

The Kaiser Permanente Research Bank (KPRB) is a core resource for research that provides data and biospecimens to investigators, enabling discoveries that contribute to health knowledge, including the prevention, diagnosis, management and treatment of disease. The KPRB collection currently comprises data and biospecimens from members recruited by the KPRB, as well as data and biospecimens from earlier KP biobanked collections, including KP Northern and Southern California’s Research Program on Genes, Environment and Health (RPGEH), and KP Northern and Southern California Men’s Health Study (CMHS).

**The specific aims of the KPRB are to:**

- ◆ Advance scientific discoveries and knowledge to improve the understanding of health and disease.
- ◆ Improve the care for Kaiser Permanente members and improve the health of society.
- ◆ Conduct research in compliance with all legal requirements and ethical best practices.
- ◆ Protect the security and privacy of individuals who contribute data and biospecimens to the KPRB resource in accordance with all applicable legal requirements.

<b>Table 1: Complete KPRB Collection<sup>1</sup></b>							
<b>(February 2021)</b>							
<b>COHORT</b>	<b>Self-Reported Health Survey<sup>3</sup></b>	<b>COVID Baseline Survey<sup>4</sup></b>	<b>Genomic Data<sup>5</sup></b>	<b>DNA<sup>6,7</sup></b>	<b>Telomere Data</b>	<b>Serum<sup>7</sup></b>	<b>Plasma<sup>7</sup></b>
<b>General<sup>2</sup></b>	273,698	124,314	113,834	323,423	109,153	165,218	23,034
<b>Pregnancy</b>	8,230	2,275	2,980	24,748	262	24,746	18,371
<b>Cancer</b>	9,798	4,645	1,466	11,015	1,409	11,000	0

<sup>1</sup> Complete collection not yet available for research, pending processing of samples, data cleaning, etc. See Table 2 for estimated availability of data and biospecimens.

<sup>2</sup> Detailed counts for pre-2015 legacy collections are available via a preparatory to research data request.

<sup>3</sup> Approximately 251K additional surveys, without accompanying samples, for pre-2015 KP Northern California legacy collection participants can be requested for research use.

<sup>4</sup> Of those participants who completed a COVID baseline survey, approximately 80K also completed one or more follow-up surveys.

<sup>5</sup> Of those participants for whom genomic data is available, approximately 103,000 were genotyped on an Axiom array as part of the GERA cohort. Imputed genotypes are also available.

<sup>6</sup> DNA source may be saliva, blood or buffy coat.

<sup>7</sup> Of those participants who submitted a sample, not all completed a survey questionnaire.

## COLLECTIONS

### General Cohort

**Recruitment Procedures:** The KPRB was formed in 2014. Enrollment of members included seven KP regions, later expanded to include KPWA when that region was added. All adult KP members are eligible to participate. Members who are registered on the KP Member Portal (kp.org) are sent an emailed invitation with a link to online enrollment. Members who consent provide a blood sample and are asked to complete the health survey.

Legacy data and biospecimens from earlier KP biobanks, including KP Northern and Southern California's Research Program on Genes, Environment and Health (RPGEH), and KP Northern and Southern California Men's Health Study (CMHS) have been added to the KPRB.

**Specimen Collection Period:** 2008-present

**Demographic Characteristics:** See Table 2

### **Pregnancy Cohort**

**Recruitment Procedures:** Participants in the pregnancy cohort were initially enrolled through voluntary contribution of an additional research blood sample at the time of their routine first and second trimester clinic visits to a KPNC facility. Women receiving prenatal care throughout KPNC were eligible to enroll. Participation rates varied by site but averaged 20% of all pregnant women.

**Survey Data Collection Period:** 2010-2017

**Specimen Collection Period:** 2010-2018

**Demographic Characteristics:** See Table 2

### **Cancer Cohort**

**Recruitment Procedures:** The cancer cohort employs the same enrollment procedures used for the general cohort; thus, the primary contact method is an email invitation with a link to online enrollment and consent. Newly diagnosed cases (identified by rapid case ascertainment (RCA) are invited to visit the Research Bank website where they complete a consent form and a brief questionnaire. The questionnaire includes the same information collected for the general cohort, with the addition of questions regarding cancer screening and family history. Cancer specific promotional materials are used to increase enrollment; thus, some cancer cases may enroll in the cohort based on conversations with their health care providers or after reading promotional materials.

**Survey Data Collection Period:** 2016 - Present

**Specimen Collection Period:** 2016 – Present

**Demographic Characteristics:** See Table 2

## **PARTICIPANT CONSENT**

Participant consent has been approved for broad use of biospecimens, survey data, and data from participants' electronic health records in studies of genetic and environmental influences on health and disease.

## **POPULATION OVERLAPS**

A limited number of cohort members submitted both a saliva and a blood sample. Some women in the pregnancy cohort also submitted a KPRB specimen prior to or following pregnancy.

## **GENOMIC AND OTHER SAMPLE ASSAY DATA**

Genomic data are available on a subset of the KPRB participants. The largest available dataset is a collection of 110K participants who were genotyped on an Axiom Array in 2010 using four custom microarrays, one for each major racial and/or ethnicity group represented in the genotyped cohort, including African Americans, East Asians (Chinese, Japanese, Korean, Southeast Asian, Filipino, and Pacific Islands), Latinos (Central and South American, Native American), and Whites (European ancestry, Middle Eastern, North African, and South Asian). Imputed genotypes and telomere data are

available on the same collection. Smaller collections of genotype and sequence data are available from separate disease-specific studies. Details of the genomic data are available upon request.

### SURVEY DATA

At enrollment, KPRB members are invited to complete a self-administered electronic or paper-based survey that covers demographic, lifestyle, and social factors. The survey focuses on factors that are not standardly captured in the EHR, such as gender identity, workplace exposures, diet, stress, social isolation, and discrimination. A complete list of KPRB survey domains is summarized in Appendix Y. To allow for comparisons between the KPRB and other study populations, the survey includes several standardized survey instruments that are widely used and validated in health research, including: PROMIS survey instruments, the Perceived Stress Index, and the Everyday Discrimination Scale.

As of February 2021, approximately 291,000 KPRB members have completed surveys, with survey completion rates among consented members ranging from 34% in the Pregnancy Cohort to 81% in the General Cohort.

Appendix Y. Kaiser Permanente Research Bank Survey Domains
Domain
Demographics
Neighborhood walkability/safety
Work exposures and shift work
General health
Social media use
Chronic pain
Diet, physical activity, & sedentary behavior
Multi-vitamin, vitamin D, calcium, & NSAIDs
Family history
Alcohol, tobacco, marijuana, & opiate drugs
Stress, isolation, support, discrimination
Reproductive history (women)
Urinary function & erectile dysfunction (men)
Health literacy

#### **COVID-19 Survey**

We conducted a series of 12 surveys in 2020 among consented participants in the KPRB to obtain information on the effects of the COVID-19 pandemic on physical and mental health, behaviors, and financial strain. These data may be used to address questions related to COVID-19. (For additional information please see the COVID-19 survey description.)

### BIOSPECIMENS

The following biospecimen types are available for research use: serum, EDTA plasma, DNA extracted from saliva, and DNA extracted from whole blood or buffy coat.

The KPRB Biorepository is ISO:2015 and College of American Pathologists (CAP) certified (CAP ID #9511943). The Biorepository adheres to industry best practices for processing, long term storage, retrieval and distribution of specimens.

### **CLINICAL (EHR) DATA**

KP has extensive clinical data generated by the integrated delivery system. The data is standardized and harmonized across regions using the KP/CESR Virtual Data Warehouse (VDW), which is the primary source of KPRB clinical data. It includes data on demographics, enrollment, encounters, diagnoses, procedures, pharmacy, vital signs, social history, laboratory results, mortality, medication orders, patient-reported outcomes, benefits, race probabilities, fracture risk scores, radiation therapy, infusion, and pregnancy data. Data are available in many content areas from the late 1990's forward and in selected cases, well before then.

The KP/CESR-VDW data source provides the advantage of high-quality standardized data achieved through a data coordinating center who work to ensure national and international coding schemes are used when possible, provide common data definitions, and perform routine data quality assessment work to monitor data quality.

### **ENVIRONMENTAL DATA**

Significant environmental data are or can easily be linked to the KPRB participants, largely by leveraging available public use datasets and geographic information system (GIS) methods.

First, selected factors, such as smoking histories, alcohol, diet and education are captured by questionnaire or in HealthConnect, the KP electronic health record (EHR). In addition, "clinical environmental" factors such as medical radiation exposure and pharmaceuticals are captured in the EHR. Second, the residential address(es) of each KPRB member can be (or has been) geocoded and the coordinates provided to link to geospatial data, typically going back a considerable amount of time in the past.

We can link to the 2000 Census and the 2010 Census via KP clinical records to derive variables of interest, but we cannot offer specific location information in a de-identified dataset, including zip code.

**Table 2: Demographic Characteristics of the KP Research Bank  
 (February 2021)**

	Participants with available surveys		Participants with available biospecimens		Participants with available genomic data	
	Number	Percent	Number	Percent	Number	Percent
<b>Gender</b>						
Females	184,557	63.82%	217,511	61.32%	58,144	56.18%
Males	104,606	36.17%	137,194	38.68%	45,350	43.82%
Other	1	<1%	1	<1%	0	0.00%
Unknown	11	<1%	14	<1%	5	<1%
<b>Age at Survey</b>						
18-29	13,740	4.75%	22,393	6.31%	1,478	1.43%
30-39	25,195	8.71%	35,695	10.06%	3,763	3.64%
40-49	29,858	10.33%	33,882	9.55%	8,710	8.42%
50-59	51,922	17.96%	61,326	17.29%	18,687	18.06%
60-69	82,242	28.44%	99,893	28.16%	30,550	29.52%
70-79	58,104	20.09%	71,799	20.24%	25,823	24.95%
80+	28,111	9.72%	29,728	8.38%	14,488	14.00%
<b>Race/Ethnicity</b>						
African American/Black	11,134	3.85%	17,375	4.90%	3,305	3.19%
American Indian/Alaska Native	1,046	0.36%	1,296	0.37%	223	0.22%
Asian & Pacific Islander	22,387	7.74%	30,993	8.74%	6,630	6.41%
Hispanic	24,007	8.30%	35,213	9.93%	7,019	6.78%
Non-Hispanic White	216,912	75.01%	253,870	71.57%	80,404	77.69%
Other/Uncertain	9,480	3.28%	11,025	3.11%	4,967	4.80%
Unknown	4,209	1.46%	4,948	1.39%	951	0.92%