

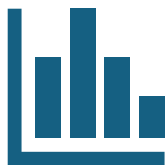
A photograph of a server room with blue ambient lighting. In the foreground, several server racks are visible, some with blue indicator lights. The background is filled with out-of-focus server racks and bright, circular bokeh light effects in shades of yellow and green. The text "Where to get data?" is centered in the image.

Where to get data?

Options



Collect your own



Biobank data



Other raw data
sources



Places to get
sumstats

Collect your own data

1. Recruit a large sample

- From clinics
- From the public


www.aaastudy.org.au



Collect your own data

2. Collect information from the participants

Survey Completion 0% 100%

 QIMR Berghofer
Medical Research Institute

Australian ASD and ADHD Study
Investigating the impact of attention and
behaviour on individuals and families

Thinking back to when you were in primary school, do you remember having problems with paying attention or controlling your behaviour?

☐ No


☐ Yes I had some problems, but I wasn't diagnosed with didn't have ADHD or ADD

☐ Yes I had some problems, and was diagnosed with ADHD or ADD

☐ I don't know

12:29

Survey Completion 0% 100%

 QIMR Berghofer
Medical Research Institute

Australian ASD and ADHD Study
Investigating the impact of attention and
behaviour on individuals and families

Thinking back to when you were in primary school, do you remember having problems with paying attention or controlling your behaviour?

