Chapter 2
Collection, Storage and Use of Human Biological Materials In The United States

As part of its analysis, the National Bioethics Advisory Commission (NBAC) sought to understand and describe the magnitude, diversity, and use of human biological material collections in the United States. To assist in this task, NBAC commissioned a study to assess the size and characteristics of the existing archives of tissues.[1] In addition, a second study was prepared for NBAC to review the historical contribution of collections of human biological materials to biomedical research.[2] This chapter, therefore, will provide information about several aspects of stored human biological materials. The first section, "Collections of Human Biological Materials," provides information about the number of specimens of human biological material stored in the United States, and the places in which these material are stored. The second section, "Definitions and Origins of Human Biological Materials," provides information about who the sources of these biological materials are, why the specimens were originally collected, what identifying information is kept with the specimens, and what type of identifying information is passed under various circumstances on to the researcher. The last section of this chapter, "Uses of Human Biological Materials," 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.

[1] These data were collected by Elisa Eiseman, Ph.D., RAND's Critical Technologies Institute, in response to a request by the NBAC Genetics Subcommittee. 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 sources of stored tissue.

[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.

December 3, 1998: This is a draft report of the National Bioethics Advisory Commission. It is being circulated for public comment. It therefore does not reflect final conclusions or recommendations of the Commission and should not be cited or referenced as such.

PART I: COLLECTIONS OF HUMAN BIOLOGICAL MATERIALS

NBAC estimates that there are over 282 million specimens from more than 176.5 million individual cases of stored human biological materials in the United States, now accumulating at a rate of over 20 million specimens per year.[3] The size and detail of collections varies considerably, ranging from formal, highly organized repositories to the storage of blood or tissue specimens in a researcher's laboratory freezer. Individual archives of human biological materials range in size from less than 200 to more than 92 million specimens.

Large collections include archived pathology specimens taken 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,[4] 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;
- pathology specimens, initially collected for clinical purposes;
- newborn screening tests accumulating in various laboratory sites;
- forensic DNA banks;[5]

[3] This estimate attempts to count both the numbers of cases from which stored tissues are derived as well as the number of specimens generated from each case. 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.

[4] The term "DNA bank" refers to a facility that stores extracted DNA, transformed cell lines, frozen blood or 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 form with individual identifiers.

[5] 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

- umbilical cord blood banks;
- organ banks;
- blood banks; and
- sperm, ovum, and embryo banks.[6]

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. Finally, 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] Three of these sources—the AFIP National Pathology Repository, GME teaching institution pathology departments, and newborn screening laboratories—represent more than 265.5 million diagnostic and therapeutic specimens from over 170 million cases. Although the tissue repositories supported by the National Institutes of Health (NIH) are not as large as those of AFIP, NIH is one of the largest funders of tissue repositories, providing over $53 million in Fiscal Year 1996. The vast majority of specimens currently in storage were originally collected for diagnostic or therapeutic reasons. Although a small percentage of these specimens may be used for research, educational, and quality control purposes, the majority is not and instead is stored for clinical and legal reasons. 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

---

addressed in this report.

[6] Due to the fact that research using human embryonic tissue is proscribed from federal funding, the use of such material was not considered in this report.

[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.

collected specifically for research are increasingly in demand, as they are more narrowly defined, are often provided with associated clinical data from individual medical records, and are more likely to have been collected with explicit consent to use for specific research purposes.

Several repositories have been established specifically for use in research. In addition, several large longitudinal studies collect and bank samples 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.

Large Tissue Banks, Repositories, and Core Facilities

Large tissue banks and repositories exist in almost every sector of the scientific and medical communities, including the military, the Federal Government, universities and academic

medical centers, commercial enterprises, and non-profit organizations. In addition, several universities have established banking facilities to support both their own research as well as collaborations with other universities. These large tissue banks, repositories, and core facilities are a major source of human biological materials used in biomedical research. Representative collections of this type are described below.[8]

Military Facilities

The military maintains two of the largest tissue repositories in the world. As mentioned previously, the National Pathology Repository and the DOD DNA Specimen Repository for Remains Identification are housed in the AFIP[9]. The AFIP is responsible for maintaining a central laboratory of pathology for consultation and diagnosis of pathologic tissue for DOD, other federal agencies, and civilian pathologists. The AFIP also conducts research in pathology, trains enlisted personnel in histopathology and related techniques, and offers over 50 pathology education courses for medical, dental, and veterinary personnel.

The National Pathology Repository,[10] located at AFIP, is the single largest and most comprehensive collection of pathology material in the world. Since 1917, the Pathology Repository has collected over 2.5 million cases and logs in approximately 50,000 cases annually. Material is stored permanently unless there is a specific request by the patient or other authorized individual to return or release the material.

Individual specimens are sent to AFIP for a variety of reasons, such as to obtain a second opinion on a diagnosis, as part of established peer-review and quality assurance programs, by DOD regulation (such as forensic cases and those subject to litigation), or because they are unusual or rare and may be useful to AFIP in its research and education missions. Pathologic specimens stored at the Pathology Repository can be used to study unusual tumors, or as part of a

[8] The complete text of the inventory appears in the commissioned paper prepared by Elisa Eiseman.
[9] Armed Forces Institute of Pathology (AFIP), http://www.afip.mil/default.html
[10] National Pathology Repository , http://www.afip.mil/repository/welcome.html

public health surveillance system to study emerging infectious diseases or trends in disease progression. For example, specimens in the Repository have been used to identify and date tissues harboring genomic material of the Human Immunodeficiency Virus (HIV) that were obtained before the availability of HIV testing and before the spread of the HIV infection. In addition, cases have been submitted over the years for specific purposes, such as to study a particular disease, or to answer current and future research questions.

