How Should Biobanks Prioritize and Diversify Biosample Collections? A 40-Year Scientific Publication Trend Analysis by the Type of Biosample

Jae-Eun Lee and Young-Youl Kim

Abstract

Biobanks are infrastructures for large-scale biology innovation. Governance of biobanks can be usefully informed by studies of publication trends, for example, on the types of biosamples employed in scientific publications. We examined trends in each of the serum, plasma, peripheral blood mononuclear cell (PBMC), buffy coat, tissue, and gut microbiome biosample-related scientific publications over the past 40 years, using data between 1977 and 2016 from the Scopus database. We found that the number of tissue-related publications was the highest in each year of our analysis than other biosamples, but was generally less than the sum of serum- and plasma-related publications. Importantly, the microbiome publications increased greatly starting in the 2010s, and currently overtook the number of publications on PBMC and buffy coat. Among serum-, plasma-, and tissue-related publications, the number of protein- and RNA-related publications was generally higher than cell-free DNA-, DNA-, and metabolite-related publications for the past 40 years. Mass spectrometry- and next-generation sequencing-related publications have increased dramatically since the 2000s and 2010s, respectively. Microbiome- and metabolite-related biosamples can help diversify future biosample collections, while tissue collections appear to maintain their importance in scientific publications. We also report here our observations on the countries that use biosample research (e.g., China, United Kingdom, United States, and others). These publication trends by the type of biosamples illuminate roadmaps by which biobanks might establish and diversify their biosample collections in the future. In addition, we note that biobanks need to secure biosamples appropriate for integrated analysis of multi-omics research data.

Keywords: biobanks, governance, technology policy, publication trends, biosamples

Introduction

BIOBANKS ARCHIVE LARGE NUMBERS OF BIOSAMPLES such as serum, plasma, peripheral blood mononuclear cell (PBMC), buffy coat, and tissues for research. They are scientific infrastructures essential for integrative biology innovation (Birch et al., 2016). Governance of biobanks and omics research can be usefully informed by studies on publication trends (Piletič and Kunej, 2017; Pirih and Kunej, 2017), including, for example, biosample-related research trends in the literature. Through trend analysis, biobanks can determine the priority of their collections portfolio among the various types of biosamples, and thus tailor their development strategy based, in part, on the utilization range of biosamples reportedly used in the literature.

Biosamples can be used for biomarker discovery for risk prediction, diagnosis, and prognosis prediction of various

human diseases and can provide information on complex and dynamic molecular mechanisms. To obtain accurate information from biosamples, it is necessary to minimize pre-analytical variations in the collection, processing, and storage stages of biosamples. The preanalytical variations can be different depending on the type of biosample, the type of analyte, and the method of analysis (Araujo et al., 2017; Betsou et al., 2016; Lee and Kim, 2017). Therefore, it is important to establish conditions for collection, processing, and storage of biosamples suitable for research purposes.

It also is necessary to collect information about conditions for collection, processing, and storage of biosamples when securing and managing biosamples; in case of blood samples, these considerations include the type of blood collection tube, precentrifuge and postcentrifuge conditions (time and temperature) of whole blood, and long-term storage conditions (period and temperature). We have previously defined this

biobanking approach as the “precision biobanking” (Lee and Kim, 2017). Analyzing the trends in scientific publications by the type of biosample can help the biobanks decide how best to prioritize different types of biosample collections.

Astrin and Betsou (2016) have reported the trends in biobank-associated publications, through the Scopus database-based analysis. The number of biobank-associated publications began to increase slowly in the 1980s and increased significantly in the 1990s.

In this study, we examined the trends in biosample-related publications registered in the Scopus database over the past 40 years (between 1977 and 2016). The beginning year for our analysis corresponds to a date (1977) before 1980 when biobank-related publications began to show an increasing trend. The Scopus database is one of the world’s largest bibliographic databases and includes >3,600,000 biosample (including serum, plasma, PBMC, buffy coat, and tissue)-related references published between 1977 and 2016. Our study thus provides information on the trends in scientific publications by the type of biosample that biobanks can consider in establishing future biobanking strategies.

