Steps in whole genome sequencing/GWAS

Jeff Leek

@jtleek

www.jtleek.com

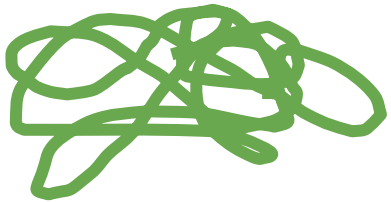
Background on WGS/GWAS

DNA ACTGACCTAGATCAGTGTAGCGATCGTATACGAGACCGATTCATCGGCAT

↓ transcription

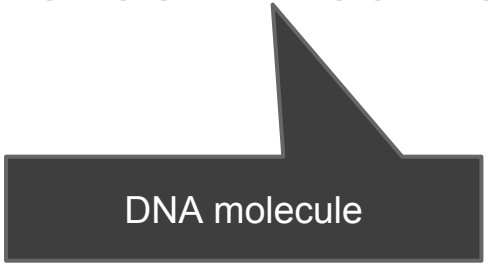
RNA AUCAGUCGAUCACCGAU

↓ translation



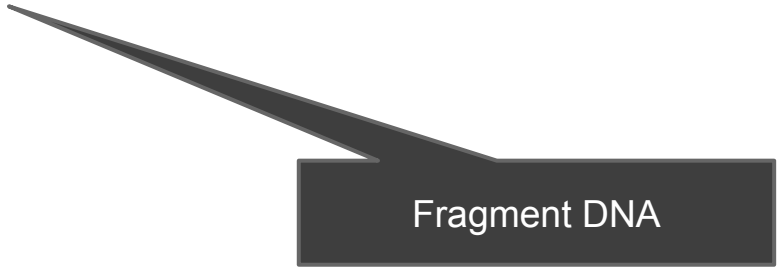
protein

ATGGGAATTCACGAATTCCTAGACCTGCCCCGGAAACCTACCGCCGCG



ACCTGCCCCGGAAACCTACC GCCGCG

ATGGGAATTCACGAATTCCTAG



ATGGGAATTCACGAATTCCTAG

ACACCTGCCCCGGAAACC

ACACCTGCCCCGGAAACC

TCCTAGACCTGCCCCGG

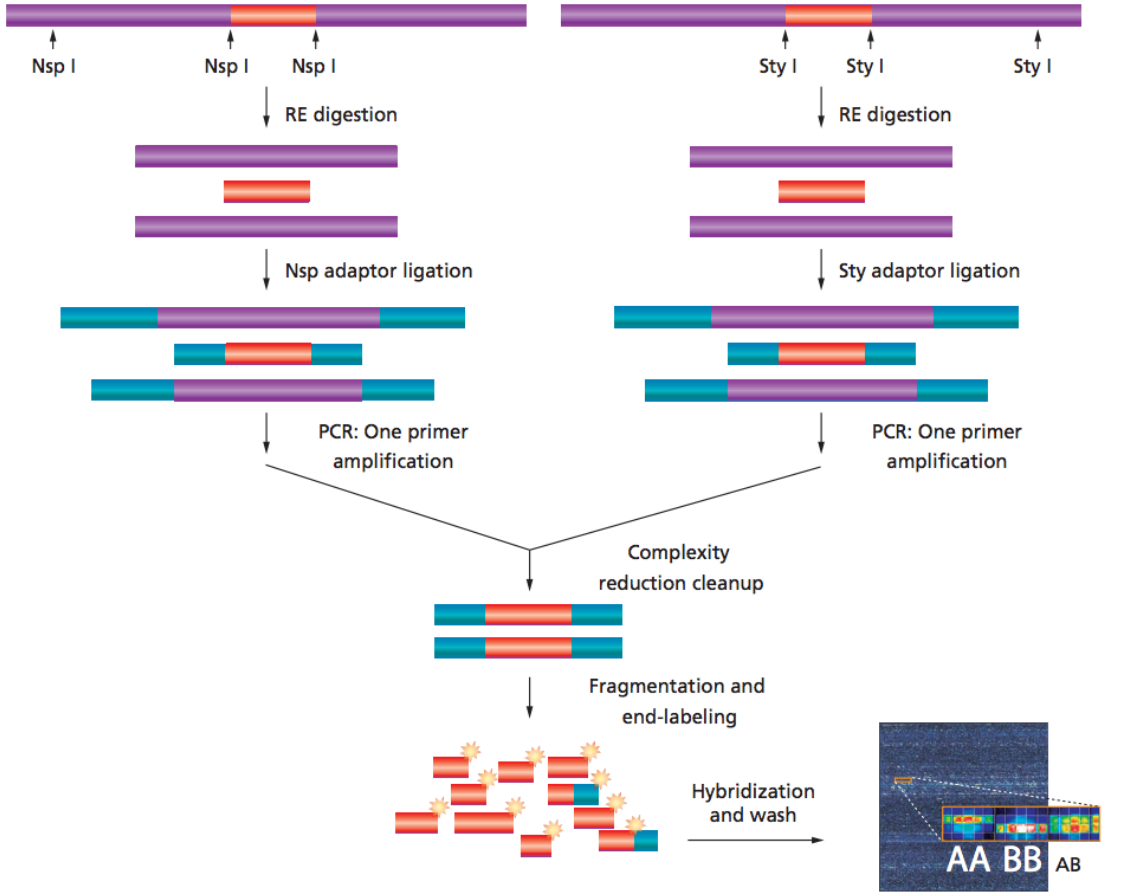
