A primer in Human Cardiovascular Genetics

dr. Sander W. van der Laan | sw
vanderlaan

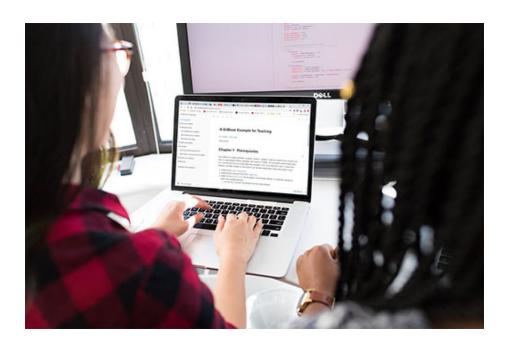
Last update: 2022-03-24

Contents

1	About this primer				
	1.1	Introduction	5		
	1.2	Background reading	6		
	1.3	Meet the Team	6		
2	Prerequisites				
	2.1	Linux, macOS, and Windows	9		
	2.2	Programs you need	9		
	2.3	The Terminal	10		
	2.4	Installing some R packages	12		
3	Basics of a Genome-Wide Association Study (GWAS)				
	3.1	Section 1	13		
	3.2	Section 2	13		
	3.3	Section 3	14		
4	WTCCC1: a GWAS on coronary artery disease (CAD)				
	4.1	Section 1	15		
	4.2	Section 2	15		
	4.3	Section 3	16		
5	Post-GWAS Analyses				
	5.1	Section 1	17		
	5.2	Section 2	17		
	5.3	Section 3	18		

6	Con	Conditional analysis 19					
	6.1	Section 1	19				
	6.2	Section 2	19				
	6.3	Section 3	20				
7	Statistical finemapping						
	7.1	Section 1	21				
	7.2	Section 2	21				
	7.3	Section 3	22				
8	Functional Mapping and Annotation of GWAS						
	8.1	Section 1	23				
	8.2	Section 2	23				
	8.3	Section 3	24				
9	Phenome-Wide Association Study (PheWAS)						
	9.1	Section 1	25				
	9.2	Section 2	25				
	9.3	Section 3	26				
10 Mendelian Randomization (MR) 2'							
	10.1	Section 1	27				
	10.2	Section 2	27				
	10.3	Section 3	28				
11 Mendelian Randomization (MR) 2							
	11.1	Section 1	29				
	11.2	Section 2	29				
	11.3	Section 3	30				
12 License your GitBook 31							
13 License your GitBook							

About this primer



1.1 Introduction

Welcome to the *A primer in Human Cardiovascular Genetics* as part of the **Genetic Epidemiology** course. In the next few days we will use this GitBook to perform quality control (QC), executing a genome-wide association study (GWAS), annotating the GWAS results, and performing further downstream analyses. We will use data from the first release of the *Welcome Trust Case-Control Consortium (WTCCC)* and focus on coronary artery disease (CAD).

Unfortunately, during this course there is no time to perform imputation, but I will provide some pointers during the course as to how to do this with minimal coding/scripting experience. And I will add a whole chapter on imputation to a future version.

1.2 Background reading

Part of this is based on four great Nature Protocols from the Zondervan group at the Wellcome Center Human Genetics.

- Zondervan KT et al. Designing candidate gene and genome-wide casecontrol association studies. Nat Protoc 2007.
- 2. Pettersson FH et al. Marker selection for genetic case-control association studies. Nat Protoc 2009.
- Anderson CA et al. Data QC in genetic case-control association studies. Nat Protoc 2010.
- Clarke GM et al. Basic statistical analysis in genetic case-control studies. Nat Protoc 2011.

An update on the community standards of QC for GWAS can be found here:

1. Laurie CC et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol 2010.

With respect to imputation you should also get familiar with the following two works:

- 1. Marchini, J. and Howie, B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010
- 2. de Bakker PIW et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 2008.
- 3. Winkler TW et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc 2014.

Importantly, follow this practical, but also review the questions asked in the presentation (on the screen or in the Dropbox) and try to answer them.

1.3 Meet the Team

We work with a team of enthusiastic lecturers with experience in bioinformatics, GWAS, genetic analyses, Mendelian randomization, and epidemiology. This year the team consists of:



Sander W. van der Laan Assistant professor Course coordinators.w. vanderlaan-2@ umcutrecht.nl | swvanderlaan



Charlotte
Onland-Moret
Associate
ProfessorN.C.
Onland@
umcutrecht.nl
Jessica van
Setten Assistant
professorj.
vansetten@
umcutrecht.nl |
j_vansetten

Getting started

Ready to start? Your first point of action is to prepare your system for this course (Chapter 2).

Prerequisites

2.1 Linux, macOS, and Windows

Most programs made to execute genetic studies are developed for the Unix environment, e.g. Linux and macOS. Therefore, they may not work as intended on window's environment. Fortunately, Windows now allow users to install a linux subsystem within Windows 10 and you can find the detail guide here. However, we highly recommend to 1) either install a linux subsystem on your Windows computer (e.g. a virtual machine with Ubuntu could work), or 2) switch to macOS in combination with homebrew as this will give you advantage of a powerful computer with a userfriendly interface and all the flexibility to use Unix-based programs for your work.

For this practical we use a Windows laptop with Ubuntu on a Virtual Machine. Therefore every command is intended for Linux/macOS.

2.2 Programs you need

You need few programs for this practical or for your (future) genetic epidemiology work for that matter.