Materials and Methods

Scientific publications (journal articles and other publications; all document types, including research articles, reviews, conference articles, book chapters, short surveys, letters, conference reviews, editorials, reports, and business articles where biosamples were mentioned) on various biosamples (serum, plasma, PBMC, buffy coat, tissue, and gut microbiome sample) were retrieved using Elsevier’s Scopus database (www.scopus.com); serum, plasma, PBMC, buffy coat, and tissue sample-related references were searched in November 2017 and microbiome sample-related references were retrieved in January 2018.

The search strings were as follows:

serum (“serum”), plasma (“plasma”), PBMC (“peripheral blood mononuclear cell*” OR “PBMC*”), buffy coat (“buffy coat*”), tissue (“tissue*”), gut microbiome (“microbiome” OR “microorganism” OR “microbiota” OR “microbial”) AND (“gut” OR “stool” OR “poo”)], cfDNA (“cell free DNA” OR “cfDNA”), DNA (“DNA” OR “SNP” OR “genomics”), RNA (“mRNA*” OR “miRNA*” OR “microRNA*” OR “gene expression” OR “transcript*” OR “transcriptome*”), protein (“protein*” OR “proteome*” OR “proteomics”), metabolite (“metabolite*” OR “metabolome*” OR “metabolomics”), next-generation DNA sequencing (“next-generation DNA sequencing” OR “next generation sequencing” OR “NGS” OR “exome sequencing” OR “whole genome sequencing”), next-generation RNA sequencing (“RNA seq” OR “RNA sequencing” OR “transcriptome sequencing” OR “microRNA sequencing” OR “miRNA sequencing” OR “miRNA-seq” OR “next generation sequencing” OR “NGS”), and mass spectrometry (“mass spectrometry” OR “SELDI-TOF” OR “MALDI-TOF” OR “MALDI-FTMS” OR “LC-MS” OR “LC-ESI-MS” OR “LC-TOF-MS” OR “UPLC-TOF” OR “GC-TOF” OR “GC-MS” OR “FIA-ESI-MS” OR “MS-MS”).

The search keywords were queried in the database of publication abstracts and asterisks were used to retrieve publications, including derivations of the keywords. Some pertinent information (the subject area, document type, source

title, author’s country, and source type) on various biosample-related references published in 2016 was investigated using the Scopus database. All publications belonged to one or more subject areas.

Statistical trend analysis

Linear regression analysis was conducted with the number of each of serum, plasma, PBMC, buffy coat, tissue, and gut microbiome biosample-related scientific publications published yearly as dependent variables and the years of publication as independent variables to assess publication trends, using SPSS, version 18.0 (SPSS, Chicago, IL, USA). A *p*-value of <0.05 was considered statistically significant.

Results

Publications related to various biosamples

We retrieved scientific publications on serum, plasma, PBMC, buffy coat, tissue, and gut microbiome samples published between 1977 and 2016, respectively, using the Scopus database for publication abstracts. As shown in Figure 1A, the number of serum-, plasma-, PBMC-, and tissue-related references showed an increasing trend for the past 40 years, but the uptrend has slowed in recent years. Among them, the number of tissue-oriented publications was annually the highest, but was less than the sum of the number of serum- and plasma-related publications, except for the last 3 years.

The differences between the number of tissue-oriented publications and other publications tended to increase gradually since the 1980s. Plasma-oriented publications were more than serum-related publications since 1982. Tissue-, plasma-, and serum-related references have been published more than 40,000 annually since 1996, 2008, and 2013, respectively. Currently, tissue-related references have been published over 90,000 per year. Publications on PBMC and buffy coat have been published in relatively fewer numbers. The number of microbiome-related references increased greatly in the 2010s (Fig. 1B) and recently overtook the number of PBMC- and buffy coat-related publications; the numbers of microbiome-, PBMC-, and buffy coat-related references published in 2016 were 3084, 2388, and 107, respectively (Table 1).

