

edge_gwas_test_run_extend_2

January 22, 2026

```
[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
#!cd ~ && git clone -b v0.1.2dev 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
Found existing installation: edge-gwas 0.1.1
Uninstalling edge-gwas-0.1.1:
  Successfully uninstalled edge-gwas-0.1.1
Cloning into 'edge-gwas'...
remote: Enumerating objects: 1283, done.
remote: Counting objects: 100% (307/307), done.
remote: Compressing objects: 100% (119/119), done.
remote: Total 1283 (delta 272), reused 188 (delta 188), pack-reused 976 (from
2)
Receiving objects: 100% (1283/1283), 3.46 MiB | 4.10 MiB/s, done.
Resolving deltas: 100% (817/817), 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  
(0.94 MB)  
    Downloading: test.bim...  
  
INFO:edge_gwas.io_handlers:Downloaded test.bim from  
https://raw.githubusercontent.com/nicenzhou/edge-gwas/main/tests/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, 30.69it/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]: info
```

```
[6]:          chrom      pos ref_allele alt_allele      MAF
variant_id
rs4475691       1   836671        C        T  0.197698
rs28705211       1   890368        G        C  0.278112
rs9777703        1   918699        T        C  0.030161
rs3121567        1   933331        G        A  0.028936
rs3934834        1   995669        C        T  0.143316
...
...      ...    ...    ...    ...
rs7418365       1   9457890       A        G  0.356196
rs7367253       1   9462280       C        T  0.088946
rs10157778       1   9465199       G        A  0.380898
rs12040542       1   9514086       T        C  0.118653
rs7411273       1   9529628       C        T  0.160077
```

[1000 rows x 5 columns]

```
[7]: # 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 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.

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

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/tmp4gh9nds_/grm.grm.*
INFO:edge_gwas.utils:Loading GRM from
/var/folders/x_/6k45gs0x2110rxc3kvrlly80000gn/T/tmp4gh9nds_/grm...
INFO:edge_gwas.utils:Loaded 3925 x 3925 GRM matrix
INFO:edge_gwas.utils:Mean diagonal: 0.999
INFO:edge_gwas.utils:Mean off-diagonal: -0.000257

```
[9]: # 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

```
[10]: # 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

```
[11]: # Cell 10: All possible QC filtering
pheno = pheno_plink
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

```

```

[12]: # 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'

```

```

[12]:    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        ...   ...  ...
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    False
914    False
915    False
916    False
917    False

[918 rows x 8 columns]

```

```
[13]: # 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

```
[14]: # Cell 12: Run EDGE analysis using full function
## USE GRM, NOT MEAN CENTERED
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,
    grm_matrix = grm_matrix,
    grm_sample_ids = sample_ids,
    mean_centered = False
)
```

INFO:edge_gwas.core:OLS optimization method: bfgs
 INFO:edge_gwas.core:Starting EDGE analysis...
 INFO:edge_gwas.core:OLS optimization method: bfgs
 INFO:edge_gwas.core:Step 1: Calculating alpha values on training data...
 INFO:edge_gwas.core:Incorporating GRM for population structure control
 INFO:edge_gwas.core:Using 'IID' column from grm_sample_ids (sample_id not found)
 INFO:edge_gwas.core:Found 2740 common samples between data and GRM
 INFO:edge_gwas.core:Analysis data samples: 2740
 INFO:edge_gwas.core:GRM samples: 3925
 INFO:edge_gwas.core:Cached GRM transformation for 2740 samples
 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:OLS optimization method used: bfgs
INFO:edge_gwas.core:Step 2: Applying alpha values on test data...
INFO:edge_gwas.core:Incorporating GRM for population structure control
INFO:edge_gwas.core:Using 'IID' column from grm_sample_ids (sample_id not found)
INFO:edge_gwas.core:Found 1175 common samples between data and GRM
INFO:edge_gwas.core:Analysis data samples: 1175
INFO:edge_gwas.core:GRM samples: 3925
INFO:edge_gwas.core:Cached GRM transformation for 1163 samples
INFO:edge_gwas.core:Cached GRM transformation for 1175 samples
INFO:edge_gwas.core:Cached GRM transformation for 1174 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1174 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1162 samples
INFO:edge_gwas.core:Cached GRM transformation for 1163 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Cached GRM transformation for 1174 samples
INFO:edge_gwas.core:Cached GRM transformation for 1174 samples
INFO:edge_gwas.core:Cached GRM transformation for 1074 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Cached GRM transformation for 1167 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1171 samples
INFO:edge_gwas.core:Cached GRM transformation for 1169 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1171 samples
INFO:edge_gwas.core:Cached GRM transformation for 1168 samples
INFO:edge_gwas.core:Cached GRM transformation for 1167 samples
INFO:edge_gwas.core:Cached GRM transformation for 1171 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1160 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
```



