conclusions. This is the case for mutation analysis [e.g., single nucleotide polymorphisms (SNP)], as well as the analysis of secondary modifications such as methylation when certain biological conditions are compared. Research will see a steady growth of available sequence information which will contribute to discovery and qualification of biomarkers and elucidation of biological processes for the benefit of patients.

BIOINFORMATICS CHALLENGES

There are now many MPS-approaches to sequencing DNA, which will continue to reduce speed and cost of sequencing. When in 2000, still in the pre-MPS era, the drafts of a human genome sequence were published, one would not have thought that already 13 years on, the cost for this undertaking would come down from around $3B to around $5-10K, and the sequencing and analysis time would shrink from 10 years "for a rough working draft" to around 3-4 wk on the average for a complete version. However, decreased sequence raw data generation time and costs mean huge challenges for IT in terms of data storage and transfer of the huge raw data files which can be in the TB range per run, and for bioinformatics data analysis capacities, including quality control, alignment, assembly, annotation, and statistical analysis. No longer is the data generation process the experimental bottleneck, but the analytical side of things. In fact, as Sboner et al[23] phrase it, there is an "unpredictable amount of extra 'human' time" which is required for the identification of the best analysis pipelines, software installation, etc. Like in the early days of microarrays experts argue about the approaches to data processing. This leads to an amount of approaches which can be even overwhelming for bioinformaticians themselves (if they would admit it): What is the most precise, fastest, aligner, assembler, normalization method, algorithm to identify SNPs, statistics for differentially expressed genes, differentially methylated sites, etc.? Some methods are listed in[2]. Evaluating which analysis pipeline suits best to which problem and to which IT environment is challenging and time consuming. The final step, the interpretation of the results, is yet another "unknown" time factor which can rarely be done automatically, but requires human intervention. In the end one needs to understand that sequencing

cannot in every case provide an immediate answer to all scientific questions. Just like in all other comparative experiments which we have become familiar with over the years, the first step in experiments involving MPS is sampling. Sampling means that individuals are selected which represent the entire group of individuals we are interested in, a process which can be attempted in a variety of statistical approaches of experimental study design, such as randomization, blocking, and randomization[24]. Many sequencing applications do not omit the need for biological replicates, a cost-factor which needs to be considered in the planning phase. Certainly this is true for transcriptome analysis, differential methylation analysis, but also for genome-wide association studies (GWAS). The latter will benefit dramatically from the increased precision and availability of whole genome information in the near future, contributing to the growing number of lead mutations in diseases (for an overview of GWAS studies, www.genome.gov/GWAstudies). MPS will allow the discovery of rare variants where commonly used SNP arrays will have to fail. Certainly, there are settings where one sample will suffice. These are occasions in which individual information about a genome is investigated, *e.g.,* in cancer-genomics or in rare diseases. This brings our discussion to the aspect of personalized medicine and MPS.

DATA MANAGEMENT AND PRIVACY

Decreasing costs and increasing availability of resources will make MPS a tool for medical research and clinical care. However, routine genome sequencing for patient care brings along important socio-ethical and legal ramifications which are heavily discussed. Crucial concerns arise around patient information to obtain informed consent, data protection and patient privacy protection, data ownership, third-party use, use of incidental findings, and how such (incidental) findings are disclosed to the patient, to name a few[25-27]. On the other hand, sequencing data can be used for a whole range of scientific and clinical applications, becoming accessible *via* databases across nations. Sequence data can be used *e.g.* for trait analysis, phylogenetic testing, and expression analysis, bringing along a wide range of possible findings which is difficult to estimate at the time of sampling. Hence, to obtain informed consent from a patient the extent of consent has to be fairly thorough, which may cause frustration and possibly unwillingness to consent, additionally posing risks of study bias due to social background. McGuire *et al*[28] proposed a tiered consent process with three levels, from intended release of data information on multiple gene loci, to single gene loci, to releasing no data. Sample donors would have to be educated about the risks and benefits of the foreseen use of their data. Data access would have to be restricted according to the intended use at the beginning of the study. Reconsideration of study purposes may enforce re-consenting.

If genomic information is released though, is it possible to fully protect the privacy of sequencing data?

Already in the pre-MPS era of 2004, Malin and his team showed that it was possible to link genomic data to named individuals in publicly available records by leveraging unique features in patient-location visit patterns[29]. With the growth of genome sequence databases it should be possible to identify individuals based on their DNA sequence (*e.g.* SNP pattern), provided a template is present. In 2004, Lin *et al*[30] published it was possible to de-identify a person by interrogating just 75 SNPs, not many when taking into consideration that SNP databases of human genomes contain hundreds of thousands per genome. Not only the patient's but also the relatives' privacy is affected, but may be affected. This has large implications not only on research, but even more importantly on health care systems and national databases. The goal of the Health Insurance Portability and Accountability Act of 1996 is to protect genomic data as personal health information (http://www.hhs.gov/ocr/privacy).