Some pertinent information (subject area, document type, source title, author’s country, and source type) on references published in 2016 was investigated (Table 1). Medicine is the major subject area (serum: 65.7%, plasma: 35.3%, PBMC: 65.4%, buffy coat: 73.8%, tissue: 55.0%, and microbiome: 55.1%) of all the retrieved references. The biochemistry, genetics, and molecular biology fields generally account for the second largest share (serum: 30.4%, plasma: 23.7%, PBMC: 39.8%, buffy coat: 20.6%, tissue: 36.4%, and microbiome: 31.2%). Document types of the retrieved publications are mostly articles (serum: 93.8%, plasma: 86.3%, PBMC: 96.1%, buffy coat: 93.5%, tissue: 82.0%, and microbiome: 68.9%).

In source title, *PLoS One* is the journal that has published the highest number of references related to serum, PBMC, and tissue, followed by *Scientific Reports*. Plasma-related publications have been most widely published by the journal *Physics of Plasma*, followed by *PLoS One* and *Scientific Reports*. Microbiome-related publications have been most widely published by the journal *Scientific Reports*, followed

FIG. 1. Annual number of biosample-related references published between 1977 and 2016. (A) Serum-, plasma-, peripheral blood mononuclear cell (PBMC)-, buffy coat-, and tissue-related references. (B) Microbiome-related references.

by *PLoS One*. In the author's country distribution, the number of serum (17,462; 40.0%)-, plasma (20,148; 40.8%)-, PBMC (1081; 45.3%)-, tissue (44,223; 46.3%)-, and microbiome (1417; 45.9%)-related publications by researchers in the United States and China is much higher compared with other countries.

The number of serum-, plasma-, and tissue-related publications from researchers in Japan and Germany is in the top five. The number of plasma-, PBMC-, tissue-, and microbiome-oriented publications from researchers in the United Kingdom is in the top five.

Publications related to different types of analytes

Publications on various analyte (cfDNA, DNA, RNA, protein, or metabolite)-related search terms were extracted from the retrieved publications on serum, plasma, PBMC, buffy coat, and tissue (Fig. 2). Among the retrieved publications related to serum, plasma, and tissue, protein-related publications were annually much more than publications related to other analytes (DNA, cfDNA, RNA, or metabolite) and showed an increasing tendency for the last 40 years. In 2016, the number of protein-related publications extracted from the retrieved publications related to serum, plasma, and tissue is 11,994, 8299, and 21,129, respectively. The number of RNA-related references extracted from the searched publications related to serum, plasma, and tissue was the second most common since 1989, 1997, and 1989, respectively, and showed an increasing tendency.

The number of these references published in 2016 is 4039, 2948, and 17,003, respectively. Metabolite-related publications extracted from the searched plasma-related references were more than those extracted from serum-related references every year. The number of publications on RNA and protein isolated from PBMC-related references increased generally steadily with a similar upward curve.

Publications related to next-generation sequencing

Publications related to next-generation sequencing were extracted from the retrieved publications related to serum, plasma, PBMC, buffy coat, and tissue; publications related to next-generation DNA sequencing were searched using the DNA (or cfDNA) and next-generation DNA sequencing search terms. Publications related to next-generation RNA sequencing were searched using the RNA and next-generation RNA sequencing search terms (Fig. 3). The total number of publications related to next-generation sequencing rose rapidly since the 2010s. Of publications related to next-generation DNA sequencing, a conference article containing the term *whole genome sequencing* in the abstract section was first published in 1999.

In 2016, the number of publications related to next-generation DNA sequencing was the greatest in tissue-related publications, followed by plasma-related publications. Of publications related to next-generation RNA sequencing, two articles containing the term *RNA sequencing* were first

TABLE 1. TOP 5 OF THE SUBJECT AREA, DOCUMENT TYPE, SOURCE TITLE, AUTHOR'S COUNTRY, AND SOURCE TYPE ON BIOSAMPLE-RELATED LITERATURES PUBLISHED IN 2016

