Governance Plan

BWH/Harvard Cohorts Biorepository

A. Name of the Custodian:
Brigham and Women's Hospital and Harvard School of Public Health (Directors: Dr. Shelley Tworoger, PhD; Dr. Rulla Tamimi, ScD; Dr. Sherilyn Sawyer, PhD)

B. Summary of the Project:
The BWH/Harvard Cohorts Biorepository (hereafter called the Biorepository) manages about 3.5 million biological specimens collected from over 200,000 participants in several large cohort studies and appropriately prepares the samples for scientific analyses.
These cohorts include:

- Nurses' Health Study (***NHS***; n=121,700)
- Nurses' Health Study II (***NHSII***; n=116,430)
- Nurses' Health Study III (***NHSIII***; currently enrolling)
- Health Professionals Follow-up Study (***HPFS***; n=51,529)
- Physician's Health Study (***PHS***; n=22,071)
- Growing Up Today Studies (***GUTS/GUTSII***; n=27,725)

Each cohort has collected or is actively collecting specimens to study multiple diseases, such as cancer, cardiovascular disease, and diabetes, among others. These samples are particularly valuable because participants already have been followed as much as 30 years since collection and they are accompanied by a wealth of data on medical histories and health-related behaviors. Sample types stored in the biorepository include plasma, white blood cells, red blood cells, whole blood, urine, DNA, toenails, saliva, stool, hair, mammograms, and paraffin-embedded human tissue. **These samples are an extremely valuable, but finite, resource; thus much effort is put forth to ensure high quality, efficient specimen handling. The goal of the Biorepository is to maximize use of samples in a cost-effective, ethical, state of the art manner.**

C. Governance Structure of the Project:
The Biorepository is a core facility jointly run by the Harvard School of Public Health (HSPH) and Brigham and Women's Hospital (BWH), an affiliate of Harvard Medical School (HMS). The financial management of the core is based at the BWH. The Biorepository benefits from the support of the two institutions as it receives space as well as administrative oversight from both the university and hospital. Faculty members from these and many other institutions, many with joint appointments, rely on services provided by the Biorepository.

The Biorepository is comparatively unique within the HSPH and BWH. Specimens, collected up to 30 years ago, are matched with information on diet, lifestyle, medication use, morbidity, mortality, or other data important for conducting health research. Each cohort (NHS, NHSII, HPFS, PHS, GUTS) represented in the Biorepository still is conducting follow-up today, most with very little attrition, and new cohorts, including the NHSIII, are actively enrolling participants. For example, the NHS is now in its 38th year of follow-up. Over 90% of women who are still alive completed the most recent questionnaire in 2012; among those providing biologic samples, follow-up is >95%. Although each cohort has slightly different approaches to maintaining contact with participants, comparable data have been collected and similar study designs used when

examining biospecimens. The two primary designs are nested case-control studies, comparing those who are diagnosed with a disease of interest during follow-up versus those without that diagnosis, and sub-cohort studies. Thus, many projects pool results from multiple Biorepository cohorts or are part of large collaborative projects pooling data from studies around the world.

The Biorepository is a mandated revenue-neutral facility. It can neither make nor lose money each year. Financial oversight is provided by the BWH to aid the core directors in determining appropriate prices for sample preparation, assisting with invoicing, and managing contracts. At both the BWH and HSPH, a financial/grants administrator has been assigned to maintain descriptions of the core for grant applications and to help in our auditing process.

The Biorepository is spread across multiple sites. We have two primary freezer farm locations, which allows the separation of samples in geographic location. These include a 2,364 square foot space located at 665 Huntington Avenue in the Francis Xavier Building (FXB) at the HSPH, which houses 47 ultra-low temperature liquid nitrogen freezers and 3 -80^{o}C mechanical freezers, with expansion space for more liquid nitrogen freezers available. Piping connects each freezer to two external 3,000-gallon liquid nitrogen tanks, which provide a continuous flow of liquid nitrogen. Freezers automatically are fed liquid nitrogen when needed. In addition, the BWH rents space in the basement of the Massachusetts College of Pharmacy (MCP; 179 Longwood Avenue) to house the other part of our freezer farm. The three rooms in MCP (B43, B44, and B34) total 2,461 square feet and contain 88 ultra-low temperature liquid nitrogen freezers and 3 -80^{o}C mechanical freezers. These freezers are connected to an external 6,000-gallon liquid nitrogen tank. Due to the large number of freezers at this location, a liquid nitrogen reservoir tank sits in the middle room to maintain a constant pressure in the system at all times. Including freezers to hold on-going projects in the main laboratory space, we have 150 freezers in total. All freezers at the BWH and HSPH are wired into the HSPH Operations Department, which provides continuous (24/7) monitoring. Laboratory staff checks every individual freezer to confirm proper temperature, liquid nitrogen levels, etc., twice a week. Three liquid nitrogen freezers (one at the HSPH and two at the MCP) and one mechanical freezer (at MCP) are kept empty, but cold. In the event of a freezer malfunction, samples can be transferred immediately to the empty freezers. Further, each freezer farm with liquid nitrogen freezers has oxygen sensors and an emergency air evacuation system. NHS and HPFS toenail and hair samples are kept in cabinets at the BWH facility at 221 Longwood Ave. and the HSPH Landmark Center in Boston, MA. Further, there is a Sakura Lab Aid Ultra cabinet for FFPE blocks, which has storage for 21,000 blocks, and steel slide cabinets for H&E slides, that are maintained in ID order. All tissue microarrays are stored in a secured, fire-safe Sentry-Safe® Safe. To date, the resource has collected specimens from over 10,000 individuals with confirmed cancer. Further, there is a Lumysis 85 laser film scanner and a high-resolution computer monitor for mammographic density measurements in a room that provides appropriate darkness for reading.

