SPECIAL ARTICLE | Genetics inMedicine

Mapping the inputs, analyses, and outputs of biobank research systems to identify sources of incidental findings and individual research results for potential return to participants

Heather R Bemmels, MS[1], Susan M Wolf, JD[2] and Brian Van Ness, PhD[3]

Progress in the debate over returning incidental findings (IFs) and individual research results (IRRs) to research participants who provide specimens and data to biobanks in genetic and genomic research requires a new tool to allow comparison across heterogeneous biobank research systems and in-depth analysis of the sources and types of findings generated for potential return. This article presents a new visual mapping tool to allow systematic and standardized depiction of (i) the specimens initially collected, (ii) the materials and data sets then created, (iii) the analyses then performed, and finally (iv) the genetic and genomic results generated, including potential IFs and IRRs. For any individual biobank research system, this sequence of four maps can be created to anticipate the sources and types of IFs and IRRs to be generated, to plan how to handle them, and then to manage them responsibly over time. We discuss how this four-map tool was created and describe its application to four national biobank systems, thereby demonstrating that this tool can provide a common platform to visualize biobank content, anticipate how IFs and IRRs will arise in a biobank research context, and inform policy development.

Genet Med 2012:14(4):385–392

Key Words: biobank; bioethics; biorepository; genetics; genomics; human subjects research; incidental findings; research ethics; return of results

INTRODUCTION

Active debate surrounds the question of whether and how to disclose to research participants the individual findings that arise in the course of research. These findings may be individual research results (IRRs) that arise in pursuing the explicit aims of the study; or they may be incidental findings (IFs) beyond the aims of the study.[1] The problem of how to manage IRRs and IFs arises in genetic and genomic research because of the increasing potential to discover information about an individual that may confer clinical benefit. Evidence also suggests that many research participants are interested in receiving information that may have health-related significance to them as well as to relatives who may share genetic traits.[2,3] The potential impact of IFs and IRRs is rapidly increasing because of advancing genome-wide technologies, including whole-genome sequencing, which may generate information relevant to disease risks and outcomes. Because data held and generated by biobank research systems may have health-related significance to the individual sources of data and specimens, biobanks are now beginning to face the question of whether and how to return IFs and IRRs. Yet progress in understanding the proper role of biobanks has faced two major barriers: (i) the tremendous heterogeneity of biobanks, which makes comparison difficult, and (ii) the complexity of biobanks and the larger associated research systems, with a range of inputs (data and specimen types), analyses run, results generated, and specific potentials for IFs and IRRs.

The debate over whether and how to return IFs and IRRs to participants has generated a significant amount of literature and a number of consensus guidelines on appropriate conditions for the return.[1,4,5] However, these guidelines have shed little light on the role and responsibility of biobanks in managing IFs and IRRs. The consensus paper by Wolf et al. in this symposium[6] is the first concerted effort to address the biobank issues and offer guidelines in the United States. The term "biobank" is used in that paper to broadly encompass organized collections of samples and data including biorepositories and databases. Consequently, here "biobank" is defined as collections of human biological samples, materials, and data sets, and the phenotypic, clinical, and outcome data that may accompany them. As a research resource, a biobank may perform four functions: (i) sample and data collection, (ii) sample processing and production of derived materials and data sets, (iii) storage (for future research), and (iv) generating and/or archiving analytical results (e.g., associations of genetic data with outcomes). Biobanks may supply samples and/or data to secondary investigators outside the biobank to perform further research. Wolf et al.[6] use the term "biobank research system" to refer to all four biobank

[1]Department of Educational Psychology, University of Minnesota, Minneapolis, Minnesota, USA; [2]University of Minnesota Law School and Medical School, Consortium on Law and Values in Health, Environment and the Life Sciences, University of Minnesota, Minneapolis, Minnesota, USA; [3]Institute of Human Genetics, University of Minnesota, Minneapolis, Minnesota, USA. Correspondence: Brian Van Ness (vanne001@umn.edu)

Submitted 11 October 2011; accepted 7 December 2011; advance online publication 1 March 2012. doi:10.1038/gim.2011.69

functions, whether performed at collection sites, the biobank itself, or secondary research sites.

Making progress in the IF/IRR debate requires an understanding of biobank content, the scope of samples and data held within the bank, the analyses performed, and the results generated. Because of the variety of biobank research systems, a common platform for defining biobank characteristics, including the pathways within those systems that result in identification of IFs and IRRs, is essential. This article presents the development and application of such a platform, a new tool for use in IF/IRR analysis. This tool represents a significant advance over current visual diagrams depicting biobank organization and governance. Large clinical trial groups such as the Eastern Cooperative Oncology Group and national repositories such as the Cancer Human Biobank of the National Cancer Institute (http://biospecimens.cancer.gov/cahub) have created diagrams of processes and governance.[7] The tool we have devised goes further, depicting the scope of the collection, derived materials, data creation, analyses, and results. Our tool shows the sources of IFs and IRRs that should be considered for return, the analyses generating them, the types of IFs and IRRs generated, and how those IFs and IRRs relate to other components of biobank research systems.