Sample type	Total	Subject area	Document type	Source title	Country	Source type			
Serum	43,698	Medicine Biochemistry, Genetics and Molecular Biology Pharmacology, Toxicology, and Pharmaceutics	28,713	Article	1120	China	9753	Journals	43,157
			13,280	Review	660	United States	7709	Books	251
			4458	Article in press	311	Japan	2978	Book series	147
Plasma	49,361	Immunology and Microbiology	4313	Conference article	308	India	2339	Conference proceedings	140
			3818	Book chapter	305	Germany	2002	Trade publications	3
		Agricultural and Biological Sciences Medicine Physics and Astronomy	17,434	Article	979	United States	11,112	Journals	45,891
			11,979	Conference article	858	China	9036	Conference proceedings	2465
			11,696	Review	683	Japan	3663	Book series	549
			7382	Article in press	471	Germany	3499	Books	430
PBMC	2388	Engineering	6279	Book chapter	384	United Kingdom	2906	Trade publications	26
			1562	Article	116	United States	609	Journals	2361
		Medicine Biochemistry, Genetics, and Molecular Biology Immunology and Microbiology Pharmacology, Toxicology, and Pharmaceutics Agricultural and Biological Sciences	950	Review	71	China	472	Book series	19
			671	Conference essay	45	Germany	182	Books	4
			236	Article in press	25	Italy	149	Conference proceedings	4
			205	Book chapter	21	United Kingdom	133	Journals	105
Buffy coat	107	Medicine Immunology and Microbiology Biochemistry, Genetics, and Molecular Biology	79	Article	15	United States	21	Books	1
			43	Review	6	Australia	8	Book series	1
		Agricultural and Biological Sciences Veterinary	22	Book chapter	3	Canada	8		
			15	Conference article	3	Iran	8		
			15	Erratum	2	Brazil	6		
									(continued)

(continued)

TABLE 1. (CONTINUED)

<i>Sample type</i>	<i>Total</i>	<i>Subject area</i>	<i>Document type</i>	<i>Source title</i>	<i>Country</i>	<i>Source type</i>
Tissue	95,474	Medicine Biochemistry, Genetics, and Molecular Biology Agricultural and Biological Sciences Engineering	52,502	<i>PLoS One</i>	2109	Journals
			34,708	<i>Scientific Reports</i>	1794	Conference proceedings
			10,724	<i>Oncotarget</i>	1435	Books
			9585	<i>Chinese Journal of Tissue Engineering Research</i>	1040	Book series
			6925	<i>Progress in Biomedical Optics And Imaging</i>	667	Trade publications
Microbio-me	3084	Medicine Biochemistry, Genetics, and Molecular Biology Immunology and Microbiology Agricultural and Biological Sciences Nursing	1698	<i>Proceedings of SPIE</i>	131	Journals
			962	<i>Scientific Reports</i>	121	Books
			737	<i>PLoS One</i>	1042	Books
			631	<i>Frontiers in Microbiology</i>	84	Book series
			198	<i>Gut Microbes</i>	38	Conference proceedings
				<i>Food and Function</i>	31	205

PBMC, peripheral blood mononuclear cell.

FIG. 2. The number of publications according to the analyte type. Publications on cell-free DNA (cfDNA)-, DNA-, RNA-, protein-, or metabolite-related search terms were extracted from the retrieved publications on serum (A), plasma (B), peripheral blood mononuclear cell (PBMC) (C), buffy coat (D), and tissue (E).

published in 1994. In 2016, the number of publications on next-generation RNA sequencing search terms was also the greatest in tissue-related publications, followed by serum-, plasma-, and PBMC-related publications. Publication on next-generation RNA sequencing search terms of buffy coat-related publications has not been published until 2016.

Publications related to mass spectrometry

Publications on mass spectrometry search terms were extracted from the searched publications related to serum, plasma, PBMC, buffy coat, and tissue (Fig. 4). The total number of publications related to mass spectrometry showed a tendency to increase greatly since the 2000s. Among publications searched using protein and mass spectrometry search strings, serum-, plasma-, and tissue-related publications showed a tendency to increase with a similar upward curve, but the growth has slowed

in recent years. Tissue-, plasma-, and serum-related publications are 738, 680, and 532 in 2016, respectively. In case of publications searched using metabolite and mass spectrometry search strings, the number of plasma-related publications was the greatest each year, followed by tissue- and serum-related publications. Plasma-, tissue-, and serum-related publications are 549, 313, and 309 in 2016, respectively.