All submitted case material is coded by pathological diagnosis, and is identified by an AFIP accession number. The source name, social security number, date of birth, age, sex, and race are stored if provided by the contributing pathologist. Any medical history provided is also stored. The source address is not routinely provided or stored but is obtained on occasion for follow-up studies. Likewise, the original consent is a matter between the patient and the clinician and is not routinely provided to AFIP by the contributing pathologist. The submitting pathologist's name and address, and the source's surgical identification numbers are also stored.

The Pathology Repository loans pathologic material for patient treatment, research, or litigation. Requests for loan of material or provision of data for research purposes requires submission and approval of a research protocol. All research protocols using stored materials or data are reviewed by the AFIP's IRB. Requests from individuals or organizations other than the original contributor must be accompanied by a properly executed authorization signed by the patient or designated representative. Research involving patient follow-up, and thus requiring identifying information, is reviewed at a full meeting of the IRB prior to approval. Other than for research protocols involving follow-up, original sources of material are not notified of research results. If an unexpected disease or abnormality is discovered, the contributing pathologist is notified, and it is then up to the pathologist to contact the patient. Otherwise, current AFIP policy requires that material be stripped of identifiers before release to outside investigators.

Since June 1992, DOD has required all military inductees, and all active duty and reserve personnel to provide blood and saliva specimens for its DNA Specimen Repository at the time of enlistment, re-enlistment, annual physical, or preparation for operational deployment (McEwen,

1997). The DNA Repository also contains specimens from civilians and foreign nationals who work with the United States military in arenas of conflict. A total of three DNA specimens are collected from each person: one bloodstain card is stored in a pouch in the service member's medical record; another bloodstain card and a buccal swab are stored at the DNA Specimen Repository. The blood is placed on special cards with the service member's Social Security number, date of birth, and branch of service designated on the front side of the card, and a fingerprint, a bar code, and signature attesting to the validity of the specimens on the reverse side. DNA will only be extracted from the specimens in the Repository when it is needed for the purpose of remains identification.

The DOD DNA Specimen Repository for Remains Identification[11] is the world's largest DNA bank. As of September 1997, the DNA Repository has received approximately 2 million DNA specimens. Specimens come into the DNA Repository at a rate of 10,000 per day, and the tally is updated every seven seconds. It is estimated that by the year 2001 the DNA Repository will contain approximately 3.5 million specimens. All DNA specimens are maintained for 50 years before being destroyed. However, donors may request that their specimens be destroyed following the conclusion of their military service obligation or other applicable relationship to DOD. The military's policy ensures that specimens can only be used for remains identification and routine quality control except where subpoenaed for the investigation or prosecution of a felony. The specimens cannot be used without consent for any other purpose, such as paternity suits or genetic testing. In addition, the specimens are considered confidential medical information, and military regulations and federal law exist to cover most concerns.

Recently, the Armed Forces DNA Identification Laboratory (AFDIL) performed mitochondrial DNA (mtDNA) analysis on specimens taken from the skeletal remains of the Vietnam Unknown, which had been exhumed from the Tomb of the Unknown at Arlington National Cemetery. This mtDNA profile was then compared to mtDNA specimens from living relatives of those deceased service members thought to have been in the area at the time. On June

[11] Armed Forces DNA Identification Laboratory, http://www.afip.mil/oafme/dna/afdil.html

30, 1998, the Pentagon announced that the remains of the memorial's Vietnam War soldier belong to Air Force pilot Michael J. Blassie, bringing closure to a 26-year ordeal for the Blassie family, who had been uncertain about the fate of their relative.

National Institutes of Health

The National Institutes of Health[12] (NIH) is one of the largest funders of tissue and data resources for basic, applied and clinical research. Some of the institutes at NIH that support tissue banks include the National Cancer Institute (NCI), the National Institute of Allergy and Infectious Disease (NIAID), the National Heart, Lung, and Blood Institute (NHLBI), the National Institute of Mental Health (NIMH), and the National Institute on Aging (NIA). Examples of tissue banking supported by NIH are described below.

The NCI Cooperative Human Tissue Network (CHTN), [13] in existence since 1987, provides biomedical researchers with access to fresh surgical or biopsy specimens of normal, benign, pre-cancerous and cancerous human tissues. The CHTN is a tissue collection system and not a tissue bank. Only rare specimens that are difficult to obtain are stored to anticipate future requests. Except for a collection of frozen tissue from rare pediatric tumors, banked specimens are generally not stored for more than one year. Normally, the specimens are obtained prospectively to fill specific researcher requests. Five member institutions coordinate the collection and distribution of tissues across the United States and Canada. Tissues are provided by the CHTN only for research purposes, and cannot be sold or used for commercial purposes.

During the first nine years of operation, the CHTN has supplied over 100,000 samples to approximately 600 investigators. Tissues obtained from the CHTN have been used in many areas of cancer research including molecular biology, immunology, and genetics. Researchers have used these tissues to study mutations of proto-oncogenes in human tumors, the role of growth factors in cancer, and to isolate new cancer genes. In order to obtain samples from the CHTN,

[12] National Institutes of Health (NIH), http://www.nih.gov/index.html
[13] NCI Cooperative Human Tissue Network (CHTN)

investigators must provide a summary of the project for which the tissue is requested, and a copy of the local IRB approval of the protocol. Over 2,000 publications have resulted from studies using tissues obtained from the CHTN.

CHTN distributes primarily coded samples (that is, identifiers are retained in the repository and samples are forwarded to investigators with a linking code). Although the samples are anonymous to the researchers, the repository maintains an identifying link. A link is maintained for quality control purposes and to ensure that the same samples are not sent when researchers ask for different samples. However, because a third party not involved with the research functions as a trustee for the coded information, the possibility of the investigator ascertaining the identity of the sample source is minimized. The repository functions as an "honest broker" or "gatekeeper" to control the flow of information. The repository determines the conditions under which specimens and data are collected and provisions for maintaining confidentiality, all of which are reviewed and approved by the repository's IRB.

