

## edge\_gwas\_test\_run\_extend\_2

December 27, 2025

```
[1]: # Cell 1: Clear cache and reinstall from main branch
!pip cache purge
# Uninstall old package
!pip uninstall edge-gwas -y

# Remove any existing clone
!rm -rf ~/edge-gwas

# Clone fresh from GitHub
!cd ~ && git clone https://github.com/nicenzhou/edge-gwas.git

# Install in editable/development mode
!pip install -e ~/edge-gwas

# Verify installation
!python -c "from edge_gwas.utils import calculate_pca_plink; print(''')
˓→Installation successful!")"
```

WARNING: No matching packages

Files removed: 0

WARNING: Skipping edge-gwas as it is not installed.

Cloning into 'edge-gwas'...

remote: Enumerating objects: 1218, done.

remote: Counting objects: 100% (436/436), done.

remote: Compressing objects: 100% (242/242), done.

remote: Total 1218 (delta 357), reused 193 (delta 193), pack-reused 782 (from 1)

Receiving objects: 100% (1218/1218), 2.61 MiB | 2.36 MiB/s, done.

Resolving deltas: 100% (774/774), done.

Obtaining file:///Users/nicen/edge-gwas

Preparing metadata (setup.py) ... done

Requirement already satisfied: numpy<2.0.0,>=1.19.0 in

/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)  
(1.26.4)