Discussion

These publication trends by the type of biosamples illuminate roadmaps by which biobanks might establish and diversify their biosample collections in the future. The number of tissue-related publications was annually the highest than others. However, the sum of the number of publications related to serum and plasma was generally more than the number of tissue-related publications. The same tendency was also found when searching

FIG. 3. The number of publications related to next-generation DNA (A) and RNA (B) sequencing.

using a Embase database (<https://www.embase.com>) (data not shown). A previous study investigated references on biobank search terms that registered in the Scopus database between 1939 and 2014 (Astrin and Betsou, 2016). When most frequent meaningful words contained in the title of biobank-related references were analyzed, the term *blood* was the most common, followed by *cell* and *tissue*. These findings suggest that the trends in biosample-related publications are associated with the trends in the biobank-related publications.

Recently, the number of publications on gut microbiome has increased dramatically and overtook the number of PBMC- and buffy coat-related publications. Microbiome- and metabolite-related biosamples can help diversify future biosample collections, while tissue collections appear to maintain their importance in scientific publications.

The gut microbiome influences human metabolic and other functions (O'Hara and Shanahan, 2006). Dysbiosis can increase the risk of diseases such as obesity (Turnbaugh et al., 2006), diabetes (Qin et al., 2012), various inflammatory bowel diseases (Gevers et al., 2014), and autoimmune diseases (Proal et al., 2013). Fecal microbiota transplantation from healthy individual to patient could become a therapeutic method for treatment of some diseases (Khoruts et al., 2010). There may be unique requirements for biobanking of gut microbiota that are distinct from biobanking of other biosamples; for example, the definition of healthy microbiota

should be established and there should be discussion about compensation for the donor (Bolan et al., 2016). Proper informed consent should be prepared for the collection of microbiome for future research and disease treatment. Information on the health status, genome, and lifestyle of the host that affects microbial function should be collected along with microorganisms.

It is also necessary to carry out studies on preanalytical variables and develop standard operation procedures (SOP) on biobanking of the gut microbiome for future research and medical use. The number of references containing the term *organoid* in the abstract section has gradually increased in the 2010s and is 264 in 2016 (data not shown). Organoids, miniature forms of organs, have been emphasizing as a model for cancer precision medicine (Cantrell and Kuo, 2015). Organoids have the potential to be used for disease modeling, drug screening, and drug safety testing (Lancaster and Knoblich, 2014). We think that it may be necessary for biobanks to consider the establishment of organoid biobanking technologies and the collection of organoids.

Researchers from the United States and China published the highest number of serum-, plasma-, PBMC-, tissue-, and microbiome-related references in 2016. The United States and China have been investing heavily in acquiring biosamples and studying in the biomedical field; for example, the United States is promoting the "All of Us" precision

FIG. 4. The number of publications related to mass spectrometry for protein (A) and metabolite (B) analysis.

medicine program (Sankar and Parker, 2017) and China has established and operates the China Kadoorie Biobank (Chen et al., 2011). These observations emphasize that the national scientific performance is influenced by investment scale, infrastructure, and stance for scientific research (Gomes et al., 2016).

Among serum-, plasma-, and tissue-related references, the number of publications on protein search terms was higher than publications on DNA (or cfDNA), RNA, and metabolite search terms every year for the past 40 years. The number of publications on RNA search terms tended to be the second largest. References on protein and RNA search strings accounted for a large proportion in PBMC- and buffy coat-related publications. Preanalytical variations depend on the analyte type (such as DNA, RNA, protein, and metabolite) as well as the processing conditions of biosamples (Lee and Kim, 2017).

To increase the utilization of biosamples, biobanks should strive to secure biosamples of appropriate quality for proteome and transcript research. Hebels et al. (2013) proposed that blood samples for proteomics should be separated ≤ 4 h at room temperature after blood collection. The National Cancer Institute Early Detection Research Network (EDRN; <https://edrn.nci.nih.gov>) SOP states that serum and plasma samples should be separated ≤ 4 h at 4°C after blood collection.