The CHTN was designed for basic research studies not requiring clinical follow-up information. Only minimal demographic data is provided with the sample to researchers. Other information routinely provided with the samples includes pathology reports and histological characterization.

The NCI-National Action Plan on Breast Cancer (NAPBC) Specimen and Data Information System[14] contains information from 14 breast tissue banks. Although this database does not represent an exhaustive national listing of all facilities holding breast cancer tissue, by centralizing information on a large number of biological specimens, it provides access to breast tissue specimens and facilitates collaboration among basic, clinical, and epidemiologic researchers. Cumulatively, the 14 breast tissue banks in the NCI-NAPBC database contain more than 130,000 cases of breast cancer-related specimens and data, with banks ranging in size from 48 cases to approximately 101,000 cases. Samples available to the research and clinical communities include

---

http://wwwicic.nci.nih.gov:80/chtn/chtnmain.html

[14] NCI-NAPBC Breast Cancer Specimen and Data Information System,

breast tissue, serum, urine, cells, and DNA from patients diagnosed with breast cancer, those at high risk, and unaffected individuals.

Research Universities and Academic Medical Centers

Research universities and academic medical centers maintain both formal human biological material banks for distributing samples throughout the research community as well as core facilities to support their own research. For example, the Harvard Brain Tissue Resource Center[15] (The Brain Bank) is a centralized repository of post-mortem human brain specimens from both diseased and normal donors. Samples from the bank are distributed for use in research on the brain and nervous system. Since the majority of research requires a very small amount of tissue, each donated brain provides a large number of samples for many researchers. Brain tissue donations are accepted by the Brain Bank from individuals or the parents, siblings and offspring of individuals with severe psychiatric or neurological disorders, as well as from unaffected individuals for purposes of comparison.

Another example, the University of California-San Francisco (UCSF) AIDS Specimen Bank, in existence since 1982, has banked over 76,000 specimens and sent out over 82,000 samples to researchers worldwide. Specimens include serum, tissue, saliva, cells, and cerebrospinal fluid from HIV-infected individuals. Specimen data are archived on a computerized database. The Bank provides investigators with samples for basic, epidemiological, and clinical research.

Commercial Enterprises

Some commercial enterprises maintain human biological material banks for their own proprietary use, while others establish banks for storage and distribution purposes. OncorMed[16]

---

http://cancernet.nci.nih.gov/breastdata/contents.htm

[15] Harvard Brain Tissue Resource Center, http://www.brainbank.mclean.org:8080/into.html

[16] OncorMed, http://www.oncormed.com

and LifeSpan Biosciences, Inc.,[17] are examples of companies that maintain proprietary tissue banks. For example, LifeSpan's Tissue and Disease Bank contains 250,000 normal and diseased human specimens. The tissue bank has over 175 different types of tissues from virtually every organ in the body, covering all ages. The tissue bank also includes over 500 different pathologic disease categories such as autoimmune diseases, infectious diseases, degenerative diseases, cancer and benign proliferative diseases, and genetic diseases.

In contrast, PathServe Human Tissue Bank [18] collects all types of human tissues and organs for sale to the research community. Tissues are obtained through post-mortem examinations, referrals from transplant banks of nontransplantable organs, and donations by next of kin. PathServe collects specimens from approximately 300 autopsies per year, and each autopsy yields approximately 100 specimens. PathServe has approximately 300 specimens stored at any one time, and distributed approximately 30,000 samples in 1996. PathServe does not maintain a centralized storage facility. Instead, specimens are stored in the morgues of different hospitals.

Non-Profit, Non-Educational Organizations

There are a variety of non-profit institutions that bank tissues for purposes of storage and distribution, such as the American Type Culture Collection (ATCC), the Coriell Institute for Medical Research, and the National Disease Research Institute (NDRI).

Since its establishment in 1925, ATCC[19] has served as an archive of living cell cultures and genetic material for researchers in the biological sciences. The mission of the ATCC is to acquire, authenticate, and maintain reference cell cultures, related biological materials, and associated data, and to distribute these to qualified scientists in government, industry, and education. The ATCC maintains approximately 2,300 human cell lines as immortalized cultures. In addition, cloned human genes are stored and supplied to the research community by ATCC.

[17] LifeSpan BioSciences, Inc., http://www.lsbio.com
[18] PathServe, http://www.tissuebank.com/

The ATCC policy on availability of cultures of human biological materials states[20]:

All ATCC cultures are publicly available with the exception of restricted patent deposits and cultures entrusted to ATCC for safekeeping. Publicly available means that the cultures are provided "on demand" to anyone meeting the eligibility requirements for receipt. This means that ATCC does not require any information on the intended research use of the cultures, and does not select recipients on the basis of research interest or affiliation.

Eligibility requirements are established by ATCC based on regulatory rules governing distribution of certain materials, and ATCC's own criteria for release of material. The requestor must be affiliated with an institution that has laboratory facilities for handling the requested cultures, and all regulatory requirements such as permits and licenses must be satisfied. ATCC will not ship cultures to an individual or a private residence or office; only to institutions such as commercial businesses, universities, and government and private laboratories.

Patent law dictates that once a patent is issued on a culture, the culture must be publicly available. At least 200 of the approximately 2,300 human cell lines at ATCC have been patented. However, publicly available does not mean that anyone who asks will be sent a culture. ATCC has in place a strict policy to ensure that cultures are distributed only to qualified organizations and researchers with legitimate and justifiable scientific uses for these materials. Parties interested in receiving cultures from ATCC must be able to verify that they have adequate facilities and expertise in working with biological materials. For agents that are classified as hazardous, or which could have serious adverse consequences for human health and safety, ATCC relies on

[19] American Type Culture Collection (ATCC), http://www.atcc.org/

[20] Personal communication from Frank Simione, Director Professional Services, ATCC, October 1998.

domestic regulations promulgated by the Secretary of Health and Human Services. Few if any cultures of human biological materials have associated identifying information. In a recent audit of cultures for examination of ownership issues, no cultures were found that contained identifiers.

