HHS Public Access

Author manuscript

Clin Genet. Author manuscript; available in PMC 2016 June 08.

Published in final edited form as:

Clin Genet. 2014 July ; 86(1): 50–55. doi:10.1111/cge.12370.

Biobanks and personalized medicine

J.E. Olson[a], **S.J. Bielinski**[a], **E. Ryu**[a], **E.M. Winkler**[b], **P.Y. Takahashi**[c], **J. Pathak**[a], and **J.R. Cerhan**[a]

[a]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

[b]Mayo Clinic Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA

[c]Department of Primary Care Internal Medicine, Mayo Clinic, Rochester, MN, USA

Abstract

We provide a mini-review of how biobanks can support clinical genetics in the era of personalized medicine. We discuss types of biobanks, including disease specific and general biobanks not focused on one disease. We present considerations in setting up a biobank, including consenting and governance, biospecimens, risk factor and related data, informatics, and linkage to electronic health records for phenotyping. We also discuss the uses of biobanks and ongoing considerations, including genotype-driven recruitment, investigations of gene–environment associations, and the re-use of data generated from studies. Finally, we present a brief discussion of some of the unresolved issues, such as return of research results and sustaining biobanks over time. In summary, carefully designed biobanks can provide critical research and infrastructure support for clinical genetics in the era of personalized medicine.

Keywords

biobanking; biobanks; biorepository; personalized medicine

There continues to be very high expectations that the sequencing of the human genome will ultimately revolutionize the practice of medicine, whereby genomics-based knowledge of the individual will lead to more effective prevention, diagnosis, and treatment, often referred to as `personalized medicine' (1, 2). Pharmacogenomics is one area within personalized medicine that could affect virtually every patient, but to date, clinical incorporation has proved slow and challenging because of many factors, including the inherent delay in the initiation of therapy when traditional reactive pharmacogenomics testing at point-of-care is used and the lack of support for commercial electronic health record (EHR) systems to integrate large-scale genomic data linked to automated clinical decision support. New research, such as the `Right Drug, Right Dose, Right Time – Using Genomic Data to Individualize Treatment' protocol (3), is testing the hypothesis that prescribers can deliver genome-guided drug therapy at the point-of-care by using pharmacogenomics data

Corresponding author: Janet E. Olson, PhD, Department of Health Sciences, Research, Mayo Clinic, 200 First Street, SW, Rochester, MN 55905, USA., Tel.: +1 (507) 284 9833;, fax: +1 (507) 266 2478;, olsonj@mayo.edu.

Conflict of interest None.

preemptively integrated in the EHR, with the goal of speeding up the incorporation of pharmacogenomics into clinical practice.

In addition to the barriers to pharmacogenomics implementation mentioned above, there are additional barriers limiting translation and impact of personalized medicine on clinical practice (4–7). Two (of many) important barriers have been the availability of high-quality specimens with phenotypic data and the robust and timely validation of predictive genetic markers, particularly in the clinical context. One solution to these issues that has been advocated is the expansion of biobanking (e.g. serum and tissue) embedded in health care systems to accelerate discovery, validation, and implementation of new genetic and other biomarkers into clinical practice (8, 9). Here, we provide a mini-review of how biobanks can support clinical genetics in the era of personalized medicine.

What is a biobank?

At the most basic level, a biobank is the systematic collection of biological specimens and health information on participants (5, 10), and for genetics research this requires the collection of a source of germline DNA, most commonly extracted from peripheral blood or buccal cells. Ideally, health information including medical, treatment, and lifestyle data at the time of specimen collection is used for accurate phenotyping and exposure assessment. Some biobanks are further enhanced by the availability of clinical follow-up to conduct outcome studies. EHRs can facilitate annotation of biobanks and ultimately can help with the development of clinical decision support as new findings are integrated into the clinical workflow (11). Biobank specimens and data must be approved for use by a human subjects review panel, and ideally includes written, informed consent (12–14), although other models exist (15). Regardless of the consent approach, ensuring the trust of the participants in a biobank is critical (16). Biobanks should also be sufficiently large, as most genetic variants individually are likely to have only a modest or small impact on phenotype; thus, larger studies are required to obtain reliable results (5). There should also be a clear understanding of the source population of a biobank's participants so that inferences derived from studies using a particular biobank are validly generalized. Finally, to be of ultimate value, a biobank must be used to address important clinical questions, which requires a mechanism to ensure access to and governance of the biobank over time, as many issues in biobanking cannot be fully anticipated when one is launched (14).

