Chapter 2
Collection, Storage and Use of Human Biological Materials in the United States

As part of its analysis, NBAC sought to understand and describe the magnitude, diversity, and use of human biological material collections in the United States. NBAC commissioned a study to assess the size and characteristics of existing collections.[1] In addition, a study was prepared for NBAC to review the historical contribution of collections of human biological materials to biomedical research.[2] This chapter provides information about several aspects of stored human biological materials, provides a schema by which NBAC classifies human biological materials, that is, the extent to which specimens are identifiable as they exist in the repository and as research samples in a scientific study, and describes some of the important purposes for which collections of human biological materials have been used in the past and may be used in the future. A detailed Appendix, "Collections of Human Biological Materials in the United States," provides information about the number of specimens of human biological material stored and the places in which these material are stored.

Collections of Human Biological Materials

In this report, the term *human biological material* is defined to encompass the full range of specimens, from subcellular structures like DNA, to cells, tissues (e.g. blood, bone, muscle, connective tissue and skin), organs (e.g., liver, bladder, heart, kidney, placenta), gametes (e.g., sperm and ova), embryos, fetal tissues, and waste (e.g., hair, nail clippings, urine, feces, and sweat, which often contain shed skin cells). At the present time, research using human embryos is prohibited from federal funding. As such, the current regulations do not apply. The use of

[1] Elisa Eiseman, Ph.D., RAND Science and Technology Policy Institute collected these data. The report, *Stored Tissue Samples: An Inventory of Sources in the United States* (available in Volume II of this report), is not meant to be a comprehensive inventory, however it does identify the major repositories or archives of stored human biological material.

[2] See David Korn, *Contributions of the Human Tissue Archive to the Advancement of Medical Knowledge and the Public Health*, a report prepared for the National Bioethics Advisory Commission, January 1, 1998, in Volume II of

human embryos in research raises special ethical concerns, which are addressed in a separate NBAC report.[3]

NBAC estimates that as of 1998 there are over 282 million specimens from more than 176.5 million individual cases of stored human biological materials in the United States, accumulating at a rate of over 20 million specimens per year.[4] The size and detail of collections varies considerably, ranging from formal, highly organized repositories to the storage of materials in a researcher's laboratory freezer. Individual collections of human biological materials range in size from less than 200 to more than 92 million specimens.

Large collections include archived pathology specimens obtained over many years during diagnostic and surgical procedures, or at autopsy, and stored cards containing blood spots from newborn screening tests (Guthrie cards) that have accumulated for a number of years. These specimens are stored at military facilities, forensic and other DNA banks,[5] government laboratories, diagnostic pathology and cytology laboratories, university- and hospital-based research laboratories, commercial enterprises, and non-profit organizations.

The collections of these materials generally fall into the following categories:

- large tissue banks, repositories and core facilities;
- materials collected as part of longitudinal studies;
- tailored collections for research studies requiring unique tissue collections;

---

this report.

[3] Cite title of stem cell report, in press, 1999.

[4] This estimate attempts to count both the numbers of cases from which stored human biological materials are derived as well as the number of specimens. For example, when a patient enters the hospital for a biopsy, the resulting tissue is accessioned in the pathology department as a single case. However, that single biopsy may generate several specimens including a number of slides, a paraffin block, and a frozen sample.

[5] The term "DNA bank" refers to a facility that stores extracted DNA, transformed cell lines, frozen blood or other tissue, or biological samples for future DNA analysis. Specimens are usually stored with some form of individual identification for later retrieval. DNA data banks are repositories of genetic information obtained from the analysis of DNA samples, sometimes referred to as "DNA profiles" The genetic information is usually stored in computerized

- pathology specimens, initially collected for clinical purposes;
- newborn screening tests accumulating in various laboratory sites;
- forensic DNA banks;[6]
- umbilical cord blood banks;
- organ banks;
- blood banks;
- sperm, ovum, and embryo banks, and
- individuals investigator's collections.

Two of the largest tissue repositories in the world, the National Pathology Repository and the DNA Specimen Repository for Remains Identification, are housed within a single institution, the Armed Forces Institute of Pathology (AFIP). These two repositories alone store more than 94 million specimens. State newborn screening laboratories collectively have archives totaling more than 13 million individual specimens. The pathology departments at Graduate Medical Education (GME) teaching institutions collectively constitute the largest and oldest stores of tissue specimens in the United States, with some over 100 years old.[7] The source of materials for these facilities is patients undergoing diagnostic or therapeutic procedures. Tissue specimens may also be taken during autopsies that are performed to establish the cause of death. In addition, volunteers may donate blood or other tissue for transplantation or research, organs for transplantation, or their bodies for anatomical studies after death. Each specimen of human tissue may be stored in multiple forms, such as slides, paraffin blocks, formalin fixed, frozen, tissue

---

form with individual identifiers.

