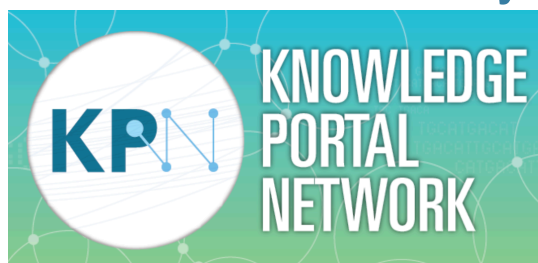# Contributing genetic association summary statistics to the

KP4CD.org

The Knowledge Portal Network is an ecosystem of interconnected open-access web portals, powered by the Human Genetics Amplifier (HuGeAMP) platform, that integrate, interpret, and present human genetic and genomic data with the goal of sparking insights into complex disease. We welcome submissions of genetic association summary statistics for traits that are relevant to the Knowledge Portals. We have analyzed the privacy risks inherent in sharing summary statistics in the Knowledge Portals and found that they are extremely low (read our white paper).

Upon receipt of your summary statistics we will integrate them into the Knowledge Portal database, making them available for browsing and querying via the interfaces, tools, and APIs of the relevant Portal(s).

At your request, we are also able to provide the summary statistic files for public download from the Portal. If you would like us to provide these files, please let us know when submitting your dataset.

We can also accept raw, individual-level data for analysis or extended processing. For more information on submitting such data, see these detailed instructions written for the Type 2 Diabetes Knowledge Portal.

We are also interested in receiving other 'omics datasets: epigenomic modifications, transcript levels and tissue-specific expression, chromatin conformation, proteomics, and more. Please contact us about these types of data.

Summary statistic file formats

Our minimum file format includes:

- variant rsID or chromosome and position (hg19; if your results are not in the hg19 genome version we can perform LiftOver in either direction)
- reference allele
- effect allele
- effect size or odds ratio
- p-value

Our preferred format includes the above values plus:

- sample size for each variant, or effective sample size for binary traits
- effect allele frequency, or if not available, minor allele frequency
- standard error for effect size

Please submit files in .tsv format, compressed if necessary.

Accompanying information

So that we can document the dataset appropriately, we would also like to have:

- a brief description of the dataset
- the total sample size, and sample size for each phenotype
- definitions of the phenotypes assayed
- ancestry of the participants
- reference to a publication describing the study, if available
- image file(s) for logo(s) of consortia involved, if you would like us to display them in the dataset documentation
- if you would like us to provide the summary statistic files for download, please also supply a README file to accompany them

Credible sets

We are interested in receiving investigator-generated credible sets for as many Knowledge Portal traits as possible. Submitted credible set files should include:

- Variant ID
- Credible set ID, or, if IDs have not been assigned, a description of the researchers' strategy for assigning variants to credible sets (*e.g.*, considering variants within 500kb of a lead variant)
- Posterior probability (preferred), or log(Bayes factor) and a suggested method for converting log(BF) to posterior probability

The process

- When you are ready to submit data, please contact us. Data files may be transferred via email, Dropbox, any other means of file transfer that you use, or via an Aspera site that we can set up.
- Let us know whether you would like us to to provide the files for download as well as integrating the results in the Portal.
- We will review the files and get back to you with any questions

More questions? Contact the team for help.