Bench space for the core is located at 221 Longwood Ave., in the Eugene Braunwald Research Center (EBRC), rooms 611, 615, and 619, owned by the BWH. The 1,834 square foot area contains a fume hood, seven liquid nitrogen freezers connected to dewars (independent tanks for transporting liquid nitrogen), and bench space for up to 12 laboratory personnel to prepare biospecimens for shipment. In addition, the staff has access to equipment shared with other investigators located on the floor including a Sorvall ultra-centrifuge. We also have bench space for up to four laboratory personnel at the HSPH at 667 Huntington Ave., room 214, with two liquid nitrogen freezers connected to dewars that store samples for active projects. The Operations director and senior laboratory manager have offices totaling 228 square feet, with computers, internet access, and a personal printer that are all linked to the cohorts' main

computer UNIX system. In addition, laboratory personnel have a desk and computer access to the freezer databases for working with the laboratory information management system. The laboratory has three dedicated computers and bench space for printing labels designed for ultra-low temperature freezing; the printers are Brady 600X-Plus II thermal-heat transfer label printers. Participant ID, volume, sample type, and a 2D barcode are printed on labels. The laboratory also has three Code Corporation 2.0 gun-format hand scanners for scanning vials going into permanent storage, and two laptops for updating freezer storage data in real-time as samples are pulled from the repository. We have additional dedicated space for the collection and processing of tumor tissue, containing four desks and bench space for handling specimens at the Landmark Center.

Administrative support at the BWH and HSPH helps manage our contract with Airgas, who maintains the external bulk tanks and supplies the liquid nitrogen that feed the freezer farm, including ensuring timely delivery of liquid nitrogen, appropriate billing and payment, and that all safety standards are met.

The Biorepository leadership relies heavily on faculty from the HSPH and BWH. Dr. Shelley Tworoger, Associate Professor of Medicine and Epidemiology, has been the Director of the Biorepository since 2006 and was the Assistant Director from 2004-2005. Dr. Tworoger is assisted by the Tissue Director, Dr. Rulla Tamimi, Associate Professor of Medicine and Epidemiology, and Operations Director, Dr. Sherilyn Sawyer.

An advisory board oversees the Biorepository (Table 1), meeting quarterly to discuss major issues, and individual members are consulted on an *ad hoc* basis. The Biorepository also is a member of the International Society for Biological and Environmental Repositories, which provides training and a wide-range of resources on repository management.

Table 1. Biorepository Advisory Board Members

Member name	Qualification
Dr. Shelley Tworoger	Director of the Biorepository
Dr. Rulla Tamimi	Tissue Director
Dr. Sherilyn Sawyer	Operations Director
Dr. Meir Stampfer	PI, NHS
Dr. Walter Willett	PI, NHSII and HPFS
Dr. Stacey Missmer	PI, GUTS
Dr. Jorge Chavarro	PI, NHSIII
Dr. Francine Grodstein	Project Director, NHS
Dr. Heather Eliassen	Project Director, NHSII
Dr. Eric Rimm	Project Director, HPFS
Dr. Immaculata DeVivo	Director, DF/HCC Genotyping Core
Dr. Shuji Ogino	Pathologist, colorectal cancer
Dr. Andrew Beck	Pathologist, breast cancer
Dr. Jonathan Hecht	Pathologist, ovarian cancer
Dr. Edwin Silverman	Chief, Channing Div. of Network Med.

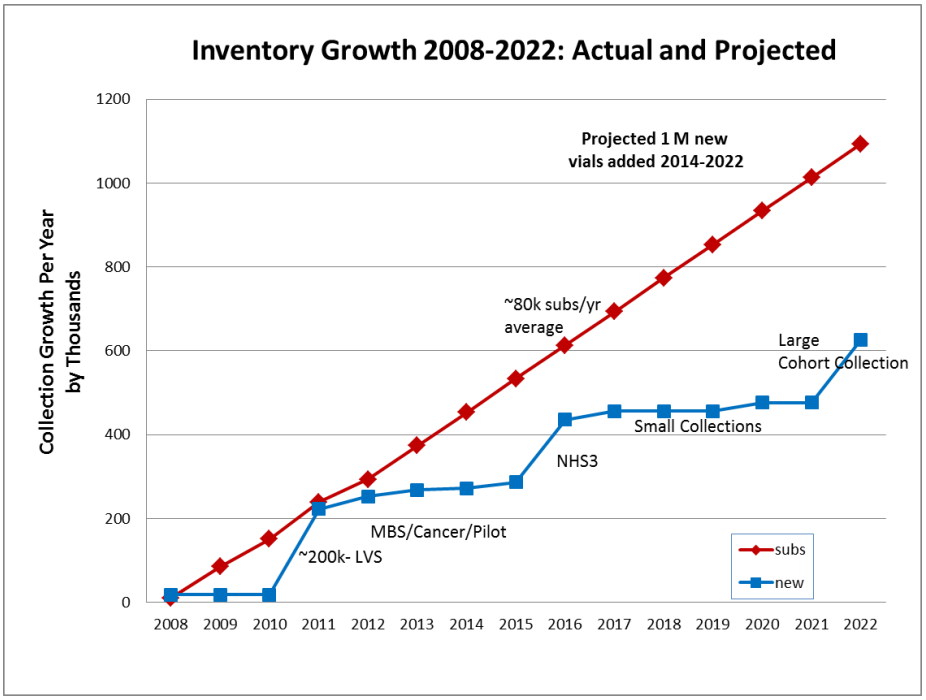
A highly qualified and experienced staff supports the Biorepository, including a PhD level Operations Director, a senior laboratory manager, project managers, data managers, laboratory information management system (LIMS) developers, and a team of research assistants. All direct laboratory personnel are paid by the Biorepository's operational budget, which also covers supplies, freezer maintenance, liquid nitrogen, and pilot studies. It uses a per aliquot charging system to equitably distribute project costs. The LIMS developers and data managers are supported through the Informatics Facility fee. A portion of the Biorepository operational budget, support for programming/LIMS infrastructure, on-going tissue collections (see section B.6) are covered

by UM1 infrastructure grants from the National Cancer Institute for the NHS (CA186107; PI: Stampfer), NHSII (CA176726; PI: Willett), and HPFS (CA167552; PI: Willett). Laboratory and data management personnel meet bi-weekly with the Biorepository Directors to discuss projects, priorities, infrastructure, and other laboratory issues