Here, we present the construction, content, and utility of this new biobank mapping tool. This article presents the tool in three parts, starting with an overview of our mapping tool, followed by the application of the tool to four diverse biobank examples, and, finally, the demonstration of the usefulness of this tool in revealing the pathways linking sources and types of IFs and IRRs that may be considered for return across a range of biobank research systems.

MATERIALS AND METHODS

Development of the mapping tool

The mapping tool was created through an inductive, expert-driven process as part of the National Institutes of Health/National Human Genome Research Institute–funded project on managing IFs and IRRs in genetic and genomic biobanks (no. R01-HG003178). The coauthors constructed a preliminary tool to map biobank inputs, analyses, and outputs and to specify the potential of a result to be an IF or IRR. The preliminary set was devised and then refined on the basis of (i) the co-authors' experience in genetic and biobanking research, genetic counseling, and IF/IRR analysis; (ii) analysis of publicly available websites presenting the structure of a range of prominent biobank research systems; and (iii) input and critique from the project's investigator team and multidisciplinary working group. The authors revised the mapping tool through an iterative process.

Publicly available websites consulted were the Framingham Heart Study (http://www.framinghamheartstudy.org), the Marshfield Clinic Personalized Medicine Research Project (http://marshfieldclinic.org/pmrp), the National Children's Study (http://www.nationalchildrensstudy.gov), and the Etiology and Early Markers Study of the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (http://prevention.cancer.gov/plco). These were used to gather additional information about the inputs, analyses, and outputs of a range of biobank research systems. These four biobank research systems include one focused on a study of development, one focused on common diseases and pharmacogenomics (for the purpose of personalized medicine), and two focused on disease-specific studies of two of the most pervasive human health conditions: cancer and cardiovascular disease. These four biobanks are all part of major, federally funded initiatives; additional details about these biobanks including their diverse populations and unique designs are presented in **Supplementary Table S1** online.

The tool was presented to project investigators at regular investigator meetings and at two meetings of the larger project working group in June and December 2010, and then revised according to the input elicited at those meetings. The working group was made up of biobank managers and experts in the fields of law, bioethics, human subject protection, patient advocacy, and genetic and genomic research.

Applying the mapping tool to four biobanks

After devising and refining the mapping tool, we then sought to demonstrate its utility through application to four biobanks that were different from those used (above) to construct it. We selected four new biobank research systems to test whether the tool is sufficiently comprehensive to work with a range of biobanks. Those four banks were the International Myeloma Foundation's Bank on a Cure (BOAC) Biobank (directed by one of the coauthors, B.V.N.), the Genetic Alliance Registry & Biobank (GARB), the NUgene Project, and the Eastern Cooperative Oncology Group (ECOG) Leukemia Biobank. The new biobank research systems were selected such that the sample included two disease-specific biobanks, a biobank focused on the study of genetic factors in common diseases, and a biobank involved in clinical trials. A list of characteristics of the four biobanks, including key characteristics discovered in applying our tool, is presented in **Supplementary Table S2** online.

Directors from the four biobanks were first contacted by e-mail to schedule an interview. The office of the Human Research Protection Program at the University of Minnesota advised that this was not human subject research, so an application to the institutional review board was not required for exemption. All directors contacted agreed to participate in the interview; three agreed to participate over the phone and one in person. Two of the four interview participants were also members of the project working group involved in providing feedback to help us revise the mapping tool.

The first author (H.R.B.) conducted the four interviews to gather information for testing the comprehensiveness and utility of the mapping and specification tool. A common, structured interview guide used in the interviews was developed. The guide included items that correlate to the components of the mapping tool. Questions asked sought to obtain information about aspects of the selected biobanks that relate

to the potential for generating and the means of handling IFs and IRRs.

RESULTS AND DISCUSSION

Part I: the mapping tool

The biobank mapping tool functions at two levels. First, the tool consists of four complementary mapping figures that sequentially detail (i) the collected biological samples, (ii) derived materials and data sets, (iii) analyses, and (iv) genetic and genomic results, including potential IFs and IRRs. This function improves and standardizes the characterization of genetic and genomic biobank research systems. Second, the series of figures creates a pathway that can be followed from collected samples to the final results, illuminating specific sources and characteristics of IFs and IRRs. An overview of the variables mapped by the tool is presented in **Table 1**.

As seen in **Supplemental Table S2** online, the four biobank systems to which we applied the tool show a range of sample collections, derived materials, research data, and analyses. It became apparent that the operation of those systems is complex. For instance, sample contribution for biobanks may come from a single site at the biobank and/or from one or more sites remote from the biobank; processes to derive biological materials may be performed on-site at the biobank and/or at outside laboratories and then transferred to the biobank; the generation of data sets may be done within and/or outside the biobank; and biobanks may perform their own analyses, provide samples and/or data to the researchers outside the biobank for study, or both. The mapping tool was designed to depict this variation and complexity.

First, the mapping tool focuses on the biological samples collected, as represented in **Figure 1** by a trapezoid. Showing collected samples in the tool is helpful because research studies are ultimately dependent on the nature of the original collection. Although noncellular collections may be banked (e.g., serum and plasma, cerebrospinal fluid), here we focus on cellular-derived samples because they provide the source from which genetic materials are derived. These are the samples most relevant to potential return of IRRs and IFs from genetic and genomic research.

