

# Assignment 3

[Re-submit Assignment](#)

**Due** 8 Aug by 23:00      **Points** 20      **Submitting** a file upload  
**Available** after 6 Aug at 12:00

## Assignment 3 – Due date: Aug 8, 11pm

Log in to Marvin.

In your home directory, create a directory called “Assignment\_3”.

In Assignment\_3, create three other directories called CODE, INPUT and OUTPUT.

Copy the following files from /tutorials/Assignment\_3/ to your INPUT directory.

SWB\_Full\_chr1.txt

SWB\_Full\_chr2.txt

...

SWB\_Full\_chr22.txt

There should now be 22 files in your Assignment\_3/INPUT directory.

Each file contains data on one chromosome. In this assignment, you will create a bash script that adds the chromosome number as a column to each file and merges the files. The final output will be the same as SWB\_Full.txt (i.e. the data you used for Assignment 1 and obtained at the end of Assignment 2.)

*The files look like this:*

MarkerName	POS	A1	A2	EAF	Beta	SE	Pval
rs6587766	57708088		T	C	0.04104	0.047	0.009
rs12143280	211547522		T	C	0.94403	0.038	0.008
rs1394396	57753150		A	G	0.03358	0.039	0.009
rs12062152	234267329		A	G	0.03918	0.035	0.008
rs12070124	234265547		T	C	0.03918	0.035	0.008
rs10923025	88579746		T	G	0.9459	-0.029	0.007
rs6686261	234266626		A	G	0.96082	-0.035	0.008
rs12096772	234265651		A	C	0.96082	-0.035	0.008
rs6671630	234266483		A	G	0.03918	0.034	0.008

- As a first step, create a text file named “YourName\_sh”. This will be a bash script, so add the relevant first line!
- The first thing you will do in this script is to create a function named “add\_chr” that will take as input the chromosome number. This function will first print the sentence “Adding chromosome number to SWB\_Full\_chrX.txt” and then add the chromosome number as second column to a file. The output file be

written into the OUTPUT directory and will have the name “SWB\_Full\_chrX.txt”, where X is the chromosome number.

3. Apply this function to all files using a for loop indexing over chromosomes 1 through 22. Run these (22) processes in the background.

-

**Note :** The **wait** command makes the shell wait until all background processes are finished. Use the **wait** command after your for loop to make sure all the background processes are finished before continuing with the next steps.

The files should now look like this:

SWB\_Full\_chr1\_chradded.txt

MarkerName	CHR	POS	A1	A2	EAF	Beta	SE	Pval	
rs6587766	1	57708088	T	C		0.04104	0.047	0.009	2.515
rs12143280	1	211547522	T	C		0.94403	0.038	0.008	6.464
rs1394396	1	57753150	A	G		0.03358	0.039	0.009	1.03e
rs12062152	1	234267329	A	G		0.03918	0.035	0.008	2.097
rs12070124	1	234265547	T	C		0.03918	0.035	0.008	3.049
rs10923025	1	88579746	T	G		0.9459	-0.029	0.007	3.052
rs6686261	1	234266626	A	G		0.96082	-0.035	0.008	3.092
rs12096772	1	234265651	A	C		0.96082	-0.035	0.008	3.099
rs6671630	1	234266483	A	G		0.03918	0.034	0.008	3.577

SWB\_Full\_chr2\_chradded.txt

MarkerName	CHR	POS	A1	A2	EAF	Beta	SE	Pval	
rs13387164	2	154144551	A	G		0.09142	-0.023	0.005	5.981e
rs11691770	2	168633442	T	G		0.03918	0.037	0.008	8.522e
rs10172421	2	204238402	A	C		0.1045	-0.021	0.005	1.107e
rs7584895	2	29229016	T	C		0.5187	-0.014	0.003	1.154e
rs12474324	2	19346130	A	G		0.94216	0.027	0.006	1.178e
rs12992085	2	168649212	A	G		0.94776	-0.033	0.008	1.325e
rs6712882	2	29233954	A	G		0.4795	0.014	0.003	1.389e
rs12995715	2	63676678	C	G		0.2388	-0.016	0.004	1.461e
rs12464722	2	19341657	T	C		0.06716	-0.026	0.006	1.56e-

...

and so on.

4. After the wait command , add a command that will display the sentence “Chromosome numbers added.” on the screen.
5. Display "Catenating per chromosome files." on the screen. Then, catenate all files (“SWB\_Full\_chrX.txt”, where  $X \in \{1, 2, \dots, 22\}$ ) using the **cat** command so that you have one file (in the OUTPUT directory) containing all chromosomes. Name the file “SWB\_Full\_chradded.txt”
6. The “SWB\_Full\_chradded.txt” file contains multiple header lines since all per chromosome files that you merged together contained a header. Remove the extra headers so that only the header that’s at the beginning of the file remains.
7. Display “Script over.” on the screen.

Congrats! You have written your first bash script! 😊

Upload your script ("**YourName**\_assignment3.sh") to Canvas.