```

INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1174 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1174 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1172 samples
INFO:edge_gwas.core:Processing variant 900/918
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1171 samples
INFO:edge_gwas.core:Cached GRM transformation for 1173 samples
INFO:edge_gwas.core:Cached GRM transformation for 1174 samples
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: bfgs
INFO:edge_gwas.core:Variants with position information: 918/918
INFO:edge_gwas.core:MAF range: 0.0221 - 0.4996
INFO:edge_gwas.core:EDGE analysis complete!

```

[15]: alpha_df

	chrom	pos	variant_id	alpha_value	ref_allele	alt_allele	eaf	\
0	1	995669	rs3934834	-0.818791	C	T	0.857090	
1	1	1011278	rs3737728	0.104772	C	T	0.708212	
2	1	1020428	rs6687776	0.776481	C	T	0.840453	
3	1	1021403	rs9651273	-0.106749	G	A	0.727356	
4	1	1038818	rs4970405	0.232136	A	G	0.895727	
..	
913	1	9455825	rs6679380	-0.957261	T	C	0.730657	
914	1	9457890	rs7418365	0.063630	A	G	0.641527	
915	1	9465199	rs10157778	0.382590	G	A	0.615863	
916	1	9514086	rs12040542	0.255909	T	C	0.881901	
917	1	9529628	rs7411273	0.332586	C	T	0.841557	
	coef_het	coef_hom	std_err_het	std_err_hom	conf_int_low_het	conf_int_high_het	\	
0	0.065655	-0.080185	0.076276	0.189068	-0.083911	0.234824		
1	0.009158	0.087410	0.062682	0.120103	-0.113750	0.210293		
2	0.039526	0.050904	0.084877	0.189973	-0.126904	0.304770		
3	-0.006279	0.058819	0.063623	0.125678	-0.131033	0.297371		
4	0.073726	0.317599	0.103311	0.257556	-0.128850	0.633707		
..		
913	-0.070803	0.073964	0.068080	0.133877	-0.204296	0.370973		
914	0.007838	0.123185	0.068623	0.125464	-0.126719	0.259152		
915	0.079462	0.207694	0.066731	0.120040	-0.051386	0.397733		
916	-0.038501	-0.150447	0.088426	0.225275	-0.211890	0.542479		

```

917  0.090912  0.273349      0.080338      0.192137      -0.066617

    conf_int_high_het  conf_int_low_hom  conf_int_high_hom  pval het \ 
0          0.215220        -0.450916        0.290546  0.389452
1          0.132066        -0.148093        0.322913  0.883849
2          0.205956        -0.321601        0.423409  0.641479
3          0.118476        -0.187615        0.305254  0.921392
4          0.276302        -0.187425        0.822623  0.475517
..
913          ...          ...
914          ...          ...
915          ...          ...
916          ...          ...
917          ...          ...

pval_hom  n_samples convergence_status
0  0.671521      2740      converged
1  0.466805      2740      converged
2  0.788755      2740      converged
3  0.639810      2740      converged
4  0.217634      2740      converged
..
913  0.580667      2740      converged
914  0.326268      2740      converged
915  0.083706      2740      converged
916  0.504295      2740      converged
917  0.154945      2740      converged

```

[918 rows x 19 columns]

[16]: gwas_df

```

[16]:   chrom      pos  variant_id ref_allele alt_allele alpha_value     coef \
0       1  995669  rs3934834        C          T   -0.818791  0.036645
1       1 1011278  rs3737728        C          T    0.104772  0.005303
2       1 1020428  rs6687776        C          T    0.776481 -0.010593
3       1 1021403  rs9651273        G          A   -0.106749 -0.268087
4       1 1038818  rs4970405        A          G    0.232136 -0.116555
..
913     ...     ...     ...     ...     ...     ...     ...
914     1  9455825  rs6679380        T          C   -0.957261 -0.013059
915     1  9457890  rs7418365        A          G    0.063630  0.030258
916     1  9465199  rs10157778        G          A    0.382590  0.007903
917     1  9514086  rs12040542        T          C    0.255909  0.151992
917     1  9529628  rs7411273        C          T    0.332586  0.160065

std_err      stat      pval  conf_int_low  conf_int_high  n_samples \
0  0.082744  0.442869  0.657944      -0.125701        0.198991      1163

```