Second, the mapping tool (**Figure 2**) shows the important relationships between the genetic and genomic data sets and the biological materials from which the data sets are generated. Although some initially collected samples are banked and studied in their original form, the significance of many samples in genetic and genomic research comes from processing

Table 1 Overview of the variables mapped by the mapping tool

Biobank structural component	Collections	Specific items, procedures, and outcomes
Collected samples	Blood	Blood spot, cord, venous (including plasma, serum)
	Cells	Breast milk, buccal/skin, feces, finger/toe nails, hair, placental material, reproductive, urine
	Tissues	Nontumor, tumor
Derived materials	Cells/cell lines	
	Nucleic acids	DNAs: nuclear DNA, mtDNA, cDNA
		RNAs: mRNA, miRNA, other RNAs
Derived data	Cytogenetic	CNVs, FISH, karyotype
	Epigenetic and imprinting	e.g., methylation profiles, chromatin
	Genetic and genomic	From DNAs: markers (e.g., SNPs, haplotypes, CNVs), modification, sequence
		From RNAs: expression profiles, cDNA sequence
Analyses	Cytogenetic	aCGH, FISH, karyotyping
	Epigenetic and imprinting	e.g., methylation, chromatin
	Genetic and genomic	Using DNAs: markers (e.g., association, linkage, haplogroup identification), modification (e.g., DNA adduct), sequencing (e.g., targeted, whole genome)
		Using RNAs: expression
Results	Cytogenetic	Neutral, deletion, expansion, sex chromosome constitution
	Epigenetic and imprinting	e.g., methylation: neutral, hypermethylation, hypomethylation
	Genetic and genomic	With expression: low, high, normal, expression of structural variants
		With linkage: evidence for linkage, linkage excluded
		With markers: ancestry, physical traits, sex chromosome constitution, variants (i.e., predictive, protective, unknown significance)
		With sequencing: benign polymorphism, deleterious mutation, neutral, variant of unknown significance

aCGH, array comparative genomic hybridization; cDNA, complementary DNA; CNV, copy number variant; FISH, florescence in situ hybridization; mRNA, messenger RNA; miRNA, microRNA; mtDNA, mitochondrial DNA; SNP, single-nucleotide polymorphism.

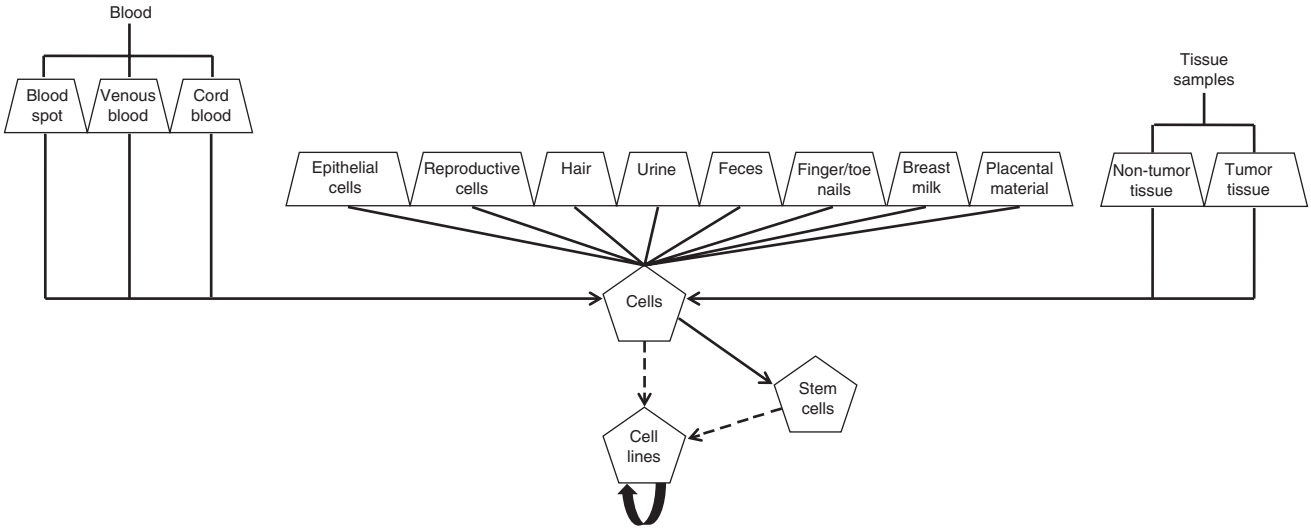

Figure 1 The first map: human biological samples collected for genetic and genomic biobanking research. When investigating whether incidental findings and individual research results are possible, one must first consider the types of samples collected and their derivatives. The trapezoid is used to symbolize the samples that are initially collected for biobanking research. The pentagon is used to represent derived materials.