Requirement already satisfied: pandas<3.0.0,>=1.2.0 in

/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)  
(2.3.3)

```
Requirement already satisfied: scipy<2.0.0,>=1.6.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(1.11.4)
Requirement already satisfied: statsmodels<1.0.0,>=0.12.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(0.14.0)
Requirement already satisfied: scikit-learn<2.0.0,>=0.24.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(1.2.2)
Requirement already satisfied: matplotlib<4.0.0,>=3.3.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(3.8.0)
Requirement already satisfied: pandas-plink<3.0.0,>=2.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(2.3.2)
Requirement already satisfied: cyvcf2>=0.30.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(0.31.4)
Requirement already satisfied: sphinx<7.0.0,>=4.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(5.0.2)
Requirement already satisfied: sphinx-rtd-theme<2.0.0,>=1.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(1.3.0)
Requirement already satisfied: sphinxcontrib-napoleon>=0.7 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(0.7)
Requirement already satisfied: numpydoc>=1.1.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(1.5.0)
Requirement already satisfied: pytest<8.0.0,>=6.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(7.4.0)
Requirement already satisfied: pytest-cov<5.0.0,>=2.10.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(4.1.0)
Requirement already satisfied: black<24.0,>=21.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(23.11.0)
Requirement already satisfied: flake8<7.0.0,>=3.8.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(6.0.0)
Requirement already satisfied: mypy<2.0.0,>=0.900 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(1.8.0)
Requirement already satisfied: isort<6.0.0,>=5.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from edge-gwas==0.1.1)
(5.9.3)
```

```
Requirement already satisfied: click>=8.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
black<24.0,>=21.0->edge-gwas==0.1.1) (8.1.7)
Requirement already satisfied: mypy-extensions>=0.4.3 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
black<24.0,>=21.0->edge-gwas==0.1.1) (1.0.0)
Requirement already satisfied: packaging>=22.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
black<24.0,>=21.0->edge-gwas==0.1.1) (23.1)
Requirement already satisfied: pathspec>=0.9.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
black<24.0,>=21.0->edge-gwas==0.1.1) (0.10.3)
Requirement already satisfied: platformdirs>=2 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
black<24.0,>=21.0->edge-gwas==0.1.1) (3.10.0)
Requirement already satisfied: coloredlogs in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from cyvcf2>=0.30.0->edge-
gwas==0.1.1) (15.0.1)
Requirement already satisfied: mccabe<0.8.0,>=0.7.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
flake8<7.0.0,>=3.8.0->edge-gwas==0.1.1) (0.7.0)
Requirement already satisfied: pycodestyle<2.11.0,>=2.10.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
flake8<7.0.0,>=3.8.0->edge-gwas==0.1.1) (2.10.0)
Requirement already satisfied: pyflakes<3.1.0,>=3.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
flake8<7.0.0,>=3.8.0->edge-gwas==0.1.1) (3.0.1)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
matplotlib<4.0.0,>=3.3.0->edge-gwas==0.1.1) (1.2.0)
Requirement already satisfied: cycler>=0.10 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
matplotlib<4.0.0,>=3.3.0->edge-gwas==0.1.1) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
matplotlib<4.0.0,>=3.3.0->edge-gwas==0.1.1) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
matplotlib<4.0.0,>=3.3.0->edge-gwas==0.1.1) (1.4.4)
Requirement already satisfied: pillow>=6.2.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
matplotlib<4.0.0,>=3.3.0->edge-gwas==0.1.1) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
matplotlib<4.0.0,>=3.3.0->edge-gwas==0.1.1) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
matplotlib<4.0.0,>=3.3.0->edge-gwas==0.1.1) (2.8.2)
```

```
Requirement already satisfied: typing-extensions>=4.1.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
mypy<2.0.0,>=0.900->edge-gwas==0.1.1) (4.15.0)
Requirement already satisfied: Jinja2>=2.10 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from numpydoc>=1.1.0->edge-
gwas==0.1.1) (3.1.3)
Requirement already satisfied: pytz>=2020.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandas<3.0.0,>=1.2.0->edge-gwas==0.1.1) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.7 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandas<3.0.0,>=1.2.0->edge-gwas==0.1.1) (2023.3)
Requirement already satisfied: cffi>=1.17 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2.0.0)
Requirement already satisfied: dask[dataframe] in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2023.11.0)
Requirement already satisfied: deprecated<2.0,>=1.2 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (1.3.1)
Requirement already satisfied: pandera[dask,mypy] in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (0.27.1)
Requirement already satisfied: tqdm<5.0,>=4.67 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (4.67.1)
Requirement already satisfied: xarray in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2023.6.0)
Requirement already satisfied: zstandard<0.24,>=0.23 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (0.23.0)
Requirement already satisfied: configparser in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pytest<8.0.0,>=6.0.0->edge-gwas==0.1.1) (1.1.1)
Requirement already satisfied: pluggy<2.0,>=0.12 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pytest<8.0.0,>=6.0.0->edge-gwas==0.1.1) (1.0.0)
Requirement already satisfied: coverage>=5.2.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
coverage[toml]>=5.2.1->pytest-cov<5.0.0,>=2.10.0->edge-gwas==0.1.1) (7.13.0)
Requirement already satisfied: joblib>=1.1.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from scikit-
learn<2.0.0,>=0.24.0->edge-gwas==0.1.1) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from scikit-
learn<2.0.0,>=0.24.0->edge-gwas==0.1.1) (2.2.0)
```

```
Requirement already satisfied: sphinxcontrib-applehelp in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (1.0.2)
Requirement already satisfied: sphinxcontrib-devhelp in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (1.0.2)
Requirement already satisfied: sphinxcontrib-jsmath in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (1.0.1)
Requirement already satisfied: sphinxcontrib-htmlhelp>=2.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2.0.0)
Requirement already satisfied: sphinxcontrib-serializinghtml>=1.1.5 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (1.1.5)
Requirement already satisfied: sphinxcontrib-qthelp in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (1.0.3)
Requirement already satisfied: Pygments>=2.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2.15.1)
Requirement already satisfied: docutils<0.19,>=0.14 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (0.18.1)
Requirement already satisfied: snowballstemmer>=1.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2.2.0)
Requirement already satisfied: babel>=1.3 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2.11.0)
Requirement already satisfied: alabaster<0.8,>=0.7 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (0.7.12)
Requirement already satisfied: imagesize in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (1.4.1)
Requirement already satisfied: requests>=2.5.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2.31.0)
Requirement already satisfied: sphinxcontrib-jquery<5,>=4 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from sphinx-rtd-
theme<2.0.0,>=1.0.0->edge-gwas==0.1.1) (4.1)
Requirement already satisfied: six>=1.5.2 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from sphinxcontrib-
napoleon>=0.7->edge-gwas==0.1.1) (1.16.0)
Requirement already satisfied: pockets>=0.3 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from sphinxcontrib-
napoleon>=0.7->edge-gwas==0.1.1) (0.9.1)
```

```
Requirement already satisfied: patsy>=0.5.2 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
statsmodels<1.0.0,>=0.12.0->edge-gwas==0.1.1) (0.5.3)
Requirement already satisfied: pycparser in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
cffi>=1.17->pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2.21)
Requirement already satisfied: wrapt<3,>=1.10 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
deprecated<2.0,>=1.2->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (1.14.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
Jinja2>=2.10->numpydoc>=1.1.0->edge-gwas==0.1.1) (2.1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
requests>=2.5.0->sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
requests>=2.5.0->sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
requests>=2.5.0->sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
requests>=2.5.0->sphinx<7.0.0,>=4.0.0->edge-gwas==0.1.1) (2025.8.3)
Requirement already satisfied: humanfriendly>=9.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
coloredlogs->cyvcf2>=0.30.0->edge-gwas==0.1.1) (10.0)
Requirement already satisfied: cloudpickle>=1.5.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2.2.1)
Requirement already satisfied: fsspec>=2021.09.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2023.10.0)
Requirement already satisfied: partd>=1.2.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (1.4.1)
Requirement already satisfied: pyyaml>=5.3.1 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (6.0.1)
Requirement already satisfied: toolz>=0.10.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (0.12.0)
Requirement already satisfied: importlib-metadata>=4.13.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (7.0.1)
Requirement already satisfied: pydantic in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (1.10.12)
```

```
Requirement already satisfied: typeguard in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (4.4.4)
Requirement already satisfied: typing_inspect>=0.6.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (0.9.0)
Requirement already satisfied: pandas-stubs in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2.3.3.251219)
Requirement already satisfied: scipy-stubs in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (1.16.3.3)
Requirement already satisfied: distributed in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (2023.11.0)
Requirement already satisfied: zipp>=0.5 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from importlib-
metadata>=4.13.0->dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(3.17.0)
Requirement already satisfied: locket in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
partd>=1.2.0->dask[dataframe]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(1.0.0)
Requirement already satisfied: msgpack>=1.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
distributed->pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(1.0.3)
Requirement already satisfied: psutil>=5.7.2 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
distributed->pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(5.9.0)
Requirement already satisfied: sortedcontainers>=2.0.5 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
distributed->pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(2.4.0)
Requirement already satisfied: tblib>=1.6.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
distributed->pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(1.7.0)
Requirement already satisfied: tornado>=6.0.4 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
distributed->pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(6.3.3)
Requirement already satisfied: zict>=3.0.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
distributed->pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(3.0.0)
Requirement already satisfied: types-pytz>=2022.1.1 in
```

```
/Users/nicen/anaconda3/lib/python3.11/site-packages (from pandas-
stubs->pandera[dask,mypy]->pandas-plink<3.0.0,>=2.0.0->edge-gwas==0.1.1)
(2025.2.0.20251108)
Requirement already satisfied: optype<0.16,>=0.14.0 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
optype[numpy]<0.16,>=0.14.0->scipy-stubs->pandera[dask,mypy]->pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (0.15.0)
Requirement already satisfied: numpy-typing-compat<20251207,>=20250818.1.25 in
/Users/nicen/anaconda3/lib/python3.11/site-packages (from
optype[numpy]<0.16,>=0.14.0->scipy-stubs->pandera[dask,mypy]->pandas-
plink<3.0.0,>=2.0.0->edge-gwas==0.1.1) (20251206.1.25)
Installing collected packages: edge-gwas
  Running setup.py develop for edge-gwas
Successfully installed edge-gwas-0.1.1
  Installation successful!
```

```
[2]: # Cell 2: Install and validate the outside tools
# Install external tools (PLINK2, GCTA, R packages)
#!echo -e "yes\n3\n2" | edge-gwas-install-tools

# Verify installation
!edge-gwas-check-tools
```

```
=====
EDGE-GWAS External Tools Check
=====
```

```
Python Packages:
```

```
-----  
numpy: Installed  
pandas: Installed  
scipy: Installed  
statsmodels: Installed  
sklearn: Installed  
matplotlib: Installed  
pandas_plink: Installed
```

```
External Tools:
```

```
-----  
PLINK2: PLINK v2.0.0-a.7 64-bit (5 Dec 2025)  
GCTA: *****
```

```
R and Packages:
```

```
-----  
R: R version 4.3.3 (2024-02-29) -- "Angel Food Cake"  
R package GENESIS: Installed  
R package SNPRelate: Installed
```

```
R package gdsfmt: Installed
```

```
=====
All tools and packages are properly installed!
```

```
You can now use all EDGE-GWAS features including:
```