```

1   0.132570  0.039999  0.968101    -0.254800    0.265405    1175
2   0.131009 -0.080860  0.935567    -0.267635    0.246448    1175
3   0.112646 -2.379895  0.017478    -0.489100    -0.047074    1175
4   0.359638 -0.324091  0.745928    -0.822167    0.589056    1174
...
913  0.048975 -0.266653  0.789784    -0.109149    0.083030    1175
914  0.113322  0.267008  0.789510    -0.192081    0.252597    1174
915  0.139954  0.056468  0.954979    -0.266688    0.282494    1175
916  0.271340  0.560152  0.575484    -0.380379    0.684363    1175
917  0.218545  0.732413  0.464064    -0.268722    0.588853    1174

          maf      eaf
0   0.143594  0.143594
1   0.283404  0.283404
2   0.150638  0.150638
3   0.280426  0.280426
4   0.096252  0.096252
...
913  0.265106  0.265106
914  0.352215  0.352215
915  0.374043  0.374043
916  0.119574  0.119574
917  0.163969  0.163969

```

[918 rows x 15 columns]

```

[17]: # Cell 12: Run EDGE analysis using full function
## USE MEAN CENTERED
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: bfgs
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: bfgs
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: bfgs
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: bfgs
INFO:edge_gwas.core:Variants with position information: 918/918
INFO:edge_gwas.core:MAF range: 0.0221 - 0.4996
INFO:edge_gwas.core:EDGE analysis complete!

```

[18]: alpha_df

	chrom	pos	variant_id	alpha_value	ref_allele	alt_allele	eaf	\
0	1	995669	rs3934834	-0.913123	C	T	0.857090	
1	1	1011278	rs3737728	-0.005884	C	T	0.708212	
2	1	1020428	rs6687776	1.051580	C	T	0.840453	
3	1	1021403	rs9651273	0.184069	G	A	0.727356	
4	1	1038818	rs4970405	0.267213	A	G	0.895727	
..	
913	1	9455825	rs6679380	-0.075699	T	C	0.730657	
914	1	9457890	rs7418365	0.030302	A	G	0.641527	
915	1	9465199	rs10157778	0.091313	G	A	0.615863	
916	1	9514086	rs12040542	-1.589936	T	C	0.881901	
917	1	9529628	rs7411273	0.076065	C	T	0.841557	
	coef_het	coef_hom	std_err_het	std_err_hom	conf_int_low_het			\
0	0.064346	-0.070468	0.038645	0.142315	-0.011431			
1	-0.000680	0.115545	0.029687	0.066546	-0.058892			
2	0.089503	0.085113	0.037049	0.120983	0.016856			
3	0.005741	0.031188	0.030277	0.071989	-0.053627			

```

4    0.092880  0.347588      0.044378      0.186461      0.005863
..
913 -0.011366  0.150150      0.031178      0.075650      ...
914  0.004339  0.143205      0.028424      0.056201      ...
915  0.011298  0.123727      0.028139      0.051524      ...
916 -0.008823  0.005549      0.042113      0.168013      ...
917  0.017243  0.226683      0.036706      0.133640      ...
               conf_int_high_het  conf_int_low_hom  conf_int_high_hom  pval het \
0                  0.140124          -0.349525        0.208588  0.096019
1                  0.057532          -0.014940        0.246030  0.981730
2                  0.162150          -0.152115        0.322341  0.015766
3                  0.065109          -0.109970        0.172345  0.849633
4                  0.179897          -0.018031        0.713207  0.036446
..
913                 ...           ...           ...           ...
914                 ...           ...           ...           ...
915                 ...           ...           ...           ...
916                 ...           ...           ...           ...
917                 ...           ...           ...           ...

      pval_hom  n_samples convergence_status
0   0.620529      2740      converged
1   0.082619      2740      converged
2   0.481798      2740      converged
3   0.664882      2740      converged
4   0.062410      2740      converged
..
913  0.047265      2740      converged
914  0.010886      2740      converged
915  0.016402      2740      converged
916  0.973655      2740      converged
917  0.089958      2740      converged

```

[918 rows x 19 columns]

[19]: gwas_df