The number of publications on mass spectrometry and next-generation sequencing search terms has rose rapidly since the 2000s and 2010s, respectively. Technology development for next-generation high-throughput sequencing has been actively progressed since the 2010s; for example, Illumina released HiSeq 2000, MiSeq, and HiSeq 2500 in the 2010s. The cost for DNA sequencing was also rapidly reduced since the 2010s (Barba et al., 2014). The technology development for biomedical research may influence the trends in scientific publications. Recently, the integrated analysis of omics data, produced internally or externally (such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics), is actively being carried out to identify complex and dynamic biological mechanisms of diseases (Greenawalt et al., 2012; Li et al., 2013; Serizawa et al., 2011; Sun and Hu, 2016).

However, the challenge remains to develop SOP for preparation of biosamples and for production, storage, management, and quality control of high-throughput data (Börnigen et al., 2015; Dona et al., 2014; Noor et al., 2015). The biobank could need to secure strategic biosample panels with high quality and specialized concept, for integrative analysis of biological big data. Publications on next-generation sequencing search terms were mostly associated with tissue of various biosamples. Tissue heterogeneity, induced by contamination of cells or adjacent tissues different from the target tissue, can cause errors in gene expression profiling (Smith et al., 2009; Zhang et al., 2017) and genome analysis (Meyerson et al., 2010; Yau et al., 2010). The biobank or researcher should try to minimize tissue heterogeneity when collecting tissue samples.

Tools for bibliometric analyses such as the Scopus database and a Derwent Innovation platform (www.derwentinnovation.com) provide information on the trends in publications by keyword, period, research field, topic, country, or journal. These information enable us to analyze changes in research trends and establish new strategies for biobanking. The biobank can also evaluate the value of the acquired biosamples, through analysis of bibliographic information about the articles and patents produced by providing biosamples to

researchers. The EuroBioBank (EBB) Network analyzed research purposes of 255 original articles specifying the use of EBB biosamples (DNA, cell, tissue, serum, and plasma) (Mora et al., 2015). Furthermore, they are developing Bioresource Research Impact Factor (BRIF), a tool for assessment on the research impact of bioresource. These activities will help the biobank's continued development.

In conclusion, our study provides information on the trends of publications by type of biosample. This analysis should help biobanks on future decisions to prioritize and diversify collections among various types of biosamples. To increase the utilization of the biosample, biobanks should collect biosamples with quality for proteome and transcript studies. As the number of publications on the latest high-throughput technologies (mass spectrometry and next-generation sequencing) has increased recently, biobanks need to secure biosamples that are appropriate for integrated analysis of multiomics data.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

- Araujo P, Bjørkkjær T, Frøyland L, and Waagbø R. (2018). Effect of storage time, temperature, antioxidant and thawing on fatty acid composition of plasma, serum and red blood cells—A pilot biobank study. *Clin Biochem* 52, 94–105.
- Astrin JJ, and Betsou F. (2016). Trends in biobanking: A bibliometric overview. *Biopreserv Biobank* 14, 65–74.
- Barba M, Czosnek H, and Hadidi A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136.
- Betsou F, Bulla A, Cho SY, et al. (2016). Assays for qualification and quality stratification of clinical biospecimens used in research: A technical report from the ISBER biospecimen science working group. *Biopreserv Biobank* 14, 398–409.
- Birch K, Dove ES, Chiappetta M, and Gürsoy UK. (2016). Biobanks in oral health: Promises and implications of post-neoliberal science and innovation. *OMICS* 20, 36–41.
- Bolan S, Seshadri B, Talley NJ, and Naidu R. (2016). Biobanking gut microbiome samples. *EMBO Rep* 17, 929–930.
- Börnigen D, Moon YS, Rahnavard G, et al. (2015). A reproducible approach to high-throughput biological data acquisition and integration. *PeerJ* 3, e791.
- Cantrell MA, and Kuo CJ. (2015). Organoid modeling for cancer precision medicine. *Genome Med* 7, 32.
- Chen Z, Chen J, Collins R, et al. (2011). China Kadoorie Biobank of 0.5 million people: Survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 40, 1652–1666.
- Dona AC, Jiménez B, Schäfer H, et al. (2014). Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem* 86, 9887–9894.
- Gevers D, Kugathasan S, Denson LA, et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392.
- Gomes JS, Minasi LB, da Cruz AD, and Rodrigues FM. (2016). Identification of trends in scientific publications related to genetic polymorphisms in gestational diabetes mellitus. *Genet Mol Res* 15, article #: 15027672.
- Greenawalt DM, Sieberts SK, Cornelis MC, et al. (2012). Integrating genetic association, genetics of gene expression, and single nucleotide polymorphism set analysis to identify sus-