Types of biobanks

For clinical genetics, some of the most useful biobanks are those based on a specific phenotype (e.g. disease-specific biobank), or groups of phenotypes (17), which is often the starting point for more complex studies. Such biobanks are useful for correlating phenotypic characteristics of a disease with genetic variation and for studying rare diseases (17–19), but in isolation are ultimately limited for understanding the genetics of risk and outcome. To address risk, case–control association studies are commonly used, where an affected case group is compared to a control group, which must be representative of the underlying population that generated the cases and does not have the phenotype of interest (5). Biobanks based on families can also be used to evaluate genetic associations, and have the

added value of also being able to address genetic transmission, penetrance, phenocopies, and anticipation. The value of disease-based biobanks is also greatly increased if the participants are followed over time, which allows linking of genetic variation with prognosis and other health-related outcomes. This may help identify new therapeutic targets, prognostic stratification, or select treatment.

While less common than disease-specific biobanks, biobanks can also be based on an exposure (e.g. occupational cohorts) or may be a `general' biobank, which means participants are selected on neither disease nor exposure history, but rather on some other selection factor, most often location of residence, membership in a defined group, or self-selected volunteers (5, 18–20). These biobanks may be population-based, embedded in a hospital, health system or insurance plan, or recruited through a variety of other mechanisms. Examples of population-based biobanks include the Marshfield Biobank (12) and the UK Biobank (13), although neither is strictly population-based according to the classic epidemiologic definition, but they are enriched for general population participants from defined geographic areas (5). Examples of biobanks embedded in hospitals or health systems include the Kaiser Research Program on Genetics, Environment and Health (Health Maintenance Organization) (21), BioVU (hospital) (15), and the Mayo Clinic Biobank (clinic) (14).

General biobanks are much more flexible as they can support a variety of studies, including cross-sectional studies of genotype–phenotype correlations; case–control studies using a biobank for cases and/or controls; and cohort studies using baseline and follow-up data in a biobank to link genetic variation with health outcomes (5). The cohort study design is particularly strong as samples and data are collected before the onset of disease, and is a particularly powerful approach for studying the interaction of genes and pre-morbid environmental exposures (5). Cohorts with comprehensive and broad-based outcomes can also be used to identify the full spectrum of disease outcomes and can also be used to identify the determinants of healthy aging. One challenge of using biobanks as cohort studies is that they must be of sufficiently large sample size in order to accrue enough cases with the phenotype(s) of interest over a reasonable timeframe and they must have nearly complete follow-up to limit bias (5, 20).

Considerations in setting up a biobank

Consenting and governance

Biobanks encompass a unique research infrastructure that requires different governance mechanisms than project-based research. Rather than providing informed consent for a specific project, participants of most biobanks provide an open consent for multiple future projects, the details of which cannot be provided at the time of enrollment. Indeed, many future projects will use biological concepts and technologies unavailable at the time of biobank consent. The governance mechanism must balance the needs of the scientific community and the participants with an emphasis on the recognition of participants, trustworthiness, and adaptive management (22).

Formal governance structures are a common and necessary component of biobanks. Although the institutional review board (IRB) is an essential component for oversight and safety, most biobanks utilize a formal access or oversight committee to approve the use of samples or data (14, 23). These committees may serve to review the science within the protocols as well as providing stewardship of finite biological samples. Skilled committee members with expertise in scientific, ethical and clinical domains provide an additional level of safety and rigor to projects using a biobank.

Informal governance structures like community advisory boards (CABs) can also be an important component of Biobank governance. Input from the community can add insight into the perspectives of participants when questions arise. Issues with return of results, academic/industry partnerships, and privacy will evolve over the life of a biobank. Participants are important stakeholders in these issues (24), and a CAB can serve as a representative voice for the community. CABs are often used in larger biobanks to solicit input and inform participants about the use of biological samples and data (25). For personalized medicine, CABs can help translate basic discovery into clinical medicine by outlining the needs of both the individual and the community.