The Biorepository uses a fee-for-service model. The standardized charging system for projects utilizing biological specimens from the cohorts equitably distributes costs across all studies and negates the need to allocate laboratory personnel and supply costs across multiple grants. The Biorepository has multiple services, including existing sample preparation, new sample collection, and consulting. The budget includes the salaries for the senior laboratory manager, project managers, and research assistants; costs for supplies (e.g., pipettes, gloves, cryovials, labels), freezer maintenance and repair; and costs of routine pilot studies (e.g., to evaluate laboratory reproducibility for a new assay). The pricing schema is per aliquot sent out to an external laboratory or per hour for sample collection and consulting. If multiple aliquots of a single sample type (e.g., plasma) from the same participant are accessed at one time, the second, third, etc. aliquots are charged a reduced price. For example, in FY2015 the costs are $42 for the first aliquot per participant accessed and $21 for the second and subsequent aliquots. With this system, we annually set per sample charges that allow us to recoup our expenditures and break even each year. These charges are reviewed and updated annually through the BWH Core Administration and are approved by the Advisory Board. Documentation regarding core services, fees, pilot requirements, and data management issues are available on-line at cohortsbio.bwh.harvard.edu.

D. Integrity of Biospecimens and Data:

Since the Biorepository protects a finite resource, substantial effort is put forth to ensure high quality and efficient specimen handling and assay methodology. The goal is to maximize research use of biospecimens in a cost-effective, ethical, state of the art manner. The aim is to collect, store, process, and distribute biological specimens to better understand the etiology of multiple human diseases.



With the pilot and quality control samples, and taking into account prior usage of samples, the Biorepository have over 3 million samples in storage at this time. Inventory growth is shown in the figure below. The Biorepository currently contains samples from 12 primary biospecimen collections across the NHS, NHSII, HPFS, GUTS, and PHS, along with specimens from 20 smaller collections in these cohorts and NHSIII (Table 2). The usage of each collection varies, with older collections having more projects.

Table 2. Collections within the Biorepository

Years	Purpose	# of Participants	Sample Type(s)
Nurses' Health Study (NHS)			
1982-1984	Main toenail cohort	62641	Toenails
1989-1990	Main blood cohort	32826	Blood
1991	Reproducibility study yr 2	227	Blood
1992	Reproducibility study yr 3	191	Blood
1996	Folate trial, baseline	685	Blood
1999	Folate trial, compliance	333	Blood
2000-2002	Main cohort, 2nd coll.	18743	Blood, Urine
2001-2004	Main cheek cohort	33040	Cheek Cells
2003	Renal function cohort	1992	Blood, Urine
2007	Cognitive function cohort	130	Blood
2007-2008	Renal function cohort	1763	Blood, Urine
2010-2011	Diet validation study	375	Blood, Urine, Saliva
Nurses' Health Study II (NHSII)			
1996-1999	Main blood cohort	29611	Blood, Urine
2000	Reproducibility study yr 2	297	Blood, Urine
2001	Reproducibility study yr 3	200	Blood, Urine
2003	Renal function cohort	1672	Blood, Urine
2006	Main cheek cohort	29392	Cheek Cells
2009	Light/melatonin study	180	24-hour Urine
2010-2011	Diet validation study	375	Blood, Urine, Saliva
2008-2011	Main cohort, 2nd coll.	16,510	Blood, Urine
2013-2014	Mind Body Study	236	Blood, Urine, Saliva, Stool, Hair, Toenails
Health Professionals Follow-up Study (HPFS)			
1987	Main toenail cohort	33737	Toenails
1993-1994	Main blood cohort	18158	Blood
1996-1997	Racial diversity study	152	Blood
2000	Reproducibility study yr 2	210	Blood
2001	Reproducibility study yr 3	194	Blood
1996	Folate trial, baseline	249	Blood
1999	Folate trial, compliance	86	Blood
2006	Main cheek cohort	13979	Cheek Cells
2012-2013	Diet validation study	450	Blood, Urine, Saliva, Stool
Physicians' Health Studies (PHS)			
1982-1984	Main blood cohort	14916	Blood
1995-2000	Main blood cohort, PHSII	1620	Blood
Growing Up Today Studies (GUTS)			
2011-2013	Stress cohort	3,000	Saliva
2013	Infertility	50	Semen
Nurses' Health Study III			
2014-	Pilot collection	250	Blood, Urine

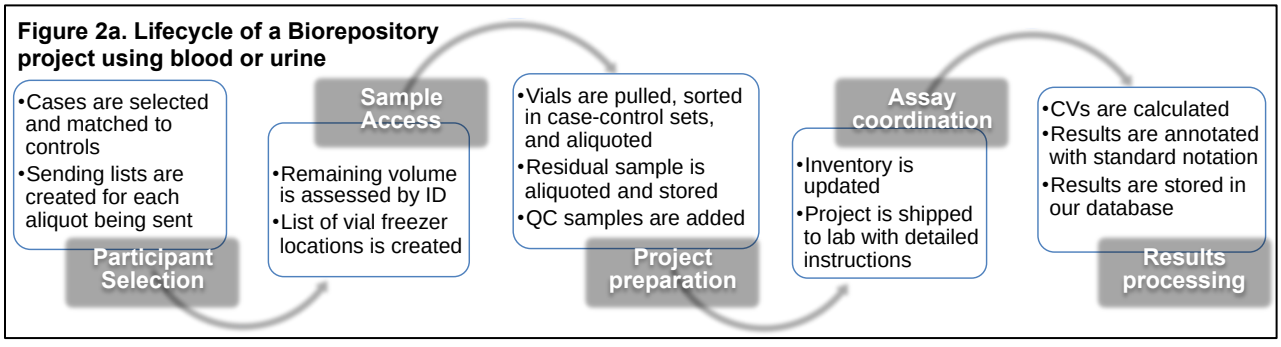
In general, most of the sample collections have followed a similar strategy. First, participants are queried (either by letter, email, or questionnaire) about their willingness to participate in a sample collection. Those who are eligible and willing are sent a kit containing the appropriate supplies to provide the specimens, depending on the sample types being collected. The kit also contains directions on how to collect the sample and a short questionnaire detailing the specifics of the sample collection (e.g., date and time of collection, number of hours since last eating, current weight, recent medication use), as well as a detailed consent form for the future use of the specimens. Participants then collect the sample(s), package them with a cold-pack, and mail them back to our laboratory via overnight courier. There, the samples are processed as appropriate. In the NHS, NHSII, and HPFS, whole blood was separated into plasma, white blood cells (WBC), and red blood cells (RBC) for storage, whereas in the PHS, plasma and whole blood were stored. Urine and saliva are stored without the use of a preservative. Cheek cell DNA was extracted on arrival and then stored.