The Coriell Institute for Medical Research[21] is a basic biomedical research institution that conducts research on the causes of genetic diseases, including cancer. The Coriell Institute houses the largest collection of human cells for research, and these cells are available to the general scientific community. Seminal research on the genes associated with Huntington's disease, cystic fibrosis, Alzheimer's disease, ataxia telangiectasia and manic depression have utilized cells from the Coriell collection. The Coriell Cell Repositories also support the Human Genome Project. Over 35,000 cell lines are currently stored representing approximately 1,000 of the 4,000 known genetic diseases, and more than 60,000 cell lines have been distributed to over 40 nations, resulting in over 8,000 research publications. In the 1970's, the Coriell Institute won contracts from the National Institute of General Medical Sciences (NIGMS) and the National Institute on Aging (NIA) to establish and maintain what have become the world's largest cell repositories for the study of genetic and aging-related diseases, respectively. In 1990, NIMH awarded the Coriell Institute a $5.7 million contract to establish a cell repository for the study of the genetic basis of Alzheimer's, manic depression and schizophrenia. New repositories have recently been set up for the study of diabetes.

The Coriell Cell Repositories have strict guidelines for submission of specimens. Each submission for inclusion in the Repository **must** be accompanied by clinical and laboratory documentation of the diagnosis and an unsigned copy of the IRB-approved consent form used to obtain the specimen. For submission to the NIGMS Human Genetic Mutant Cell Repository, a model informed consent form is available from the Repository.[22] This model informed consent has been reviewed by OPRR and approved by the NIGMS Human Genetic Mutant Cell

[21] Coriell Institute for Medical Research
http://arginine.umdnj.edu/info.html

[22] NIGMS Human Genetic Mutant Cell Repository Model Informed Consent Form,
http://locus.umdnj.edu/nigms/comm/submit/model.html

Repository IRB. In addition, the Office for Protection from Research Risks (OPRR) has provided guidance on Protections for Human Subjects in the NIGMS Human Genetic Mutant Cell Repository[23] and Submission of Non-Identifiable Materials to the Repository.[24] This guidance on the Protection for Human Subjects states: ". . . research material may only be utilized in accordance with the conditions stipulated by the cell repository IRB. Any additional use of this material requires prior review and approval by the cell repository IRB and, where appropriate, by an IRB at the recipient site, which must be convened under an applicable OPRR-approved Assurance."

The Coriell Cell Repositories do not consider its collection of human cell lines to be publicly available. Cell cultures and DNA samples are distributed only to qualified professional persons who are associated with recognized research, medical, educational, or industrial organizations engaged in health-related research or health delivery. Before cell cultures or DNA samples can be ordered, to ensure compliance with the federal regulations for the protection of human subjects (45 CFR Part 46), the principal investigator must provide the Repository with a description of the research to be done with the cell cultures or DNA samples ("Statement of Research Intent"). The principal investigator and the institutional official who can make legal commitments on behalf of the institution must also sign an "Assurance Form" detailing the terms and conditions of sale. Both the Assurance Form and the Statement of Research Intent must accompany each order placed with the Repository.

The National Disease Research Institute (NDRI), founded in 1980, was initially established as a network to obtain human tissue for diabetes research. Since then, it has grown into a center for retrieving and distributing a full range of normal and diseased cells, tissues and

[23] Office for Protection from Research Risks Guidance on Protections for Human Subjects in the National Institute of General Medical Sciences Human Genetic Mutant Cell Repository, May 21, 1997, http://locus.umdnj.edu/nigms/submit/gg.html

[24] Office for Protection from Research Risks Guidance on Protections for Human Subjects in the NIGMS Human Genetic Mutant Cell Repository, Submission of Non-Identifiable Materials to the Repository, May 21, 1997, http://locus.umdnj.edu/nigms/submit/snimr.html

organs for biomedical research. NDRI currently provides 140 different types of human tissues obtained from autopsies, eye banks, surgical procedures, and organ retrieval programs. More than 130,000 tissues have been retrieved and delivered to over 2000 scientists throughout the United States for use in research on more than 100 different diseases.

PATHOLOGY SPECIMENS

Large quantities of human biological materials are collected for diagnostic or therapeutic reasons. These specimens are usually sent to a clinical, diagnostic, or pathology laboratory for examination. These laboratories may be located at GME teaching institutions, physicians' offices, community hospitals, or independent laboratories. Most patients sign a general consent stating that after completion of any diagnostic tests, some of the specimen may be saved for research purposes. Although samples are made available for research, educational, and quality control purposes, the vast majority is never used for these purposes.

To be accredited, these pathology laboratories are required to keep pathological specimens for a minimum length of time. The Clinical Laboratory Improvement Amendments of 1988 (CLIA)(42 CFR 493) set forth the conditions that laboratories must meet to be certified to perform testing on human specimens. CLIA stipulates that laboratories must retain cytology slides for a minimum of 5 years, histopathology slides for a minimum or 10 years, and paraffin blocks for a minimum of 2 years (Clinical Laboratory Improvement Amendments, 1996). In addition, some states have regulations that require retention of pathology specimens for longer periods of time. Once the regulated length of time for storage is met, institutions may continue to store pathology specimens based on the policies of the institution.

Pathology Departments at Graduate Medical Education Teaching Institutions

Collectively, pathology departments at GME teaching institutions constitute the largest and oldest stores of human biological materials in the United States. GME teaching institutions include medical schools, Armed Forces hospitals, Veterans Affairs medical centers, the Public

Health Service, state, county and city hospitals, non-profit institutions, and health maintenance organizations. In 1997, there were 1,687 accredited GME teaching institutions (i.e., sites for clinical training) in the United States (American Medical Association, 1997). Combined, the GME pathology residency programs accumulate well over 8 million cases per year.. Most medical school pathology departments store specimens indefinitely. As a result, some tissues have been archived from 20 to over 100 years. A conservative estimate is that there are more than 160 million cases stored at GME teaching institutions with pathology residency programs, and several million more stored at those institutions without pathology residency programs.