- Basic EDGE analysis
- PCA calculation with PLINK2
- GRM calculation with GCTA
- PC-AiR analysis with R/GENESIS

```
[3]: # Cell 3: Load the functions from package
from edge_gwas import *
import pandas as pd
from edge_gwas import EDGEAnalysis
from edge_gwas.utils import *
from edge_gwas.visualize import manhattan_plot, qq_plot, plot_alpha_distribution
from IPython.display import Image, display
```

```
[4]: # Cell 4: Download test files to default 'tests' directory
download_test_files(version='main', overwrite=True)
```

```
INFO:edge_gwas.io_handlers:Downloading test files to tests
```

```
Downloading test files to: tests
```

```
    Downloading: test.bed...
```

```
INFO:edge_gwas.io_handlers:Downloaded test.bed from
```

```
https://raw.githubusercontent.com/nicenzhou/edge-gwas/main/tests/test.bed
```

```
INFO:edge_gwas.io_handlers:Downloaded test.bim from
```

```
https://raw.githubusercontent.com/nicenzhou/edge-gwas/main/tests/test.bim
```

```
    (0.94 MB)
```

```
    Downloading: test.bim... (0.03 MB)
```

```
    Downloading: test.fam...
```

```
INFO:edge_gwas.io_handlers:Downloaded test.fam from
```

```
https://raw.githubusercontent.com/nicenzhou/edge-gwas/main/tests/test.fam
```

```
    (0.07 MB)
```

```
    Downloading: test.phen...
```

```
INFO:edge_gwas.io_handlers:Downloaded test.phen from
```

```
https://raw.githubusercontent.com/nicenzhou/edge-gwas/main/tests/test.phen
```

```
    (0.09 MB)
```

```
    Downloading: test.vcf...
```

```
INFO:edge_gwas.io_handlers:Downloaded test.vcf from
```

```
https://raw.githubusercontent.com/nicenzhou/edge-gwas/main/tests/test.vcf
```

(0.23 MB)

```
=====
```

Summary:

Downloaded: 5  
Skipped: 0  
Failed: 0

```
=====
```

[4]: {'downloaded': ['test.bed', 'test.bim', 'test.fam', 'test.phen', 'test.vcf'],  
'skipped': [],  
'failed': []}

[5]: # Cell 5: Load data  
geno, info = load\_plink\_data('tests/test.bed', 'tests/test.bim', 'tests/test.  
↳fam')  
geno\_vcf, info\_vcf = load\_vcf\_data('tests/test.vcf')

```
INFO:edge_gwas.utils:Loading PLINK data from tests/test.bed  
Mapping files: 100% | 3/3 [00:00<00:00, 34.04it/s]  
INFO:edge_gwas.utils:Loaded 3925 samples and 1000 variants  
INFO:edge_gwas.utils:Checking and recoding to ensure minor allele is ALT...  
INFO:edge_gwas.utils:Flipped 991 variants to ensure minor allele is ALT  
INFO:edge_gwas.utils:Example flipped variants: ['rs4475691', 'rs28705211',  
'rs9777703', 'rs3121567', 'rs3934834']  
INFO:edge_gwas.utils:MAF range: 0.0228 - 0.5000  
INFO:edge_gwas.utils:Loading VCF data from tests/test.vcf  
[W::bcf_hrec_check] Invalid tag name: "1000Genomes.ASN_AF"  
[W::bcf_hrec_check] Invalid tag name: "1000Genomes.AMR_AF"  
[W::bcf_hrec_check] Invalid tag name: "1000Genomes.AFR_AF"  
[W::bcf_hrec_check] Invalid tag name: "1000Genomes.EUR_AF"  
[W::bcf_hrec_check] Invalid tag name: "UCSC.Alignability/Uniqueness"  
INFO:edge_gwas.utils:Loaded 1 samples and 902 variants  
INFO:edge_gwas.utils:Using dosages (DS field)  
INFO:edge_gwas.utils:Checking and recoding to ensure minor allele is ALT...  
INFO:edge_gwas.utils:Flipped 54 variants (dosage)  
INFO:edge_gwas.utils:Example flipped variants: ['chrM:73', 'chrM:150',  
'chrM:152', 'chrM:195', 'chrM:410']  
INFO:edge_gwas.utils:MAF range: 0.0000 - 0.5000
```