1. Biospecimen management and quality control

a. Identification of cases and controls. Confirmed cases of the disease of interest diagnosed after sample collection are matched to 1-3 controls at risk of disease (e.g., alive) when the case occurred. Controls are selected at random and matched on factors such as cohort, year of birth (±1yr), menopausal status at blood draw and diagnosis (premenopause, postmenopause, unknown), fasting (≥8hr, <8hr), and date and time of day (±1mo, ±2hr).

b. Sample storage/preparation. During initial processing of blood and urine samples, vials were filled as much as possible to minimize air contact, and gasketed screw tops were attached to prevent water vapor exposure. Vials of plasma, whole blood, buffy coat, red blood cells, saliva, and urine are stored in the vapor phase of liquid nitrogen freezers (<–130^{0}C). Each participant's samples were aliquoted into multiple vials, affixed with labels rated for ultra-cold storage, and split between several freezers so that only part of a participant's sample would be lost if a freezer were to thaw.

The Biorepository has standard operating procedures for conducting projects using blood or urine samples (Figure 2a). For each project, specifications outlining which samples should be included and the relevant assays is approved by the investigator. Then a list of eligible IDs is generated and available specimen volume assessed. We reserve at least 500uL of plasma or urine for each participant to have sample for future, highly important assays. Next, we generate a list of cryovials and freezer locations from our database. After pulling samples from the freezer, cases and matched control samples will be placed next to one another, in random order, and handled together for all remaining steps. Samples will be re-aliquotted if needed, using the smallest aliquot providing adequate volume for the assay(s) of interest. Samples are aliquoted into appropriate volumes for each project; the remaining volume will be aliquotted into 0.1, 0.2, and 0.3 ml volumes to minimize future freeze-thaw cycles, returned to the freezer, and the database is updated. Samples are labeled by participant ID and a unique ID to maintain blinding. Matched case-control sets are indicated by an empty space between sets in the box so they are assayed in the same batch.



Because of the necessity of limiting freeze-thaw cycles to protect sample integrity, the Biorepository is always growing as new subvials are made. We specifically do not intersperse the originally collected vials and the subvials to further ensure safekeeping of the sample (i.e., in case one freezer thaws, the subvials are in another freezer). For two oldest cohorts, PHS and NHS, we have consolidated original vials to gain additional freezer space. However, substantial usage of the samples is necessary to gain enough space to justify the time and resources to conduct this process; thus we assess annually whether consolidation is warranted.

c. Quality control. Before a set of samples is sent to a laboratory, QC pool replicates from at least 3 pools are randomly interspersed throughout the project to assess assay variability. We have sufficiently large pools to use the same QCs for many years, allowing for comparison of assay results across projects. We include 10% QCs, which are labeled by a fake ID number and placed in the box in "sets" so that they are indistinguishable from the case-control sample sets. To further facilitate comparison of results across disease sites, we include an identical set of 20 samples from local volunteers in each project (i.e., "drift samples"). The project is double-checked for correct ID and volume and then shipped via overnight courier in a Styrofoam box packed with sufficient dry ice to keep the specimens frozen for two days, in case of a delay. For local laboratories, we hand-courier the samples on dry ice. Detailed instructions accompany each project outlining the number of samples in the project, a listing of sample IDs and locations, the assays to be conducted, and how results should be returned to the Biorepository with a formatted excel file and requests for definitions of samples that have values above or below the limit of detection.

To further protect the samples and their use, three pilot studies are required for each assay (in the Biomarker Laboratory) to ensure that the assay can be appropriately conducted on the samples. Also, assays with volumes of 500uL or more must be approved by the advisory board prior to piloting. The pilots are outlined below in Table 3. Costs of pulling pilot samples and domestic shipping are covered by the Biorepository. The investigator is responsible for assay and international shipping costs, as well as negotiating with the assay lab.The Biorepository director, or designee, will determine the acceptability of the results. All projects must meet the following guidelines:

(a) Stability of the biomarker for 24-48 hours in whole blood

Blood samples from participants were received in the Laboratory as whole blood. Once received, they were centrifuged and aliquoted into plasma, buffy coat, and RBC components and archived in liquid nitrogen freezers. All samples were processed on the day they were received. The vast majority of blood samples were received and processed within 24-30 hours of collection; <3% were received more than 48 hours after collection. The majority of samples were kept cool (with a chill pack) during transport to our laboratory.

(b) Appropriateness of using samples collected in sodium heparin or EDTA tubes

Samples were collected using sodium heparin or EDTA as the anticoagulant. The laboratory must confirm that these tubes are routinely accepted for the analysis of interest, otherwise a pilot study will need to be conducted to establish that the anticoagulant will not interfere with assay performance.

(c) Laboratory assay to be used

All assays must be conducted using the best available technology to ensure that the appropriate parameter is assayed, the sample volume required is minimized, and the assay reproducibility is maximized. The definition of "acceptable" volume will be determined on a study-by-study basis and will depend in large part on the importance/priority of the study hypothesis. In the proposal, the investigator must be clear in describing the various assay methods currently available and their rationale for using the specific assay being proposed.