Clinical Service and Diagnostic Laboratories

The majority of clinical service and diagnostic laboratories are not associated with GME teaching institutions. These include laboratories within physicians' offices or community hospitals, and independent laboratories. In 1991, there were approximately 640,000 clinical laboratories and other facilities that perform laboratory tests on human specimens (Department of Health and Human Services, 1991). The number of tissues stored at these laboratories varies greatly, but the minimum storage time is determined by CLIA and state regulations.

NEWBORN SCREENING LABORATORIES

Archives of newborn screening cards for inborn errors of metabolism (Guthrie Cards) represent an enormous source of banked DNA. Guthrie cards are special filter paper that contain dried blood spots from newborn babies, and contain identifying information, such as the mother's name and address, hospital of birth, baby's medical record number, baby's doctor's name and address. Guthrie cards are used to test newborns for several different diseases, including congenital hypothyroidism, phenylketonuria, galactosemia, hemoglobinopathies (e.g., sickle cell anemia), biotinidase deficiency, homocystinuria, Maple Syrup Urine disease, and cystic fibrosis. Interest in using Guthrie cards for population-wide genetic epidemiological studies has grown,

given the stability of DNA in dried blood, and the ability to analyze the DNA in these samples (McEwen and Reilly, 1994).

A 1994 survey of all newborn-screening programs in all 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands revealed that the majority of laboratories have accumulated less than 500,000 Guthrie Cards over the years. However, one laboratory reported a collection of more than 6 million Guthrie cards. The number of cards currently collected over a 1-year period ranged from less than 10,000 in 4 labs to more than 500,000 in 2 populous states (McEwen and Reilly, 1994). The trend in most states is to save Guthrie cards for longer and longer periods of time. Forty of the state newborn screening laboratories retain—at least for a short period of time—all the Guthrie cards that they receive through their newborn-screening programs, including those cards that test negative (McEwen and Reilly, 1994). The length of time that Guthrie cards are stored range from several weeks or months to indefinitely (McEwen and Reilly, 1994).

A growing recognition of the epidemiological utility of Guthrie cards for HIV seroprevalence surveys and DNA analysis has highlighted issues regarding retention, storage, and use of residual blood specimens from Guthrie cards. However, even though all states participate in some form of newborn screening, few have issued regulations that explicitly define the scope of permissible use of Guthrie card specimens (Andrews, 1995). While most laboratories would decline to release individually identifiable Guthrie cards to third parties without a written release or other explicit authorization, a large number would at least consider sharing anonymous cards for research purposes (McEwen and Reilly, 1994).

FORENSIC DNA BANKS

In 1989, the Virginia Division of Forensic Science[25] became the first state laboratory to offer DNA analyses to law enforcement agencies, and the first to create a DNA databank of previously convicted sex offenders. Subsequently, all 50 states have laws authorizing blood

[25] Virginia Division of Forensic Science, http://www.state.va.us/~dcjs/forensic/

sampling of some convicted felons to obtain DNA profiles. In addition, the Federal Bureau of Investigation[26] (FBI) is exploring ways to create a Forensic DNA bank for the District of Columbia (Finn, 1997).

In addition to collecting specimens from sex offenders and violent felons, a number of states also require specimens from juvenile offenders, non-violent felons, such as drug or white collar offenders, and those convicted of misdemeanors (McEwen, 1997). South Dakota requires specimens from people merely arrested (not convicted) for a sex offense (Finn, 1997), with several other states considering similar bills (McEwen, 1997). There is also a proposal to establish a federal DNA data bank that would include profiles of people convicted of offenses similar to those covered by most state laws in federal or military courts (McEwen, 1997).

Convicted offenders are required to provide blood, or in some cases, saliva, either at sentencing or before release from prison. Some states also require specimens from people already incarcerated before the law's effective dates. The DNA from these specimens is analyzed for its unique identification characteristics. Nationwide, specimens from about 380,000 offenders have been collected, mostly in Virginia and California, and about 116,000 specimens (30 percent) have been analyzed (McEwen, 1997). These DNA identification profiles are stored, along with the specimens themselves, to help identify suspects by matching biological evidence found at crime scenes to state DNA databases.

The DNA Identification Act of 1994 (P.L. 103-322, 1994 H.R. 3355, 108 Stat. 1796, §210304), a federal law enacted in the fall of 1994 as part of the Omnibus Crime Control Law, created a national oversight committee to develop guidelines for DNA forensics, established a grant program to assist state and local crime laboratories in developing or improving forensic DNA testing capabilities, and authorized the FBI to establish the Combined DNA Index System (CODIS) for law enforcement identification purposes. CODIS is a software system developed by the FBI and currently installed in 94 laboratories in 41 states and Washington, D.C. Using

[26] Federal Bureau of Investigations (FBI), http://www.fbi.gov/

CODIS, federal, state, and local law enforcement agencies are able to compare DNA profiles from crime scenes to DNA profiles of felons in the CODIS database.

On October 13, 1998, the FBI introduced a new tool for fighting violent crime, the National DNA Index System (NDIS). [27] NDIS is the final level of the CODIS system and serves as a repository for DNA profiles submitted by participating states. NDIS allows forensic laboratories throughout the United States to exchange and compare DNA profiles electronically, thereby linking unsolved serial violent crimes to each other and to known sex offenders.