Program	Link	Description
PLINK	[https:/	/www.cog- genomics.org/plink2/](https://www.cog- genomics.org/plink2/){target="_blank"} PLINK is a free, open-source genetic analysis toolset, designed to perform a range of basic data parsing and quality control, as well as basic and large-scale analyses in a computationally efficient manner.
R	[https:/	/cran.r-project.org](https://cran.r-project.org/){target="_blank"} Progam to perform statistical analysis and visualisations.
RStudio	[https:/	/www.rstudio.com](https: //www.rstudio.com){target="_blank"} A userfriendly R-wrap-around for code editing, debugging, analyses, and visualisation

Table 1: Programs needed for genetic epidemiology.

All genetic analyses can be done in PLINK, even on your laptop, but with large datasets, e.g. UK-biobank size, it's better to switch to a high-performance computing cluster like we have at the Uithof.

Nowadays, a lot of people also use programs like SNPTEST, BOLT-LMM, and GCTA as alternatives to run GWAS and downstream analyses, e.g. heritability estimation, Fst-calculation, etc.

Mendelian randomisation can be done either with the SMR or GSMR function from GCTA or with R-packages, like TwoSampleMR.

2.3 The Terminal

For all the above programs, except RStudio, you will need the Terminal. This comes with every major operating system. And you will (start to) make your own scripts. The benefit of using scripts is that each step is clearly layed out and easily replicable, thus allows for greater reproducibility, easier troubleshooting and can be run on high-performance computer clusters.

Download the data you need to your Desktop: LINK.

Now open the terminal, it should be on the left in the toolbar as a little black computer-monitor-like icon. Mac uses can type command + space and type terminal, a terminal screen should open.

From now on we will use little code blocks like the example to indicate a code you should type/copy-paste and hit enter. If a code is followed by a comment, it is indicated by a # - you don't need to copy-paste and execute this.

```
CODE BLOCK
```

```
CODE BLOCK # some comment here
```

You can navigate around the computer through the terminal by typing cd <path>; cd stands for "change directory" and means "some_file_directory_you_want_to_go_to".

```
# For Linux/macOS Users
cd ~ # will bring you to your home directory
cd ../ # will bring you to the parent directory (up one level)
cd XXX # will bring you to the XXX directory
```

Let's navigate to the folder you just downloaded.

```
cd ~/Desktop/practical
```

Let's check out what is inside the directory, by listing (ls) its contents.

```
ls -lh
```

```
# For Linux/macOS Users
ls -l # shows files as list
ls -lh # shows files as list with human readable format
ls -lt # shows the files as list sorted by time edited
ls -lS # shows the files as list sorted by size
```

Adding the flags -lh will get you the contents of a directory in a list (-l) and make the size 'human-readable' (-h).

You can also count the number of files.

```
ls | wc -1
```

2.4 Installing some R packages

I tested this Virtual Machine and everything should be fine, except some libraries weren't there. We need to in stall them.

To be able to install certain **r**-packages, we need to install some Linux (Ubuntu) software. Type the following:

```
sudo apt-get install libcurl4 libcurl4-openssl-dev -y
sudo apt-get install libssl-dev
```

Now close the terminal window - really making sure that the terminal-program has quit.

Open a new terminal window and open r by simply typing R. You should install the following packages, and then you're good to go!

```
install.packages(c("httr", "usethis", "data.table", "devtools", "qqman", "CMplot", "tible", "pidevtools::install_github("kassambara/ggpubr")
library("ggpubr")
```

All in all this may take some time, good moment to relax, review your notes, stretch your legs, or take a coffee.

Basics of a Genome-Wide Association Study (GWAS)

3.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.2 Section 2

WTCCC1: a GWAS on coronary artery disease (CAD)

4.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4.2 Section 2

esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?

4.3 Section 3

Post-GWAS Analyses

5.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.2 Section 2

Conditional analysis

6.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6.2 Section 2

Statistical finemapping

7.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

7.2 Section 2

Functional Mapping and Annotation of GWAS

8.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

8.2 Section 2

Phenome-Wide Association Study (PheWAS)

9.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

9.2 Section 2

Mendelian Randomization (MR)

10.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

10.2 Section 2

Mendelian Randomization (MR)

11.1 Section 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

11.2 Section 2

License your GitBook

In the spirit of Open Science, it is good to think about making your course materials Open Source. That means that other people can use them. In principle, if you publish materials online without license information, you hold the copyright to those materials. If you want them to be Open Source, you must include a license. It is not always obvious what license to choose.

The Creative Commons licenses are typically suitable for course materials. This GitBook, for example, is licensed under CC-BY 4.0. That means you can use and remix it as you like, but you must credit the original source.

If your project is more focused on software or source code, consider using the GNU GPL v3 license instead.

You can find more information about the Creative Commons Licenses here. Specific licenses that might be useful are:

- CC0 ("No Rights Reserved"), everybody can do what they want with your work.
- CC-BY 4.0 ("Attribution"), everybody can do what they want with your work, but they must credit you. Note that this license may not be suitable for software or source code!

For compatibility between CC and GNU licenses, see this FAQ.

License your GitBook

In the spirit of Open Science, it is good to think about making your course materials Open Source. That means that other people can use them. In principle, if you publish materials online without license information, you hold the copyright to those materials. If you want them to be Open Source, you must include a license. It is not always obvious what license to choose.

The Creative Commons licenses are typically suitable for course materials. This GitBook, for example, is licensed under CC-BY 4.0. That means you can use and remix it as you like, but you must credit the original source.

If your project is more focused on software or source code, consider using the GNU GPL v3 license instead.

You can find more information about the Creative Commons Licenses here. Specific licenses that might be useful are:

- CC0 ("No Rights Reserved"), everybody can do what they want with your work.
- CC-BY 4.0 ("Attribution"), everybody can do what they want with your work, but they must credit you. Note that this license may not be suitable for software or source code!

For compatibility between CC and GNU licenses, see this FAQ.