```

[19]:   chrom      pos variant_id ref_allele alt_allele alpha_value     coef \
0       1  995669  rs3934834         C          T   -0.913123  0.064335
1       1 1011278  rs3737728         C          T   -0.005884  0.046101
2       1 1020428  rs6687776         C          T    1.051580 -0.002344
3       1 1021403  rs9651273         G          A    0.184069 -0.128329
4       1 1038818  rs4970405         A          G    0.267213 -0.036155
..
913      1  9455825  rs6679380         ...          ...      ...
914      1  9457890  rs7418365         A          G    0.030302  0.063996

```

```

915      1  9465199  rs10157778          G      A  0.091313  0.054538
916      1  9514086  rs12040542          T      C -1.589936 -0.024760
917      1  9529628  rs7411273          C      T  0.076065  0.070691

      std_err      stat     pval conf_int_low conf_int_high n_samples \
0    0.067344  0.955313  0.339620   -0.067796    0.196466    1163
1    0.106614  0.432415  0.665520   -0.163076    0.255279    1175
2    0.060648 -0.038645  0.969180   -0.121334    0.116647    1175
3    0.115349 -1.112534  0.266139   -0.354644    0.097986    1175
4    0.216991 -0.166618  0.867700   -0.461893    0.389584    1174
..
913   0.101197 -0.630084  0.528763   -0.262312    0.134786    1175
914   0.090056  0.710626  0.477459   -0.112695    0.240687    1174
915   0.089514  0.609264  0.542469   -0.121090    0.230166    1175
916   0.043112 -0.574318  0.565864   -0.109346    0.059826    1175
917   0.190100  0.371861  0.710064   -0.302287    0.443669    1174

      maf      eaf
0    0.143594  0.143594
1    0.283404  0.283404
2    0.150638  0.150638
3    0.280426  0.280426
4    0.096252  0.096252
..
913   0.265106  0.265106
914   0.352215  0.352215
915   0.374043  0.374043
916   0.119574  0.119574
917   0.163969  0.163969

```

[918 rows x 15 columns]

```

[20]: # Cell 12: Run EDGE analysis using full function
## PLAIN
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 = False
)

```

```

INFO:edge_gwas.core:OLS optimization method: bfgs
INFO:edge_gwas.core:Starting EDGE analysis...
INFO:edge_gwas.core:OLS optimization method: bfgs
INFO:edge_gwas.core:Step 1: Calculating alpha values on training 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:Alpha calculation complete. Processed 918 variants.
INFO:edge_gwas.core:Skipped 0 variants due to convergence issues.
INFO:edge_gwas.core:OLS optimization method used: bfgs
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: bfgs
INFO:edge_gwas.core:Variants with position information: 918/918
INFO:edge_gwas.core:MAF range: 0.0221 - 0.4996
INFO:edge_gwas.core:EDGE analysis complete!

```

[21]: alpha_df

	chrom	pos	variant_id	alpha_value	ref_allele	alt_allele	eaf	\
0	1	995669	rs3934834	-1.217846	C	T	0.857090	
1	1	1011278	rs3737728	0.022409	C	T	0.708212	
2	1	1020428	rs6687776	1.035492	C	T	0.840453	
3	1	1021403	rs9651273	-0.005678	G	A	0.727356	
4	1	1038818	rs4970405	0.301028	A	G	0.895727	
..	
913	1	9455825	rs6679380	-0.101376	T	C	0.730657	
914	1	9457890	rs7418365	0.213514	A	G	0.641527	
915	1	9465199	rs10157778	0.330241	G	A	0.615863	
916	1	9514086	rs12040542	3.043454	T	C	0.881901	
917	1	9529628	rs7411273	0.086551	C	T	0.841557	
	coef_het	coef_hom	std_err_het	std_err_hom	conf_int_low_het	conf_int_high_het		\
0	0.073954	-0.060726	0.044740	0.144159	-0.013773	0.232876		
1	0.002664	0.118886	0.040245	0.071878	-0.076250	0.220383		