UMBILICAL CORD BLOOD BANKS

Umbilical cord blood contains stem cells (progenitor cells that produce all other blood cells) which can be used to treat patients with blood diseases, certain genetic disorders, and patients receiving chemotherapy and/or radiation treatment for cancer. In 1988, the first successful human cord blood transplant was performed in a child with Fanconi Anemia using cord blood from a sibling (Gluckman et al., 1989). Since then, over 500 autologous and allogeneic umbilical cord blood transplants have been performed worldwide, with the majority done in the past two to three years (Perdahl-Wallace, 1997). Nonetheless, the Working Group on Ethical Issues in Umbilical Cord Blood recently concluded that "until additional data are obtained regarding safety and efficacy, umbilical cord blood banking and use ought to be considered an investigational technology rather than a proven treatment" (Sugarman et al., 1997).

ORGAN BANKS

Organ and tissue banks recover, process, store and distribute for transplantation human organs, bone, and tissue. Donations are from people who agree to donate upon their death and families who consent on behalf of the deceased. Some organ and tissue banks may also have tissue available for educational and research purposes. However, the demand for organs, bone

[27] http://www.fbi.gov

and tissue usually exceeds the current supply. Therefore, usually only organs and tissues not suitable for transplantation are available for research.

Under a NIH-sponsored program, cord blood is now being collected and stored at several large public banks around the United States. When parents donate their baby's cord blood to a public bank, they generally pay no fees, but they give up all rights to the sample to help build the public supply of cord blood for use in transplantation and research. In the last few years, privately owned companies have also begun offering umbilical cord blood banking services to individuals and families. When dealing with private storage companies, users pay a one-time fee for the collection, testing, and freezing of the blood, and an annual fee for storing the blood. The stored blood may be withdrawn if illness occurs later in life. Collectively, private and public cord banks store more than 18,000 units of cryopreserved umbilical cord blood, the majority of which is used for transplants. However, both private and public banks do supply come cord blood for research purposes.

BLOOD BANKS

The American Red Cross[27] collected approximately 5.8 million blood donations in 1996, about half of all U.S. blood donations. The American Red Cross usually maintains about a 3-day supply of fresh blood as well as approximately 20,000 units of frozen blood at any one time. The American Red Cross also maintains the world's largest registry of frozen rare blood.

Fresh red blood cells have a shelf life of 21 to 42 days depending on the preservative used, and platelets have a shelf life of 5 days. Plasma can be stored frozen for 1 to 5 years, and frozen whole blood can be stored for at least 10 years. Plasma that can not be transfused is used for making blood derivatives, such as Factor VIII for hemophiliacs, or for making diagnostic reagents.

[27] American Red Cross, http:// www.redcross.blood

Platelets and red cells that expire are sold for research purposes. Researchers are informed that the specimens have been found negative for all FDA-required tests, and only by special request, may be provided with the donor's age and gender.

PART II: DEFINITIONS AND ORIGINS OF HUMAN BIOLOGICAL MATERIALS

In this report, ***human biological material*** is defined as including everything from subcellular structures like DNA, to cells, tissue (bone, muscle, connective tissue and skin), organs (e.g., liver, bladder, heart, kidney), blood, gametes (sperm and ova), embryos, fetal tissue[28], and waste (urine, feces, sweat, hair and nail clippings, shed epithelial cells, placenta). By far, the most common source of such material 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 culture, or extracted DNA.

Identifiability of Specimen Sources

In the debate about research use of human biological materials, the language used to describe the identifiability of research samples varies. Previous guidelines and reports have categorized specimens by the conditions under which they are stored (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 4 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

[28] Due to the unique and ethically complex nature of research on gametes, embryos and fetal tissue, their use in research is not addressed in this report.

(i.e., the **specimen**) and sometimes refer to the material (i.e., the **sample**) and the accompanying information that is sent forward 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 has 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).

Research samples are often considered to fall within one or the other of two categories: 1) ***identifiable samples*** are those for which, one way or the other, the source individual can be identified, which means the sample can be connected, or linked, to the person from whom it came; and 2) ***unidentifiable samples*** are those for which the source individual cannot be identified by either the investigator or the repository. The reason one refers to the former as "one way or the other" identifiable, is because the information content of the research sample varies, from very little identifying information that, nevertheless, could allow one (perhaps with some difficulty) to link the sample to the person, to a sample that contains information allowing very easy identification of the person—with or without a name attached—from whom the sample was obtained.

For purposes of clarity and to facilitate discussion, NBAC adopted the following definitions of the diverse status 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.

Most repositories contain identified materials 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. An example of such a repository, however, is the following:

A repository might have collections of 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. Another example is 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:

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

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 impossible for the investigator, the repository, or a third party.

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 very small group of identifiable persons.

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, with 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 and use of unidentified samples follows:

A researcher studying malaria needs O^+ blood to grow the malaria parasite. The researcher recruits 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 fairly 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 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.

The last category of research samples is "identified." 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 stratify populations 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 (see chapters 4 and 5).

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.

PART III: RESEARCH USES OF HUMAN BIOLOGICAL MATERIALS

Once removed, human biological materials may serve many beneficial purposes, including clinical care, forensic determinations, identification of individuals, and research use. The most familiar and widespread use of such materials is in the diagnosis and treatment of illness. Another common use of human biological materials is for quality control purposes in diagnostic and pathology laboratories. Other uses include the identification of a person, such as in paternity testing and cases of abduction or soldiers missing in action, and forensic purposes in crime cases where biological evidence is available. The focus of this section, however, is the use of human biological materials in research.

In the examples described below, 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.

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.[30] For example,

> In 1953 autopsies of young 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

[30] For a survey of such benefits, see David Korn, "Contributions of the Human Tissue Archive to the Advancement of Medical Knowledge and the Public Health," a report to the National Bioethics Advisory Commission, January 1, 1998, in Volume II of this report.

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; Falk, 1981; Dannaher, 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 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 each of the many 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. In contrast, being able to apply such new technologies to archival materials, where clinical course, therapeutic response and outcome are already known, can save time and money, to say nothing of human suffering.

Recent progress in elucidating the initiation and progression of cancer has been most dramatic and gratifying in the area of colorectal cancer (Lenauer, 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).