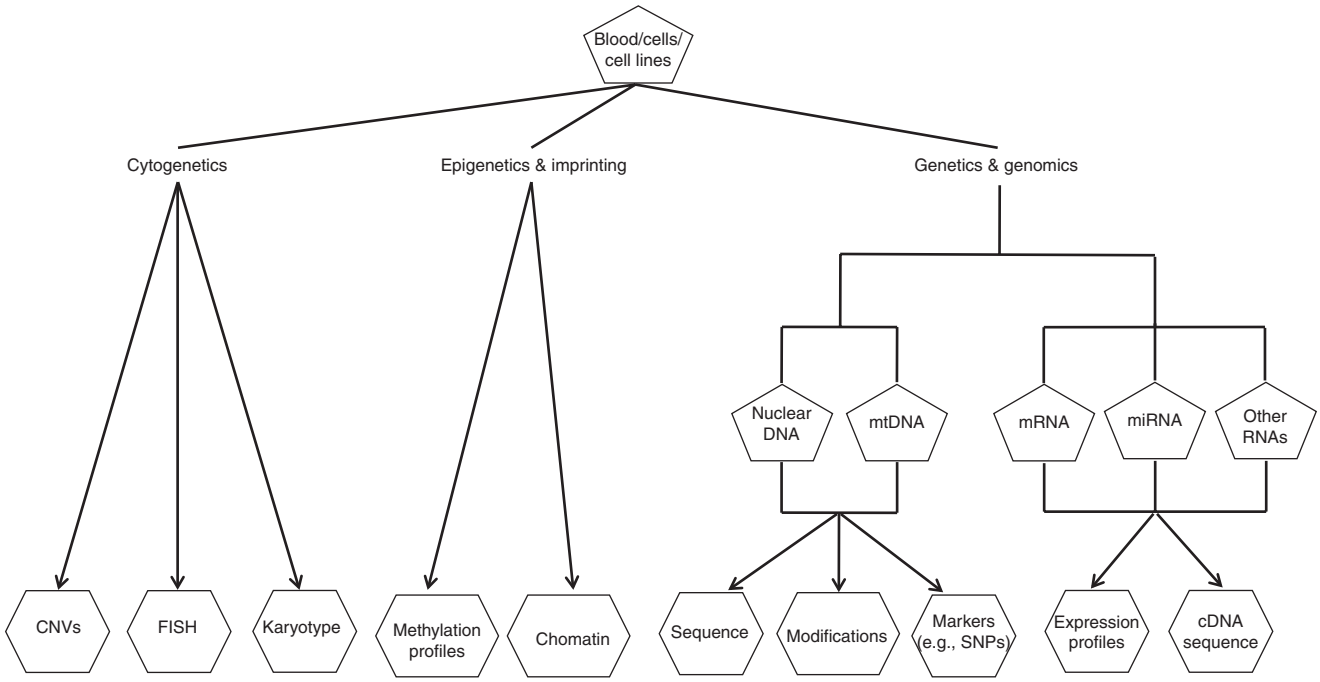

Figure 2 The second map: biological data sets stored in genetic and genomic biobanking research. The second consideration necessary for management of potential incidental findings and individual research results is the data sets created and stored. The hexagon is used to represent biological data sets. Derived biological materials are symbolized by the pentagon. cDNA, complementary DNA; CGH, comparative genomic hybridization; FISH, florescence in situ hybridization; mRNA, messenger RNA; miRNA, micro RNA; mtDNA, mitochondrial DNA; SNPs, single-nucleotide polymorphisms.

them into secondary biological materials—referred to here as "derived materials." Those newly derived materials are also then stored in biobanks and used in research. The pentagon denotes the derived biological materials. Derived materials for genetic and genomic research include cells (including stem cells), transformed cell lines, and subsequently isolated nucleic acids such as nuclear or mitochondrial DNA, RNAs (i.e., messenger RNA, microRNA, and other RNAs such as small-interfering RNA), and laboratory generated, complementary DNA.

In **Figure 2**, hexagons are used to represent data sets commonly used and stored in contemporary genetics and genomics. Thus, we have included derived information (genomic data sets) as a component of the archived materials in a biobank. Indeed, some biobanks now store genomic information only in the form of coded data sets (e.g., the Database of Genotypes and

Phenotypes, available at http://www.ncbi.nlm.nih.gov/gap), in order to supply electronic data for analysis by outside researchers. **Figure 2** provides detail on information stored as cytogenetic data sets, such as copy number variants, fluorescence in situ hybridization results, or karyotype results; epigenetic data sets including methylation profiles or other modifications (e.g., chromatin); data sets for nuclear or mitochondrial DNA results including sequence, genetic marker (e.g., single-nucleotide polymorphisms (SNPs), haplotypes, copy number variants), and DNA modification (e.g., DNA adducts); expression profiles for messenger RNA, microRNA, and other RNAs; and complementary DNA sequence.