- ceptibility Loci for type 2 diabetes mellitus. *Am J Epidemiol* 176, 423–430.
- Hebels DG, Georgiadis P, Keun HC, et al. (2013). Performance in omics analyses of blood samples in long-term storage: Opportunities for the exploitation of existing biobanks in environmental health research. *Environ Health Perspect* 121, 480–487.
- Khoruts A, Dicksved J, Jansson JK, and Sadowsky MJ. (2010). Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent *Clostridium difficile*-associated diarrhea. *J Clin Gastroenterol* 44, 354–360.
- Lancaster MA, and Knoblich JA. (2014). Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science* 345, 1247125.
- Lee JE, and Kim YY. (2017). Impact of preanalytical variations in blood-derived biospecimens on omics studies: Toward precision biobanking? *OMICS* 21, 499–508.
- Li Q, Seo JH, Stranger B, et al. (2013). Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* 152, 633–641.
- Meyerson M, Gabriel S, and Getz G. (2010). Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 11, 685–696.
- Mora M, Angelini C, Bignami F, et al. (2015). The EuroBioBank Network: 10 years of hands-on experience of collaborative, transnational biobanking for rare diseases. *Eur J Hum Genet* 23, 1116–1123.
- Noor AM, Holmberg L, Gillett C, and Grigoriadis A. (2015). Big data: The challenge for small research groups in the era of cancer genomics. *Br J Cancer* 113, 1405–1412.
- O'Hara AM, and Shanahan F. (2006). The gut flora as a forgotten organ. *EMBO Rep* 7, 688–693.
- Piletić K, and Kunej T. (2017). Minimal standards for reporting microRNA:target interactions. *OMICS* 21, 197–206.
- Pirih N, and Kunej T. (2017). Toward a taxonomy for multi-omics science? Terminology development for whole genome study approaches by omics technology and hierarchy. *OMICS* 21, 1–16.
- Proal AD, Albert PJ, and Marshall TG. (2013). The human microbiome and autoimmunity. *Curr Opin Rheumatol* 25, 234–240.
- Qin J, Li Y, Cai Z, et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60.
- Sankar PL, and Parker LS. (2017). The Precision Medicine Initiative's All of Us Research Program: An agenda for research on its ethical, legal, and social issues. *Genet Med* 19, 743–750.
- Serizawa RR, Ralfkiaer U, Steven K, et al. (2011). Integrated genetic and epigenetic analysis of bladder cancer reveals an additive diagnostic value of FGFR3 mutations and hypomethylation events. *Int J Cancer* 129, 78–87.
- Smith MJ, Culhane AC, Donovan M, et al. (2009). Analysis of differential gene expression in colorectal cancer and stroma using fluorescence-activated cell sorting purification. *Br J Cancer* 100, 1452–1464.
- Sun YV, and Hu YJ. (2016). Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv Genet* 93, 147–190.
- Turnbaugh PJ, Ley RE, Mahowald MA, et al. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031.
- Yau C, Mouradov D, Jorissen RN, et al. (2010). A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biol* 11, R92.
- Zhang JD, Hatje K, Sturm G, et al. (2017). Detect tissue heterogeneity in gene expression data with BioQC. *BMC Genomics* 18, 277.

Address correspondence to:

Jae-Eun Lee, PhD

Division of Biobank for Health Sciences

Center for Genome Science

Korea National Institute of Health

Korea Centers for Disease Control and Prevention

Osong Health Technology Administration Complex

187, Osongsaengmyeong2-ro

Osong-eup

Heungdeok-gu

Cheongju-si

Chungcheongbuk-do, 28159

Korea

E-mail: eunwin21@daum.net;

eunwin21@korea.kr

Abbreviations Used

BRIF = Bioresource Research Impact Factor

cfDNA = cell-free DNA

EBB = EuroBioBank

EDRN = National Cancer Institute Early Detection Research Network

PBMC = peripheral blood mononuclear cell

SOP = standard operation procedures