Table 4. Pilot studies required by the Biorepository

Pilot name/Description	Approx. number of samples	Notes
1. Split : blinded duplicate samples to assess CVs of the QCs and participant samples.	▪ 16 participants (8 Heparin, 8 EDTA) x 2 replicates; ▪ 2 QC pool samples x 2 replicates *TOTAL: 36*	Split pilots are required when: ▪ assay has never been used in cohort samples ▪ pilot results are >3 yrs old before cohort samples are used ▪ PI will use a different assay lab or preservative type than that used in original pilot study ▪ assay lab personnel has changed significantly ▪ assay methodology has changed
2. Processing Method (PM): samples processed immediately, versus kept chilled *as whole blood* with a cool pack and processed 24 and 48 hours after blood draw to assess CVs of the QCs, and CVs, Spearman correlations, and ICCs across processing method times.	▪ 14 participants (7 heparin, 7 EDTA) x 3 time points; ▪ 2 QC pool samples x 2 replicates *TOTAL: 46*	▪ External data from samples collected in a similar fashion may be substituted for this pilot (contact Mary Townsend or Shelley Tworoger to determine acceptability of external data)
3. Within-person Stability (WPS): repeated samples from the same person, 1-2 years apart, to determine whether one measure represents long-term exposure; CVs of the QCs, and CVs, Spearman correlations, and ICCs across draws are calculated. ** *Some older assays may not have WPS pilot results. PIs are encouraged to complete this pilot, even for assays used successfully in previous projects. The Biorepository will send the pilot samples along with the PI's project samples to avoid any delay in completing the project.*	▪ 40 participants (all heparin *or* EDTA) x 2 time points; ▪ 4 QC pool samples x 2 replicates *TOTAL: 88*	▪ Available datasets for plasma assays include: postmenopausal women (NHS), premenopausal women with luteal and follicular samples (NHSII), men (HPFS) ▪ Available datasets for urine assays include: postmenopausal women (NHS) or premenopausal women with luteal urine samples (NHSII) ▪ Larger WPS studies (e.g., more time points or participants) can be conducted for measurement error correction purposes. After the first 80 participant samples, PIs will be charged for additional samples at the usual BLOB price. ▪ External data supporting analyte stability over time may be substituted for this pilot (contact Mary Townsend or Shelley Tworoger to determine acceptability of external data)

(d) Reproducibility of the laboratory assay

The assay laboratory must be able to conduct the assay with a high degree of precision (i.e., low coefficient of variation or high reliability coefficient). This information must be obtained through a blinded evaluation of the laboratory. Unfortunately, coefficients of variation provided by laboratory investigators are not sufficient, as, in our experience, these data do not always reflect the true magnitude of laboratory error. The evaluation

must be recent and, if at all possible, should have been performed by the same technician who will be conducting the study analyses.

(e) Range of the biomarker in the cohort

For many biomarkers of interest knowledge of a usual range in an adult population will be sufficient (e.g., plasma antioxidant levels); in this instance, the usual range and how this range was determined (i.e., in what population) should overlap with the expected range in the cohort being used. However, for other assays, where the range may vary substantially by population (e.g., levels of DDE/PCBs), a pilot study to determine levels observed in the cohort may need to be conducted prior to receiving final approval for conducting a project.

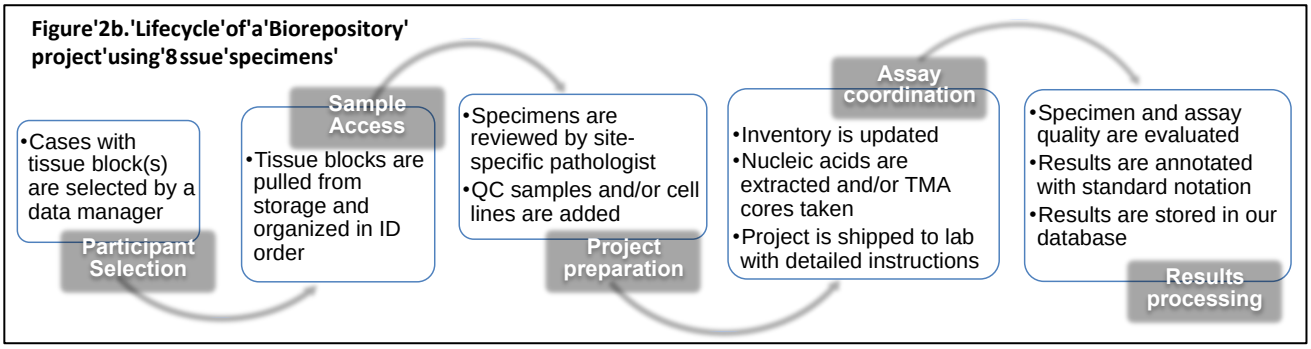
(f) Stability of the biomarker over time

In most of the cohorts, only one blood sample per participant has been collected. Thus data must indicate that assay of a single blood sample will provide an integrated measure of longer-term exposure that an association between the biomarker and disease could reasonably be detected, if indeed one exists. An example of an assay that would not be appropriate would be luteinizing hormone; its pulsatile secretion results in large peaks and valleys that a single sample will provide an extremely misclassified exposure measure. If these data are not already available, applicants are advised to consider conducting a pilot study to assess stability over a minimum of a 4-week period.

2. Tissue management and quality control

a. Identification of cases and controls. Individuals of specific cancers or benign diseases that have appropriate tissue blocks or slides are identified by the data manager. In general, we use prospective analyses for tissue characteristics and thus do not select controls.

b. Sample storage/preparation. The Biorepository has dedicated space for the collection and processing of tumor tissue, with four desks and bench space for handling specimens (see Facilities and Resources). Expert pathologists for each tumor site (see Table 1) will select appropriate blocks for TMAs and DNA/RNA extraction and circle epithelial areas of tumor and normal tissue on the corresponding H&E slide for each selected block. Blocks then are put into ID order and prepared for creation of TMAs or DNA/RNA extraction (Figure 2b).



TMAs are created at the DF/HCC Specialized Histopathology Core using a tissue arrayer (Beecher Instruments, Sun Prairie, WI), which removes small, circular cores from tissue blocks and arrays up to 450 cores on the same block. Besides increasing throughput, other benefits include conservation of tissue, improved internal experimental control, and reduced reagents consumption. The TMA core receives matching blocks and annotated slides, along with a corresponding spreadsheet with tumor details. Three cores are punched from each block and are arrayed in the recipient TMA; each spot has a unique X and Y coordinate. Cores are 0.6mm in diameter. A corresponding spreadsheet is created for assay results. For RNA and DNA extraction, blocks will be sent to the DF/HCC High Throughput Genotyping Core (Core PI, Dr.

DeVivo, is on our Advisory Board, Table 1). The first step is to extract total RNA from 3 core samples taken from FFPE tumor tissue (and adjacent normal tissue as appropriate); note that paired normal and tumor tissues will be handled together throughout the analytic process. RNA will be isolated using the Qiagen AllPrep RNA isolation kit (Hilden, Germany). The AllPrep kit allows for the collection of a DNA pellet from the specimen, which will be extracted and used for PIK3CA profiling in Project 1 and WES in Project 2 (the pellet will be stored for future use for Project 3). Total RNA yield will be measured by ribogreen or Nanodrop Technologies ND-1000 Spectrophotometer (Wilmington, DE). DNA concentration will be measured using picogreen. Resultant RNA and DNA specimens will be labeled by participant ID and stored in -80^{o}C mechanical, alarmed freezers.

c. Quality control. The Biorepository has protocols for ensuring high quality assay results for IHC and tissue RNA or DNA profiling assays. For IHC, we optimize antibodies on known positive and negative tissue. Then we perform the assay on a "test" TMA that uses extra cores from participant tissue to assess staining on our samples and the distribution of the stain. Slides of the test array are cut to optimize antibody assays and to ensure that the assay works on samples collected from a variety of hospitals around the United States that vary in age from 2-30 years. Further each cancer endpoint is required to have a dedicated pathologist who has been vetted and approved by the Advisory Board. For nucleic acid extraction, total RNA (and DNA) yields are assessed to ensure that enough sample exists to conduct the relevant assay. We also conduct pilot experiments of 10 technical replicates using discarded FFPE tissue for each platform and tumor type to be used. For the primary assays, we will include technical replicates (n=4-6) on each plate. QC RNA will be obtained from discarded FFPE tumor tissue of a similar age/histology as our tumors and from cell lines.

In addition, mammograms are collected to measure breast density. Films of the cranio-caudal views of each breast are digitized at 261 microns/pixel with a Lumysis 85 laser film scanner, which covers a range of 0 to 4.0 optical densities. The software for the computer-assisted thresh-holding was developed at the University of Toronto.

Overall both arms of the Biorepository have multiple functions, with standardized operating procedures in place to ensure the accurate and best use of the biological specimens that have been and will be collected. To ensure the integrity of the human research participants' data that accompany the biospecimens, all projects must receive approval from the BWH Human Subjects Committee prior to implementation (note that HSPH accepts the BWH IRB approval). Further, as analyses of genetic susceptibility to disease are associated with complex ethical considerations, a full discussion of the ethical implications of these analyses must be conducted. The senior cohort investigators and/or the Advisory Board can consult with the appropriate Ethical Advisory Committee prior to seeking approval from the BWH Human Subjects Committee. Investigators are aware that analyses, which identify women at very high risk of disease, are particularly problematic in this regard. Further all data are stored on a protected server that sits behind the BWH firewall and are password protected. Participant data and their biospecimens are stored by ID number only, and the link between participant ID numbers and their HIPPA information are stored separately with access limited to those who are responsible for following participants.

E. Access to Biospecimens and Data:

Both internal and external investigators can request the use of specimens within the Biorespository. We have an extensive history of successful collaborations with investigators external to the BWH and HSPH community. We welcome new collaborations and strive to make their establishment as simple and transparent as possible while also maximizing the scientific yield. To this effect, we have developed the following policies and guidelines for external access

to archived data and biospecimens with the approval of the cohorts' external Advisory Boards and in compliance with our NIH approved plan for data-sharing. The procedures are below.

Internal Investigators: Internal investigators must go through a multi-tiered system to obtain approval to submit a grant to use the Biorepository. First, the investigator must contact the Biorepository to obtain an estimate of the charges to be included in the budget, by completing the appropriate forms. Further the assays will be reviewed to determine whether any pilots are needed. Once the project has been approved by the appropriate Biorepository manager, then the investigator must submit the project for review by at least one member of the Advisory Board. The purpose of this review is to provide constructive feedback, avoid presentation of proposals that would not be approved by the larger group because of feasibility issues, internal consideration to maximize fundability and feasibility, and provide a global view on costs and resource utilization. Once approval is obtained from this committee, the grant aims must be presented at an investigator meeting for the appropriate cohort(s), which occur every other week. Here the scientific merit of the study will be discussed and approved.

External Investigators: Any investigator wishing to develop a collaboration with one of the cohorts and to use the biospecimens from the Biorepository must send a one to two page description of the proposed analyses ("letter of intent") to the appropriate cohort Director (see http://www.channing.harvard.edu/nhs/?page_id=476 for specific details of the process). If a project is judged feasible (given existing resources), of substantial scientific interest, and is not currently under consideration by another investigator, the investigator will be invited to submit a detailed proposal. Letters of intent can be submitted at any time throughout the year. Within about 14 days, the applicant will either be contacted to discuss the proposal further (e.g., to better determine feasibility) or will be notified whether submission of a full grant proposal would be appropriate.

The reasons for proposing use of the Biorepository sample archive, rather than another data source, must be clearly described. Although the Biorepository is a unique resource, it is also finite. Therefore, archived samples will be used only for analyses where other, less precious, blood collections cannot provide adequate or similar information. For example, the assessment of markers of disease prognosis generally will not be considered an appropriate use of the archive. In addition, proposals to evaluate highly speculative hypotheses are not considered appropriate and will not be approved. Finally, laboratory analyses which are either already funded or have been proposed by other investigators will not be considered for approval.

Full study proposals will be reviewed by a member of the Advisory Board (as for internal investigators). The proposal's format should be similar to an NIH grant (i.e., specific aims, background and significance, preliminary studies and methods) but should be no longer than 10 pages in length. It is anticipated that decisions will be made four to eight weeks after proposal submission. The Advisory Board will decide to accept, accept pending revisions, or reject a proposal. For either of the latter two outcomes, a summary of the reasons for this decision will be provided. An "accept pending revisions" will be given if the proposal has considerable scientific merit, yet one or more issues need to be addressed before the project can proceed. Arrangements will be made to provide an expedited review of a revised proposal, which addresses the concerns raised. An appeals process is available and described in detail at http://www.channing.harvard.edu/nhs/?page_id=476.

For proposals that will require the development of funding outside the proposing organization, the approval process described above must be factored into the timing of any grant application.

Once a proposal has been approved the investigator must submit a form to the Biorepository to determine budget/costs for the project and assess the need for pilot studies.

