

Laboratory book

Dylan Lawless

École polytechnique fédérale de Lausanne



Start Aug 2020

Publication Statement

This lab book concerns two projects are explicitly labelled by color as **projectA** and **projectB** as well as several other non-specified projects, and day-to-day notes and ideas.

The author confirms that the work submitted is their own, except where work which has formed part of jointly-authored publications has been included. The author confirms that appropriate credit has been given within this document where reference has been made to the work of others.

This copy has been produced on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgements.

Contents

Publication Statement	iii
1 Lab book 2020	1
Introduction	1
Sat, 8 Aug 2020 projectA	1
Mon, 10 Aug 2020 projectB	2
Fri 14 Aug projectB	2
Processing viral data	2
Fri, 27 Aug Lab meeting	3
Second gen P-value	3
2 Methods and tools	7
Imputation	7
Burden test / SKAT	7
.	7
Go to End	7
Bibliography	8
Index	9

Chapter 1

Lab book 2020

Introduction

Stuff about eigenvectors and eigenvalues which are being indexed at the end of this book. eigenvector
eigenvalue

The current classification of primary immunodeficiencies (PIDs) was compiled by the Expert Committee of the International Union of Immunological Societies [1] in *The Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity*. Over approximately 50 years the list of inborn errors of immunity has grown to over 350 disorders. I use vim snippets in my .vimrc: for producing formatting quickly including, bold face, italics, headings, and code.

```
1 Computer code like this is writtin using the command. begin{\lstlisting} and  
end{\lstlisting}.
```

Footnote example¹.

This is a margin
note using the
geometry
package, set at
3cm vertical
offset to the line
it is typesetted.

Sat, 8 Aug 2020 **projectA**

I will add margin note after some placeholder text. Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

¹ Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

The R script
used is in "viral
sequence / phylo
/ tree.r"

Mon, 10 Aug 2020 [projectB](#)

Working on the phylogenetic trees. I am producing phylogenetic trees and then splitting based on the obvious clade groups, add color, and add all labels. I used ggtree and ape for this work.

<https://yulab-smu.github.io/treedata-book>

<https://4va.github.io/biodatasci/r-ggtree.html>

phylogenetic tree
ggtree
ape

Index: phylogenetic tree, ggtree, ape.

Fri 14 Aug [projectB](#)

Here are some specific notes on a recent task.

Processing viral data

For [projectB](#) I used ncbi-tools-bin to convert to fasta format (note the software "Sequin" does the same job but it GUI only). <https://www.huge-man-linux.net/man1/asn2fsa.html>

ncbi-tools

```
1 sudo apt-get install ncbi-tools-bin
2 for file in rsv.*.sqn; do asn2fsa -a a -i $file -v $file.fa; done
3 dir: ./fa
```

Each viral genome has 11 segments that are sequenced. With awk, I split each fasta to a new file and labelled with the read name. i.e. segment 1 for all samples is in the file seg_1, etc. used \$split.sh dir: ./seg redo this step and split based on whole line name per fasta (remove spaces) I used the macOS command line version of muscle to align, printing out clustal format

```
1 $ for file in seg_*; do ~/Desktop/muscle3.8.31_i86darwin64 -in $file -clw -
   out clw.$file; done
2 dir: ./clw
```

I had ncbi-tools-bin installed since it runs on Debian. I will try on my local macOS. https://ftp.ncbi.nlm.nih.gov/toolbox/ncbi_tools/

```

1 mv mac.asn2fsa ./tools/
2 mkdir rsv_viral_sequence_local
3 sync sqn dir from scitas
4
5 To extract the nucleotide from sqn.
6 l.export_nucleotide.sh
7 #!/bin/bash
8 # output options -v for amino, -o for nucleotide.
9 cd sqn
10 for file in rsv.*.sqn;
11     do ~/tools/mac.asn2fsa -a a -i $file -o ${file%.*}.nt.fa;
12 done
13 mv /*.nt.fa ../nt

```

Then make a simple header.

```

1 $ cat 2.simplify_header.sh
2 #!/bin/bash
3 cd ../nt
4 for file in *.nt.fa;
5     do awk -F" " ' />/{ $0=$1 } 1' $file > ${file%.*}.nt_smp.fa ;
6     done
7
8 mv /*.nt_smp.fa ../nt_simple/
9
10 move all to post/inspire/viral_sequence/rsv_viral_seq/

```

Fri, 27 Aug Lab meeting

Second gen P-value

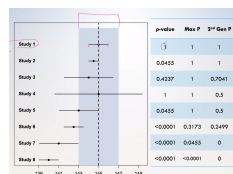


Figure 1.1: Second gen P val.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed

accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Chapter 2

Methods and tools

Imputation

Genipe Imputation pipeline tutorial <http://pgxcentre.github.io/genipe/index.html>

Imputation

Burden test / SKAT

SKAT Lee et al. [2] and SKAT-O Lee et al. [3] (I believe this is correct but check). Author homepage <https://www.hsph.harvard.edu/skat/> Some info on the methods from Golden helix https://doc.goldenhelix.com/SVS/tutorials/variant_analysis/skatAndSkato.html. Package in R <https://cran.r-project.org/web/packages/SKAT/SKAT.pdf> A different SKAT paper on burden testing <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6174288/>.

Go to End

Bibliography

- [1] Capucine Picard, H Bobby Gaspar, Waleed Al-Herz, Aziz Bousfiha, Jean-Laurent Casanova, Talal Chatila, Yanick J Crow, Charlotte Cunningham-Rundles, Amos Etzioni, Jose Luis Franco, et al. International union of immunological societies: 2017 primary immunodeficiency diseases committee report on inborn errors of immunity. *Journal of clinical immunology*, 38(1):96–128, 2018.
- [2] S. Lee, M. C. Wu, and X. Lin. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*, 13(4):762–775, Jun 2012. ISSN 1468-4357. doi: 10.1093/biostatistics/kxs014. URL <http://dx.doi.org/10.1093/biostatistics/kxs014>.

- [3] Seunggeun Lee, Mary J. Emond, Michael J. Bamshad, Kathleen C. Barnes, Mark J. Rieder, Deborah A. Nickerson, David C. Christiani, Mark M. Wurfel, and Xihong Lin. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *The American Journal of Human Genetics*, 91(2):224 – 237, 2012. ISSN 0002-9297. doi: <https://doi.org/10.1016/j.ajhg.2012.06.007>. URL <http://www.sciencedirect.com/science/article/pii/S0002929712003163>.

Index

ape, [2](#)

eigenvalue, [1](#)

eigenvector, [1](#)

ggtree, [2](#)

Imputation, [7](#)

ncbi-tools, [2](#)

phylogenetic tree, [2](#)

SKAT, [7](#)