[6] Only forensic DNA banks set up through state and federal legislation are discussed in this report. The use of human biological materials in other repositories for forensic purposes also raises several ethical issues and is not addressed in this report. See, add NRC report, OTA report references.

[7] Graduate Medical Education (GME) programs are the primary means of medical education beyond the four-year medical school training received by all physicians. Usually called residency programs, they are based in hospitals or other health care institutions, some of which do and some of which do not have formal relationships with medical schools. GME teaching institutions include medical schools, the Armed Forces hospitals, Veterans Affairs medical centers, the Public Health Service, state, county and city hospitals, non-profit institutions, and health maintenance organizations.

culture, or extracted DNA. Three of these sources—the AFIP National Pathology Repository, GME teaching institution pathology departments, and newborn screening laboratories—contain more than 265.5 million diagnostic and therapeutic specimens from over 170 million cases. Although the repositories supported by the National Institutes of Health (NIH) are not as large as those of AFIP are, NIH is one of the largest financial supporters of repositories, providing over $53 million in funding in Fiscal Year 1996.

The vast majority of specimens currently stored in the United States were originally collected for diagnostic or therapeutic reasons with varying levels of specificity about future uses provided in the informed consent process. A small percentage of these specimens may be used for research, educational, and quality control purposes. The majority of specimens are stored for clinical and legal reasons (e.g., confirmatory diagnoses, malpractice purposes). Most of these collections are generally referred to as "pathology specimens" and have been the primary source of human biological materials used to date in research. However, samples collected specifically for particular research purposes increasingly are in demand as biomedical research requires more precisely categorized samples with associated clinical data. In these cases, research samples are more likely to have been collected with explicit consent to use for specific research purposes.

As a result of these research needs, special repositories have been established specifically for research purposes. In addition, investigators conducting large, longitudinal studies collect and bank specimens from study participants over considerable periods of time. Likewise, a fair amount of current research is simultaneously engaged in creating special collections and contributing to existing banks of human biological material. Collectively, these special research collections now contain more than 2.3 million specimens.

Other than for diagnostic, therapeutic (e.g., transplantation or transfusion), or research purposes, human biological materials are collected and stored for a variety of other reasons.

Blood banks collect approximately 12 million units of blood a year, but only about 20,000 to 40,000 units are stored at any one time. Also, most of the blood collected is used for transfusions, and very little is used for other purposes, such as research and quality control. Organ banks do not collect the same volume of tissue as do blood banks, but are similar in that most of the organs and tissues collected are used for transplants, and very little is available for research purposes. Forensic DNA banks collect and store tissues for use in criminal investigations. The Department of Defense (DOD) DNA Specimen Repository and some commercial DNA banks store DNA specimens for remains identification. Sperm, ovum and embryo banks store specimens for anonymous donation or for later use by the individual storing the material. Umbilical cord blood banks also store blood for anonymous donation and later use by families banking their newborn's cord blood. Table 1 summarizes sources of stored specimens in the United States.

[insert table here]

Identifiability of Human Biological Materials

A key consideration in deciding whether the federal regulations apply and whether IRB review and consent is required is determining whether a human subject is involved. This determination may be conditioned by the extent to which biological material can be linked to the person from whom it was obtained. The debate about research use of human biological materials has been at times confounded both by the fact that the language used varies, and that it is often at odds with the categories used in the applicable regulations. For example, previous guidelines and reports often categorize specimens by the identifying conditions under which they are stored in repositories (with or without identifiers), although current federal regulations permit investigators to access stored specimens, make them anonymous by removing identifiers, and then use them in research without seeking consent of the donor (see chapter 3 for further discussion).

Part of the confusion around the term "identifiable" arises from the fact that people sometimes refer to the state of the information attached to the biological material in the repository (i.e., the **specimen**) and sometimes refer to the material (i.e., the **sample**) and the accompanying information that is provided to the researcher. For example, the specimen might be identified in the repository but no identifying information is forwarded with the research sample sent to the scientist. This distinction is of considerable importance because the potential for both benefit and harm is greater when the sample is directly or easily linked to the donor, placing the burden of protection in different places, depending on who has access to the information (e.g., the researcher or the pathologist, or both). If samples are identifiable, the potential exists for the investigator or a third party to contact the subject or act in some way that might effect the subject.

NBAC adopted the following definitions of human biological materials, depending on whether they are sitting in storage in a repository, or whether some of the material from a repository has been selected for research purposes.

Repository collections of human biological materials (i.e., specimens) are one of two types:

Unidentified specimens are those for which identifiable personal information was not collected or, if once collected, is not maintained and cannot be retrieved by the repository.

Identified specimens are those linked to personal information, such that the person from whom the material was obtained could be identified by name, patient numbers, or clear pedigree location (i.e., their relationship to a family member, whose identity is known).