Biospecimens

To be successful, biobanks must pay close attention to collecting high-quality specimens that will be useful long into the future. Depending upon the type of biobank, specimens may include blood and blood derivatives (including dried blood spots), urine, saliva, stool, and surgical tissue (e.g. tumor/normal). Some important variables to collect surrounding these specimens include the date/time of the sampling, collection method, details of processing and final storage, and data about delays (19). An important goal is to ensure sufficient sample handling information such that the experience of each sample is completely retraceable. Quality metrics are key indicators of the usefulness of biobank specimens. As DNA is fundamental to genetic studies, numerous methods to estimate quality have been developed, including total DNA yield and DNA amplification by polymerase chain reaction (PCR) (26), and RNA quality assessment by the success of reverse transcription and product length of quantitative real-time PCR products (27). The quality of fluid biospecimens may require assessment in relation to a specific analyte (27). Best practice guidelines have been published, including the Office of Biorepositories and Biospecimen Research (http://biospecimens.cancer.gov/bestpractices/), the NCI Best Practices for Biospecimen Resources (available at http://biospecimens.cancer.gov/practices/) as well as others (27). Collection, processing and storage of specimens have rapidly evolved into biospecimens science (10, 28).

Risk factor data

Because common, multifactorial diseases such as cardiovascular disease or cancer are hypothesized to be caused by a large number of small, often additive, genetic and environmental effects or modest gene–environment interactions (29), the usefulness of biobanks are greatly enhanced when they also have lifestyle and other risk factor data available. With risk factor data, studies can model the independent and joint effects of genetics and environmental factors on disease risk and outcome at both the individual and

the population level. Commonly collected risk factor data includes demographic information, general health and functioning, personal and family medical history, health behaviors (e.g. diet, physical activity, and smoking), and medication use (both prescription and over the counter medications). Systematic collection of family history data can also be highly useful, as it is generally stored as unstructured text in the EHR, making it difficult to cost-effectively retrieve for research studies (30). Risk factors not commonly included in EHRs are particularly useful. This, however, further increases the need for even larger studies and more extensive data collection or pooling of multiple biobanks to achieve the needed sample size, which has led to the era of large consortia (20).

Biobank informatics

The value of biobank samples is enhanced by the presence of a high-quality informatics system to track over time data concerning enrollment and consent; sample acquisition, processing, storage, and distribution; quality assurance/quality control; collection and/or linkage to subject data (such as clinical data); data security and access; and reporting functions (31). These systems play a critical role in providing sample and data accountability and tracking a sample from collection to processing, storage, use and final disposal. Use of barcodes to enhance the tracking is strongly encouraged. Equally important are robust informatics that can manage the vast quantities of data generated from samples. A good informatics system will be able to integrate large volumes of data from multiple sources, including both clinical and research data. Use of recognized standards, including commonly used data elements, enhance the ability to harmonize with other biobanks for pooling projects. A minimum set of informatics system requirements has been outlined and should be incorporated when developing or selecting informatics systems to support biobanking activities (31).

EHR-driven phenotyping

Biobanks that can be linked to EHRs have an especially rich resource from which to draw a wealth of data. However, with the increase in the size of biobanks, a barrier to utilization of these data is the time-consuming and onerous task of manual retrieval of EHR data. Because of this, the development of methods to rapidly extract phenotype data from the EHR is an active area of investigation (32). In particular, the Electronic Medical Records and Genomics (eMERGE) (33) consortium, a network of nine academic medical centers, has demonstrated the effectiveness of EHR-derived phenotyping algorithms for cohort identification to conduct genomeand phenome-wide association studies (30, 34–36).

However, one must acknowledge several challenges in leveraging EHRs for research, including data that are often incomplete, inaccurate, conflicting, highly complex, and potentially biased. Furthermore, electronic data will be increasingly available from what have been considered non-clinical sources such as patient behavior/activity or social networks, and these can be combined with EHR-derived data to create more comprehensive ecological views of patients. These opportunities will naturally uncover issues and challenges around integration, analysis, interpretation and sharing of `big data'.