[6]: # Cell 6: PCA on genotyped calls/raw genotyped data using sklearn, plink, PC-AiR  
pca\_df\_sklearn = calculate\_pca\_sklearn(geno, n\_pcs=10)  
pca\_df\_plink = calculate\_pca\_plink('tests/test', n\_pcs=10)  
pca\_df\_pcair = calculate\_pca\_pcair('tests/test', n\_pcs=10)

```
INFO:edge_gwas.utils:Calculating 10 principal components using scikit-learn...  
INFO:edge_gwas.utils: genotype_df basic validation passed: 3925 samples, 1000  
variants
```

```
INFO:edge_gwas.utils:Explained variance ratio: [0.01817612 0.01467451 0.01316931  
0.01167561 0.01104282]  
INFO:edge_gwas.utils:Total variance explained by 10 PCs: 0.117  
INFO:edge_gwas.utils:PCA complete for 3925 samples  
INFO:edge_gwas.utils:Calculating 10 PCs using PLINK2 (exact method)...  
INFO:edge_gwas.utils:Input format: bfile  
INFO:edge_gwas.utils:MAF threshold: 0.01  
INFO:edge_gwas.utils:LD pruning: window=50, step=5, r^2<0.2  
INFO:edge_gwas.utils:Step 1: LD pruning...  
INFO:edge_gwas.utils:Step 2: Calculating PCA...  
INFO:edge_gwas.utils:Variance explained by first 5 PCs: [0.11810096 0.10897235  
0.10476602 0.10282947 0.09930652]  
INFO:edge_gwas.utils:Total variance explained: 1.000  
INFO:edge_gwas.utils:PCA complete. Found 3925 samples.  
INFO:edge_gwas.utils:Calculating 10 PCs using PC-AiR...  
INFO:edge_gwas.utils:Calculating kinship matrix using GCTA...  
INFO:edge_gwas.utils:Calculating GRM using GCTA (method: grm)...  
INFO:edge_gwas.utils:MAF threshold: 0.01, Threads: 1  
INFO:edge_gwas.utils:GRM calculation complete. Output:  
/var/folders/x_/_6k45gs0x2110rxc3kvrlly80000gn/T/tmp3n7rxocp/pcair_grm.grm.*  
INFO:edge_gwas.utils:Running PC-AiR in R...  
INFO:edge_gwas.utils:Variance explained by first 5 PCs: [0.11964859 0.10842273  
0.10597959 0.10297332 0.10025663]
```

```
[7]: # Cell 7: GRM calculation use GCTA  
grm = calculate_grm_gcta('tests/test')
```

```
INFO:edge_gwas.utils:Calculating GRM using GCTA (method: grm)...  
INFO:edge_gwas.utils:MAF threshold: 0.01, Threads: 1  
INFO:edge_gwas.utils:GRM calculation complete. Output:  
/var/folders/x_/_6k45gs0x2110rxc3kvrlly80000gn/T/tmpmxlgiz51/grm.grm.*
```

```
[8]: # Cell 8: Prepare phenotype (adjust column names as needed)  
# First, check what columns are in test.pheno  
pheno_raw = pd.read_csv('tests/test.phen', sep='\t')  
  
# Then prepare phenotype with correct column names  
pheno = prepare_phenotype_data(  
    'tests/test.phen',  
    outcome_col='disease',  
    covariate_cols=[], # Empty list - no covariates  
    sep=' '
```

```
INFO:edge_gwas.utils:Loading phenotype data from tests/test.phen  
INFO:edge_gwas.utils:Loaded 3925 samples with columns: ['IID', 'disease']  
INFO:edge_gwas.utils: phenotype_df validation passed  
INFO:edge_gwas.utils: Prepared phenotype data for 3925 samples
```

```
[9]: # Cell 9: Add the calculated PCs to pheno table
#pheno_sklearn = attach_pcs_to_phenotype(pheno, pca_df_sklearn, n_pcs=10)
#pheno_plink = attach_pcs_to_phenotype(pheno, pca_df_plink, n_pcs=10)
pheno_pcair = attach_pcs_to_phenotype(pheno, pca_df_pcair, n_pcs=10)
```

INFO:edge\_gwas.utils:Attaching 10 PCs to phenotype data  
 INFO:edge\_gwas.utils:Phenotype samples: 3925  
 INFO:edge\_gwas.utils:PCA samples: 3925  
 INFO:edge\_gwas.utils:Samples with PCs after merge: 3925

```
[10]: # Cell 10: All possible QC filtering
pheno = pheno_pcair
geno_filtered, pheno_filtered = filter_genotype_data(
    geno,
    pheno,
    min_maf=0.01,
    max_missing_per_variant=0.1,
    min_call_rate_per_sample=0.95
)

#geno = filter_variants_by_maf(geno, min_maf=0.01)
#geno = filter_variants_by_missing(geno, max_missing=0.01)
#geno, pheno = filter_samples_by_call_rate(geno, pheno, min_call_rate=0.05)
#hwe = calculate_hwe_pvalues(geno)
#geno = filter_variants_by_hwe(geno, hwe_threshold=1e-6)

#geno_vcf = filter_variants_by_maf(geno_vcf, min_maf=0.01)
#geno_vcf = filter_variants_by_missing(geno_vcf, max_missing=0.01)
#geno_vcf = filter_samples_by_call_rate(geno, pheno, min_call_rate=0.95)
```