Figure 3 shows the range of analyses performed on the biological materials and derived data sets. The tool uses circles to represent the analyses. The tool maps genetic and genomic analyses based on three structural domains: cytogenetic, epigenetic and imprinting, and molecular genetic and genomic information. IFs and IRRs are possible in any of those domains. The analyses in the cytogenetic domain consist of karyotyping, fluorescence in situ hybridization, and comparative genomic hybridization. The analyses in the epigenetic and imprinting domain include methylation and other chromatin studies. The analyses in the molecular genetic and genomic domain are commonly done using different platforms: genotyping for analyses of genetic markers (e.g., association, haplogroup identification, linkage), biochemical for analyses of genetic modifications (e.g., DNA adduct), expression, and sequencing. **Figure 3** also represents the scope of analyses by demonstrating that the numbers of sites or loci involved in a study vary from a single genetic site to whole exome and genome. Pentagons and hexagons are reproduced in **Figure 3** for cases in which derived materials and/or data sets can result in multiple analyses.

The mapping tool finally identifies possible results from the genetic and genomic biobank research studies. These results are represented in **Figure 4** by rectangles and categorized into four groups based on the possible meaning(s) of the finding. The categories are (i) information such as ancestry or phenotypic traits (e.g., earwax type, eye color) that may not be considered directly related to health, (ii) results considered neutral (e.g., variations or copy number variants considered benign or normal variation), (iii) inherited and acquired genetic variants that are typically deleterious—but can be protective—and cause known health consequences, and (iv) variants for which the significance is uncertain. **Figure 4** maps the results that may be of interest to research participants or the individual sources of data and specimens, especially due to health significance.

As shown in **Figure 4**, symbols are used to signify which of the four categories each result may represent (as detailed in the figure legend). **Figure 4** thus organizes results into types based on their potential for being an IF or IRR of health significance

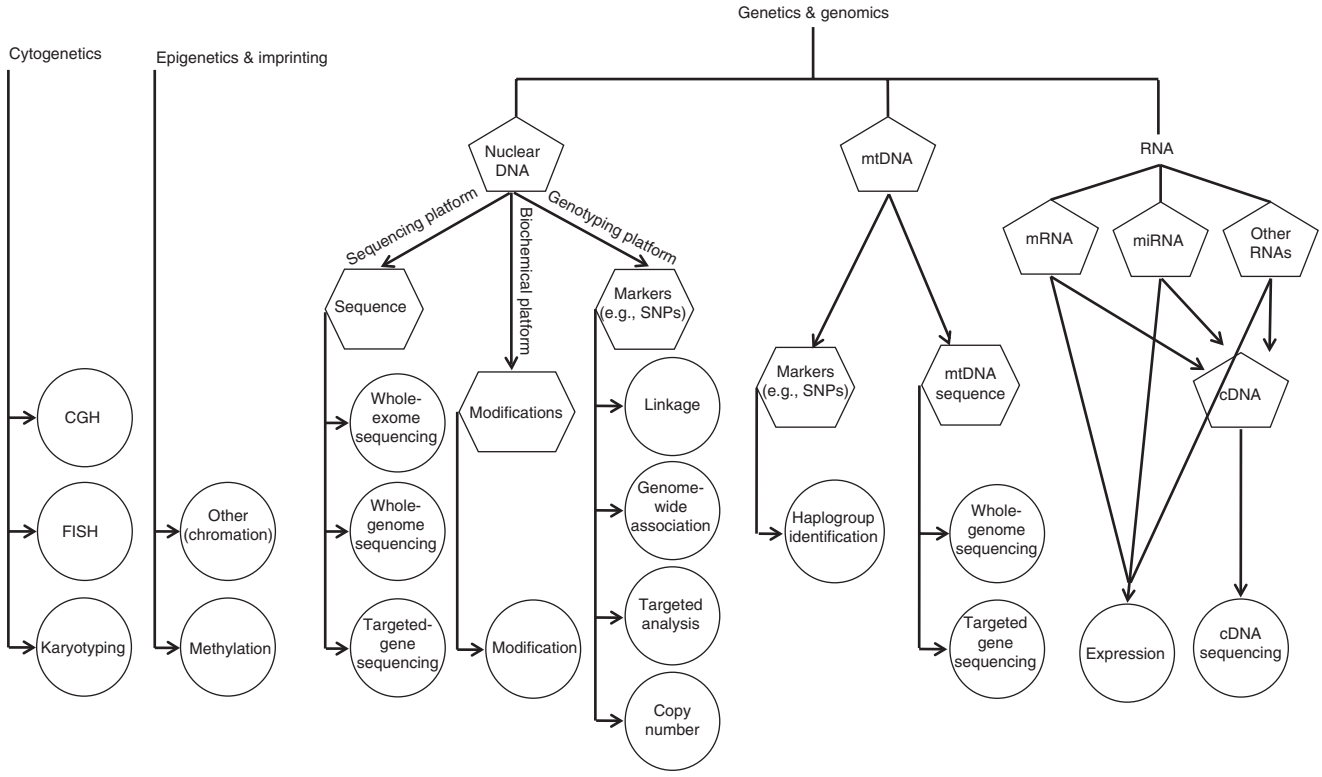

Figure 3 The third map: genetic and genomic analyses. The third consideration necessary for management of potential incidental findings and individual research results is the types of analyses performed on the biological samples, materials, and data sets. The analyses are represented by circles. Where data sets and derived materials can result in multiple analyses, the hexagons and pentagons are reproduced, respectively. cDNA, complementary DNA; CGH, comparative genomic hybridization; FISH, florescence in situ hybridization; mRNA, messenger RNA; miRNA, micro RNA; mtDNA, mitochondrial DNA; SNPs, single-nucleotide polymorphisms.