Most repositories contain identified specimens by virtue of the fact that the vast majority of human biological materials in storage were originally collected with identifying information for diagnostic or therapeutic reasons. Examples of repositories containing identified materials include pathology laboratories and newborn screening laboratories where specimens are collected and stored with identifying information such as the patient's name, hospital identification number and/or social security number. In addition to identifying information, clinical and demographic information are often available with these specimens. In contrast, there are relatively few collections of human biological materials that contain unidentified specimens. Consider the following examples of such a repository:

- *A repository that collects specific blood types such as O-positive (O^+) or AB-negative (AB^-). Donors who have these blood types are asked to contribute to the bank based on having these specific blood types, but no information about the donor is recorded when the sample is collected except for the blood type.*

- *A repository that collects human biological materials, such as brain, pancreas or kidney, that were originally collected by a hospital, but are submitted to the repository with no identifying information. These specimens may be contributed with some corresponding clinical and demographic information, but any information provided with the specimen is not sufficient, either directly or indirectly, to identify the individual from whom the specimen was originally collected.*

Research samples are the collections of human biological materials provided to investigators by repositories. Such materials are of at least four types, which are differentiated by the amount of information that is conveyed to the investigator about the person from whom the sample comes. NBAC defines the different types as follows:

1. Unidentified samples—sometimes termed "anonymous"—are those supplied by repositories from an unidentified collection of human biological materials.

2. Unlinked samples—sometimes termed "anonymized"—are those supplied by repositories from identified human biological materials without identifiers or codes such that the ability to identify particular individuals via clinical or demographic information supplied with the sample, or biological information derived from the research would be extremely difficult for the investigator, the repository, or a third party.

3. Coded samples—sometimes termed "linked" or "identifiable"—are those supplied by repositories from identified materials with a code rather than a name or any other personal identifier such as a patient number, where the repository (or its agent) retains information linking the code to particular human materials or where the extent of the clinical or demographic information provided with the sample is sufficient that the investigator, the repository, or a third party could link the biological information derived from the research with material from a particular person or a small group of identifiable persons.

4. Identified samples are those supplied by repositories from identified materials with a personal identifier (such as a name or patient number) sufficient to allow the biological information derived from the research to be linked directly, by the researcher, to the particular person from whom the material was obtained.

By definition, **unidentified samples** can only come from collections of unidentified materials. Because of the scarcity of truly anonymously collected human biological materials, few research samples are unidentifiable. An example of a researcher's collection of unidentified samples follows:

- *A researcher studying malaria needs O^+ blood to grow the malaria parasite. The researcher has someone recruit donors with O^+ blood to donate a unit of blood. The researcher only needs to know the blood type of the donors and needs no identifying information from the donors. When the blood is collected, the researcher gives each vial a number, but keeps no record of which unit of blood came from each donor. The researcher places all of the blood that is collected in storage until there is enough blood stored to perform the planned experiments.*

On the other hand, repository collections of identified materials may be provided to researchers as unlinked, coded, or identified samples. The use of **unlinked samples** in research is a common occurrence. Unlinked samples are used when there is a one-time need for tissue and clinical/demographic information. Because there is no link maintained between the sample and the individual from whom it came, neither the researcher nor the repository knows which sample came from which source. Therefore there is no way to go back to get more information about the source or to get another piece of the same sample. For example:

- *A researcher at a university is studying a mutation of a gene that may be associated with prostate cancer. The researcher needs 100 samples of prostate tumors with accompanying clinical information, such as the size of the tumor. The researcher does not need any other information about the individual from whom the tumor was removed. The researcher contacts the pathology department at the university and requests the samples. The pathologist pulls 100 specimens from the pathology archives, records in a separate file the medical records number of the selected samples, removes any identifying information, gives each specimens a new unique identifier, and gives the samples to the researcher. There is no link maintained between the samples and the individual from whom it came. This means that neither the researcher nor the*

pathologist knows which sample came from which patient, although the pathologist may retain a record of the group of 100 samples used.

Another common category of samples used in research is those that are coded. **Coded samples** may be used, for example, when a researcher anticipates the need to obtain additional medical information about the source, to provide information to the source, or to get additional samples over time. For coded samples, the identification of the individual is not provided. Instead, each sample is given a unique identifier, and the repository for quality control or other purposes keeps a link. The link also provides the potential for one-way flow of information from the repository to the researcher and at times reverse flow of information from the researcher to the repository. Thus, coded samples could allow researchers to obtain follow-up data on treatment, recurrence, and survival, and may allow researchers to communicate research findings to subjects or their physicians. An example of the use of coded samples in research follows:

- *A researcher studying systemic lupus erythematosis (SLE) wants to know if there is some way to predict if a patient will go on to need a kidney transplant. The researcher uses frozen serum from patients with SLE that have been coded for research purposes. During the course of this research, a unique (e.g., serological) marker is found that may be predictive of rapidly progressive kidney disease. The researcher wants to determine if there is a connection between the newly discovered marker and patients requiring a kidney transplant. Therefore, the researcher wants to receive follow-up information about each patient, particularly information relating to time to renal failure and need for dialysis and/or kidney transplant.*