Uses of biobanks and ongoing considerations

Research using genotype-driven recruitment

As the genetic data from biobanks accumulate, another opportunity arises to advance the cause of personalized medicine in the form of genotype-driven recruitment (GDR) to assist with the functional characterization of disease-associated genetic variants. Unlike the traditional approach to research in which subjects are selected based on phenotype, the wealth of pre-existing genetic data makes it increasingly possible to select the study groups based on genotype (37). In this setting, the frequency of various phenotypes can be compared in genotype-defined groups to determine a likely phenotype for the genotype under study. GDR has been described as a very useful design to elucidate the influence of geno-type on health-related outcomes such as disease risk, treatment response and outcome (38).

However, this type of study design is not without its ethical concerns (38, 39). A primary concern is the potential for disclosure of genetic results during study invitation, in which genetic results may be either explicitly or implicitly disclosed, thus compromising a subjects' right to *not* know their genetic information. On the other hand, if genetic results are not disclosed at the time of recruitment, then a subject is not fully informed on their reason for eligibility. While a recent workshop on the ethical conduct of GDR studies addressed some of these issues (39), the assumption of the workshop was that the original genetic data resulted from traditional research studies with an informed-consent document and IRB oversight. However, genetic data are increasingly available from other sources, such as clinical care and from direct-to-consumer genetic testing services (38). Thus, more work is needed to generate guidelines for all the varied circumstances under which genetic data may become available for researchers.

Gene by environment interaction

Biobanks can also support genetic research in the era of personalized medicine by the discovery and validation of interactions between genetic variants and environmental factors, including medical, treatment, lifestyle, and related data. Many complex diseases are known to be the result of the combined effect of genes, environmental factors, and their interactions (5, 40). Thus, understanding the impact of gene–environment interactions may allow us to provide individualized preventive services or offer personalized treatment after a disease has been diagnosed (41). For instance, studies of differential antidepressant response have suggested that neither genes nor environmental factors (e.g. stressful life events) alone were a sufficient cause, but the interplay between the two was predictive for both depression causation and treatment response (42). While it is crucial to integrate the genetic and environmental factors for the successful identification of gene-environmental interactions, many genetic studies have not collected extensive environmental risk factor data. Furthermore, many environmental and related lifestyle exposures tend to change over time, and therefore these exposures need to be periodically reassessed (40). Thus, data collection needs to be carefully considered for both the initial enrollment and during follow-up to support our ability to discover and validate interactions between genes and other possible risk factors.

Reuse of data

Another source of data in biobanks can be from completed studies conducted using biobank samples that return the research data to the biobank at the completion of the study. This makes possible the secondary use of existing research data and can provide many opportunities for new discoveries beyond the scope of the original study, and reduce the burden on patients who may be at risk from repeated data gathering intrusion into their lives (43). Analysis of data from a different perspective than the original study creates an opportunity for a deeper understanding of the original study finding (43, 44). For example, the Mayo Clinic Biobank requires all the data generated using their materials to be deposited into a secure central database for future use (14); as safeguards to protect privacy and confidentiality, any study using returned data will need an approval from the Access Committee and a separate IRB approval. Popular data types for reuse are genome-wide association and whole-exome sequencing data, which can be used to reduce genotyping costs for subsequent studies and improve the characterization of genetic variants that are clinically relevant and actionable.