There are countless examples in which investigators have used archival collections of human tissues to search for specific chromosomal and genetic abnormalities of pathogenetic interest. For example, a recent effort is attempting to decipher the genetics of prostate cancers, the most common cancer in American men and a significant cause of cancer morbidity and mortality (Smith, 1996). The goal of this new multi-institutional project is to differentiate the various forms of prostate cancer, determine the most effective methods of treatment for each, and eventually find a cure. The research is dependent on the availability of carefully characterized

tissue samples of prostate cancers and close correlation with clinical data to establish the natural history of the tumors and their responses to different therapeutic strategies.

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. Other large longitudinal studies include the Health Professionals Follow-up Study, Mr. Fit, and the Family 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.

The NIH-sponsored Bogalusa Heart Study,[31] at the Louisiana State University, has been ongoing since 1972 and is the longest and most detailed study of children in the world. The purpose of the study is to understand the environmental and hereditary aspects of early coronary artery disease, essential hypertension, and cardiovascular risk factors in African American and Caucasian children in the semi-rural community of Bogalusa, Louisiana. In addition, over 160 substudies have been conducted including special studies on socioeconomic evaluations, blood pressure, lipid levels, genetics, exercise, heart murmurs, and pathology. Knowledge gained in the study has been applied to develop, test and evaluate methods for cardiovascular risk intervention. The research involves longitudinal observations of more than 14,000 children and young adults, some of whom will be followed until 38 years of age.

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 human biological diversity. 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.[32] Researchers search for genes by constructing finer and finer maps of known gene locations and functions or by

[31] Bogalusa Heart Study, http://mcl.tulane.edu/cardiohealth/bog.htm

[32] 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 samples used in the study.

comparing DNA of individuals 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, 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. At this stage, having valid and definitive criteria that accurately determine a particular diagnosis or trait may make the difference between success and failure. The actual research designs selected in molecular genetics studies and the selected participants are closely allied.

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 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.

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.

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.[33] 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.

Diagnosis, family history, and DNA samples were collected using identical procedures across multiple sites. The collecting researchers were given a 12-month proprietary period for analyzing their data, at the end of which the data were made available to other qualified investigators. The repository contains information on 862 individuals with Alzheimer's disease, 432 individuals with bipolar disorder, and 270 individuals with schizophrenia.

These researchers founded a resource that is now in high demand. Requesting investigators receive a file of demographic and diagnostic variables necessary for genetic analysis, with accompanying documentation, access to DNA samples, a code manual listing additional clinical and demographic data, and pedigree drawings.

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

Although there are numerous additional investigator-initiated studies, some have not been able to recruit the necessary number of participants. Determining the necessary number is problematic since such estimates are specific to the underlying mode of genetic transmission, which is unknown. The more complex the transmission pattern, the larger the study must be. Researchers who began collecting 10 years ago would have thought that 100 to 200 affected individuals and relatives would have been adequate. Now that multiple susceptibility genes are hypothesized, much larger samples than previously expected are necessary.

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 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 are the health examination surveys conducted by the Centers for Disease Control and Prevention (CDC). Since 1960, the National Center for Health Statistics (NCHS) of the CDC 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[34], 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 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.

Community-Based Studies to Determine Gene Frequency

Certain diseases, particularly those with strong genetic components, are often found to be more common in groups that share similar characteristics, whether they be genes, environmental exposures, or lifestyles. For example, in the category of genetic disorders, Sickle Cell Anemia is predominantly found in African Americans, Cystic Fibrosis in Caucasians, particularly of European descent, Tay Sachs in individuals of Ashkenazi Jewish descent, and thalassemia in Mediterranean populations. These are all autosomal recessive disorders, requiring two defective genes for manifestation of the disorder, meaning otherwise healthy carriers (people with one

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

defective gene, and one normal gene) can only pass the disorder to their children by mating with another carrier (and even then the odds in each pregnancy of passing on the disorder are 1 in 4). The likelihood of two carriers producing affected offspring is greater in populations that are geographically, politically, socially, or culturally isolated or segregated.

In the 1980s there was growing evidence that there might be a genetic component to breast cancer. In 1990, researchers had determined that mutations in a gene, labeled BRCA1, and later another gene, BRCA2, cause inherited forms of breast and ovarian cancer. Knowing that breast cancer runs in families, investigators collected data on women whose mothers, grandmothers, or sisters had the disease (Easton, 1993; Tonin, 1995). Characteristic mutations were found in Ashkenazi Jews. In one study, investigators aimed to estimate the risk of breast and ovarian cancer in the Ashkenazi Jewish population through relatively simple assays to determine the frequency of these mutations (Struewing, 1997). They enlisted the participation of 5,331 Jewish men and women over the age of 20 living in the Washington, D.C. area. Participants provided family histories and blood samples. Participants were told at the beginning of the study that they would not be informed of the results of the test. The scientists found that over 2 percent of Ashkenazi Jews in the study population carried mutations in the BRCA1 or BRCA2 gene, conferring increased risks of breast, ovarian, and prostate cancer (Struewing, 1997). In comparison, less than one percent of the non-Jewish population carry a mutated BRCA1 and BRCA2 gene (Whittemore, 1997).

CONCLUSIONS

This chapter described the large volume of human biological materials that exists in the United States at this time, and it also provided examples of how these 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.

Many 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 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

REFERENCES

1. American Medical Association. 1997. *Graduate Medical Education Directory 1997-1998*. American Medical Association, Chicago.

2. Andrews LB. 1995. State laws and regulations governing newborn screening. American Bar Foundation, Chicago.

3. 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.

4. 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.

5. 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.

6. Clinical Laboratory Improvement Amendments. 1996. Clinical Laboratory Improvement Amendments of 1988 (CLIA). 42 CFR 493 (10-1-96 Edition); pp. 796-921.