Identified samples are used when the research involves continual sample collection and/or clinical follow-up or when the researcher has direct contact with the research subject. With

identified research samples, the investigator can go back directly to the source of the sample and request additional information. For example:

- *A researcher is investigating the genetic causes of psoriasis. The researcher identifies patients with psoriasis or psoriatic arthritis through medical records and requests samples of skin biopsies from the pathology laboratory. After the researcher completes the experiments on the skin biopsies, the patients and their families are contacted to further participate in the research by providing blood samples. This allows the researcher to perform linkage analysis to try to localize genes that may play a role in psoriasis.*

Need to Identify Source for Research or Clinical Purposes

For research samples that are identified or coded, there are several possible reasons for an investigator to want to go back to the source either to gather additional clinical or biological information or to provide potentially valuable therapeutic information to the individual.

Increasingly genetic research requires that there be sufficient phenotypic (i.e., clinical) information accompanying the genotypic (i.e., DNA-based) information obtained from the biological material. Thus, investigators identify those individuals of interest according to the requirements of their research protocol and then intensively investigate a smaller subset. As smaller subpopulations of interest are identified, clinical investigators are likely to need more clinical information about the population being studied. This will require some mechanism for ongoing information retrieval. With coded research samples, the "trustee" of the sample retains the ability to gather more data for the investigator. With identified research samples, the investigator can go back directly and request additional information. The possibility that the investigator, or an agent of the investigator, will contact the source (or the source's physician) for additional information should be discussed in the consent process.

There might also be circumstances in which an investigator wants to provide information to the sample source, whether directly or indirectly. An example is an investigator who discovers new information that leads to a better diagnosis of a clinical condition, an effect of a previously administered therapy, or a misdiagnosis that might have important implications for the health of an individual source. Another example is the discovery of an infectious agent and its public health implications. In both of these examples, there have been compelling arguments made supporting the investigator's duty to contact the source. In cases where the implications of a finding are not as clear, that is, where findings are preliminary or where there is no effective intervention available, contact is less desirable and more controversial because of the possibility that people could act on these findings, however tentative and conditional, in a way that may result in harm.

Past Research Use of Human Biological Materials

Historically, the science of pathology has led the way in the investigation of the mechanisms of disease causation by proceeding progressively from whole organs and tissues to cells, and then from the subcellular to the supramolecular and molecular manifestations of disease expression (Rosai, 1997).

The range of medical benefits already obtained through the use of stored biological samples is impressive. For example,

- *In 1953 autopsies of American soldiers killed in the Korean conflict revealed that atherosclerosis begins at a much earlier age than was previously thought and that blockage of arteries can be far advanced in the absence of symptoms; this research contributed to findings concerning diet and exercise which have had a major public health impact in this country, evidenced by a significant reduction in coronary artery disease (Enos, 1953; 1955; Solberg,1983; Strong, 1986).*

- *In the late 1960s the study of samples of tissue from an unusual tumor of the vagina led to the discovery that a non-steroidal estrogen hormone diethylstilbestrol (DES), then commonly given to women during pregnancy, is carcinogenic (Herbst 1970;1971;1974;1981).*

- *Thirty years ago a series of studies on tissue samples of precancerous lesions of the uterine cervix led to the routine use of Pap smears, which have played an important role in the early diagnosis and more successful treatment of cervical cancer. (Herbst 1970;1971;1974;1981; Younge, 1949).*

- *Analysis of tissue from autopsies of persons in certain occupations, such as chemical manufacturing and uranium mining, have established causal links between exposure to environmental substances and certain diseases, including a cancer of the liver known as hepatic angiosarcoma and cancer of the bronchial epithelium (Creech, 1974; Dannaher, 1981; Falk, 1981; Popper, 1978; Regelson, 1968; Roth, 1957).*

- *The analysis of autopsied lung tissue from smokers played a major role in establishing that smoking causes lung cancer, that the risk of cancer increases with the duration of exposure to the chemicals contained in cigarette smoke, and that precancerous changes in the bronchial epithelium can be reversed by cessation of smoking (Auerbach, 1962; 1979; Flehinger, 1984; Frost, 1984).*

As the science and knowledge of human disease have progressed, researchers using human biological materials have developed or co-opted in steady succession the newest in scientific tools and methodologies. Novel insights and expanded knowledge of agents and mechanisms of disease causation have attracted a broader representation of the biomedical

research community, including immunologists, virologists, and geneticists, to the vast and valuable resource of human biological materials for investigating human disease.

The Value of Human Biological Materials to Current Research

Biomedical research routinely relies on the availability of stored human biological materials as well as the willingness of individuals to participate in research protocols by donating blood, tissue, or DNA samples to research. Research in cancer, infectious disease, mental disorders is advanced by access to such materials. In addition, large, longitudinal studies that aim to study the causes of diseases in certain populations over time depend on a continuous source of biological materials for study. Some examples are provided below.