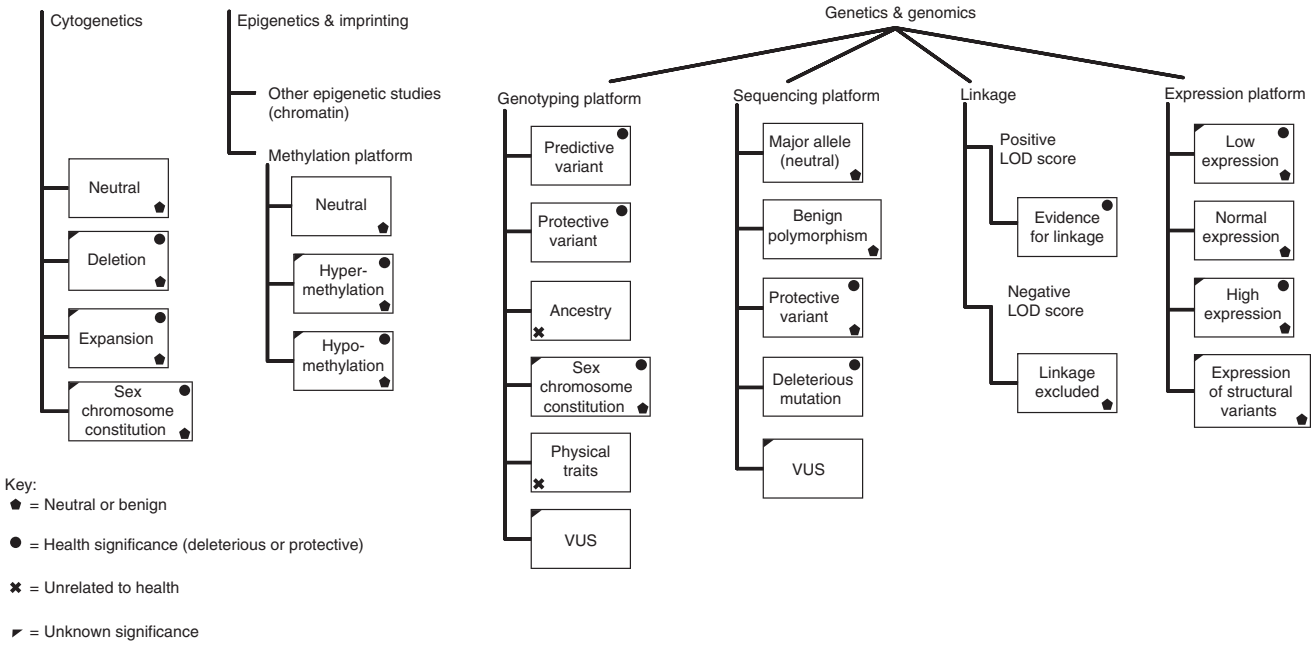

Figure 4 The fourth map: results of genetic and genomic analyses and associated meanings. The results are coded with symbols to represent the following categories: (i) information, such as ancestry or traits, that is not generally thought of as related to health (X), (ii) results typically considered "normal" (i.e., no detected variations or changes such as polymorphisms or copy number variants that are generally considered benign or normal variation) (circle), (iii) inherited and acquired variants that are typically considered abnormal and have associated health or reproductive consequences (solid hexagon), and (iv) variants for which the clinical significance is unknown (upper left corner shaded). LOD, log of the odds; VUS, variant of unknown significance.

and thus potentially returnable. Note that this is a narrower interpretation of returnability than some authors propose; Fabsitz et al., for example, would allow a researcher discretion to return a wider set of IFs and IRRs.[5] This fourth mapping tool, with its indication of what IFs and IRRs are potentially returnable based on health significance, provides a focal point for developing governance policies on return.

Part II: applying the mapping tool to four biobanks

The value of the mapping tool is in its ability to be applied to diverse biobank research systems to illuminate IF/IRR issues. To demonstrate utility, the tool was applied to four new test biobanks selected for their diverse aims, designs, and contents. The characteristics that informed our choice of the four test biobank research systems are presented in **Supplementary Table S2** online. The **Supplementary Figures S1–S4** online show our application of the mapping tool to each biobank. The unique components of each biobank research system are highlighted in yellow. Our intent was to determine whether the mapping tool was sufficiently comprehensive to accommodate a new set of diverse biobank systems, revealing sources of potential IFs and IRRs that might be considered for return.