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

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

9. Department of Health and Human Services. 1991. Clinical Laboratory Improvement Amendments (CLIA). November 29, 1992. Press Release 1991.11.29.

10. Enos WF, Holmes RH, Beyer J. Coronary disease among United States soldiers killed in action in Korea. *Journal of the American Medical Association* 1953; 152:1090-1093.

11. Enos WF, Beyer JC, Holmes RH. Pathogenesis of coronary disease in American soldiers killed in Korea. *Journal of the American Medical Association* 1955; 152:912-914.

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

13. Fearon ER, Hamilton SR, Vogelstein B. Clonal analysis of human colorectal tumors. *Science* 1987 Oct. 9; 238(4824):193-197.

14. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990 Jun 1; 61(5):759-767.

15. Finn P. 1997. Revolution Underway in Use of DNA Profiles. Bid to Link U.S. Databanks is Crime-Solving Edge. *The Washington Post*. Sunday, November 16, 1997; B4.

16. Flehinger BJ, Melamed MD, Zaman MB, 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 1984; 130:555-560.

17. Frost JK, Ball WCJr, Levin ML, et. al: Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study. *Am Rev Respir Dis* 1984; 130:549-554.

18. Gluckman E, Broxmeyer HE, Auerbach AD et al., 1989. Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical cord blood from an HLA-identical sibling. *New England Journal of Medicine* 321: 1174-1178.

19. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethlation of DNA from benign and malignant human colon neoplasms. *Science* 1985 Apr 12; 228(4696):187-190.

20. Herbst AL, Scully, RE. Adenocarcinoma of the vagina in adolescence. *Cancer*, 1970; 25:745-757.

21. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. *NEJM* 1971; 284:878-881.

22. Herbst AL, Robboy SJ, Scully RE, Poskanzer DC. Clear-cell adenocarcinoma of the vagina and cervix in girls: Analysis of 170 registry cases. *Am J Obstet Gynecol* 1974; 119:713-724.

23. Herbst AL. Clear cell adenocarcinoma and the current status of DES - exposed females. *Cancer* 1981 July 15 Supplement; 48:484-4881.

24. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* 1996 Oct 18; 87(2):159-170.

25. Kinzler KW, Nilbert MC, Su LK, et. al: Identification of FAP locus genes from chromosome 5q21. *Science* 1991 Aug 9; 253(5020):661-665.

26. Kinzler HW, Nilbert MC, Vogelstein B, et. al: Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* 1991 Mar 15; 251(4999):1366-1370.

27. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature* 1997 Apr 10; 386(6625):623-627.

28. Masliah E, Terry RD, Alford M, et. al: Cortical and subcortical patterns of synaptophysin-like immunoreactivity in Alzheimer's disease. *Am J Path* 1991; 138:235-246.

29. McEwen JE and Reilly PR. 1994. Stored Guthrie Cards as DNA "Banks". *American Journal of Human Genetics* 55: 196-200.

30. McEwen JE. 1997. DNA Data Banks. In *Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era*, MA Rothstein, ed., Chapter 11.

31. National Heart, Lung, and Blood Institute. 1996. Blood Specimen Repository: 1996 Catalog.

32. Perdahl-Wallace EB. 1997. Placental cord blood transplantation. *Transplant Forum* 4(2): 4-5.

33. Popper H, Thomas LB, Telles NC, et. al: Development of hepatic angiosarcoma in man induced by vinyl chloride, thorotrast, and arsenic - comparison with cases of unknown etiology. *Am J Pathol* 1978; 92:349-376.

34. Raine CS. The Norton lecture: A review of the oligodendrocyte in the multiple sclerosis lesion. *J Neuroimmunol* 1997; 77:135-152.

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

36. Rosai J. Pathology. A Historical Opportunity. *Am J Path* 1997; 151: 3-6.

37. Roth F. The sequelae of chronic arsenic poisoning in Moselle vintners. *German Med Monthly* 1957; 2:172-175.

38. Smith JR, Freije D, Carpten JD, et. al: Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* 1996; 274:1371-1374.

39. Solberg LA, Strong JP. Risk factors and atherosclerotic lesions. A review of autopsy studies. *Arteriosclerosis* 1983; 3:187-198.

40. Strong JP. Coronary atherosclerosis in soldiers. A clue to the natural history of atherosclerosis in the young. *Journal of the American Medical Association* 1986; 256:2863-2866.

41. Struewing JP, Hargte P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, Brody LC, and Tucker MA. 1997. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med* 336:1401-1408.

42. Sugarman J, Kaalund V, Kodish E, Marshall MF, Reisner EG, Wilfond BS, and Wolpe PR. 1997. Ethical issues in umbilical cord blood banking. *Journal of the American Medical Association* 278(11): 938-943.

43. Technical Working Group on DNA Analysis Methods (TWGDAM). 1989. The Combined DNA Index System (CODIS): A theoretical model. In *DNA Fingerprinting: An Introduction*, LT Kirby, ed., New York: Stockton Press, pp. 279-317.

44. Vogelstein B, Fearon ER, Kern SE, et. al: Allelotype of colorectal carcinomas. *Science* 1989 Apr 14; 244(4901):207-211.

45. Vogelstein B, Fearon ER, Hamilton SR, et. al: Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988 Sep 1; 319(9):525-532.

46. Will RG, Ironside JW, Zeidler M, et. al: A new variant of Creutzfeldt-Jakob disease in the UK. *Lancet* 1996; 347:921-925.

47. Whittemore AS, Gong G, and Itnyre J. 1997. Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: Results from three U.S. population-based case-control studies of ovarian cancer. *Am J Hum Genet* 60:496-504.

48. Wrobel, S., "Serendipity, science and a new Hantavirus," *FASEB J* 1995; 9:1247-1254.

49. Younge PA, Hertig AT, Armstrong D. A study of 135 cases of carcinoma in situ of the cervix at the Free Hospital for Women. *Am J Obstet Gynecol* 1949; 58:867-892.

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