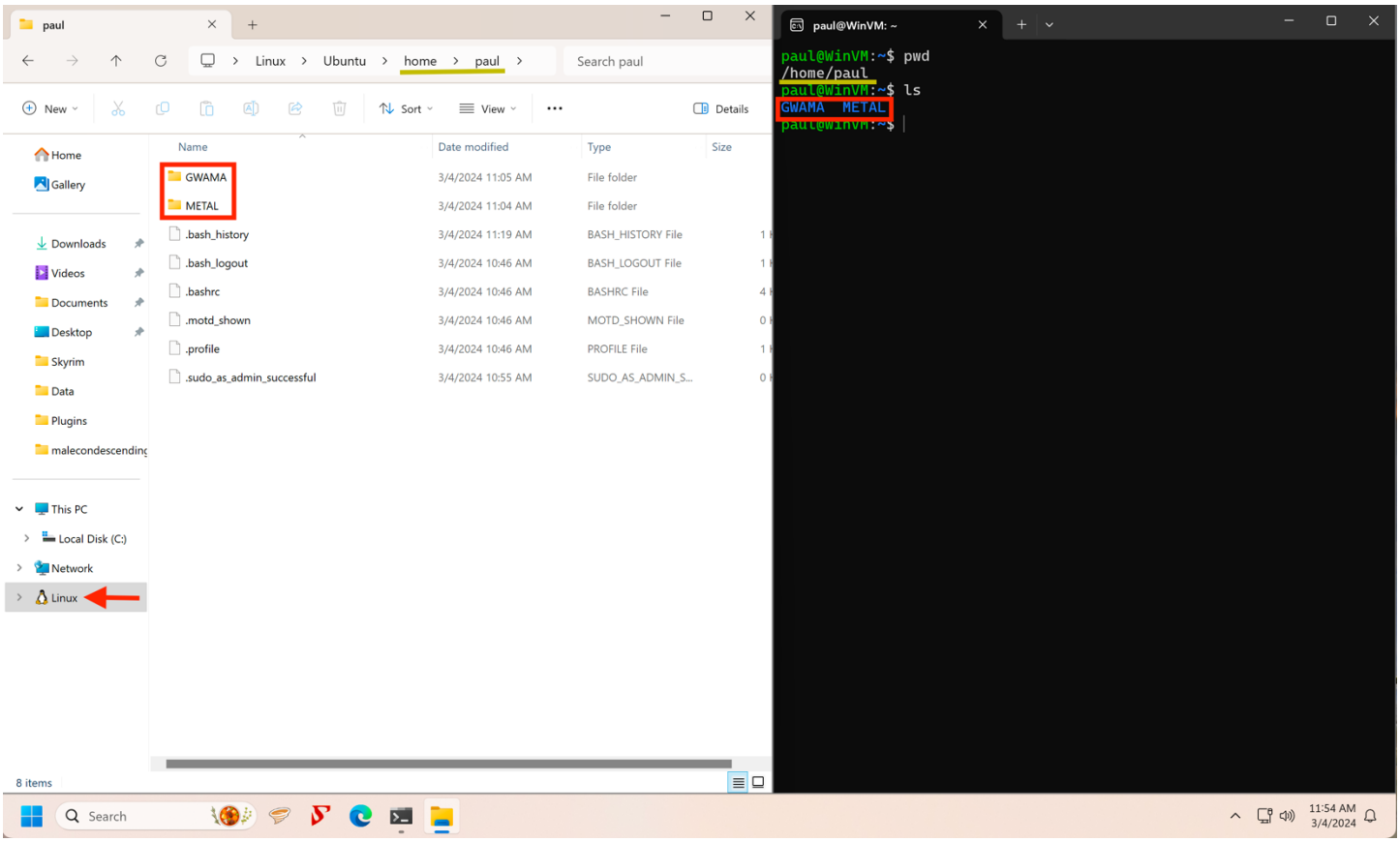
METAL documentation: https://genome.sph.umich.edu/wiki/METAL_Documentation
GWAMA documentation: https://genomics.ut.ee/en/tools

Navigate to your METAL and GWAMA folders (for Windows users):

1. **Open the File Explorer.** At the bottom of the sidebar you should see a "Linux" penguin icon since you installed WSL (Windows Subsystem for Linux). This is where your Linux files are stored. **Navigate to `Linux > Ubuntu > home > (your Linux username)`**. This is your home folder. Mine is named "paul".
   Inside, you should see the METAL & GWAMA folders you installed earlier.