INFO:edge\_gwas.utils:=====
=====  
 INFO:edge\_gwas.utils:Starting genotype QC filtering  
 INFO:edge\_gwas.utils:=====
=====  
 =====  
 INFO:edge\_gwas.utils:Input: 3925 samples x 1000 variants  
 INFO:edge\_gwas.utils:  
 [1/3] Filtering variants by MAF >= 0.01  
 INFO:edge\_gwas.utils:Filtered variants by MAF >= 0.01  
 INFO:edge\_gwas.utils:Kept 1000/1000 variants (100.0%)  
 INFO:edge\_gwas.utils:MAF range in filtered data: 0.0228 - 0.5000  
 INFO:edge\_gwas.utils:  
 [2/3] Filtering variants by missing rate <= 0.1  
 INFO:edge\_gwas.utils:Filtered variants by missing rate <= 0.1  
 INFO:edge\_gwas.utils:Kept 918/1000 variants (91.8%)  
 INFO:edge\_gwas.utils:Missing rate range in filtered data: 0.0000 - 0.0991  
 INFO:edge\_gwas.utils:  
 [3/3] Filtering samples by call rate >= 0.95

```

INFO:edge_gwas.utils:Filtered samples by call rate >= 0.95
INFO:edge_gwas.utils:Genotype samples: kept 3915/3925 (99.7%)
INFO:edge_gwas.utils:Phenotype samples: kept 3915/3925 (99.7%)
INFO:edge_gwas.utils:Call rate range: 0.9521 - 1.0000
INFO:edge_gwas.utils:
=====
INFO:edge_gwas.utils:FILTERING SUMMARY
INFO:edge_gwas.utils:=====
=====
INFO:edge_gwas.utils:Variants: 1000 → 918 (91.8% retained)
INFO:edge_gwas.utils:Samples: 3925 → 3915 (99.7% retained)
INFO:edge_gwas.utils:Phenotype samples: 3915
INFO:edge_gwas.utils:=====
=====

      step   criterion   before   after   removed
MAF filter     >= 0.01     1000     1000       0
Variant missing    <= 0.1     1000     918      82
Sample call rate    >= 0.95    3925     3915      10
=====
```

```
[11]: # Cell 10: Validate the geno, pheno, and both with alignment (remove
      ↵individuals who are only in geno or pheno)
#validate_genotype_df(geno_filtered)
#validate_phenotype_df(pheno_filtered, outcome_col='disease',
      ↵covariate_cols=get_pc_covariate_list(n_pcs=10,pc_prefix='PC'))
geno_fixed, info_fixed, final_report = validate_and_fix_encoding(geno_filtered,
      ↵info)
geno, pheno = validate_and_align_data(geno_fixed, pheno_filtered)

final_report
```

```

INFO:edge_gwas.utils: genotype_df basic validation passed: 3915 samples, 918
variants
INFO:edge_gwas.utils:Checking encoding (minor allele as ALT)...
INFO:edge_gwas.utils:Encoding validation: PASS=918, FAIL=0, WARNING=0
INFO:edge_gwas.utils:Overall: PASSED
INFO:edge_gwas.utils:MAF range: 0.0228 - 0.5000
INFO:edge_gwas.utils: No fixes needed
INFO:edge_gwas.utils:Validating and aligning genotype and phenotype data...
INFO:edge_gwas.utils:Sample overlap:
INFO:edge_gwas.utils: Genotype samples: 3915
INFO:edge_gwas.utils: Phenotype samples: 3915
INFO:edge_gwas.utils: Common samples: 3915
INFO:edge_gwas.utils: Data validated and aligned: 3915 samples
INFO:edge_gwas.utils: Genotype shape: (3915, 918)
INFO:edge_gwas.utils: Genotype index: 'sample_id'
INFO:edge_gwas.utils: Phenotype shape: (3915, 11)
```

```

INFO:edge_gwas.utils: Phenotype index: 'sample_id'

[11]:    variant_id  n_valid  alt_freq      maf  minor_is_alt  status  issue  \
0      rs3934834    3885  0.143115  0.143115      True  PASS  None
1      rs3737728    3915  0.289272  0.289272      True  PASS  None
2      rs6687776    3914  0.156873  0.156873      True  PASS  None
3      rs9651273    3913  0.274981  0.274981      True  PASS  None
4      rs4970405    3912  0.101866  0.101866      True  PASS  None
..
913     rs6679380    3915  0.268072  0.268072      True  PASS  None
914     rs7418365    3912  0.356595  0.356595      True  PASS  None
915     rs10157778   3911  0.381105  0.381105      True  PASS  None
916     rs12040542   3910  0.118542  0.118542      True  PASS  None
917     rs7411273   3910  0.160102  0.160102      True  PASS  None

      was_fixed
0        False
1        False
2        False
3        False
4        False
..
913       ...
914       False
915       False
916       False
917       False

[918 rows x 8 columns]

```

```

[12]: # Cell 11: Split data for alpha calculation and alpha applying sets
from edge_gwas.utils import stratified_train_test_split
train_g, test_g, train_p, test_p = stratified_train_test_split(
    geno, pheno, 'disease',
    test_size=0.3,
    random_state=42,
    geno_id_col='sample_id',
    pheno_id_col='sample_id',
    is_binary=False
)

```

```

INFO:edge_gwas.utils:Splitting data into train/test (70%/30%)
INFO:edge_gwas.utils:Found 3915 common samples (dropped 0 from genotype, 0 from phenotype)
INFO:edge_gwas.utils:Training set: 2740 samples
INFO:edge_gwas.utils:Test set: 1175 samples

```

```
[13]: # Cell 12: Run EDGE analysis using full function
edge = EDGEAnalysis(outcome_type='continuous', n_jobs=-1)
alpha_df, gwas_df = edge.run_full_analysis(
    train_g, train_p, test_g, test_p,
    outcome='disease',
    covariates=get_pc_covariate_list(n_pcs=10, pc_prefix='PC'),
    variant_info = info,
    mean_centered = True #??? NEED TO ASK MOLLY IF WE WANT TO DO THIS IN THE ↴ALPHA CALCULATION
)
```