```

2    0.116093  0.112114    0.043600    0.123208    0.030600
3   -0.000144  0.025302    0.040209    0.076708   -0.078986
4    0.109805  0.364766    0.049335    0.187752    0.013068
..
...
913 -0.014813  0.146117    0.043000    0.083220   -0.099128
914  0.038679  0.181153    0.043252    0.066757   -0.046132
915  0.057628  0.174504    0.043243    0.062842   -0.027165
916 -0.021096 -0.006932    0.047536    0.169522   -0.114306
917  0.019851  0.229358    0.043393    0.135753   -0.065236

      conf_int_high_het  conf_int_low_hom  conf_int_high_hom  pval_het \
0            0.161682        -0.343398     0.221946  0.098449
1            0.081578        -0.022056     0.259827  0.947227
2            0.201586        -0.129476     0.353704  0.007798
3            0.078699        -0.125110     0.175714  0.997149
4            0.206542        -0.003384     0.732917  0.026115
..
...
913          0.069503        -0.017063     0.309297  0.730511
914          0.123490        0.050254     0.312052  0.371264
915          0.142421        0.051280     0.297727  0.182757
916          0.072114        -0.339337     0.325473  0.657227
917          0.104938        -0.036831     0.495547  0.647368

      pval_hom  n_samples convergence_status
0    0.673613       2740      converged
1    0.098246       2740      converged
2    0.362925       2740      converged
3    0.741539       2740      converged
4    0.052142       2740      converged
..
...
913  0.079236       2740      converged
914  0.006697       2740      converged
915  0.005526       2740      converged
916  0.967387       2740      converged
917  0.091233       2740      converged

[918 rows x 19 columns]

```

[22]: gwas_df

```

[22]:   chrom      pos variant_id ref_allele alt_allele alpha_value      coef \
0      1  995669  rs3934834         C          T    -1.217846  0.048402
1      1 1011278  rs3737728         C          T     0.022409  0.040711
2      1 1020428  rs6687776         C          T     1.035492 -0.002149
3      1 1021403  rs9651273         G          A    -0.005678 -0.158919
4      1 1038818  rs4970405         A          G     0.301028 -0.023449
..
...

```

```

913      1  9455825  rs6679380          T      C  -0.101376 -0.065436
914      1  9457890  rs7418365          A      G   0.213514  0.081464
915      1  9465199  rs10157778         G      A   0.330241  0.087808
916      1  9514086  rs12040542         T      C   3.043454  0.015465
917      1  9529628  rs7411273          C      T   0.086551  0.076165

      std_err      stat      pval conf_int_low conf_int_high n_samples \
0    0.051905  0.932518  0.351264   -0.053437     0.150241    1163
1    0.107922  0.377225  0.706075   -0.171033     0.252455    1175
2    0.061526 -0.034932  0.972140   -0.122863     0.118564    1175
3    0.112787 -1.409014  0.159098   -0.380209     0.062371    1175
4    0.201765 -0.116217  0.907500   -0.419312     0.372415    1174
..
913   0.099311 -0.658904  0.510088   -0.260284     0.129412    1175
914   0.096942  0.840331  0.400896   -0.108738     0.271666    1174
915   0.094291  0.931243  0.351921   -0.097191     0.272807    1175
916   0.023273  0.664494  0.506506   -0.030197     0.061127    1175
917   0.189315  0.402321  0.687522   -0.295271     0.447602    1174

      maf      eaf
0    0.143594  0.143594
1    0.283404  0.283404
2    0.150638  0.150638
3    0.280426  0.280426
4    0.096252  0.096252
..
913   0.265106  0.265106
914   0.352215  0.352215
915   0.374043  0.374043
916   0.119574  0.119574
917   0.163969  0.163969

```

[918 rows x 15 columns]

```

[23]: # Cell 13: Run EDGE analysis using two steps
## PLAIN, SEPERATED 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 = False)
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'))

```

INFO:edge_gwas.core:OLS optimization method: bfgs
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:OLS optimization method used: bfgs
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: bfgs
INFO:edge_gwas.core:Variants with position information: 918/918
INFO:edge_gwas.core:MAF range: 0.0221 - 0.4996