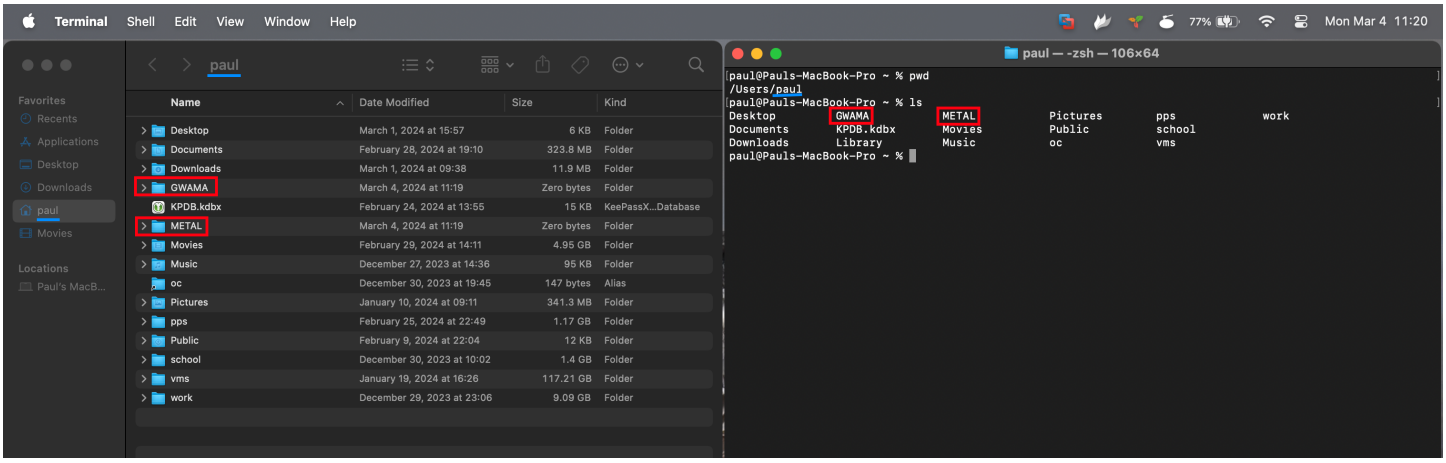
2. **Open WSL** (you can type "WSL" in the search bar to launch it).
   a. Type `pwd` to print the folder your terminal is currently looking at. It should be your home folder.
   b. Type `ls` to list files/folders in the current folder. You should see the same ones as in File Explorer. (Except those starting with a ".", you can ignore those.)



Navigate to your METAL & GWAMA folders (for Mac users):

1. **Open Finder** (the file explorer).
   Finder should place you in your home folder, named after your username (mine is “paul”).
   You should see the METAL & GWAMA folders you installed earlier.

2. **Open the Terminal app**. This is pre-installed on all Macs, but does not appear in your applications by default. To find it, use the spotlight search (⌘**+space**).
   a. Type **`pwd`** to print the folder your terminal is currently looking at. It should be your home folder.
   b. Type **`ls`** to list the files/folders in the current folder. You should see the same ones as in Finder.



METAL & GWAMA Demo (for all users)

1. **Download data and METAL code from https://paulhanson2000.github.io/.**
   Place the files in your home folder.
   Unzip the data files, but keep them in a folder called `data`.
   (If you want, you can unzip from the terminal by typing `unzip data.zip`)

2. **Inspect the data**: type `zless a_data_file.txt.gz`

```
paul@Pauls-MacBook-Pro ~ % zless data/DIAMANTE-EAS.txt.gz
```

   You should see something like this. Press `j` & `k` to scroll up and down.
   Press `q` to quit.

```
CHR position(b37) chrposID MARKERNAME EA NEA EAF BETA SE P N
2 11336 chr2:11336 rs113656530 c g 0.9256 0.0079 0.0251 0.7528 139705
2 11343 chr2:11343 rs114705059 t c 0.9971 -0.7277 0.6970 0.2964 139705
2 11357 chr2:11357 rs111385029 a g 0.9272 0.0078 0.0247 0.7518 139705
2 11486 chr2:11486 rs73138514 a g 0.9255 0.0079 0.0251 0.7528 139705
2 11594 chr2:11594 rs114792740 t g 0.0154 -0.2206 0.2658 0.4066 139705
2 11607 chr2:11607 rs73138516 t c 0.9257 0.0080 0.0246 0.7446 139705
2 11834 chr2:11834 rs73910134 a g 0.9271 0.0076 0.0245 0.756 139705
```

3. Inspect `metal_script.txt`, again by using `zless`, or using a text editor like Notepad.

```
# Meta-analysis weighted by standard error does not work well
# when different studies used very different transformations.
# In this case, some attempt was made to use similar trait
# transformation and you can request a standard error based
# analysis by uncommenting the following line:
# SCHEME   STDERR

# Usually, it is a good to apply genomic control to each
# input file. However, in this example, all the markers being
```

Running METAL on DIAMANTE data

1. From your home folder, type:
   `METAL/build/metal/metal metal_script.txt > metal.log`

2. Inspect `metal.log` for errors. (Use `zless` or a text editor to read it.)

```
################################################################################
## Executing meta-analysis ...
## Complete results will be stored in file 'METAANALYSIS1.TBL'
## Column descriptions will be stored in file 'METAANALYSIS1.TBL.info'
## Completed meta-analysis for 879496 markers!
## Smallest p-value is 2.787e-43 at marker 'rs1913657'
```

3. Inspect the results and description in `METAANALYSIS1.TBL` and `METAANALYSIS.TBL.info`

MarkerName	Allele1	Allele2	Weight	Zscore	P-value	Direction
rs16826408	t	c	432181.00	-1.781	0.07484	--+
rs908551	t	c	432181.00	1.891	0.05863	+++
rs12468764	a	g	432181.00	-1.854	0.06379	---
rs115329555	a	g	432181.00	-1.194	0.2325	--+
rs250924	t	c	432181.00	-1.319	0.1871	--+
rs74691018	t	c	432181.00	-0.380	0.7042	++-
rs58254652	t	c	432181.00	-0.566	0.5717	+--

Heterogeneity analysis with METAL

4. Edit `metal_script.txt` and change:
   `ANALYZE` to `ANALYZE HETEROGENEITY`

5. Run metal again the same way as in step 1.

6. Inspect the results. You should notice some new columns, like HetISq and HetPVal. When **in `zless`, press slash ("/") to search the file**. Look up the rs ID which was most significant and look at its HetPVal. Is there evidence that this variant is heterogenous?

Running GWAMA on DIAMANTE data

1. Create a file `gwama_input.txt` in your home folder, and put the paths to the DIAMANTE datasets inside like so:
   `data/DIAMANTE-EAS.txt.gz`
   `data/DIAMANTE-EUR.txt.gz`
   `data/DIAMANTE-SAS.txt.gz`

2. From your home folder, type:
   `GWAMA/GWAMA --quantitative --filelist gwama_input.txt`
   (`--quantitative` tells GWAMA that our data contains beta coefficients from a quantitative association analysis. Otherwise, GWAMA expects odds ratios from a case-control analysis.)

```
Quantitative trait (BETA+SE)
File gwama_input.txt contained 3 studies.
------------------------------------
Reading file: data/DIAMANTE-EAS.txt.gz
Strand column missing! Expecting always positive strand.
Marker count: 760565 Markers passing sanity check (and filters): 760559
Strand problems: 0 Wrong alleles: 0
Effect problems: 0 Multiple occurances: 6
------------------------------------
Reading file: data/DIAMANTE-EUR.txt.gz
Strand column missing! Expecting always positive strand.
Marker count: 879506 Markers passing sanity check (and filters): 879496
Strand problems: 59416 Wrong alleles: 0
Effect problems: 0 Multiple occurances: 10
------------------------------------
Reading file: data/DIAMANTE-SAS.txt.gz
Strand column missing! Expecting always positive strand.
Marker count: 874777 Markers passing sanity check (and filters): 874768
Strand problems: 694 Wrong alleles: 0
Effect problems: 0 Multiple occurances: 9
------------------------------------
Preparing output...
------------------------------------
GWAMA program finished current job successfully!
Please check gwama.log.out for full information.
Analysis finished.
```

3. Examine `gwama.log.out` and `gwama.err.out` for information about any errors.

4. Inspect the results in `gwama.out`.
   Do these results seem to agree with METAL's? Check rs1913657, which was METAL's top variant.

Random-effects analysis with GWAMA

Our datasets are from different ancestries and many of our top variants had high heterogeneity, so adjusting for random effects is appropriate.

5. Run GAWMA again as in step 1, except adding the `--random` option this time.
   If you don't want to overwrite your old results, add `--output <output_name>` too.

6. Inspect the results.
   Now that we're adjusting for random effects, how have the results changed?

Running METAL on MAGIC data

1. Let's try a different dataset.
   Make a copy of `metal_script.txt` and edit it to use the other three data files instead. I.e.
   `DGI_three_regions.txt.gz`,
   `magic_SARDINIA.txt.gz`,
   and `MAGIC_FUSION_results.txt.gz`

2. Run METAL as before.
   rs560887 should be the most significant variant.

Allele frequencies

GWAMA detected allele frequency discrepancies automatically, but METAL can also output some information to help diagnose these issues.

3. Uncomment these lines in `metal_script.txt`:
   `AVERAGEFREQ ON`
   `MINMAXFREQ ON`

4. Run METAL again, and take note of the new FreqSE, MinFreq, and MaxFreq columns in the results. Based on this, do you think the significant variant rs560887 might have strand issues?

Inverse-variance weighting

Instead of weighing our studies by their sample size, we can instead weigh them by the inverse of their variance for their effect estimates. This way, studies with a more precise estimate for a certain variant will be weighed more heavily.

5. Uncomment this line in `metal_script.txt`:
   `SCHEME STDERR`

6. Run METAL and find rs6716049 in the results.
   Why do you think METAL decided to give it a positive effect estimate despite two of the three input files reporting a negative effect? What is the standard error (SE) in each of the input files?

Genomic correction

7. Below are the genomic control lambdas computed on the full studies In `metal_script.txt`, add `GENOMICCONTROL <number>` before each `PROCESS` command.
   DGI_three_regions: 1.044 magic_SARDINIA: 1.061 MAGIC_FUSION_Results: 1.008

8. Run METAL. Is top variant rs560887 as significant as it was before? Why?

Naïve genomic correction

9. Comment out the `GENOMICCONTROL <number>` lines you just added, and instead uncomment `GENOMICCONTROL ON` near the top of the script.
10. Run METAL, and compare
    Why is it not a good idea to use `GENOMICCONTROL ON` in our case? (Hint: the MAGIC data we're using is not genome-wide data, it is a subset of the most significant regions.)