☐ No

☐ Yes I had some problems, but I wasn't diagnosed with didn't have ADHD or ADD

☐ Yes I had some problems, and was diagnosed with ADHD or ADD

☐ I don't know

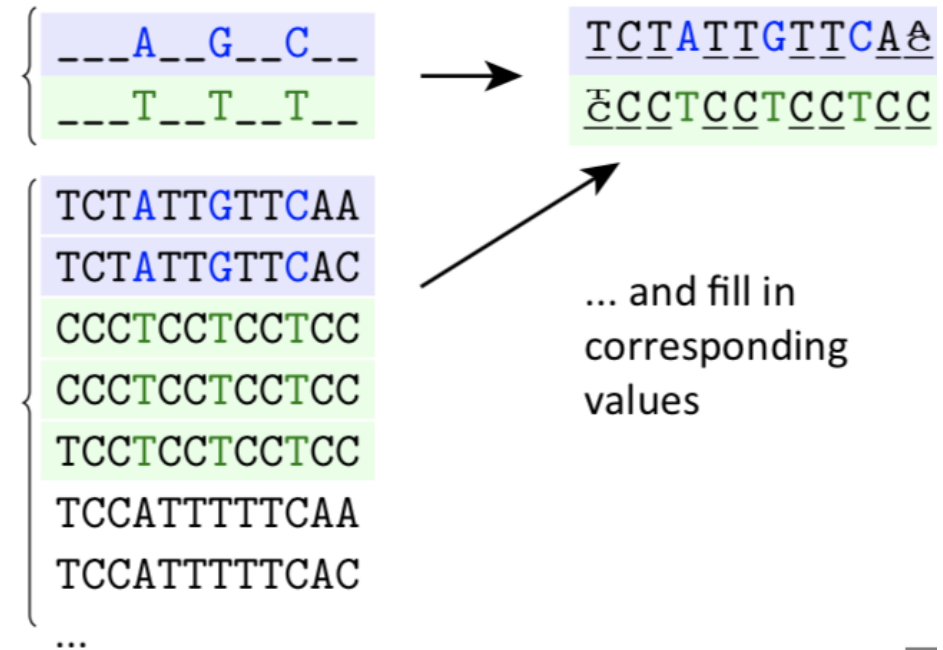
Collect your own data

3. Collect a DNA sample



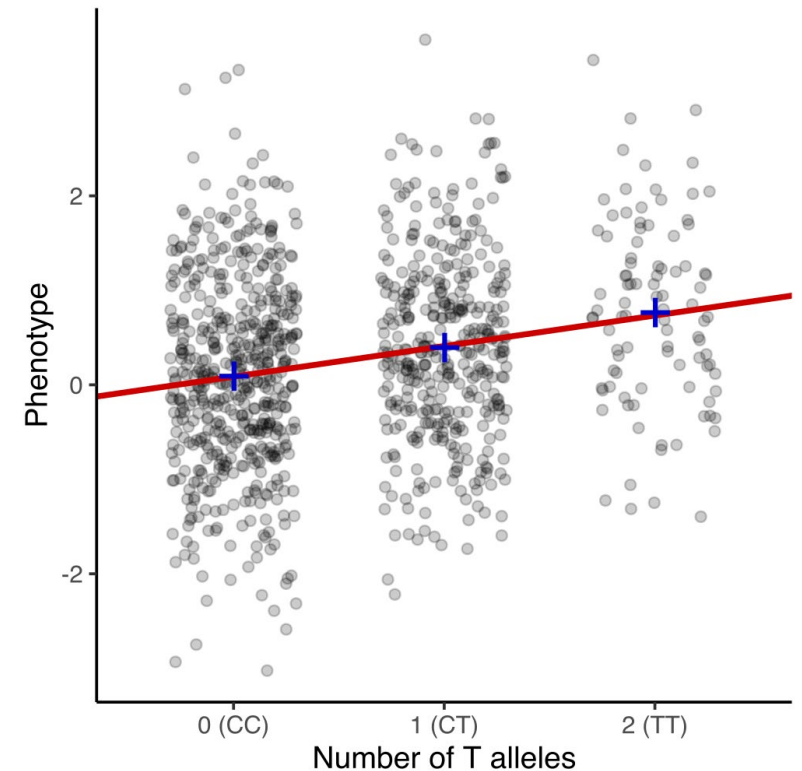
Collect your own data

4. Genotype the samples



Collect your own data

5. Look for genetic variants that can predict trait (regression)



‘Open’ Biobanks

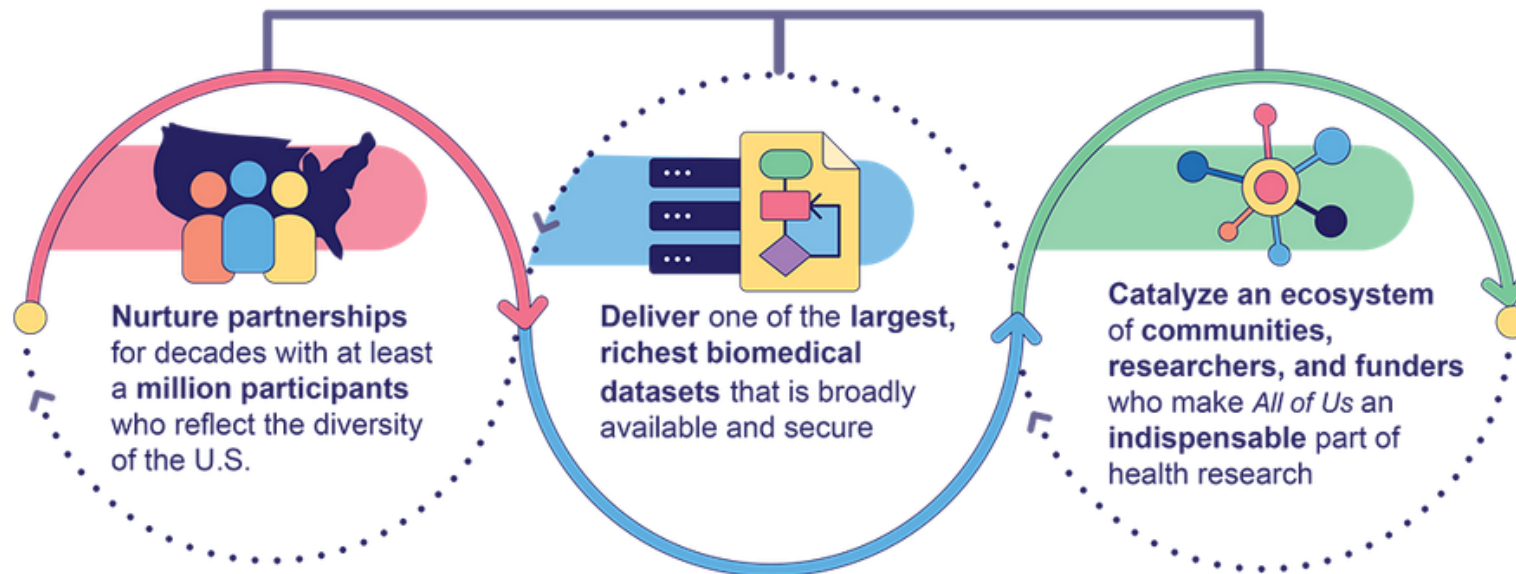
- Until recently you could apply to download the phenotypes and genotypes
- From now on you need to use the Research Analysis Platform (UKB-RAP)