INFO:edge\_gwas.core:OLS optimization method: bfsgs  
 INFO:edge\_gwas.core:Starting EDGE analysis...  
 INFO:edge\_gwas.core:Using mean-centered codominant model (no intercept)  
 INFO:edge\_gwas.core:OLS optimization method: bfsgs  
 INFO:edge\_gwas.core:Step 1: Calculating alpha values on training data...  
 INFO:edge\_gwas.core:Using mean-centered codominant model (no intercept)  
 INFO:edge\_gwas.core:Processing variant 100/918  
 INFO:edge\_gwas.core:Processing variant 200/918  
 INFO:edge\_gwas.core:Processing variant 300/918  
 INFO:edge\_gwas.core:Processing variant 400/918  
 INFO:edge\_gwas.core:Processing variant 500/918  
 INFO:edge\_gwas.core:Processing variant 600/918  
 INFO:edge\_gwas.core:Processing variant 700/918  
 INFO:edge\_gwas.core:Processing variant 800/918  
 INFO:edge\_gwas.core:Processing variant 900/918  
 INFO:edge\_gwas.core:Alpha calculation complete. Processed 918 variants.  
 INFO:edge\_gwas.core:Skipped 0 variants due to convergence issues.  
 INFO:edge\_gwas.core:Mean-centered codominant model was used (no intercept)  
 INFO:edge\_gwas.core:OLS optimization method used: bfsgs  
 INFO:edge\_gwas.core:Step 2: Applying alpha values on test data...  
 INFO:edge\_gwas.core:Processing variant 100/918  
 INFO:edge\_gwas.core:Processing variant 200/918  
 INFO:edge\_gwas.core:Processing variant 300/918  
 INFO:edge\_gwas.core:Processing variant 400/918  
 INFO:edge\_gwas.core:Processing variant 500/918  
 INFO:edge\_gwas.core:Processing variant 600/918  
 INFO:edge\_gwas.core:Processing variant 700/918  
 INFO:edge\_gwas.core:Processing variant 800/918  
 INFO:edge\_gwas.core:Processing variant 900/918  
 INFO:edge\_gwas.core:EDGE GWAS complete. Analyzed 918 variants.  
 INFO:edge\_gwas.core:Skipped 0 variants.  
 INFO:edge\_gwas.core:OLS optimization method used: bfsgs  
 INFO:edge\_gwas.core:MAF range: 0.0221 - 0.4996  
 INFO:edge\_gwas.core:EDGE analysis complete!

```
[14]: # Cell 13: Run EDGE analysis using two steps
#edge = EDGEAnalysis(outcome_type='continuous', n_jobs=-1)
#alpha_df_inv = edge.calculate_alpha(train_g, train_p, variant_info = info, □
    ↪outcome='disease', □
    ↪covariates=get_pc_covariate_list(n_pcs=10,pc_prefix='PC'), mean_centered = □
    ↪True)
#gwas_df_inv = edge.apply_alpha(test_g, test_p, variant_info = info, □
    ↪outcome='disease', covariates=get_pc_covariate_list(n_pcs=10,pc_prefix='PC'))
```

```
[15]: # Cell 14: Add fake chr and pos to the gwas results for testing the fucntions □
    ↪for plotting
```

```
# Human genome chromosome lengths (GRCh38/hg38) in base pairs
```

```
chr_lengths = {
    1: 248956422,
    2: 242193529,
    3: 198295559,
    4: 190214555,
    5: 181538259,
    6: 170805979,
    7: 159345973,
    8: 145138636,
    9: 138394717,
    10: 133797422,
    11: 135086622,
    12: 133275309,
    13: 114364328,
    14: 107043718,
    15: 101991189,
    16: 90338345,
    17: 83257441,
    18: 80373285,
    19: 58617616,
    20: 64444167,
    21: 46709983,
    22: 50818468
}
```

```
# Calculate total genome length
total_length = sum(chr_lengths.values())
```

```
# Calculate probability for each chromosome based on its length
chr_probs = []
chr_nums = []
for chr_num in range(1, 23):
    chr_nums.append(chr_num)
    chr_probs.append(chr_lengths[chr_num] / total_length)
```

```

# Calculate cumulative probabilities for chromosome assignment
cumsum_probs = []
running_sum = 0
for prob in chr_probs:
    running_sum += prob
    cumsum_probs.append(running_sum)

# Add chr and pos columns to alpha_df
import numpy as np

# Add chr and pos columns to gwas_df
n_variants_gwas = len(gwas_df)

chr_assignments_gwas = []
pos_assignments_gwas = []

np.random.seed(43)  # Different seed for test data
random_values_gwas = np.random.random(n_variants_gwas)

for rand_val in random_values_gwas:
    # Find which chromosome this random value falls into
    assigned_chr = 22 # default to last chr
    for i in range(len(cumsum_probs)):
        if rand_val < cumsum_probs[i]:
            assigned_chr = chr_nums[i]
            break

    chr_assignments_gwas.append(assigned_chr)

    # Assign random position within that chromosome
    max_pos = chr_lengths[assigned_chr]
    pos = np.random.randint(1, max_pos + 1)
    pos_assignments_gwas.append(pos)

# Add to dataframe
gwas_df['chr'] = chr_assignments_gwas
gwas_df['pos'] = pos_assignments_gwas

# Sort by chr then pos
gwas_df = gwas_df.sort_values(['chr', 'pos']).reset_index(drop=True)

print(f"\nAdded chr and pos columns to {len(gwas_df)} GWAS variants")
print(f"\nChromosome distribution in GWAS results:")
for chr_num in range(1, 23):
    count = sum(1 for c in chr_assignments_gwas if c == chr_num)
    print(f"  Chr {chr_num}: {count} variants")

```

```
# Display first few rows to verify
print(f"\nFirst few rows of gwas_df with chr/pos:")
print(gwas_df.head(10))
```

Added chr and pos columns to 918 GWAS variants

Chromosome distribution in GWAS results:

```
Chr 1: 78 variants
Chr 2: 77 variants
Chr 3: 55 variants
Chr 4: 56 variants
Chr 5: 56 variants
Chr 6: 53 variants
Chr 7: 50 variants
Chr 8: 43 variants
Chr 9: 53 variants
Chr 10: 42 variants
Chr 11: 43 variants
Chr 12: 45 variants
Chr 13: 31 variants
Chr 14: 34 variants
Chr 15: 38 variants
Chr 16: 27 variants
Chr 17: 26 variants
Chr 18: 32 variants
Chr 19: 15 variants
Chr 20: 26 variants
Chr 21: 16 variants
Chr 22: 22 variants
```

First few rows of gwas\_df with chr/pos:

	variant_id	snp	alpha_value	coef	std_err	stat	\
0	rs242052	rs242052	-0.589834	-0.165771	0.077663	-2.134493	
1	rs950601	rs950601	-0.434689	-0.075307	0.068533	-1.098841	
2	rs2651929	rs2651929	-1.892338	-0.036256	0.026378	-1.374456	
3	rs7521197	rs7521197	-0.320352	-0.436127	0.240919	-1.810264	
4	rs2493215	rs2493215	0.169081	-0.098508	0.085215	-1.156000	
5	rs2063313	rs2063313	-4.069402	0.014333	0.016434	0.872199	
6	rs707593	rs707593	0.163850	-0.159159	0.094489	-1.684417	
7	rs428001	rs428001	-0.040393	-0.115290	0.080408	-1.433819	
8	rs16838750	rs16838750	0.645575	-0.075819	0.120667	-0.628332	
9	rs3007421	rs3007421	-1.453162	0.014122	0.046861	0.301356	

	pval	conf_int_low	conf_int_high	n_samples	...	StdErr	P-value	\
0	0.033013	-0.318148	-0.013395	1167	...	0.077663	0.033013	

```

1 0.272065 -0.209769 0.059155 1173 ... 0.068533 0.272065
2 0.169566 -0.088010 0.015499 1172 ... 0.026378 0.169566
3 0.070513 -0.908812 0.036558 1175 ... 0.240919 0.070513
4 0.247919 -0.265701 0.068684 1175 ... 0.085215 0.247919
5 0.383280 -0.017910 0.046576 1173 ... 0.016434 0.383280
6 0.092370 -0.344547 0.026229 1174 ... 0.094489 0.092370
7 0.151893 -0.273051 0.042470 1175 ... 0.080408 0.151893
8 0.529910 -0.312569 0.160931 1174 ... 0.120667 0.529910
9 0.763197 -0.077819 0.106063 1173 ... 0.046861 0.763197

```

	MAF	EAF	chr	pos	ref_allele	alt_allele	maf	\
0	0.205227	0.205227	1	2209687	A	G	0.205227	
1	0.311168	0.311168	1	2414425	C	A	0.311168	
2	0.334898	0.334898	1	5373929	A	G	0.334898	
3	0.055745	0.055745	1	13442680	T	C	0.055745	
4	0.460851	0.460851	1	15414235	A	G	0.460851	
5	0.148764	0.148764	1	32090525	A	G	0.148764	
6	0.341993	0.341993	1	34069586	C	T	0.341993	
7	0.379574	0.379574	1	39143445	G	A	0.379574	
8	0.088586	0.088586	1	39367428	C	T	0.088586	
9	0.123188	0.123188	1	49856442	G	A	0.123188	

```

eaf
0 0.205227
1 0.311168
2 0.334898
3 0.055745
4 0.460851
5 0.148764
6 0.341993
7 0.379574
8 0.088586
9 0.123188

```

[10 rows x 26 columns]

```

[16]: # Cell 15: Visualize with alpha distribution, gwas results in manhattan plot, ↵
      ↵and genetic inflation in qq plot
alpha_distribution = plot_alpha_distribution(alpha_df, 'alpha.png', xlim=(-3,3))
display(Image('alpha.png'))

manhattan_plot(gwas_df, 'manhattan.png')
display(Image('manhattan.png'))

lambda_gc = qq_plot(gwas_df, 'qq.png')
print(f"Lambda GC: {lambda_gc:.3f}")
display(Image('qq.png'))

```

Note: 60 alpha values outside range [-3, 3] excluded from plot

Alpha distribution plot saved to alpha.png

Total alpha values: 918

Values plotted (within [-3, 3]): 858

Alpha statistics:

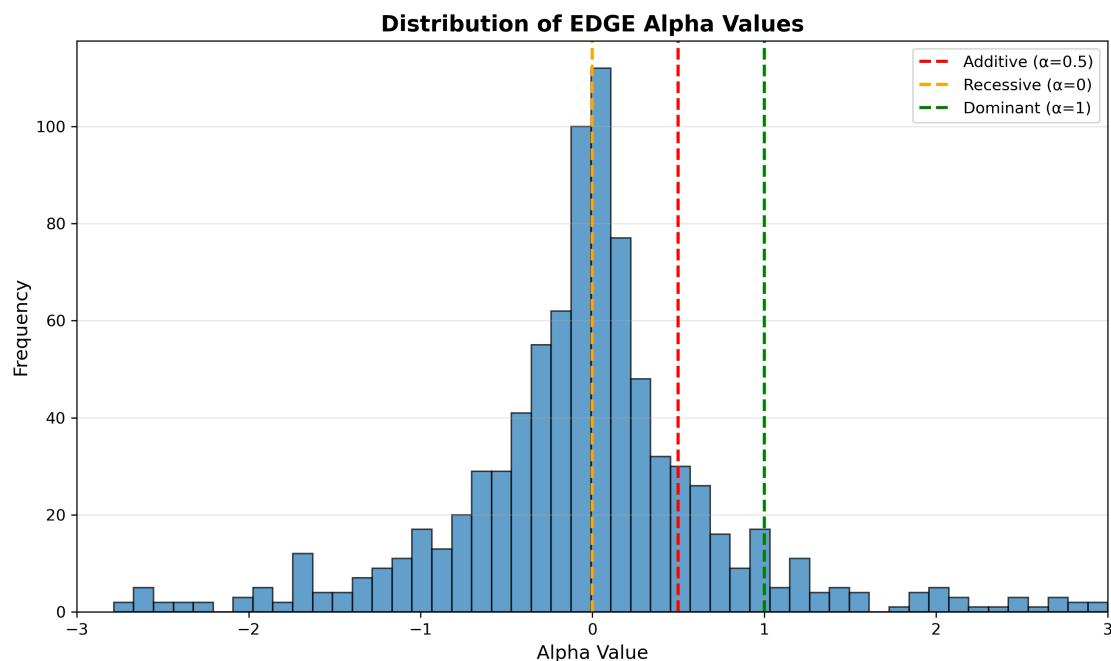
Mean: 0.399

Median: -0.025

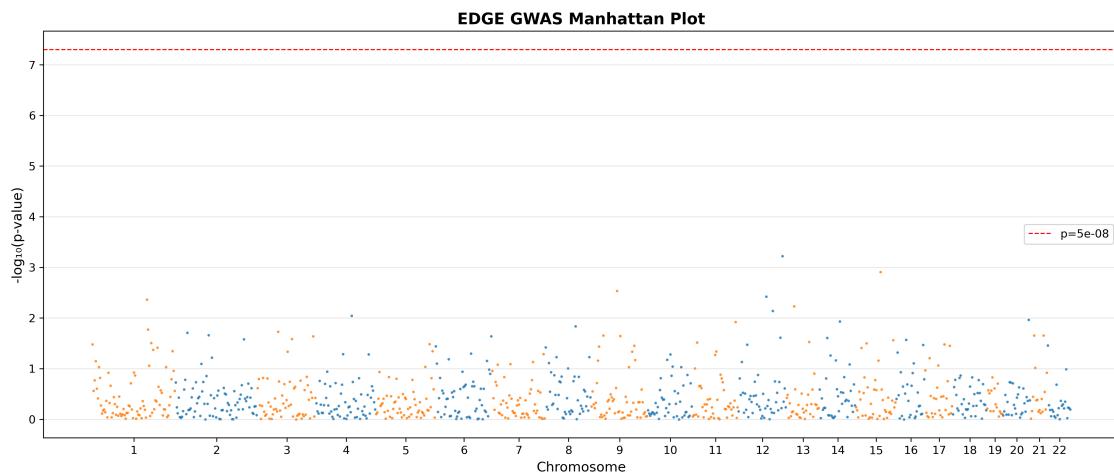
Std: 12.795

Min: -97.780

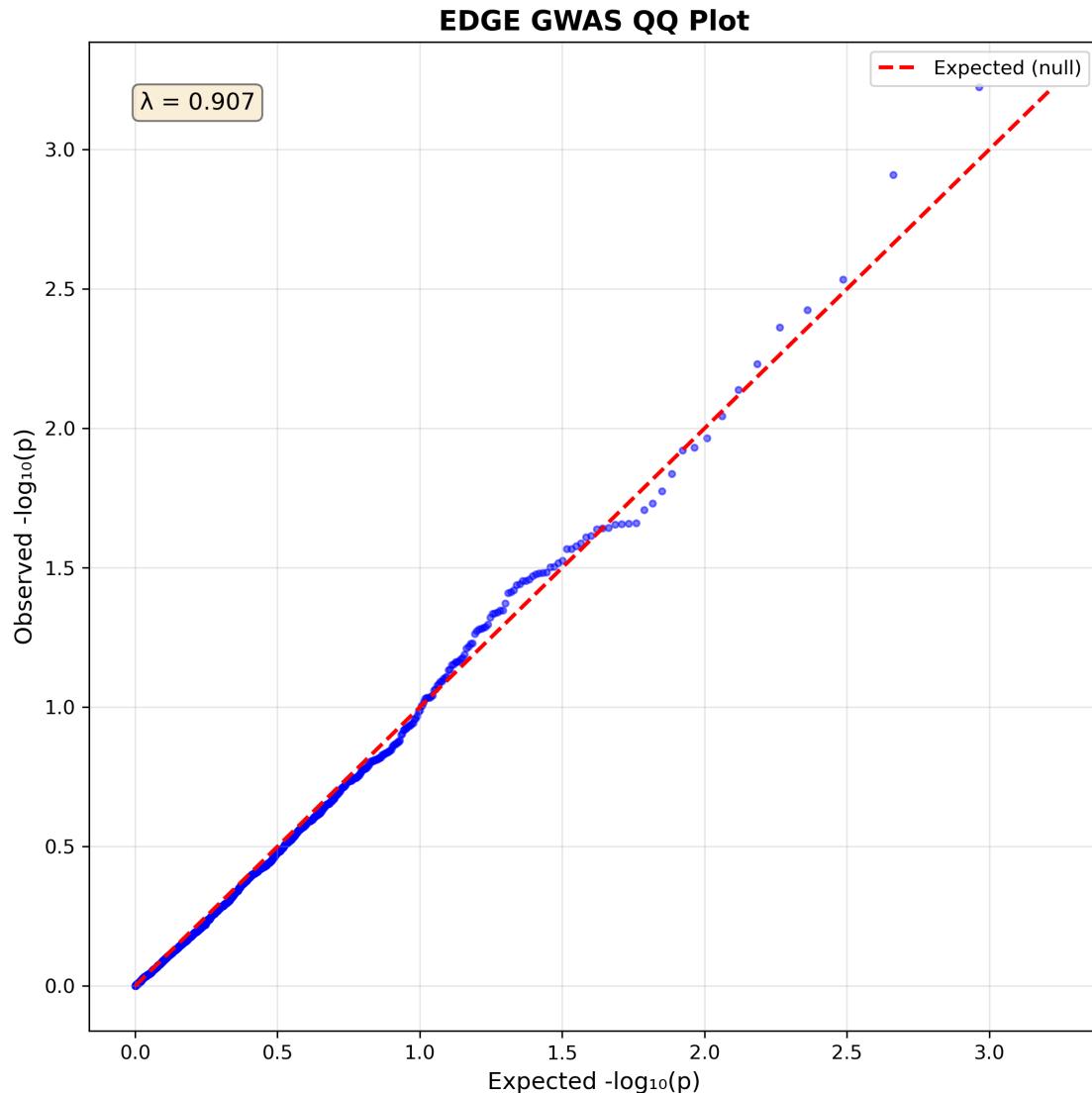
Max: 244.898



Manhattan plot saved to manhattan.png



QQ plot saved to qq.png  
Genomic inflation factor ( ): 0.907  
Lambda GC: 0.907



[ ]:

[ ]: