The NHGRI Sample Repository for Human Genetic Research

CORIELL INSTITUTE
FOR MEDICAL RESEARCH

X. Jian, B. Kerr, E. Horton, A. Liautaud, M. Mitchell, L. Scheinfeldt.
Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103

Background

The NHGRI Sample Repository for Human Genetic Research (NHGRI Repository)), established by NHGRI in 2006 at the Coriell Institute for Medical Research, facilitates studies of human genetic and genomic variation by characterizing and distributing biospecimens from thousands of participants living around the world (n>3,700), including the 1000 Genomes Project samples, to the research community. NHGRI Repository DNA and lymphoblastoid cell lines (LCLs) are publicly available, renewable, and are characterized with publicly available whole genome sequencing data and other large-scale genomic data. NHGRI Repository cell lines are a unique resource that can be used for in vitro modeling, and as a reliable and renewable source of DNA and RNA; NHGRI Repository cell lines and DNAs provide a gold standard reference set for testing new assays and validating new platforms, pipelines and workflows; and as sequencing technologies change over time, NHGRI Repository samples can be used to confirm reproducible results, and enable comparison between different technologies and data analysis methods.

Hundreds of thousands of NHGRI Repository DNA and cell culture samples (>200,000) have been distributed to researchers (Figure 1) in 56 countries around the world supporting a wide range of research studies. The most commonly cited research topics include use as reference materials in genetic/genomic assay development and clinical genetic testing, mutation discovery and characterization, functional studies including a wide range of functional genomic studies, and evolutionary studies. A more detailed summary of the types of research using NHGRI Repository samples is shown in Figure 2. Finally, thousands of resulting scientific publications have used HapMap and 1000 Genomes Project data and samples. The NHGRI Repository website can be accessed at https://catalog.coriell.org/NHGRI.

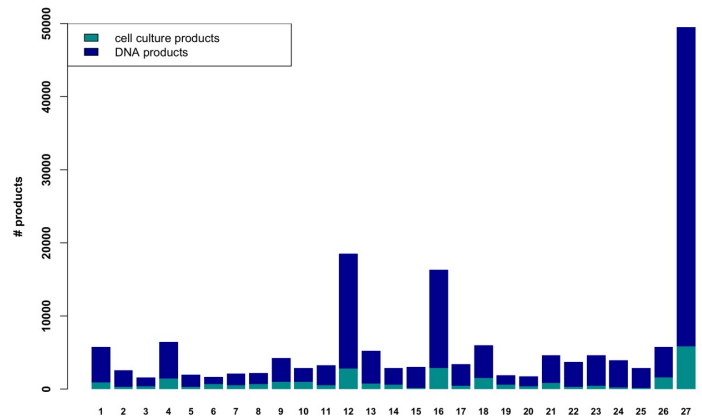## Distribution by Population Sample

Figure 1. NHGRI Repository DNA and cell product distribution by population sample. Population IDs correspond to Table 1.

Population Sample Information

Table 1: NHGRI Repository Population Samples

Population ID	Population Descriptor	Population Sample Size
1	African Ancestry in SW, USA	62
2	African Caribbean in Barbados	120
3	Bengali in Bangladesh	144
4	British from England and Scotland	100
5	Chinese Dai in Xishuangbanna, China	102
6	Chinese in Metropolitan Denver, CO, USA	129
7	Colombian in Medellin, Colombia	136
8	Esan in Nigeria	173
9	Finnish in Finland	103
10	Gambian in Western Division, Mandinka	179
11	Gujarati Indians in Houston, TX, USA	109
12	Han Chinese in Beijing, China	120
13	Han Chinese South	163
14	Iberian Populations in Spain	157
15	Indian Telugu in the UK	118
16	Japanese in Tokyo, Japan	120
17	Kinh in Ho Chi Minh City, Vietnam	124
18	Luhya in Webuye, Kenya	120
19	Maasai in Kinyawa, Kenya	205
20	Mende in Sierra Leone	128
21	Mexican Ancestry in Los Angeles, CA, USA	71
22	Peruvian in Lima, Peru	122
23	Puerto Rican in Puerto Rico	139
24	Punjabi in Lahore, Pakistan	158
25	Sri Lankan Tamil in the UK	128
26	Toscani in Italia	114
27	Yoruba in Ibadan, Nigeria	120

Distribution by Research Category
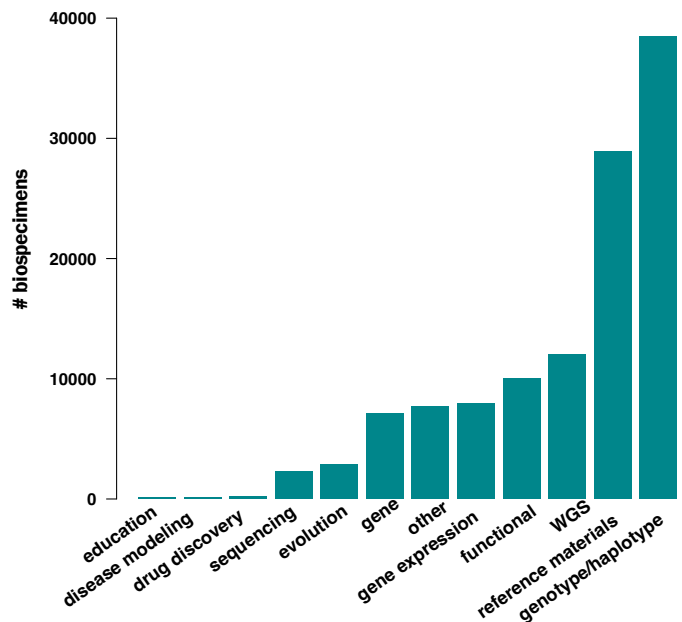
Figure 2. NHGRI Repository biospecimen distribution by research category.