```

[24]: alpha_df_inv

	chrom	pos	variant_id	alpha_value	ref_allele	alt_allele	eaf	\
0	1	995669	rs3934834	-1.217846	C	T	0.857090	
1	1	1011278	rs3737728	0.022409	C	T	0.708212	
2	1	1020428	rs6687776	1.035492	C	T	0.840453	
3	1	1021403	rs9651273	-0.005678	G	A	0.727356	
4	1	1038818	rs4970405	0.301028	A	G	0.895727	
..	
913	1	9455825	rs6679380	-0.101376	T	C	0.730657	
914	1	9457890	rs7418365	0.213514	A	G	0.641527	
915	1	9465199	rs10157778	0.330241	G	A	0.615863	
916	1	9514086	rs12040542	3.043454	T	C	0.881901	
917	1	9529628	rs7411273	0.086551	C	T	0.841557	
	coef_het	coef_hom	std_err_het	std_err_hom	conf_int_low_het	conf_int_high_het		\
0	0.073954	-0.060726	0.044740	0.144159	-0.013773	0.231882		
1	0.002664	0.118886	0.040245	0.071878	-0.076250	0.255971		
2	0.116093	0.112114	0.043600	0.123208	0.030600	0.249702		
3	-0.000144	0.025302	0.040209	0.076708	-0.078986	0.223794		
4	0.109805	0.364766	0.049335	0.187752	0.013068	0.387537		
..	

```

913 -0.014813 0.146117      0.043000      0.083220      -0.099128
914  0.038679 0.181153      0.043252      0.066757      -0.046132
915  0.057628 0.174504      0.043243      0.062842      -0.027165
916 -0.021096 -0.006932     0.047536      0.169522      -0.114306
917  0.019851 0.229358      0.043393      0.135753      -0.065236

    conf_int_high_het  conf_int_low_hom  conf_int_high_hom  pval_het  \
0            0.161682        -0.343398       0.221946  0.098449
1            0.081578        -0.022056       0.259827  0.947227
2            0.201586        -0.129476       0.353704  0.007798
3            0.078699        -0.125110       0.175714  0.997149
4            0.206542        -0.003384       0.732917  0.026115
..
913          ...           ...           ...           ...
914          ...           ...           ...           ...
915          ...           ...           ...           ...
916          ...           ...           ...           ...
917          ...           ...           ...           ...

pval_hom  n_samples convergence_status
0  0.673613      2740      converged
1  0.098246      2740      converged
2  0.362925      2740      converged
3  0.741539      2740      converged
4  0.052142      2740      converged
..
913  0.079236      2740      converged
914  0.006697      2740      converged
915  0.0005526     2740      converged
916  0.967387      2740      converged
917  0.091233      2740      converged

```

[918 rows x 19 columns]

[25]: gwas_df_inv

```

[25]:   chrom      pos  variant_id ref_allele alt_allele alpha_value      coef  \
0      1  995669  rs3934834        C          T   -1.217846  0.048402
1      1  1011278  rs3737728        C          T    0.022409  0.040711
2      1  1020428  rs6687776        C          T   1.035492 -0.002149
3      1  1021403  rs9651273        G          A  -0.005678 -0.158919
4      1  1038818  rs4970405        A          G   0.301028 -0.023449
..
913    ...    ...    ...    ...    ...    ...    ...
914    1  9455825  rs6679380        T          C  -0.101376 -0.065436
915    1  9457890  rs7418365        A          G   0.213514  0.081464
916    1  9465199  rs10157778        G          A   0.330241  0.087808
917    1  9514086  rs12040542        T          C   3.043454  0.015465

```

```

917      1  9529628    rs7411273          C          T      0.086551  0.076165

      std_err      stat      pval conf_int_low conf_int_high n_samples \
0   0.051905  0.932518  0.351264     -0.053437      0.150241    1163
1   0.107922  0.377225  0.706075     -0.171033      0.252455    1175
2   0.061526 -0.034932  0.972140     -0.122863      0.118564    1175
3   0.112787 -1.409014  0.159098     -0.380209      0.062371    1175
4   0.201765 -0.116217  0.907500     -0.419312      0.372415    1174
..
913   0.099311 -0.658904  0.510088     -0.260284      0.129412    1175
914   0.096942  0.840331  0.400896     -0.108738      0.271666    1174
915   0.094291  0.931243  0.351921     -0.097191      0.272807    1175
916   0.023273  0.664494  0.506506     -0.030197      0.061127    1175
917   0.189315  0.402321  0.687522     -0.295271      0.447602    1174

      maf      eaf
0   0.143594  0.143594
1   0.283404  0.283404
2   0.150638  0.150638
3   0.280426  0.280426
4   0.096252  0.096252
..
913   0.265106  0.265106
914   0.352215  0.352215
915   0.374043  0.374043
916   0.119574  0.119574
917   0.163969  0.163969