Return of research results/incidental findings

Biobanks have the capacity to generate a large amount of genetic data, some of which may have health implications for participants, raising the need to address return of both primary research results as well as incidental findings. Although the recent American College of Medical Genetics Policy Statement calling for clinical genetic testing laboratories to seek and report a defined list of incidental findings was not intended to apply to research settings, it has sparked much debate about the obligation for biobanks to return findings to participants (45). Many recommend that results be returned to biobank participants if certain criteria are met (46–49), although the problem of what to do with genes with pleiotrophic effects is yet to be resolved (50). The Presidential Commission for the Study of Bioethical Issues recommends that researchers should address how incidental findings will be handled during the consent process and that a plan should be developed for managing incidental findings that may arise (51). The process of returning results to participants incurs significant costs to biobanks and must be taken into account when planning for result disclosure (52). When evaluating research results for potential return to biobank participants, it is important to consider established health implications, actionable interventions, test validity, and participant interest in receiving results (49). Input from a CAB can also provide important insights when developing return of results policies and procedures. As most tests are performed in a research laboratory, it is necessary that the results be validated in a Clinical Laboratory Improvement Amendments (CLIA)-certified clinical laboratory. It is also necessary that biobank participants receive appropriate clinical follow-up with a geneticist and/or genetic counselor after receiving research results, which may incur costs to participants (53).

In the Mayo Clinic Biobank, we have opted to return potentially meaningful research results (including incidental findings) to participants (14). For example, we have returned information concerning risk of venous thromboembolism to women due to variation in Factor V Leiden and prothrombin with the rationale that this may impact their decision to use exogenous hormones. During the sample request process, the Mayo Clinic Biobank

Access Committee works with the researcher, the CAB (as needed), and other subject matter experts to determine which research results to return to participants. The working group considers clinical utility (proven therapeutic or preventative interventions possible), personal utility (repercussions for reproductive decision making or life planning), and test validity (analytical validity of the test and ability to replicate results). If it is determined that the results should be returned, participants are offered the opportunity to receive results. If the participant opts to receive his or her personal results, CLIA confirmation of testing is recommended and facilitated. Particular procedures must be flexible enough to evolve, expand, and change over time as personalized medicine continues to grow and evolve.

Sustainability

Biobanks are often launched without a long-term plan for sustaining them (10, 19). While a large component of the cost is the upfront collection and processing of samples, there are significant costs to maintaining samples, data, and access to a biobank. Cost recovery models vary from institutional support to complete support through user fees, although the latter are hard to set, given large initial costs and a life cycle of a biobank over decades (19). More recently, it has been suggested that biobanking in clinical settings might be incorporated into the cost of business and embedded in the fee and insurance reimbursement structure (8, 54).

Conclusions

Biobanks can provide critical research and infrastructure support for clinical genetics in the era of personalized medicine. Studies using biobanks can effectively support discovery and validation of genetic associations and gene–environment interactions, which will inform general biological insights on disease pathogenesis and can ultimately be translated for risk assessment/stratification, new diagnostics, pharmacogenomics, and drug development, all are important to supporting the practice of clinical genetics. This will drive new clinical tests and incorporation of results into the clinical workflow and the regulatory environment (e.g. FDA), which ultimately needs to include clinical decision support. However, the time frame is long term and continuous, and requires deep integration into the clinical environment; indeed, we must also be willing to recognize that actual implementation into clinical care can take a decade or more (55). Ultimately, biobanks embedded in a clinical practice should encourage a learning environment and fundamentally change the `fabric' or `ecosystem' of health care systems to ensure rapid and valid translation of genetics results to patients and the population (8).

Acknowledgements

This project was funded, in part, by the Mayo Clinic Center for Individualized Medicine.

References

1. Guttmacher AE, Collins FS. Welcome to the genomic era. N Engl J Med. 2003; 349:996–998. [PubMed: 12954750]
2. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. Nat Rev Genet. 2007; 8:151–157. [PubMed: 17230201]