Cancer Research

Pathology specimens have been invaluable resources for much cancer research. The availability of large archives of carefully documented and clinically correlated specimens permits the direct, much more rapid and less expensive approach of applying new detection technologies directly to existing specimens. To try to initiate new prospective studies for each new promising candidate gene for many of the varieties of human cancer would not only be extraordinarily costly in dollars and human effort, but would require study periods of many years, or even decades.

Recent progress in elucidating the initiation and progression of cancer has been most dramatic and gratifying in the area of colorectal cancer (Lengauer, 1997). During the past decade at least five specific genetic changes have been found that seem to constitute a progressive pathway from normal to neoplastic colon tissues. Some of these revelations have been derived in subsets of patients with known hereditary forms of colorectal cancer, while others appear more generally to be present in those without known patterns of familial inheritance. At least one of these genetic changes, the inactivation of the p53 gene, is known to occur, at least at times, in the germline, while the others appear to be exclusively of somatic origin (Kinzler, 1991a; 1991b; 1996).

Research on the role of the p53 gene was enabled by the availability of a large human tissue repository containing various forms and stages of colorectal cancers, as well as blood specimens from the same patients. The tissue archive consisted largely of typically fixed and embedded specimens, but in addition the scientists benefited immensely from a large collection of frozen samples (Fearon, 1987; 1990; Goelz, 1985; Vogelstein, 1988; 1989).

Screening Human Biological Materials Archives to Track Viruses

Stored biological specimens can be valuable resources during public health emergencies, when investigators are trying to identify or track an emerging virus. For example, in 1993 healthy young people began mysteriously dying in the Four Corners area of the American Southwest from a form of pneumonia. Within months the Hantavirus was identified as the culprit. The rapid solution of this public health mystery can be attributed to many sources, including a suspicious clinician, an epidemiologist, observant Navajo elders, and two human tissue archives. One archive was that of the Centers for Disease Control and Prevention (CDC), containing vast libraries of viruses, viral proteins, and serum specimens from around the world. The second archive held pulmonary tissues from the autopsied victims of this strange new disease. The CDC archive permitted initial serological screening tests, from which arose the first suggestion that a Hantavirus might be involved. The initial screens were followed by tests of autopsy tissue specimens with specific Hantavirus monoclonal antibodies, and ultimately, the tissue samples were exposed to Hantavirus genetic probes that revealed the presence and tissue distribution of viral genetic material. These molecular tools permitted identification of the local deer mouse as the host of the pathogenic Hantavirus. Studies of older human autopsy tissue established that the virus was, in fact, not a new variant but a fairly old virus with a well-established symbiotic relationship with the mice in the region that must have been disturbed in some way so as to initiate human infections (Wrobel, 1995).

Human Tissue as a Singular Resource in Brain Research

Sometimes use of biological materials is the only way to study certain aspects of human disease, for example, in studies of certain diseases of the brain and central nervous system. Currently there are no accurate animal or tissue culture models for many common diseases of the human brain, including brain tumors and most of the primary neurodegenerative diseases (e.g., Alzheimer's disease, Parkinson's disease, Amyotrophic Lateral Sclerosis, or Multiple Sclerosis). Moreover, neurological specimens, particularly of the brain, are often inaccessible.

Until relatively recently, most brain tumor research was conducted with animal models, or with cultured immortalized brain cell lines. Over the last five years, several studies have correlated genetic alterations in human brain tumors with the degree of malignancy and prognosis. These studies relied on frozen samples and specially fixed samples of human brain cancers to assess gene amplification, gene deletions, gene mutations, and cell cycle parameters. Many insights into the pathobiology of brain tumors are emerging from these studies (Blessed, 1968; Masliah, 1991; Raine, 1997; Will, 1996).

Longitudinal Studies

Longitudinal studies, in which the same group of individuals is studied at intervals over a period of time, often collect large numbers of specimens that can be used for both retrospective (i.e., looking back at data and trends over time) and current or future research. Several well-known longitudinal studies have been conducted over the years, including the Physicians' Health Study, the Nurses' Health Study, and the Framingham Heart Study.

As an example, the NIH Women's Health Initiative (WHI) is a 15-year research program, concluding in the year 2005, which focuses on the major causes of death, disability and impaired quality of life in postmenopausal women. The overall goal of WHI is to reduce coronary heart disease, breast and colorectal cancer, and osteoporosis in postmenopausal women through

prevention, intervention, and risk factor identification. The study will involve over 164,500 women of all races and socioeconomic backgrounds ages 50 to 79. The women are enrolled in either a clinical trial or an observational study and will be followed for 8 to 12 years, during which they will provide multiple blood samples. Participants sign a consent form that states that the collection of blood samples is for use in future research, which may include genetic research, and participants will not be informed of any test results. Participants may opt out of having their samples used for genetic research, if they so desire. Participants' charts contain identifying information including name, Social Security number, address and telephone number, and are bar-coded. Blood samples are labeled with matching barcodes to link them back to the charts. All study records are kept indefinitely for analysis and follow-up.