```

[918 rows x 15 columns]

```

[26]: # 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: 82 alpha values outside range [-3, 3] excluded from plot

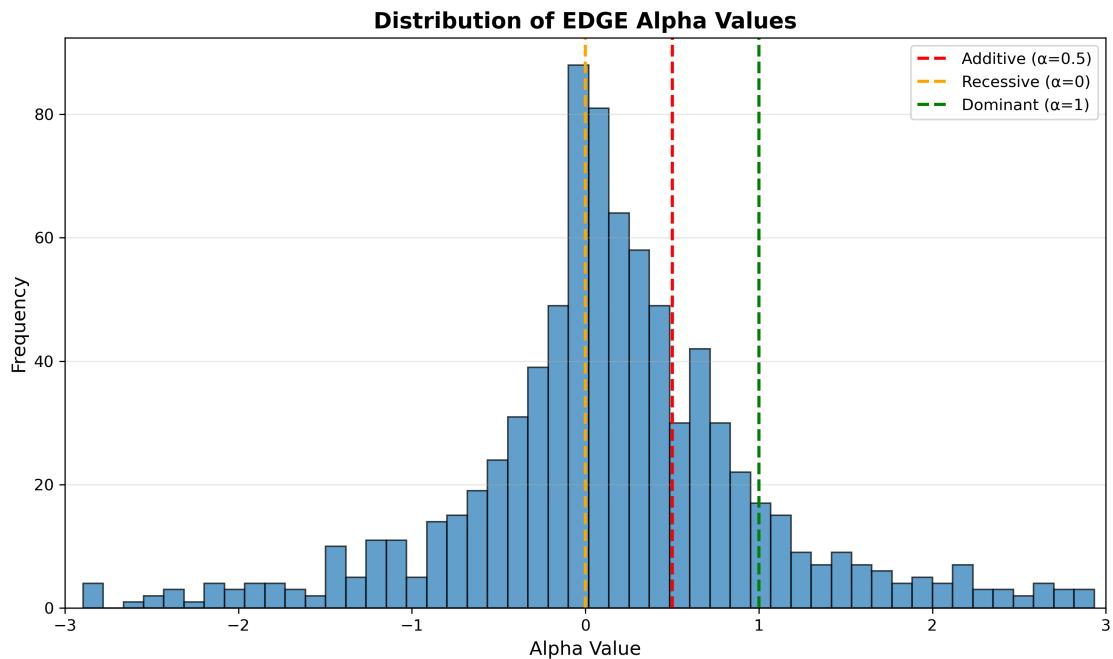
Alpha distribution plot saved to alpha.png

Total alpha values: 918

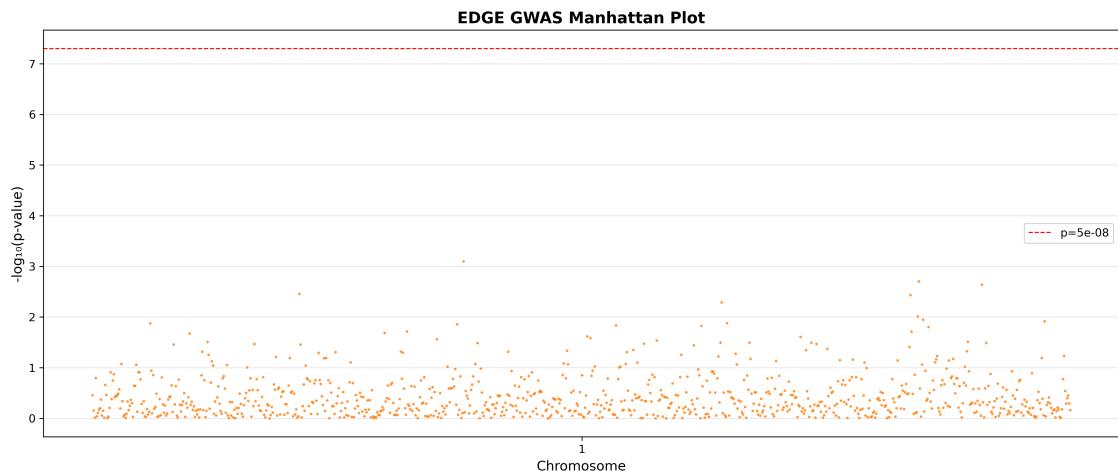