3. Bielinski SJ, Olson JE, Pathak J, et al. Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time-using genomic data to individualize treatment protocol. Mayo Clin Proc. 2014; 89:25–33. [PubMed: 24388019]
4. Burke W, Psaty BM. Personalized medicine in the era of genomics. JAMA. 2007; 298:1682–1684. [PubMed: 17925520]
5. Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR. Genetic epidemiology and public health: hope, hype, and future prospects. Lancet. 2005; 366:1484–1498. [PubMed: 16243094]
6. Hamburg MA, Collins FS. The path to personalized medicine. N Engl J Med. 2010; 363:301–304. [PubMed: 20551152]
7. Farrugia G, Weinshilboum RM. Challenges in implementing genomic medicine: the Mayo Clinic Center for Individualized Medicine. Clin Pharmacol Ther. 2013; 94:204–206. [PubMed: 23588321]
8. Fiore LD, D'Avolio LW. Detours on the road to personalized medicine: barriers to biomarker validation and implementation. JAMA. 2011; 306:1914–1915. [PubMed: 22045770]
9. Khleif SN, Doroshow JH, Hait WN. AACR-FDA-NCI Cancer Biomarkers Collaborative consensus report: advancing the use of biomarkers in cancer drug development. Clin Cancer Res. 2010; 16:3299–3318. [PubMed: 20501613]
10. Hewitt RE. Biobanking: the foundation of personalized medicine. Curr Opin Oncol. 2011; 23:112–119. [PubMed: 21076300]
11. Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet. 2011; 12:417–428. [PubMed: 21587298]
12. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. Pers Med. 2005; 2:49–79.
13. Ollier W, Sprosen T, Peakman T. UK Biobank: from concept to reality. Pharmacogenomics. 2005; 6:639–646. [PubMed: 16143003]
14. Olson JE, Ryu E, Johnson KJ, et al. The Mayo Clinic Biobank: a building block for individualized medicine. Mayo Clin Proc. 2013; 88:952–962. [PubMed: 24001487]
15. Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008; 84:362–369. [PubMed: 18500243]
16. Gaskell G, Gottweis H. Biobanks need publicity. Nature. 2011; 471:159–160. [PubMed: 21390108]
17. Terry SF, Horn EJ, Scott JA, Terry PF. Genetic alliance registry and biobank: a novel disease advocacy-driven research solution. Pers Med. 2011; 8:207–213.
18. Asslaber M, Zatloukal K. Biobanks: transnational, European and global networks. Brief Funct Genomic Proteomic. 2007; 6:193–201. [PubMed: 17916592]
19. Riegman PH, Morente MM, Betsou F, de Blasio P, Geary P. Biobanking for better healthcare. Mol Oncol. 2008; 2:213–222. [PubMed: 19383342]
20. Manolio TA, Weis BK, Cowie CC, et al. New models for large prospective studies: is there a better way? Am J Epidemiol. 2012; 175:859–866. [PubMed: 22411865]
21. Scott CT, Caulfield T, Borgelt E, Illes J. Personal medicine – the new banking crisis. Nat Biotechnol. 2012; 30:141–147. [PubMed: 22318029]
22. O'Doherty KC, Burgess MM, Edwards K, et al. From consent to institutions: designing adaptive governance for genomic biobanks. Soc Sci Med. 2011; 73:367–374. [PubMed: 21726926]
23. Yuille M, Dixon K, Platt A, et al. The UK DNA banking network: a “fair access” biobank. Cell Tissue Bank. 2010; 11:241–251. [PubMed: 19672698]
24. Haga SB, Zhao JQ. Stakeholder views on returning research results. Adv Genet. 2013; 84:41–81. [PubMed: 24262096]
25. McCarty CA, Garber A, Reeser JC, Fost NC. Study newsletters, community and ethics advisory boards, and focus group discussions provide ongoing feedback for a large biobank. Am J Med Genet A. 2011; 155A:737–741. [PubMed: 21572889]

26. Betsou F, Gunter E, Clements J, et al. Identification of evidence-based biospecimen quality-control tools: a report of the International Society for Biological and Environmental Repositories (ISBER) Biospecimen Science Working Group. J Mol Diagn. 2013; 15:3–16. [PubMed: 23195791]
27. International Society for Biological and Environmental Repositories (ISBER). Best practices for repositories: collection, storage, retrieval, and distribution of biological materials for research. Biopreserv Biobank. 2012; 10:79–161. [PubMed: 24844904]
28. Moore HM, Compton CC, Alper J, Vaught JB. International approaches to advancing biospecimen science. Cancer Epidemiol Biomarkers Prev. 2011; 20:729–732. [PubMed: 21430299]
29. Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. Science. 1997; 278:1580–1581. [PubMed: 9411782]
30. Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE consortium. Sci Transl Med. 2011; 3:79re71.
31. National Cancer Institute. Best practices for biospecimen resources. 2011. from http://biospecimens.cancer.gov/bestpractices/2011-NCIBestPractices.pdf
32. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc. 2013; 20:117–121. [PubMed: 22955496]
33. McCarty CA, Chisholm RL, Chute CG, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics. 2011; 4:13. [PubMed: 21269473]
34. Denny JC. Chapter 13: mining electronic health records in the genomics era. PLoS Comput Biol. 2012; 8:e1002823. [PubMed: 23300414]
35. Ritchie MD, Denny JC, Zuvich RL, et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. Circulation. 2013; 127:1377–1385. [PubMed: 23463857]
36. Denny JC, Crawford DC, Ritchie MD, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am J Hum Genet. 2011; 89:529–542. [PubMed: 21981779]
37. McGuire SE, McGuire AL. Don't throw the baby out with the bathwater: enabling a bottom-up approach in genome-wide association studies. Genome Res. 2008; 18:1683–1685. [PubMed: 18974262]
38. Budin-Ljosne I, Soye KJ, Tasse AM, Knoppers BM, Harris JR. Genotype-driven recruitment: a strategy whose time has come? BMC Med Genomics. 2013; 6:19. [PubMed: 23702358]
39. Beskow LM, Fullerton SM, Namey EE, Nelson DK, Davis AM, Wil-fond BS. Recommendations for ethical approaches to genotype-driven research recruitment. Hum Genet. 2012; 131:1423–1431. [PubMed: 22622788]
40. Aschard H, Lutz S, Maus B, et al. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. Hum Genet. 2012; 131:1591–1613. [PubMed: 22760307]
41. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005; 6:287–298. [PubMed: 15803198]
42. Keers R. Will gene-environment interactions explain differential antidepressant response? Pers Med. 2012; 9:319–322.
43. Law, M. Reduce, reuse, recycle: issues in the secondary use of research data. 2005. from http://www.iassistdata.org/downloads/iqvol291law.pdf
44. Dale, A.; Arber, SL.; Procter, M. Doing secondary analysis. Allen & Unwin; London: 1988.
45. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013; 15:565–574. [PubMed: 23788249]
46. National Bioethics Advisory Commission. Research involving human biological materials: ethical issues and policy guidance. Vol. Vol. 1. National Bioethics Advisory Commission; Rockville, MD: 1999.
47. Wolf SM, Lawrenz FP, Nelson CA, et al. Managing incidental findings in human subjects research: analysis and recommendations. J Law Med Ethics. 2008; 36:219–248. 211. [PubMed: 18547191]

48. Caulfield T, McGuire AL, Cho M, et al. Research ethics recommendations for whole-genome research: consensus statement. PLoS Biol. 2008; 6:e73. [PubMed: 18366258]

49. Fabsitz RR, McGuire A, Sharp RR, et al. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. Circ Cardiovasc Genet. 2010; 3:574–580. [PubMed: 21156933]

50. Kocarnik JM, Fullerton SM. Returning pleiotropic results from genetic testing to patients and research participants. JAMA. 2014 Epub ahead of print.

51. U.S. Department of Health and Human Services. Presidential Commission for the Study of Bioethical Issues, Anticipate and Communicate: ethical management of incidental and secondary findings in the clinical, research, and direct-to-consumer contexts. U.S. Department of Health and Human Services; Washington, DC: 2013. from www.bioethics.gov

52. Black L, Avard D, Zawati MH, Knoppers BM, Hebert J, Sauvageau G. Funding considerations for the disclosure of genetic incidental findings in biobank research. Clin Genet. 2013; 84:397–406. [PubMed: 23662709]

53. 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. [PubMed: 22436882]

54. McDonald SA, Watson MA, Rossi J, Becker CM, Jaques DP, Pfeifer JD. A new paradigm for biospecimen banking in the personalized medicine era. Am J Clin Pathol. 2011; 136:679–684. [PubMed: 22031304]

55. Rosenberg RN. Translating biomedical research to the bedside: a national crisis and a call to action. JAMA. 2003; 289:1305–1306. [PubMed: 12633194]