The International Myeloma Foundation's BOAC Biobank (see **Supplementary Figure S1** online) receives venous blood, buccal cells, and both tumor and nontumor samples from multiple clinical trials, in addition to obtaining those types of samples directly from patients and controls through a mail-in mouthwash sample. The BOAC Biobank derives cells and isolates DNA from the samples, but does not derive cell lines. The BOAC laboratory generates SNP data that are held in the databank. Clinical trial groups in which patients are enrolled provide access to clinical outcomes. The investigators at the BOAC Biobank analyze disease risk and therapeutic outcomes associated with targeted SNPs (3,400 variants). Those analyses can be traced through the mapping tool to reveal that the research at the BOAC Biobank could generate results with any of the four meanings (normal, abnormal, non–health related, or uncertain significance). For example, researchers at the BOAC Biobank may identify a predictive deleterious variant for outcomes related to myeloma (IRRs) or health impacts beyond myeloma (IFs). Yet, because all SNPs investigated in BOAC analyses are limited to targeted, single-nucleotide variants selected for their possible association with cancer, the likelihood of finding IFs beyond the scope of the research study is more limited, as compared with studies that employ a broader (e.g., genome-wide, whole exome) methodology.

Application of the mapping tool to the GARB and associated research system is presented in **Supplementary Figure S2** online. The GARB is a cooperative model with multiple member organizations and a registry and biobank that aggregate clinical data, medical records, and biological samples and materials. The clinical information collected by individual member organizations covers a common set of data elements, with the addition of disease-specific information. The samples collected by individual member organizations currently include venous blood, cord blood, plasma and serum (for nongenetic analyses including biomarker studies), peripheral blood mononuclear cells, buccal cells, and tissues including nonaffected nontumor

organ tissue samples, breast tumor tissue, and tissues from organ harvest. The GARB, however, is a flexible system in which samples collected are directed by the member organization, so the types of samples collected change as the research aims of member organizations evolve. The GARB frequently derives and stores isolated DNA and transformed cell lines from the blood samples. The GARB generates karyotype, sequence, and marker (SNPs) data from karyotyping, targeted gene sequencing, and SNP-based GWAS and linkage studies, respectively. By following the path from analyses to results, the mapping and specification tool shows that research by the GARB and member organizations could generate results with health-related significance of different kinds. Unlike the BOAC Biobank, the broad genome-wide methodology of the GWAS analysis in the GARB increases the potential to generate IFs as compared with the potential of more targeted analyses.

The mapping of the biobank research system for the NUgene Project is demonstrated in **Supplementary Figure S3** online. The NUgene Project collects only venous blood. The NUgene Project biobank does not store the blood, but instead creates and stores isolated DNA. From the DNA, researchers associated with the NUgene Project obtain SNP and copy number data on subsets of NUgene samples through GWAS analysis. Affiliated collaborators apply to investigate the DNA samples using additional methodologies. Researchers must re-deposit their genotyping data in the NUgene Project biobank for storage and enrichment of the available NUgene resources. By following the pathway from analyses to results, our mapping tool again demonstrates the types of results obtained with the NUgene Project, which can include IFs and IRRs of possible health-related significance that might be considered for return to participants. This is similar to the other biobanks using marker studies and genome-wide methodologies. Although not visually represented in the mapping tool, the NUgene Project also collects and stores phenotypic data that may contain IFs, including electronic medical record data, demographic and participant data, and survey data on environmental exposures. The mapping tool shows that analysis of multiple data sets results in potentially important health-related information that might be considered for return.

The outcome of applying the mapping tool to the ECOG Leukemia Biobank and research system is presented in **Supplementary Figure S4** online. The complex, cooperative ECOG Leukemia Biobank is governed by a set of researchers who collect samples and send them to the biobank that serves almost exclusively as a storage facility for biological materials, and several separate research labs perform nearly all the analyses for the project. The tool maps the samples initially collected through the larger ECOG group collection sites. The ECOG Leukemia Biobank then derives cells from both blood and marrow samples and often creates transformed cell lines. The ECOG Leukemia Biobank itself generates no biological data sets and conducts no genetic studies. Rather, secondary researchers apply to obtain and analyze samples stored in the ECOG Leukemia Biobank. Thus, the major potential for IFs and IRRs from this project is generated by the secondary research labs. Certainly, in order to complete the entire mapping tool, it is necessary that information on all aspects of a biobank research system be collected, including information derived from secondary research sites. In this case, the informant for the ECOG Leukemia Biobank emphasized that there were so many secondary research projects that providing information on all of them would be difficult. As a result, the fourth tool mapping the ECOG Leukemia Biobank research system may be incomplete and dependent on policies governing how data are collected and returned to the central biobank.

Part III: implications for IFs and IRRs

Comparing across the four applications of this new tool shows the utility of the tool. As shown in the **Supplementary Figure S2** online, of these four biobanks, Genetic Alliance faces the broadest set of results and the most diverse set of potentially returnable IFs and IRRs. This suggests the greatest need to anticipate and set up processes and policies to manage these IFs and IRRs. Yet it is not simply the fourth map (showing results) that illuminates this. In **Supplementary Figure S1** online, map 1 for Genetic Alliance shows the broadest range of inputs into the biobank research system, map 2 shows the diverse materials and data sets generated, and map 3 shows the broadest range of analyses performed. This sequential picture of how the diverse set of ultimate results originates helps signal that managing IFs and IRRs will involve not just analyzing final results, but also potentially intervening at earlier points. For example, if the biobank itself conveys materials to secondary researchers for analysis, they will want to consider addressing in material use agreements or data use agreements with those researchers how the IF and IRR issues will be handled, if the secondary analyses yield potentially returnable IFs or IRRs.