Relying on Stored Materials for Locating Genes

The human genome is the complete set of genetic instructions that set in motion the development of an individual. Though the DNA of any two people is roughly 99.9 percent identical, the variation in this last tenth of a percent is the source of considerable genetic variation. Inherited susceptibility to various diseases—which occurs when a gene fails to give correct instructions for a trait or function—is one small part of this diversity[8] Researchers search for genes by constructing finer and finer maps of known gene locations or by comparing DNA of individuals (or, more commonly, of families) with a given disease or trait to those who do not have that disease or trait.

The first phase of identifying a disease-related gene is the collection of diagnostic information and blood samples from an appropriate set of affected individuals and their relatives. Typically, blood samples are drawn from family members, and the blood cells are immortalized so they can be grown continuously in the laboratory. These immortalized cells, called cell lines,

[8] Some research aims specifically to document human genetic variation, such as the Human Genetic Diversity Project of the National Institutes of Health. This project relies on stored blood samples collected as part of the National Health and Nutrition Examination Survey (NHANES). No identifying information is provided with the blood

can then be used to make DNA in unlimited quantities, allowing many different researchers access to this resource. The art of this collection phase is in identifying appropriate families.

Linkage studies are widely used to detect and locate genes that determine susceptibility to certain disorders, and are often based on the identification of large, densely affected families so that the inheritance patterns of known sections of DNA (called "markers") can be compared to the family's transmission of the disorder. If a known marker can be correlated with the presence or absence of the disorder, this finding narrows the location of the suspect gene. Great strides in linkage analysis, including laboratory and statistical methods, are increasing the power of this method and decreasing its cost.

Linkage-disequilibrium studies in isolated populations capitalize upon the likelihood that the susceptibility genes for a particular disorder probably came from one or a few founding members. Whether the isolation of the population is geographic or cultural, there are fewer individuals in the community's original founding genealogies and therefore fewer variations of the disease genes within the population. This limited variation makes the search for genetic association with a disease easier. In addition, the groups of markers that surround each of these susceptibility genes are likely to have the same limited variation, which further simplifies gene identification. (See Box for description of such a study in an isolated population [insert Iceland box here].)

Association studies depend on the investigator hypothesizing that a specific gene or genes may influence the disorder. In this type of study, the investigator examines whether those people with the disorder have a different version of the gene than those without the disorder among related or unrelated individuals. Unlike linkage studies that usually focus on large groups of related family members, association studies can be done using unrelated individuals.

---

samples used in the study. See the National Research Council report, xxxxxxx, 1997.

Pinpointing the likely genetic anomaly in linkage and linkage-disequilibrium studies can only occur once an investigator narrows the search to a fairly small region in the genome. That "small" region, however, may still be large enough to contain DNA that codes for dozens of traits, and the investigator must now choose which parts of the region to study further. Because the Human Genome Project is well on the way to identifying the location of all genes, this mapping of the human genome will greatly simplify the identification of possible susceptibility genes. Once the genes in a narrow DNA region are cataloged, they may each be tested and the susceptibility gene identified.

An example of use of DNA repositories in linkage studies is the National Institute of Mental Health's (NIMH) Genetics Initiative, begun in 1989.[9]. The goal of this special, large-scale initiative in molecular genetics is to collect data from enough families to find the genes that influence the onset of selected mental disorders. In addition, the initiative enabled the establishment of a national repository of demographic, clinical, diagnostic, and genetic data from individuals with bipolar disorder, schizophrenia, or Alzheimer's disease to aid researchers in identifying factors responsible for these disorders.

Research Requiring Unique Tissue Collections

Most researchers using human biological materials have relied on specimens from pathology laboratories or existing tissue banks. However, some research studies require specialized samples, i.e., with specific biological, clinical, or demographic characteristics, and therefore must create a unique collection, which might have limited appeal to the broad research community but high value to a small group of investigators.

For example, the University of Southern California AIDS-Malignancy Clinical Trials Consortium (AM-CTC) helps design, develop, and conduct clinical trials of novel agents to be

[9] See the National Institute of Mental Health at http://www.nimh.gov/

used against AIDS-related malignancies. In addition, the AM-CTC stores tumor tissue and other relevant biologic materials that have been obtained from patients participating in their trials. As another example, Stanford University is investigating the role of environmental toxicants and genetic susceptibility factors in the etiology of Amyotrophic Lateral Sclerosis (ALS). It has a specialized collection of samples from patients with ALS.

Another example is the health examination surveys conducted by the CDC. Since 1960, the CDC's National Center for Health Statistics (NCHS) has conducted 7 health examination surveys of the population of the United States, the National Health Examination Surveys (NHES) Cycles 1, 2 and 3, the National Health and Nutrition Examination Surveys (NHANES) I, II and III, and the Hispanic Health and Nutrition Examination Survey (HHANES). The surveys are designed to assess periodically the health and nutritional status of children and adults in the United States through interviews and direct physical examinations. The surveys employ interviews to answer questions about demographics, socioeconomic status, dietary habits and health-related issues, and physical and dental examinations, which include physiologic assessments and laboratory tests. Blood samples are collected as part of the physiologic assessments, and placed in storage banks after laboratory tests are completed.