'Open' Biobanks



- Access via RAP
 - Obtaining an account can be difficult if you are not a US person
 - No access/approval costs
 - Running costs \$\$\$



Working on a RAP

- RAP costs \$\$\$\$
- Remember to stop processes that are not needed
- Deleting data not needed
- Timing can be difficult to work out
- Share code and knowledge to help others
- Downloading data/results files may require approval
 - ALLofUS – approval to download sumstats took about 2 days

Smaller biobanks

- Health and Retirement Study – <https://hrs.isr.umich.edu/about>
- Nurses health study - <https://nurseshealthstudy.org/>
- ROSMAP - <https://dss.niagads.org/cohorts/religious-orders-study-memory-and-aging-project-rosmap/>
- ALSPAC – <https://www.bristol.ac.uk/alspac/>

‘Closed’ Biobanks

- Access via collaboration - often via a data-buddy system
 - 23andMe
 - Million Veterans Study
 - BioVU
 - MOBA
 - Kadoorie Biobank
 - etc...



Biobank Resource Centre

Developed by the University of British Columbia Office of Biobank Education and Research and the Canadian Tissue Repository Network

[Home](#) [About Us](#) [Education](#) [Registration](#) [Certification](#) [Products and Services](#) [Sign Up](#) [Contact Us](#)

[Home](#) / [Products and Services](#) / [Biobank Locator](#)

Biospecimen Cohort Locator

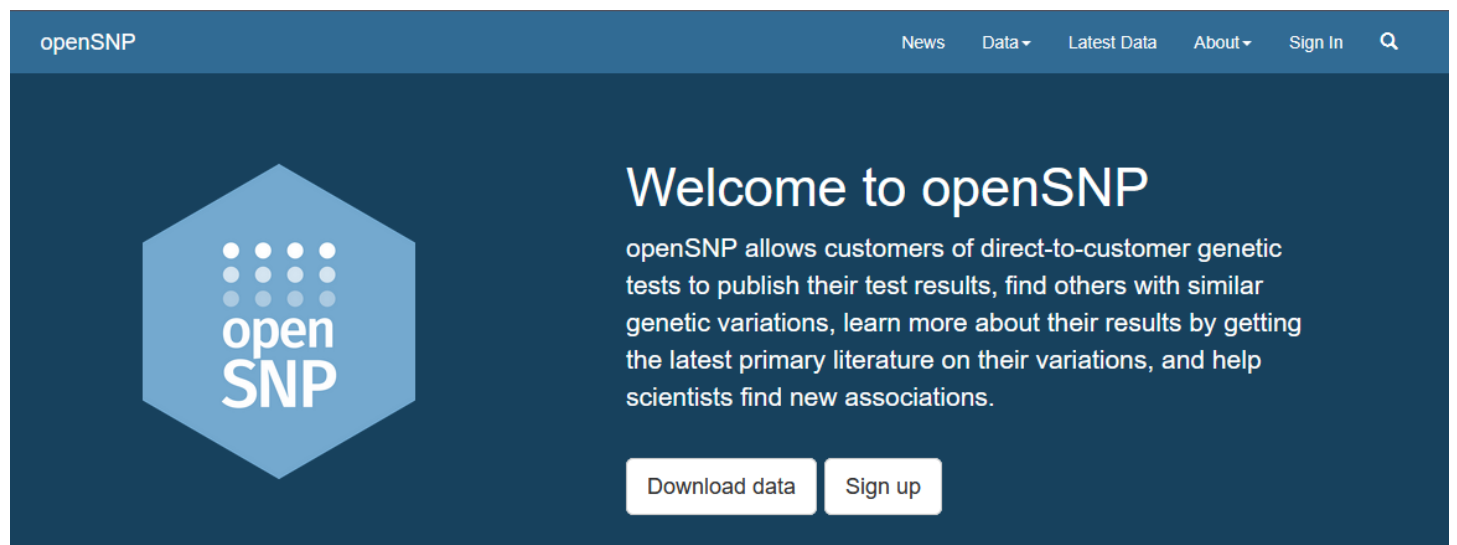
[Click here for the Biospecimen Cohort Locator](#)

Biobank Locator

There are currently 396 registered biobanks that have given permission to be listed. Please note that the list includes both Canadian and International biobanks.