The extent of result disclosure poses another issue. How much does a patient need to learn about the results, especially incidental findings, which were not part of the original study. What are results and who is interpreting them? As Sharp pointed out in a detailed discussion[31], the amount of data and potential findings with all their false positives and negatives, is equally overwhelming for the practitioner as it will be for the study participant. Many mutations may be harmless, and a result-interpretation may again be interpreted as a result by a study participant[31].

These are only a few critical concerns that have to be addressed urgently. The scientific community needs to ensure that the legal and ethical framework which makes social discrimination based on genetic information impossible is appropriate for the developing technology. International databases and cloud computing impose the necessity of international legislation which puts the patient rights first. By ensuring privacy protection, study participation has a chance to be beneficial for the individual, not a potential risk for social exclusion.

OUTLOOK

Over the next years prices per sequenced nucleotide will continue to fall, sequencing machines will become smaller, cheaper and easier to use, eventually making genomic sequencing a standard tool in research and clinics. Despite growing databases, MPS data interpretation will remain a challenge. The legal and ethical frameworks for using MPS data need to be defined on an international level, granting respect to sample-providing individuals as well as the research goals of scientists and clinicians. International consortia need to address the possibility that the current speed of genome research may outrun the pace of legal regulation, and impose adjustments.

REFERENCES

1   **Margulies M**, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z,

Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; **437**: 376-380 [PMID: 16056220 DOI: 10.1038/nature03959]

2   **Xuan J**, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: Promises and challenges. *Cancer Lett* 2013; **340**: 284-295 [PMID: 23174106 DOI: 10.1016/j.canlet.2012.11.025]

3   **Wang M**, Ye Y, Tang H. A de Bruijn graph approach to the quantification of closely-related genomes in a microbial community. *J Comput Biol* 2012; **19**: 814-825 [PMID: 22697249 DOI: 10.1089/cmb.2012.0058]

4   **Martens-Uzunova ES**, Olvedy M, Jenster G. Beyond microRNA - Novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett* 2013; **340**: 201-211 [PMID: 23376637 DOI: 10.1016/j.canlet.2012.11.058]

5   **Sengupta S**, Ruotti V, Bolin J, Elwell A, Hernandez A, Thomson J, Stewart R. Highly consistent, fully representative mRNA-Seq libraries from ten nanograms of total RNA. *Biotechniques* 2010; **49**: 898-904 [PMID: 21143212 DOI: 10.2144/000113556]

6   **Nepomnyashchaya YN**, Artemov AV, Roumiantsev SA, Roumyantsev AG, Zhavoronkov A. Non-invasive prenatal diagnostics of aneuploidy using next-generation DNA sequencing technologies, and clinical considerations. *Clin Chem Lab Med* 2013; **51**: 1141-1154 [PMID: 23023923 DOI: 10.1515/cclm-2012-0281]

7   **MacPhee IA**, Fredericks S, Tai T, Syrris P, Carter ND, Johnston A, Goldberg L, Holt DW. The influence of pharmacogenetics on the time to achieve target tacrolimus concentrations after kidney transplantation. *Am J Transplant* 2004; **4**: 914-919 [PMID: 15147425 DOI: 10.1111/j.1600-6143.2004.00435.x]

8   **Fanta S**, Niemi M, Jönsson S, Karlsson MO, Holmberg C, Neuvonen PJ, Hoppu K, Backman JT. Pharmacogenetics of cyclosporine in children suggests an age-dependent influence of ABCB1 polymorphisms. *Pharmacogenet Genomics* 2008; **18**: 77-90 [PMID: 18192894 DOI: 10.1097/FPC.0b013e3282f3ef72]

9   **Levy D**, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Köttgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FU, Smith AV, Taylor K, Scharpf RB, Hwang SJ, Sijbrands EJ, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JC, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, van Duijn CM. Genome-wide association study of blood pressure and hypertension. *Nat Genet* 2009; **41**: 677-687 [PMID: 19430479 DOI: 10.1038/ng.384]

10   **D'Alessandro LCA**, Mital S. Pediatric tranplantion: opportunities for pharmacogenomics and genomics. *PERS MED* 2013; **10**: 397-404

11   **Kurzawski M**, Droździk M. Pharmacogenetics in solid organ transplantation: genes involved in mechanism of action and pharmacokinetics of immunosuppressive drugs. *Pharmacogenomics* 2013; **14**: 1099-1118 [PMID: 23837483 DOI: 10.2217/pgs.13.89]

12   **De Santis D**, Dinauer D, Duke J, Erlich HA, Holcomb CL, Lind C, Mackiewicz K, Monos D, Moudgil A, Norman P, Parham P, Sasson A, Allcock RJ. 16(th) IHIW : review of HLA typing by NGS. *Int J Immunogenet* 2013; **40**: 72-76 [PMID: 23302098 DOI: 10.1111/iji.12024]

13   **Erlich RL**, Jia X, Anderson S, Banks E, Gao X, Carrington M, Gupta N, DePristo MA, Henn MR, Lennon NJ, de Bakker PI. Next-generation sequencing for HLA typing of class I loci. *BMC Genomics* 2011; **12**: 42 [PMID: 21244689 DOI: 10.1186/1471-2164-12-42]