After a project is funded (internal and external investigators): A grant must be funded before it can be entered into the Biorepository queue for projects (*funded*, means having a fundable score or a NOGA). Once this has occurred, a Get-In-Queue form must be completed and sent to the appropriate Biorepository manager, and should reflect revisions made to the aims since the time of grant submission. Approximately 3 months before a project comes up in the Biorepository queue, a senior data manager will contact the principal investigator regarding project specifications. The investigator will be asked to complete a selection template, which outlines the specific set of samples to be included in the project. The principal investigator of the grant funding the project must sign off on the specifications after conferring with data management before the selection can begin. The investigator must ensure that the needed files (e.g., disease files) are completed prior to starting the selection process. If substantial changes are made to the selection after it has begun, such that they require significant reprogramming time, the project may be deferred so as to keep the selection projects moving forward without delay. This will be reviewed on a "project-by-project" basis.

When the selection files are completed and provided to the Biorepository by the data manager, the PI will be contacted to confirm the details of the specific assays and sample types. A form must be completed further detailing the account for payment of the project and discuss any possible revisions to the assay plans (assay additions/deletions, etc.). The PI will be given a final estimate of the charges. Further, the PI must approve the costs and the Biorepository must receive confirmation from a Grants Administrator that sufficient funds exist to pay for the project. The PI will be kept updated via email as to the progress of the projects and be notified when samples are collected and/or sent to outside labs. Once the Biorepository has started collecting or accessing samples, if changes are made in the specifications (e.g., such that samples already prepared will not be sent), the investigator will be charged for any work done on the project to that point.

To the extent possible, all analyses will be conducted as a single batch with appropriate masked QC samples added to the batch. If, as is frequently the case, a large number of samples are being assayed in a study, the precision of the assay must be monitored on an ongoing basis. Results from these QC samples will be reported on a batch-by-batch basis to the investigator who will be responsible for monitoring reproducibility. Any sample remaining after the completion of the approved laboratory assays must be returned promptly to the Biorepository. Further, samples will only be sent with an ID number. No identifying or other covariate information will be sent to outside collaborators without express permission of the Advisory Board and the advent of a data use or material transfer agreement.

An invoice will be sent only after the project is complete. In instances where the project is very large, spanning many months, an interim invoice may be sent, reflecting the work done to date. Investigators will be charged the per aliquot fee at the time the work is done. We reserve the right to change the charges to cover the actual costs of running the laboratory.

External collaborators must agree to keep the cohort investigators updated on the progress of the study by providing either a written or verbal report at least every 6 months. Failure to adhere to a reasonable progress schedule (as assessed by the Advisory Board) could lead to termination of the collaborative relationship with no further data tables or analyses provided.

F. Release of Research Results:

The Biorepository triages and evaluates all results obtained from the biological specimens sent for assay. Thus the outside lab will be requested to send results and related documentation regarding the assay(s) conducted directly to Biorepository. Final results files will be prepared by the data management staff and put onto the central, password-protected computer system. The CVs of the masked QCs (or other appropriate quality control measures) will be calculated and documented online. The investigator will be directed to the paths for these files for review and analysis. A preliminary, brief review of the QC results will be done by one of the faculty members working with the Biorepository, but a more thorough review of the QC data is the investigator's responsibility. Genetic analyses are stored in an Oracle database or if they are GWAS data, on a dedicated HSPH server.

All data analyses must be conducted on the Channing Cohorts computer system, unless previously approved by the Advisory Board with a data use agreement in place. For external investigators, the most efficient way for these analyses to be accomplished is to develop a collaboration with an internal investigator and agree on the analysis plan in advance (to whatever extent possible). Once the laboratory assays are complete and results sent to the Biorespository, the external collaborating investigator can provide to a local statistician a set of data analysis requests and a series of empty tables that indicate how the results are to be presented. For any paper using Biorepository data, at least one member of the cohort's investigative team should be a coauthor on and will need to sign-off on any manuscript prior to its submission for publication.

Any dispute regarding data interpretation may be brought to the Advisory Board for consideration. Where appropriate, the Board will seek additional consultation from independent experts. Since the Advisory Board meets as a group only four times per year, some delay in coming to a resolution could occur. Therefore, it behooves all collaborating investigators to work together closely in resolving any dispute.

Before an analysis using cohort data can begin, an analysis proposal must be approved. For NHS, the proposed topic is submitted and they are periodically sent around to inform other investigators; the investigator may be asked to present the proposal at a cohort meeting. For NHSII, NHSIII, HPFS, PHS, and GUTS, all analysis proposals must be presented at a cohort meeting. Upon completion of data analyses or when feedback from the research group is needed, investigators for all cohorts must present analysis results in the form of data tables at a cohort meeting. These data tables must have previously been reviewed and approved by all co-investigators on the manuscript. When analyses and a draft of the manuscript are completed, the investigator must fill out a form that lists the computer programs that generated all of the data results presented in tables and text. These computer programs are then checked by one of the computer staff in a program review, after which a coauthor (usually the second author) checks the data in tables and text with the statistical program output for accuracy and consistency. This review of programs is not needed if the data are analyzed offsite with appropriate data use agreements. After successful completion of these reviews, the manuscript goes through a final review in which it is read by Dr. Meir Stampfer, Associate Director of the Channing Division of Network Medicine; this process allows tracking of all manuscripts and of external collaborations as mandated by the National Cancer Institute. A manuscript checklist, ensuring IRB approval and listing co-authors contributions, as well as the manuscript data ID form signed by the program reviewer must be submitted for this review. The manuscript may only be submitted to a journal after approval from Dr. Stampfer is received.

Cohort participants are informed of results in several ways. All participants receive an annual or biennial newsletter regarding new results coming from the cohort studies. Specifically, updates

are included regarding the use of biospecimens as well as information about informed consent, particularly with respect to genetics studies. Further, a public website contains information related to each cohort study. Each cohort has a phone line and email address by which participants can ask specific questions or raise particular concerns.