If you can not find the biobank or types of biospecimens you are looking for, [click here](#) for a list of other online resources.

Other sources of raw (real) data



About Us

openSNP is a community-driven and community-owned project which lives through the contributions of its members. It is not associated with any institution and the only funding source for openSNP is [continuous crowdfunding, which pays for the hosting costs](#).

Who's running the openSNP.org infrastructure?

The bulk of the coding was done by [Bastian Greshake Tzovaras](#), [Helge Rausch](#) and [Philipp Bayer](#), who are also taking care of running and maintaining the live website.

Bastian is working on his PhD in bioinformatics at the University of Frankfurt, while Philipp already successfully took that step and is now a bioinformatics post-doc at the University of Western Australia in Perth. Helge is working and living as a webdesigner in Berlin and probably the biggest reason why this website still works.

Getting in Touch

There are several ways to get in touch with us - the code is hosted at [GitHub](#) where we of course accept pull requests, there's a [Gitter channel](#) if you want to chat to us, and you can reach out to us at the above Twitter handles or [@openSNPorg](#).

In addition to that openSNP thrives through the kindness of dedicated contributors who have made numerous contributions. [Each and every one of them is celebrated in our humans.txt](#).

Philipp Bayer



Helge Rausch



Bastian Greshake Tzovaras



Open SNP

For genotyping users



Upload Your Genotyping

Upload your raw genotyping or exome data from [23andMe](#), [deCODEme](#) or [FamilyTreeDNA](#) to the openSNP database to make it available for everybody.



Share Your Phenotypes

Share your phenotype with other openSNP users and find others with similar characteristics and traits. Your data may help scientists discover new genetic associations!



Share your stories

With openSNP you can share stories about your genetic variations and phenotypes, and discover the stories of other users.



Find literature on genetic variation

openSNP gets the latest open access journal articles on genetic variations from the [Public Library of Science](#). Popular articles are indexed via the social reference manager [Mendeley](#) and summaries are provided by [SNPedia](#).

For Scientists



Search for phenotypes

Many diseases and traits are suspected to have genetical components. Genome Wide Association Studies are a simple tool find genetic markers. Easily find people with the variation you are interested in via openSNP. The variation you are looking for is not entered yet? Just add it to openSNP



Easily download datasets

The mass download-function of openSNP allows you to easily download the full genotyping raw-data in the file formats that are provided by 23andMe, deCODEme and FamilyTreeDNA. As the files can be grouped by their variations for specific phenotypes it is easy to get datasets that are already usable for association studies.



Get notified about new data

openSNP delivers a RSS feed for each phenotype. So you can easily get all new datasets that get available for the phenotypes of your interest, without the need to check for new entries by hand. For all data junkies that need more data: There is also a feed that carries all new datasets.

Places to get sumstats

- All of US phenotypes (run by the Broad) -
you need a RAP account to be able to access
- UK Biobank phenotypes (run by the Broad) -
<https://pan.ukbb.broadinstitute.org/downloads>
- FinGen - https://www.finngen.fi/en/access_results
<https://mvp-ukbb.finngen.fi/>
- Million Veterans - <https://www.mvp.va.gov/pwa/discover-mvp-data>

Places to get sumstats

- PGC - <https://pgc.unc.edu/for-researchers/download-results/>
- GIANT - https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
- ENIGMA - <https://enigma.ini.usc.edu/research/download-enigma-gwas-results/>
- GWAS Catalogue - <https://www.ebi.ac.uk/gwas/> (patchy)
- SSGAC - <https://thessgac.com/> (requires account)
- Authors of papers



Questions, comments, concerns...