References

Albert Selva-O'Callaghan, a. M.-C.-F.-A.-H.-J. (2018). Statin-induced myalgia and myositis: an update on pathogenesis and clinical recommendations. *Expert Rev Clin Immunol*, 215–224.

Barnett GC, Thompson D, Fachal L, et. al. (2014). A genome-wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity. *Radiotherapy and Oncology*, 178-185.

Donald Clark, D. B. (2016). Clinical Characteristics Associated With Complete Statin Intolerance: The Cleveland Clinic Experience. *Journal of the American College of Cardiology*.

Ruaño G, W. A. (2010). Mechanisms of statin-induced myalgia assessed by physiogenomic associations. *Atherosclerosis*, 451-6.

Satish Ramkumar, A. R. (2016). Statin Therapy: Review of Safety and Potential Side Effects. *Acta Cardiol Sin*, 631-639.

SEARCH Collaborative Group, Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R. (2008). SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med.* , 789-99.

Shilpa Bhardwaj, S. S. (2013). Muscular effects of statins in the elderly female: a review. *Clinical Interventions in Aging*, 47–59.

Understanding Pharmacogenomics. (2018, 02). Retrieved from Cancer.Net: https://www.cancer.net/navigating-cancer-care/how-cancer-treated/personalized-and-targeted-therapies/understanding-pharmacogenomics

What are genome-wide association studies? (n.d.). Retrieved from Genetics Home Reference: https://ghr.nlm.nih.gov/primer/genomicresearch/gwastudies.

What is pharmacogenomics? (n.d.). Retrieved from Genetics Home Reference: https://ghr.nlm.nih.gov/primer/genomicresearch/pharmacogenomics

Steinbaum, S. R. (2018, July 17). *Side Effects of Cholesterol-Lowering Statin Drugs.* Retrieved from WebMD Medical Reference: https://www.webmd.com/cholesterol-management/side-effects-of-statin-drugs#2

Elam MB, M. G. (2017). Patients experiencing statin-induced myalgia exhibit a unique program of skeletal muscle gene expression following statin re-challenge. *PLoS ONE 12(8): e0181308*, 1-26.

Thomas SELLKE, M. J. BAYARRI, and James O. BERGER, Calibration of p Values for Testing Precise Null Hypotheses, The American Statistician, February 2001, Vol. 55, No. 1

Macwan, J. (2013). Model-Based Approaches to Characterize Clinical Pharmacokinetics of Atorvastatin and Rosuvastatin in Disease State. (Open Access Dissertations). University of Rhode Island, Kingston, USA.

Genetic Science Learning Center. (2016, February 1) Making SNPs Make Sense. Retrieved July 01, 2019, from https://learn.genetics.utah.edu/content/precision/snips/

Susa ST, Preuss CV. Drug Metabolism. [Updated 2019 Jun 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2019 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK442023/

Canner, D 2010, Reductase pathway, viewed 12 August 2019, Available from: https://proteopedia.org/wiki/index.php/Image:HMG-CoA_reductase_pathway.png.

National Center for Biotechnology Information. PubChem Database. Atorvastatin, CID=60823, https://pubchem.ncbi.nlm.nih.gov/compound/Atorvastatin (accessed on Aug. 12, 2019).

M. Whirl-Carrillo, E.M. McDonagh, J. M. Hebert, L. Gong, K. Sangkuhl, C.F. Thorn, R.B. Altman and T.E. Klein. "Pharmacogenomics Knowledge for Personalized Medicine" Clinical Pharmacology & Therapeutics (2012) 92(4): 414-417.

ADR1 - Regulatory protein . (n.d.). From UniProt: https://www.uniprot.org/uniprot/P07248#function

Cathryn M. Lewis, J. K. (2012). Introduction to Genetic Association Studies. *Cold Spring Harbor Laboratory Press*, 297-305. From Cold Spring Harbor Laboratory Press: http://cshprotocols.cshlp.org/content/2012/3/pdb.top068163.full

Christian P. Strassburg, S. K. (2008). Variability and Function of Family 1 Uridine-5'-Diphosphate Glucuronosyltransferases (UGT1A). *Critical Reviews in Clinical Laboratory Sciences*, 485-530.
Data Analysis - Pearson's Correlation Coefficient. (n.d.). From University of the West of England: http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1442

Galarnyk, M. (2018, September 12). *Understanding Boxplots*. From Towards Data Science: https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51

How to Correctly Interpret P-values. (2014, April 17). From The Minitab Blog: https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-correctly-interpret-p-values

Joanna Somers, T. P. (2013). A perspective on mammalian upstream open reading frame function. *The International Journal of Biochemistry & Cell Biology Volume 45*, 1690-1700.

Julia M. Barbarino, C. E. (2014). PharmGKB summary: very important pharmacogene information. *Pharmacogenetics and Genomics*, 177–183.

Kenton, W. (2019, June 25). *T-test definition*. From Investopedia: https://www.investopedia.com/terms/t/t-test.asp

Lynch, S. S. (2017, March). *Overview of Response to Drugs*. From MSD Manuals Staff: https://www.msdmanuals.com/en-sg/home/drugs/factors-affecting-response-to-drugs/overview-of-response-to-drugs

M.J.De Groot, D. S. (2007). Molecular Modeling and Quantitative Structure–Activity Relationship of Substrates and Inhibitors of Drug Metabolism Enzymes. *Comprehensive Medicinal Chemistry II Volume 5*, 809-825.

Pablo Cingolani, F. C. (2018, January). *Variant annotations in VCF format.* From SnpEff: http://snpeff.sourceforge.net/VCFannotationformat_v1.0.pdf

Ramji DP, T. M. (1991). The transcription factor LF-A1 interacts with a bipartite recognition sequence in the promoter regions of several liver-specific genes. *Nucleic Acids Res*, 1139-46.

Roland, J. (2019, February 11). *Which Statin Is Best for People with Diabetes?* From Healthline: https://www.healthline.com/health/statins-for-diabetics-which-is-best

Single-nucleotide polymorphism. (2018, November 30). From International Society of Genetic Genealogy Wiki: https://isogg.org/wiki/Single-nucleotide_polymorphism

Strassburg CP, K. S. (2008). Variability and function of family 1 uridine-5'-diphosphate glucuronosyltransferases (UGT1A). *Crit Rev Clin Lab Sci.*, 485-530.

Szumilas, M. (2010). Explaining Odds Ratios. *J Can Acad Child Adolesc Psychiatry*, 227-229.

T-box transcription factor TBX22. (n.d.). From UniProt: https://www.uniprot.org/uniprot/Q9Y458#function

T-box transcription factor TBX8. (n.d.). From UniProt: https://www.uniprot.org/uniprot/Q22292

Tests, B. (2014, August). *Investigating myalgia in patients taking statins.* From bpac better medicine: https://bpac.org.nz/BT/2014/August/docs/BT24-myalgia.pdf

Yolanda Smith, B. (2019, February 27). *Drug Metabolism*. From News Medical Life Sciences : https://www.news-medical.net/health/Drug-Metabolism.aspx

Appendix

How to do the variant calling and association study

Below are the main scripts necessary to run the pipeline. Directories where the tools were installed were based on the folders used in LCFG Lab, and therefore need to be changed and specified accordingly.

1. Run variant calling for each samples with GATK's HaplotypeCaller one by one (***HaplotypeCaller***).
2. Merge GVCF files into a single VCF file with GATK's GenotypeGVCFs (***Genotypegvcfs***).
3. Variant Hard Filter with GATK's VariantFiltration (***VariantFiltration***) to generate high quality VCF with GATK's default parameter.
4. Select variants from VCF with GATK's SelectVariants (***SelectVariants***) to exclude the low quality SNPs.
5. Select only variants with GATK's SelectVariants (***SelectVariants***) to exclude the insertion and deletions.
6. Run variant annotation using SnpEff (***SnpEff***)
7. Run PLINK Allelic association Fisher's Exact Test (**PLINK_*Fisher***), then run the multiple correction test.
8. Run PLINK Linear and Logistic Models (**PLINK_*Linear***), then run the multiple correction test.
9. Run PLINK Linear and Logistic Models and the covariates (**PLINK_*Covariates***).
10. Run PLINK Linkage Disequlibrium (***PLINK_LD***)

- **HaplotypeCaller**

```
gatk --java-options "-Xmx4g" HaplotypeCaller -R
/media/ubuntu/STATINBAM/RefGene/human_g1k_v37_decoy.fasta –I
$i.final.bam -L 2:234494085-234681945 -O $i.indels.g.vcf -ERC
GVCF
```

- **GenotypeGVCFs**

```
java -jar /home/raechell/Downloads/GenomeAnalysisTK.jar -T
GenotypeGVCFs -R
/media/raechell/Master2/RefGene/human_g1k_v37_decoy.fasta –
variant $i -O UGT1A.vcf
```

- **VariantFiltration**

```
java -jar /home/raechell/Downloads/GenomeAnalysisTK.jar -T
VariantFiltration -R
/media/raechell/Master2/RefGene/human_g1k_v37_decoy.fasta –V
UGT1A.vcf --filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0
|| MQRankSum < -12.5 || ReadPosRankSum < -8.0" --filterName "low
quality" -o UGT1A_filtered_snps.vcf
```

- **SelectVariants (exclude lowqual)**

```
java -jar /home/raechell/Downloads/GenomeAnalysisTK.jar -T
SelectVariants -R
/media/raechell/Master2/RefGene/human_g1k_v37_decoy.fasta –V
UGT1A_filtered_snps.vcf --excludeFiltered -o UGT1A_PASS.vcf
```

- **SelectVariants (exclude indels)**
  ```
  java -jar /home/raechell/Downloads/GenomeAnalysisTK.jar -T
  SelectVariants -R
  /media/raechell/Master2/RefGene/human_g1k_v37_decoy.fasta -V
  UGT1A_PASS.vcf -xlSelectType INDEL -o UGT1A_SNPs.vcf
  ```

- **SnpEff Annotation**
  ```
  java -Xmx4g -jar /home/raechell/snpEff/snpEff.jar -v GRCh37.75
  UGT1A_SNPs.vcf > UGT1A_SNPs.ann.vcf
  ```

- **PLINK_Fisher & Multiple Test Correction**
  ```
  plink --ped mydata.ped --map autosomal.map --pheno pheno.txt --
  fisher
  plink --ped mydata.ped --map autosomal.map --pheno pheno.txt --
  fisher --adjust
  ```

- **PLINK_Linear & Multiple Test Correction**
  ```
  plink --ped mydata.ped --map autosomal.map --pheno pheno.txt --
  linear
  plink --ped mydata.ped --map autosomal.map --pheno pheno.txt --
  linear –adjust
  ```

- **PLINK_Covariates**
  ```
  plink --ped mydata.ped --map autosomal.map --pheno pheno.txt --
  covar covar.txt --linear
  ```

- **PLINK_LD**
  ```
  plink --ped mydata.ped --map autosomal.map --r2 --ld-window
  #ofSNPs
  ```