G. Legacy and Contingency Plans:

One major advantage of the Biorepository organization is that it pools resources from multiple cohorts and grants together to ensure long-term funding. This mitigates the possibility that loss of funding for one or two grants will lead to the closing of the Biorepository. Further, because participants are actively being followed in each of the cohorts and the informed consent allowed for future use of the specimens (without specific assays in mind), the scientific use of the samples increases with time. The Biorepository is consistently evaluating new assays and hypotheses for their appropriateness, with requests for new grant aims reaching more than 20 per year.

The Biorepository also has outlined rules for dealing with depletion of biospecimens. Specifically, justification must be proposed for the amount of specimen needed, and the Biorepository work with laboratory investigators to minimize the amount used, while maintaining high reproducibility and accuracy of the assay. The most precious specimen types (e.g., plasma) will not be sent once reaching a low volume threshold without express permission from at least two Advisory Board members that the project is of high enough scientific interest to warrant use of the “last drop”. Further, volume will be protected for cases of a disease diagnosed within 2 years of sample collection as these can be used for prospectively evaluating early detection markers. For tissue, all daughter samples and any derivatives are stored for later use, and any extra tissue that is received is saved for future research.

H. Retention of Biospecimens, Data, and Records:

All participants provide informed consent before biospecimens are collected and stored by the Biorepository; in some older collections, informed consent was implied by the receipt of blood questionnaires and samples (note that a letter explaining the collection discussed important issues related to informed consent). The consent forms discuss that a participant can withdraw his or her samples at any time; however if assays have been conducted previously on those samples, the data may have been used for statistical analysis. If a participant requests that his or her sample be removed from the Biorepository, this information is updated in our database and the samples are either destroyed, or for tumor tissue, can be sent back to the hospital at the participant’s request. To date, we have had less than 10 participants withdraw their samples.

With respect to managing participant data, all data from the cohorts are stored at the Channing Division of Network Medicine on networked computer systems that exist to facilitate the group’s mission of collaborative scientific research. The computer systems are financially supported by all research grants and contracts using cohort data, and in turn the computer systems may be used only for work related to those supporting projects. Individuals' computer accounts are valid only during the period of employment or collaboration on such projects. Accounts will be discontinued upon termination of employment or collaboration.

Computer users should be mindful that the computing resources they use are shared by others. Users shall work in such a way as not to inconvenience others or prevent others from working. This means, for example, not monopolizing shared resources such as disk space, printers or batch queues.

Each user is issued a login name and password which grants access to specific computer

resources after providing documentation of having completed Human Subjects Training (generally whatever is required by the individual's institution is acceptable). Each user must choose a reasonably secure password, and never share it with anyone else and must take reasonable precautions to prevent unauthorized persons from accessing computer accounts or data. Users must not attempt to circumvent security restrictions, and must report any observed breach or flaw in computer security to their supervisor or to a computer system administrator. Further additional security restrictions are implemented with respect to protecting the link between a participant ID and his or her HIPPA information (e.g., name, address).

The computing system is designed to facilitate collaboration among a diverse group of researchers. Access restrictions on data, documentation, and software are often kept to a minimum for this reason. Users should refrain from reading others' files except for legitimate collaborative purposes. In cases where higher security is required, certain UNIX commands may be used to restrict read or write access for sensitive files. Software, data, documents and other files on the computer systems, even if they are publicly readable, must not be copied or otherwise appropriated without proper permission. No data can ever be copied to an individual's personal computer or laptop. In the case of locally developed software, files or other documents, permission must be obtained from the author and, where applicable, the author's supervisor. In cases where ownership or authorship is uncertain, a senior system administrator should be consulted. In ongoing collaborative projects, permission to copy may be granted for an entire class of files rather than for each file individually. Employees working on the same project may be expected by their supervisor to share programs and files freely among one another. In many cases of copying within the local computer systems, verbal permission will be adequate, but written permission is preferred when there is a possibility of misunderstanding.

Written permission is required before the user may copy or transfer data, software, research results or other information to a different computer system or location. Principal investigators may institute a more specific policy regarding copying and access restrictions for each study. Users must follow any study-specific policy in addition to this general policy.

All files on the Channing Laboratory computer system, and all work, inventions, and software created on the system, remain the property of the institution administering the grant or contract under which the work was created. In most cases this institution is the BWH or HSPH. Exceptions to this must be discussed in advance with an administrative representative of the administering institution and with the PI of the cohort.

I. Sharing of Resources:

We are committed to sharing the data collected from participants in the various cohorts. The NHS and NHSII share a common External Advisory Committee; and the other cohorts have a separate External Advisory Committee. Our data-sharing plan over the past 5 years has facilitated access to the cohorts by more than 75 outside users, with many submitting R01 applications. We have an active resource-sharing program that has been developed with input from our External Advisory Committee and the NCI. Briefly, with approval of our External Advisory Committee and institutional IRB, we have adopted a data enclave approach to data sharing. For both the blood and tissue repositories and the questionnaire data, guidelines are available on our web site for outside users to access the resources of the NHS (www.channing.harvard.edu/nhs/questionnaires/index.shtml). Typically, an outside user prepares a brief proposal that is reviewed by the NHS investigator group to identify a local investigator to facilitate access to the data once approval is obtained. If a project is judged feasible (given database resources), of substantial scientific interest, is not currently being pursued by another investigator, and is not currently under consideration (typically listed as a

specific aim of a submitted or funded grant), the letter of intent for collaboration will be approved. If uncertainty arises as to the merit of a request, then we consult with the NHS External Advisory Committee for guidance. We are aware that NIH policies on data sharing are currently evolving, and we will be in regular communication through our program officers to be sure we are fully compliant, and consistent with the spirit of making the maximal scientific use of data that we collect, while at the same time protecting the privacy of information provided by our participants.

J. Conflict of Interests (COI):

The Biorepository uses the BWH and HSPH COI policies to ensure that investigators running and using the Biorepository do not have a conflict with specific projects or the overall mission of the Biorepository. All faculty are required to update their COI at each institution annually as well as sign COI for each grant that a particular faculty person participates in. More information about COI policies for the BWH can be found at http://hms.harvard.edu/public/coi/policy/integritypolicy.html and for the HSPH at http://www.hsph.harvard.edu/administrative-offices/faculty-affairs/outside-activities/outside-professional-activities.html. External investigators must complete COI at their specific institution.