Cumulatively, all of the CDC's health examination surveys have analyzed and banked samples from more than 85,000 participants. The most recent survey, NHANES III[10], conducted between 1988 and 1994, performed laboratory tests on approximately 29,314 people of all races aged one year and older from 81 counties in 26 states. Some of the 30 topics investigated in the NHANES III included high blood pressure, high cholesterol, obesity, second-hand smoking, lung disease, osteoporosis, HIV/AIDS, hepatitis, *Helicobacter pylori*, immunization status, diabetes, allergies, growth and development, anemia, dietary intake, antioxidants, and nutritional blood

[10] National Health and Nutrition Examination Survey (NHANES), http://www.cdc.gov/nchswww/about/major/nhanes/nhanes.htm

measures. The NHANES I analyzed blood and urine samples from 23,808 study participants, and NHANES II analyzed 20,322 samples. The HHANES was a one-time survey conducted from 1982 to 1984 that provided data on 11,653 people of Hispanic origin.

Summary

This chapter provides examples of how human biological materials have been and continue to be invaluable resources for a wide variety of studies aimed at understanding the etiology and progression of disease, the effects of viral and environmental impacts on health, and finding genes that might be responsible for the underlying mechanisms of disease. There is tremendous variability in the identifiability of the samples used depending on the source of the material and the research purpose. In some cases, such as the study of the Hantavirus, it was not necessary to identify the individuals who served as the sources of the stored samples. For other types of research, such as the studies of families with a high prevalence of mental illness where extensive information on demographics, diagnosis, and family history was crucial, the ability to identify the source of the sample may be necessary.

Most of the specimens sitting in repositories will never be used in research. Many research studies will rely on large numbers of unidentified or unlinked research samples to investigate the basic mechanisms of health and disease, or to screen samples for evidence of disease, environmental insult, or responsiveness to potential therapeutic agents. Other studies, however, will rely on coded or identifiable samples. That is, an investigator might initially request samples with no linking data and later request additional clinical data linked to the sample. In still other cases, the research might require that the investigator know who provided the sample, or the sample source might even be a patient, as well as a research subject, of the scientist. How human biological materials are used in research and the extent to which research samples can be linked to their sources are critical considerations when trying to determine risks and necessary protections of the persons who are the sources of the material.

Chapter 2 References

1. Auerbach O., Hammond, E.C., Garfinkel, L., "Changes in bronchial epithelium in relation to cigarette smoking,1955-1960 vs. 1970-1977," *New England Journal of Medicine* 300:381-386, 1979.

2. Auerbach, O., Stout, A.P., Hammond, E.C., Garfinkel, L., "Changes in bronchial epithelium in relation to sex, age, residence, smoking and pneumonia," *New England Journal of Medicine* (67):111-119, 1962.

3. Blessed, G., Tomlinson, B.E., Roth, M., "The association between quantitative measures of dementia and the senile change in the cerebral gray matter of elderly subjects," *British Journal of Psychiatry* 114:797-811, 1968.

4. Creech, J.S., Johnson, M.N., "Angiosarcoma of liver in the manufacture of polyvinyl chloride," *Journal of Occupational Medicine* 16:150-151, 1974.

5. Dannaher, C.L., Tamburro, C.H., Yam, L.T., "Occupational carcinogenesis: The Louisville experience with vinyl chloride-associated hepatic angiosarcoma." *American Journal of Medicine* 70:279-287, 1987.

6. Enos, W.F., Holmes, R.H., Beyer, J. "Coronary disease among United States soldiers killed in action in Korea," *Journal of the American Medical Association* 152:1090-1093, 1953.

7. Enos, W.F., Beyer, J.C., Holmes, R.H. "Pathogenesis of coronary disease in American soldiers killed in Korea," *Journal of the American Medical Association* 152:912-914, 1955.

8. Falk, H., Herbert, J., Crawley, S., et al., "Epidemiology of hepatic angiosarcoma in the United States: 1964-1974," *Environmental Health Perspectives* 41:107-113, 1981.

9. Fearon, E.R., Hamilton, S.R., Vogelstein, B., "Clonal analysis of human colorectal tumors," *Science* 238(4824):193-197, October 9, 1987.

10. Fearon, E.R., Vogelstein, B., "A genetic model for colorectal tumorigenesis," *Cell* 61(5):759-767, June 1, 1990.

11. Flehinger, B.J., Melamed, M.D., Zaman, M.B., et al., "Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Memorial Sloan-Kettering study," *Am Rev Resp D*is 130:555-560, 1984.

