

## Resources

<https://skingeneticsconsortium.org/resources.html>

On our website we have provided the following files:

- python script ('upload\_gwas.py') needed to upload GWAS
- An example JSON file
- Example batch array script to automate upload of a large number of GWAS

Please download the python script before beginning.

## Instructions to upload GWAS:

1. Go to: <https://kpndataregistry.org/sgc>
2. Find your cohort and press edit cohort button
3. Click on GWAS Metadata tab and complete the form

Upload Files/Edit Cohort

Phase 0 Upload GWAS Metadata

**GWAS Cohort Metadata**  
Provide cohort-level information required for GWAS submission. This metadata applies to all datasets submitted for this cohort.

**General Information**

Country/Countries of Ascertainment \* Data Freeze Date \*

Contact Person and Email \* Lead Collaborator \*

**Contributors & Publication**

Other Contributors to Include as Authors Other Publication Requirements

**Acknowledgements and Funding Information \***

Summary Stats Publicly Available? \* Relevant References (PMIDs or URLs) \*

4. As outlined in our analysis plan you will need to upload JSON files alongside GWAS summary statistic data. All fields below are required. You may include additional fields. Please ensure you have generated these JSON files (1 for every GWAS file) before attempting upload.

```
{"cohort_name": "My Cohort", # use exact cohort name as registered in the portal.  
"phenotype": "ATOPIC_DERM", # phenotype code  
"ancestry": "EUR", # ancestry code (EUR, AFR, AMR, EAS, SAS, Combined)  
"sex": "All", # or "Male" or "Female"  
"codes_used": "L20, L21", # diagnosis codes for case definition  
"cases": 1000, # number of cases (integer)  
"controls": 10000, # number of controls (integer)
```

```

"male_proportion_cases": 0.5,      # proportion of cases that are male (0-1)
"male_proportion_controls": 0.5,   # proportion of controls that are male (0-1)
"assoc_test_software_and_version": "regenie v4.1", # GWAS software
"assoc_test_model_and_settings": "mixed model logistic" # GWAS model
}

```

Note: The portal currently only accepts the ancestry codes listed above. We are happy to add others - contact us if you need us to include a different ancestry that reflects your data, so we can decide on a code and ensure harmonisation across the consortium.

5. Create a file called `.credentials` in the same directory as the python script. This file should be 2 lines with the first containing your username (i.e email) and the second line containing your password that you use to access our portal. Note: the credentials of the uploader of GWAS summary stats **must** match the credentials of the uploader of the respective Phase 0 upload. Please use the same email as the one used in the Phase 0 upload, and if necessary, indicate who the uploader is in the GWAS meta-data form.
6. For single GWAS uploads: Run the provided python script using the below code, replacing `file.txt` and `metadata.json` with paths to your files and `my_dataset` with a unique identifier for the results being uploaded (used for tracking and to flag any files that have issues)

```

./upload_gwas.py \
--dataset my_dataset \ #unique identifier
--metadata metadata.json \ #Path to JSON file
--validate \ # This can be removed, but you must then validate via the website
file.txt #Path to GWAS file

```

Alternatively: If you are doing bulk uploads, we provide an additional example batch array uploader script on our website that is designed to schedule all uploads (but only 10 at any one time). This script may need to be tweaked according to your HPC scheduler & needs to be amended to give a sensible name to the `--dataset` flag on line 25 (i.e something representing your cohort). The script can then be run using the below code:

```

#1: collect paths for all GWAS files
ls ../all_gwas/*gz > upload_list.txt

#2: run initial bulk upload (script refers to upload_list.txt as default)
sbatch upload_array.sh

```

7. Once running, you can check upload progress on our website here: <https://kpndataregistry.org/sgc>
8. If you removed the `--validate` flag, please ensure you validate your files online afterwards, to confirm all were processed correctly.
9. Our recommendation is you **validate at least one file successfully before uploading all GWAS**, to save deleting/re-uploading everything if the portal has an issue with your format.

Thank you for uploading your data!

If you encounter any issues, please email us at [info@skingeneticsconsortium.org](mailto:info@skingeneticsconsortium.org) for assistance.