14   **Grigoryev YA**, Kurian SM, Nakorchevskiy AA, Burke JP, Campbell D, Head SR, Deng J, Kantor AB, Yates JR, Salomon DR. Genome-wide analysis of immune activation in human T and B cells reveals distinct classes of alternatively spliced genes. *PLoS One* 2009; **4**: e7906 [PMID: 19936255 DOI: 10.1371/journal.pone.0007906]

15   **Sibbritt T**, Patel HR, Preiss T. Mapping and significance of the mRNA methylome. *Wiley Interdiscip Rev RNA* 2013; **4**: 397-422 [PMID: 23681756 DOI: 10.1002/wrna.1166]

16   **Ku CS**, Naidoo N, Wu M, Soong R. Studying the epigenome using next generation sequencing. *J Med Genet* 2011; **48**: 721-730 [PMID: 21825079 DOI: 10.1136/jmedgenet-2011-100242]

17   **Rodriguez BA**, Frankhouser D, Murphy M, Trimarchi M, Tam HH, Curfman J, Huang R, Chan MW, Lai HC, Parikh D, Ball B, Schwind S, Blum W, Marcucci G, Yan P, Bundschuh R. Methods for high-throughput MethylCap-Seq data analysis. *BMC Genomics* 2012; **13** Suppl 6: S14 [PMID: 23134780 DOI: 10.1186/1471-2164-13-S6-S14]

18   **Woese CR**, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 1977; **74**: 5088-5090 [PMID: 270744]

19   **Diaz PI**, Hong BY, Frias-Lopez J, Dupuy AK, Angeloni M, Abusleme L, Terzi E, Ioannidou E, Strausbaugh LD, Dongari-Bagtzoglou A. Transplantation-associated long-term immunosuppression promotes oral colonization by potentially opportunistic pathogens without impacting other members of the salivary bacteriome. *Clin Vaccine Immunol* 2013; **20**: 920-930 [PMID: 23616410 DOI: 10.1128/CVI.00734-12]

20   **Borewicz K**, Pragman AA, Kim HB, Hertz M, Wendt C, Isaacson RE. Longitudinal analysis of the lung microbiome in lung transplantation. *FEMS Microbiol Lett* 2013; **339**: 57-65 [PMID: 23173619 DOI: 10.1111/1574-6968.12053]

21   **Jaramillo A**, Fernández FG, Kuo EY, Trulock EP, Patterson GA, Mohanakumar T. Immune mechanisms in the pathogenesis of bronchiolitis obliterans syndrome after lung transplantation. *Pediatr Transplant* 2005; **9**: 84-93 [PMID: 15667618 DOI: 10.1111/j.1399-3046.2004.00270.x]

22   **Wen J**, Chen J, Ji SM, Cheng D, Liu ZH. Evaluation of vascular lesions using circulating endothelial cells in renal transplant patients. *Clin Transplant* 2012; **26**: E344-E350 [PMID: 22515202 DOI: 10.1111/j.1399-0012.2012.01620.x]

23   **Sboner A**, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! *Genome Biol* 2011; **12**: 125 [PMID: 21867570 DOI: 10.1186/gb-2011-12-8-125]

24   **Grass P**. Experimental Design. In: Scherer A, editor Batch Effects and Noise in Microarray Experiments: Sources and Solutions: Wiley and Sons, 2009: 19-31

25   **Greenbaum D**, Sboner A, Mu XJ, Gerstein M. Genomics and privacy: implications of the new reality of closed data for the field. *PLoS Comput Biol* 2011; **7**: e1002278 [PMID: 22144881 DOI: 10.1371/journal.pcbi.1002278]

26   **Mathieu G**, Groisman IJ, Godard B. Next generation sequencing in psychiatric research: what study participants need to know about research findings. *Int J Neuropsychopharmacol* 2013; **16**: 2119-2127 [PMID: 23725748 DOI: 10.1017/S1461145713000527]

27   **Ong FS**, Grody WW, Deignan JL. Privacy and data management in the era of massively parallel next-generation sequencing. *Expert Rev Mol Diagn* 2011; **11**: 457-459 [PMID: 21707452 DOI: 10.1586/ERM.11.34]

28   **McGuire AL**, Gibbs RA. Genetics. No longer de-identified. *Science* 2006; **312**: 370-371 [PMID: 16627725 DOI: 10.1126/science.1125339]

29   **Malin B**, Sweeney L. How (not) to protect genomic data pri-
      vacy in a distributed network: using trail re-identification to
      evaluate and design anonymity protection systems. *J Biomed
      Inform* 2004; **37**: 179-192 [PMID: 15196482 DOI: 10.1016/
      j.jbi.2004.04.005]
30   **Lin Z**, Owen AB, Altman RB. Genetics. Genomic research

and human subject privacy. *Science* 2004; **305**: 183 [PMID:
15247459 DOI: 10.1126/science.1095019]
31   **Sharp RR**. Downsizing genomic medicine: approaching the
      ethical complexity of whole-genome sequencing by starting
      small. *Genet Med* 2011; **13**: 191-194 [PMID: 21311340 DOI:
      10.1097/GIM.0b013e31820f603f]