Comparison across the four applications of this new tool also shows that although several biobanks may generate similar results for potential return (e.g., the results of marker studies indicated on **Supplementary Figures S1–S3** online), the specific IFs or IRRs questions they face may be different. This is in part because different predictive variants, for example, (as shown on the fourth map in each set) may be generated using different analyses (as shown in the third map in each set), generating different levels of confidence in the results. Most guidelines on the return of IFs and IRRs require a high level of confidence in the results as a precondition for offering them to individual sources of data and specimens. As this suggests, applying all four sequential maps to each biobank research system illuminates the IF and IRR issues in greater depth than simply applying the fourth map to show expected types of results.

The mapping tool presented here thus serves as a broadly applicable launching pad for future development, implementation, and assessment of plans for managing and returning IFs and IRRs from genetic and genomic biobank research systems. Comprised of four figures, the tool sequentially and visually maps the inputs, analyses, and results of biobank research systems, including potential IFs and IRRs. We developed the

tool using a set of four biobanks and have demonstrated that the tool is comprehensive enough to map an additional four biobanks, providing visual access to contents and potential IFs and IRRs of health significance. Because the tool is visual, the maps generated are easy to understand, but reveal the complexity of the content (samples, data sets, analyses, and results). The tool can help address core issues of analytic validity, clinical utility, clinical significance, and actionability—issues that the guidelines published to date have deemed central to decisions about potential return.[1,5,6]

Our mapping tool has limitations. The content and processes of biobank systems will likely change as new technologies are developed, and this tool will need to evolve to depict these innovations accordingly. In addition, the tool does not map all aspects of biobank research systems germane to the IF/IRR debate. For instance, the tool is not comprehensive of all possible sources of IFs and IRRs in genetic and genomic biobanking research. IFs and IRRs can be identified in the course of quality control processes, and in clinical and/or phenotypic data. Also, the mapping tool does not map biobank systems for other sets of human (e.g., proteomes, immunonomes) and nonhuman biological structures (e.g., microbiomes) because the focus of this initial work to develop a mapping tool is specifically the analysis of an individual's own genetic and genomic structure. In addition, the tool does not specifically address sample identification, CLIA regulatory issues, or the complexity of who returns IFs or IRRs and how they do so.

Regardless of the limitations, this new tool has multiple important uses and represents a significant advance. The tool depicts in a standardized way the contents and processes of biobank research systems that produce IFs and IRRs for potential return. This allows those constructing a biobank research system to anticipate systematically how IFs and IRRs may arise. Similarly, when a biobank research system is already up and running but is now being analyzed to determine what IFs and IRRs are being generated, our tool allows a systematic mapping to answer that question. A biobank research system can also be mapped at multiple points in time, to depict clearly the changes over time that affect its production of IFs and IRRs. The mapping tool will be especially helpful in analyzing biobank research systems in which the analyses are so numerous that the potential for generating IFs and IRRs is highly complex. Once the mapping tool is used to identify the sources and types of IFs and IRRs that a biobank research system can generate, then analysis of how to manage these IFs and IRRs can proceed, including determination of which IFs/IRRs, if any, to offer back to individual sources of data and specimens. The mapping tool can also be used to facilitate comparison and collaboration among biobank research systems to develop common policies, governance approaches, and mechanisms for identifying and returning IFs and IRRs to individuals.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

ACKNOWLEDGMENTS

Preparation of this article was supported by National Institutes of Health (NIH)/National Human Genome Research Institute (NHGRI) grant award no. 2-R01-HG003178-03 on "Managing Incidental Findings and Research Results in Genomic Biobanks & Archives" (S.M. Wolf, principal investigator; B. Van Ness, F. Lawrenz, and J. Kahn, co-investigators). The views expressed in this article are those of the authors and not necessarily those of the NIH or the NHGRI. We thank E. Horn, E. Paietta, B. Van Ness, S. Terry, and W. Wolf for providing information about their biobanks, and the Working Group of the Project on Managing Incidental Findings and Research Results in Genomic Biobanks and Archives for feedback on the mapping tool.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Wolf SM, Lawrenz FP, Nelson CA, et al. Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Ethics* 2008;36:219–48, 211.
2. Kaufman D, Murphy J, Scott J, Hudson K. Subjects matter: a survey of public opinions about a large genetic cohort study. *Genet Med* 2008;10: 831–839.
3. Wendler D, Emanuel E. The debate over research on stored biological samples: what do sources think? *Arch Intern Med* 2002;162:1457–1462.
4. National Bioethics Advisory Commission. Research Involving Human Biological Materials: Ethical Issues and Policy Guidance. Rockville, MD, 1999. http://bioethics.georgetown.edu/nbac/human/oversumm.pdf.
5. Fabsitz RR, McGuire A, Sharp RR, et al. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from an NHLBI working group. *Circ Cardiovasc Genet* 2010;3:574–580.
6. Wolf SM, Crock, BN, Van Ness, B, et al. Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet Med*, 2012;14:361–384.
7. Winickoff D, Neumann LB. Towards a social contract for genomics: property and the Public in The 'Biotrust' Model. *Genomics Soc Policy* 2005;1:8–21.