Values plotted (within [-3, 3]): 836

Alpha statistics:

Mean: -2.713
 Median: 0.098
 Std: 58.895
 Min: -1686.506
 Max: 40.958

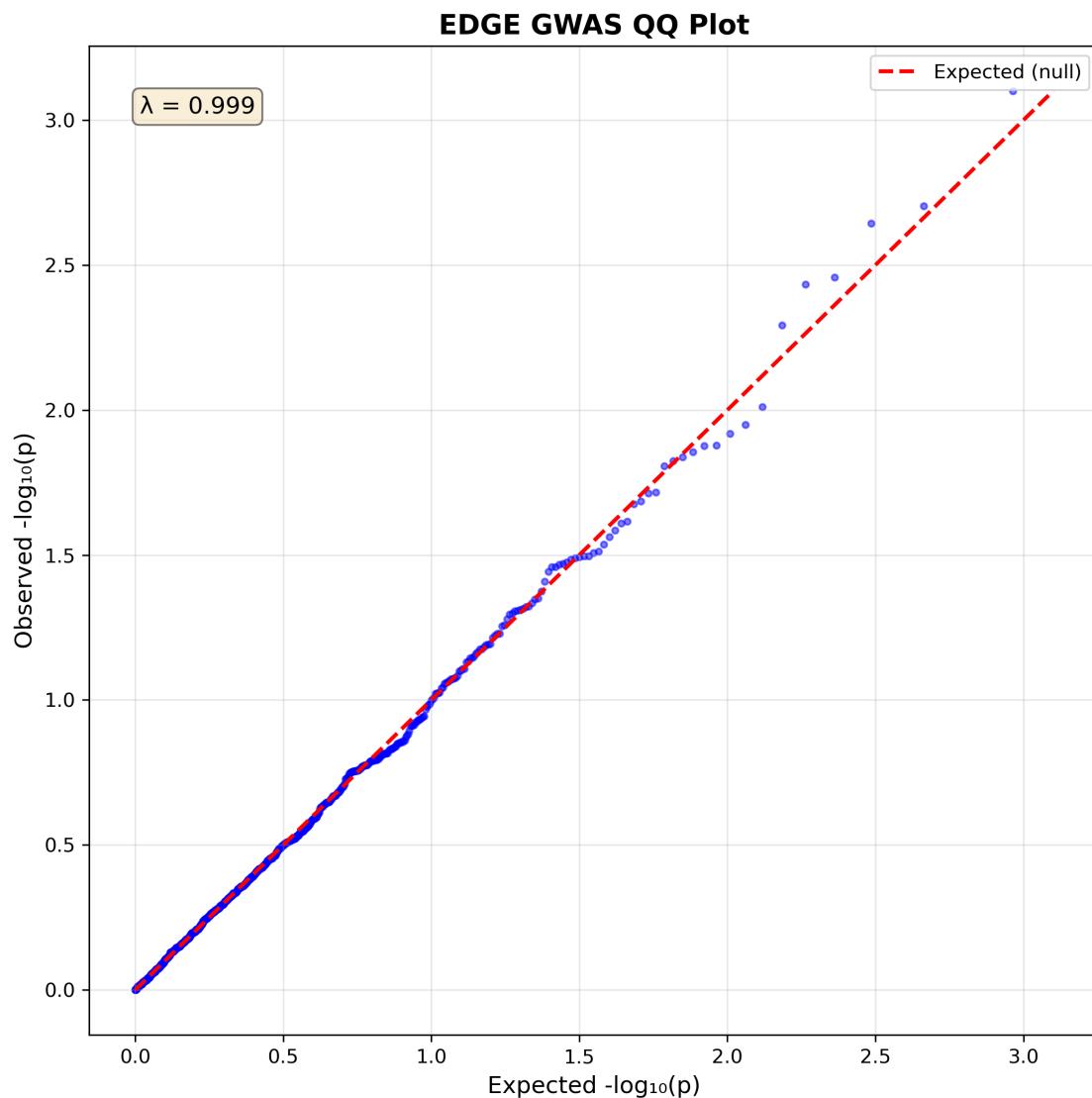


Manhattan plot saved to manhattan.png



QQ plot saved to qq.png
 Genomic inflation factor (): 0.999

Lambda GC: 0.999



[]: