The NHGRI Sample Repository for Human Genetic Research: biospecimens and a new genomic data search tool

A. Liautaud, D. Kusic, G. Calendo, J. Madzo, B. Kerr, S. Sampson, L. Willis, E. Horton, M. Mitchell, N. Gharani, L. Scheinfeldt.

Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103

CORIELL INSTITUTE FOR MEDICAL RESEARCH

Background

The NHGRI Sample Repository for Human Genetic Research (NHGRI Repository) housed at the Coriell Institute for Medical Research facilitates studies of human genetic and genomic variation by establishing, characterizing and distributing a large (n>3,700) and diverse publicly available collection of renewable biospecimens obtained from people living around the world, including biospecimens associated with the 1000 Genomes Project collection. Participants who generously donated to the NHGRI Repository consented to their biospecimens and associated data being used for a wide range of general research and to broad data sharing of largescale genomic data collected from their biospecimens. Through the 1000 Genomes Project, the majority of this collection has been characterized with publicly available whole genome sequencing data and other large-scale genomic data.

We recently launched a user-friendly and integrated search tool (https://catalog.coriell.org/1/Browse/Genomic-Data-Search) that allows catalog users to dynamically query the 1000 Genomes Project Phase 3 whole genome sequencing data[1] by individual SNP (rsid), or by gene (HUGO symbol) and to search the GEUVADIS gene expression data[2] by gene to more easily identify biospecimens with variation profiles of interest for a range of research applications.

[1] The 1000 Genomes Project Consortium, Nature 526: 68–74 (2015).
[2] Lappalainen et al., Nature 501: 506–511 (2013).

SNP Search

rs4244285

Organism:	*Homo sapiens*
Chromosome:	10
Position:	96541616
Alleles:	G/A

Click here to view dbSNP entry
Click here to enter a new search

Figure 1. Search by rsid to return an interactive table of each subject, subject affiliation, genotype, and sex that can be filtered or exported to an Excel file.

Gene Search

CYP2C19

Chromosome:	10

Click here to view NCBI entry
Click here to return to the search page

Figure 2. Search by gene to return an interactive table of each variant, chromosome and position, alleles, and functional annotation that can be filtered or exported to an Excel file

Table 1: 1000 Genomes Project Population Samples

Population	Individual Samples	Population	Individual Samples
African Ancestry in SW, USA	62	Japanese in Tokyo, Japan	120
African Caribbean in Barbados	120	Kinh in Ho Chi Minh City, Vietnam	124
Bengali in Bangladesh	144	Luhya in Webuye, Kenya	120
British from England and Scotland	100	Maasai in Kinyawa, Kenya	205
Chinese Dai in Xishuangbanna, China	102	Mende in Sierra Leone	128
Colombian in Medellín, Colombia	136	Mexican Ancestry in Los Angeles, CA, USA	71
Esan in Nigeria	173	Peruvian in Lima, Peru	122
Finnish in Finland	103	Puerto Rican in Puerto Rico	139
Gambian in Western Division, Mandinka	179	Punjabi in Lahore, Pakistan	158
Gujarati Indians in Houston, TX, USA	109	Sri Lankan Tamil in the UK	128
Han Chinese in Beijing, China	120	Toscani in Italia	114
Han Chinese South	163	Yoruba in Ibadan, Nigeria	120
Iberian Populations in Spain	157	Chinese in Metropolitan Denver, CO, USA	129
Indian Telugu in the UK	118	CEPH Collection *	186

* Samples available from the NIGMS Repository

Gene Expression Search
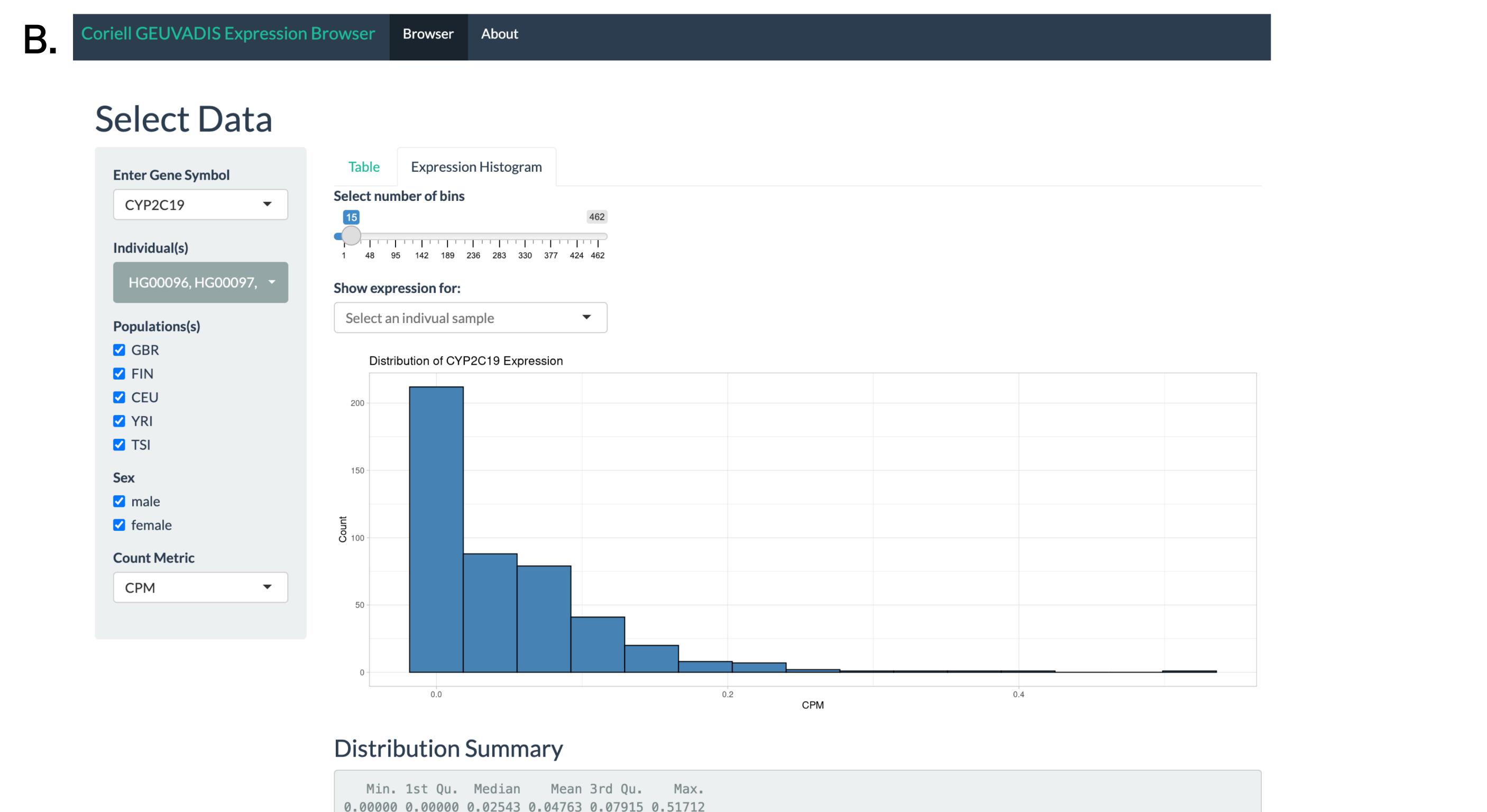
A.

B.

Figure 3. Search by gene to dynamically display A) a table containing gene expression information for the selected samples that can be exported to and Excel file or B) an interactive histogram of the gene expression data across the selected samples