**Module 1**

High-throughput inference and false discovery rates

1–3

Foci: FDR; inference; gene expression

Concepts:

p-values, standard scores, Z and t scores, standard error, False Discovery Rate (FDR)

Algorithms:

Lab:

P-value correction with False Discovery Rate (FDR) on gene expression data with python or R

P-values …. Various corrections and p-values Bill Nobles paper.

Links:

StatQuest: FDR and the Benjamini-Hochberg Method clearly explained <https://youtu.be/K8LQSvtjcEo>

Papers:

1. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289-300.

2. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98(9):5116-5121.

3. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-9445.

**Module 2**

Phenotypic variation, genome-wide association, and heritability

Foci: Phenotype, heritability, GWAS

4–6

Concepts:

SnpChip, Complex Trait, Population stratification

Algorithms:

Bonferonni GWAS

Simple GWAS – Package exists

Manhattan Plots, calculating association, Mixed Model ANOVA, ANOVA, Principal components analysis, eigen vectors and values

Population correction

Measuring heritiability

Linear mixed models

Lab:

Use GCTA software

Links:

5D - Genome-wide association studies, part 1 https://youtu.be/bNpYzOr7I94

Mixed effects models with R https://youtu.be/VhMWPkTbXoY

StatQuest: Principal Component Analysis (PCA), Step-by-Step <https://youtu.be/FgakZw6K1QQ>

Eigenvectors and eigenvalues | Essence of linear algebra, chapter 14 <https://youtu.be/PFDu9oVAE-g>

Papers:

4. Manolio TA, Collins FS, Nancy J Cox, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-753.

5. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82.

6. Yang J, Bakshi A, Zhu Z, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47(10):1114-1120.

7. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904-909.

**Module 3**

Phenotypic variation and environment

8–10

Foci: Phenotype, exposures, EWAS

Concepts:

linear regression, regularization of regression, standardized coefficients, interpreting regression coefficients, social epidemiology, correlation globes, variable importance, variable importance plots, Vibration of Effects (VoE); model specification;

Algorithms:

Vibration of Effects (VoE)

Lab:

Replicate Chirags’ paper in a notebook. Patel CJ, Burford B, Ioannidis JPA. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol. 2015;68:1046-1058.

Links:

MIT CompBio Lecture 23 - Multi-Phenotype analyses <https://youtu.be/IRN-_8S35CI>

Chirag Patel "Building a search engine to identify environmental factors... <https://youtu.be/Y1DAPyXroBY>

Papers:

8. Ioannidis JPA, Loy EY, Poulton R, Chia KS. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. Sci Transl Med. 2009;1(7):7ps8-7ps8.

9. Patel CJ, Ioannidis JPA, Cullen MR, Rehkopf DH. Systematic assessment of the correlations of household income with infectious, biochemical, physiological, and environmental factors in the United States, 1999-2006. Am J Epidemiol. 2015;181(3):171-179.

10. Patel CJ, Burford B, Ioannidis JPA. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol. 2015;68:1046-1058.

**Module 4**

Metabolome-metagenome interactions and disease

11–14 or 15–17

Foci: Host-microbiome interactions through metabolome, chronic disease risk

Concepts:

Hypothesis testing, ANOVA, Chi-square tests, probability distributions, Survival analysis, Kaplan-Meier estimate, confidence intervals

Algorithms:

Equations for various statistics

Lab:

Shotgun sequencing, microbiome

Mapping

Stool sequence … fraction of taxa different bacteria

Simple enrichment

1. ANOVA lab
2. Survival analysis lab

Links:

What is Survival Analysis | Kaplan-Meier Estimation | Time to Event Model <https://youtu.be/Th0_SHH9zZI>

Hypothesis testing and p-values | Inferential statistics | Probability a... <https://youtu.be/-FtlH4svqx4>

Papers:

11. Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med. 2013;19(5):576-585.

12. Tang WHW, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med. 2013;368(17):1575-1584.

13. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature. 2011;472(7341):57-U82.

14. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448-453.

OR

15. Pedersen HK, Gudmundsdottir V, Nielsen HB, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. Nature. 2016;535(7612):376-381.

16. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2013;490(7418):55-60.

17. Wang J, Jia H. Metagenome-wide association studies: fine-mining the microbiome. Nat Rev Microbiol. 2016;14(8):508-522.

**Module 5**

Major Theme: diagnostics and disease classification

Unsupervised clustering and disease classification

18–20

Foci: Clustering, disease; precision medicine

Concepts:

Unsupervised learning, clustering

Algorithms:

Hierarchical clustering, linkage methods, distance versus similarity

Lab:

Hierarchical clustering of gene expression data.

Links:

StatQuest: Hierarchical Clustering <https://youtu.be/7xHsRkOdVwo>

Papers:

18. Sørlie T. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences. 2001;98(19):10869-10874.

19. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863-14868.

20. Butte AJ, Kohane IS. Creation and implications of a phenome-genome network. Nat Biotechnol. 2006;24(1):55-62.