12. Frost, J.K., Ball, W.C. Jr., Levin, M.L., et al., "Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study," *Am Rev Respir Dis* 130:549-554, 1984.

13. Goelz, S.E., Vogelstein, B., Hamilton, S.R., Feinberg, A.P., "Hypomethylation of DNA from benign and malignant human colon neoplasms," *Science* 228(4696):187-190, April 12, 1985.

14. Herbst, A.L., Scully, R.E., "Adenocarcinoma of the vagina in adolescence," *Cancer* 25:745-757, 1970.

15. Herbst, A.L., Ulfelder, H., Poskanzer, D.C., "Adenocarcinoma of the vagina." New England Journal of Medicine 284:878-881, 1971.

16. Herbst, A.L., Robboy, S.J., Scully, R.E., Poskanzer, D.C., "Clear-cell adenocarcinoma of the vagina and cervix in girls: Analysis of 170 registry cases," *American Journal of Obstetrics and Gynecology* 119:713-724, 1974.

17. Herbst, A.L., "Clear cell adenocarcinoma and the current status of DES - exposed females," *Cancer* 15 Supplement; 48:484-4881, July 15, 1981.

18. Kinzler, K.W., Vogelstein, B., "Lessons from hereditary colorectal cancer," *Cell* 87(2):159-170, October 18, 1996.

19. Kinzler, K.W., Nilbert, M.C., Su, L.K., et al., "Identification of FAP locus genes from chromosome 5q21," *Science* 253(5020):661-665, August 9, 1991.

20. Kinzler, H.W., Nilbert, M.C., Vogelstein, B., et al., "Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers," *Science* 251(4999):1366-1370, March 15, 1991.

21. Lengauer, C., Kinzler, K.W., Vogelstein, B., "Genetic instability in colorectal cancers," *Nature* 386(6625):623-627, April 10, 1997.

22. Masliah, E., Terry, R.D., Alford, M., et al., "Cortical and subcortical patterns of synaptophysin-like immunoreactivity in Alzheimer's disease," *American Journal of Pathology* 138:235-246, 1991.

23. Popper, H., Thomas, L.B., Telles, N.C., et al., "Development of hepatic angiosarcoma in man induced by vinyl chloride, thorotrast, and arsenic - comparison with cases of unknown etiology," *American Journal of Pathology* 92:349-376, 1978.

24. Raine, C.S., "The Norton lecture: A review of the oligodendrocyte in the multiple sclerosis lesion," *Journal of Neuroimmunology* 77:135-152, 1997.

25. Regelson, W., Kin, U., Ospimam, J., et al., "Hemangioendothelial sarcoma of liver from chronic arsenic intoxication by Fowler's Solution," *Cancer* 21:514-522, 1968.

26. Rosai, J., "Pathology: A Historical Opportunity," *American Journal of Pathology* 151: 3-6, 1997.

27. Roth, F., "The sequelae of chronic arsenic poisoning in Moselle vintners," *German Medical Monthly* 2:172-175, 1957.

28. Solberg, L.A., Strong, J.P., "Risk factors and atherosclerotic lesions. A review of autopsy studies," *Arteriosclerosis* 3:187-198, 1983.

29. Strong, J.P., "Coronary atherosclerosis in soldiers. A clue to the natural history of atherosclerosis in the young," *Journal of the American Medical Association* 256:2863-2866, 1986.

30. Vogelstein, B., Fearon, E.R., Kern, S.E., et al., "Allelotype of colorectal carcinomas," *Science* 244(4901):207-211, April 14, 1989.

31. Vogelstein, B., Fearon, E.R., Hamilton, S.R., et al., "Genetic alterations during colorectal-tumor development," *New England Journal of Medicine* 319(9):525-532, September 1, 1988.

32. Will, R.G., Ironside, J.W., Zeidler, M., et al., "A new variant of Creutzfeldt-Jakob disease in the UK," *Lancet* 347:921-925, 1996.

33. Wrobel, S., "Serendipity, science and a new Hantavirus," *FASEB Journal* 9:1247-1254, 1995.

34. Younge, P.A., Hertig, A.T., Armstrong, D., "A study of 135 cases of carcinoma in situ of the cervix at the Free Hospital for Women," *American Journal of Obstetrics and Gynecology* 58:867-892, 1949.

Table 1. Stored Human Biological Materials in the United States

Type of Repository	# of cases	# of specimens	Cases/Year
Large Tissue Banks, Repositories, and Core Facilities	>2.6 million	>96 million	364,825
Longitudinal Studies	>263,500	>263,500	
Pathology Specimens	>160 million	>160 million	>8 million
Newborn Screening Laboratories	>13.5 million	>13.5 million	<10,000 to >50,000
Forensic DNA Banks	380,000	380,000	
Umbilical Cord Blood Banks	18,300	18,300	
Organ Banks		>75,500	>75,500
Blood Banks		~12 million	~12 million
Grand Total	>>176.5 million	>>282 million	>20 million