AATTCCTAGACCTGCCCC

CGAATTCCTACACCTG

ATGGGAATTCACGAATTCCTAGACCTGCCCCGGAAACCTACCGCCGCG

Genome

http://www.affymetrix.com/estore/catalog/131533/AFFY/Genome-Wide+Human+SNP+Array+6.0#1_3



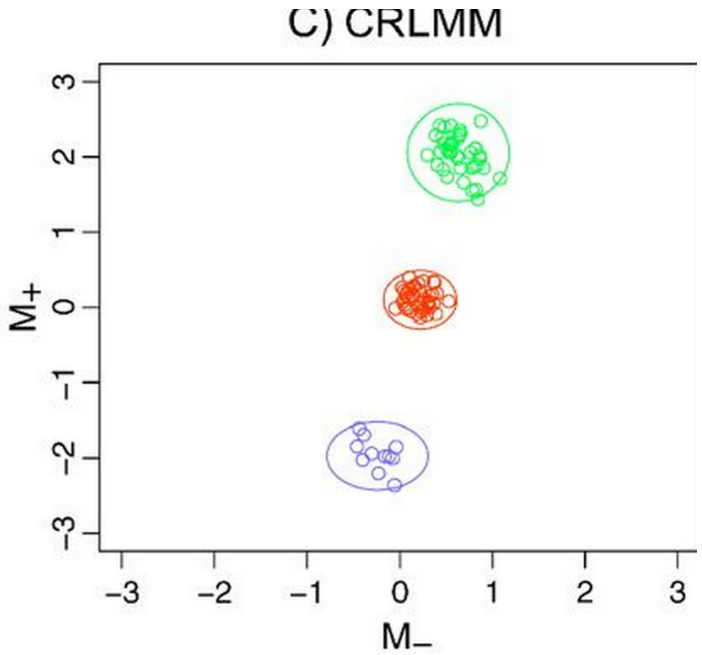
Steps

1. Variant identification
2. Population stratification correction
3. Statistical tests
4. Examining local region
5. Annotation

Step 1: Variant identification (SNP chip)

Software:

- crlmm

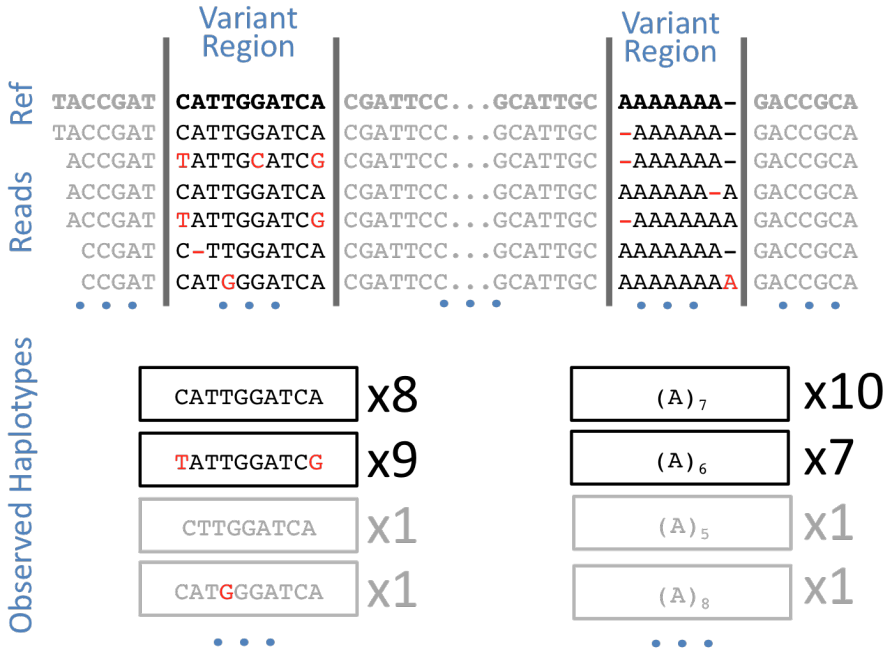


http://www.ncbi.nlm.nih.gov/pubmed/17189563

Step 1: Variant identification (sequencing)

Software:

- freeBayes
- GATK

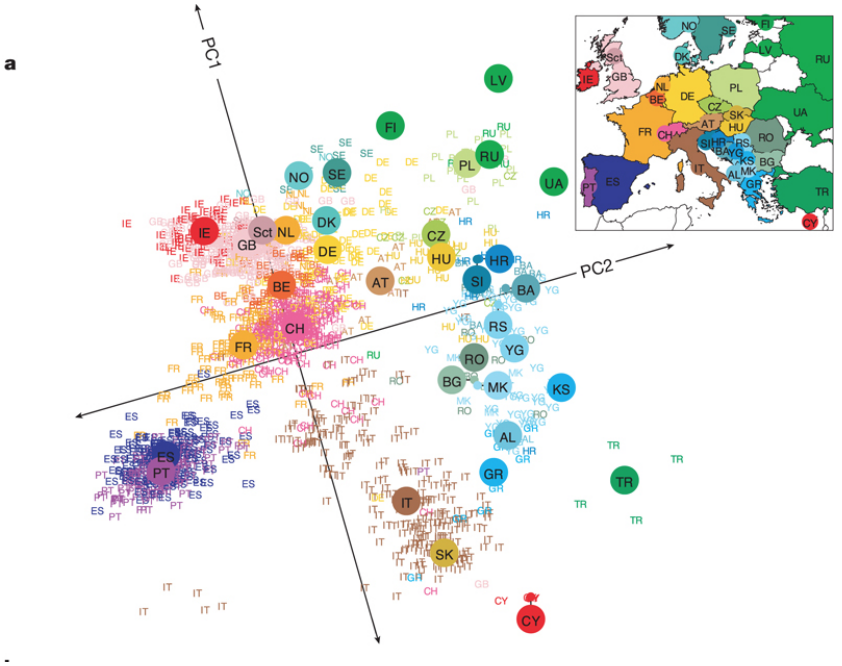


https://github.com/ekg/freebayes

Step 2: Population stratification

Software:

- EIGENSOFT
- snpStats



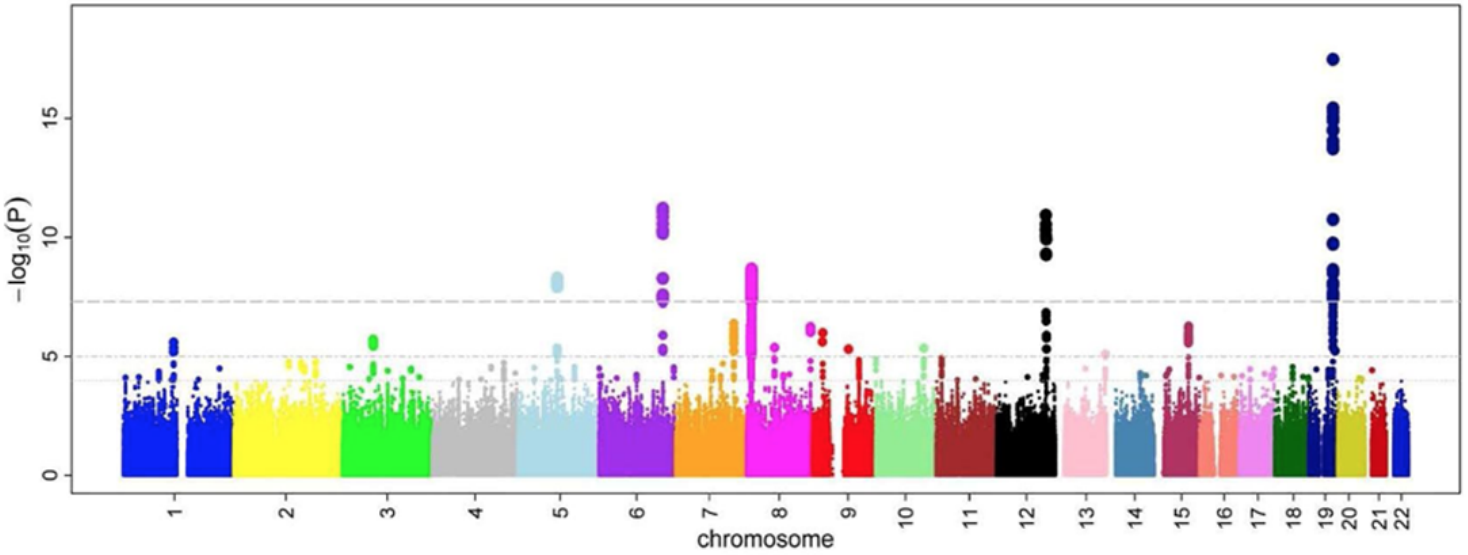
http://www.nature.com/nature/journal/v456/n7218/full/nature07331.html

Step 3: Statistical tests

https://en.wikipedia.org/wiki/Manhattan_plot

Software:

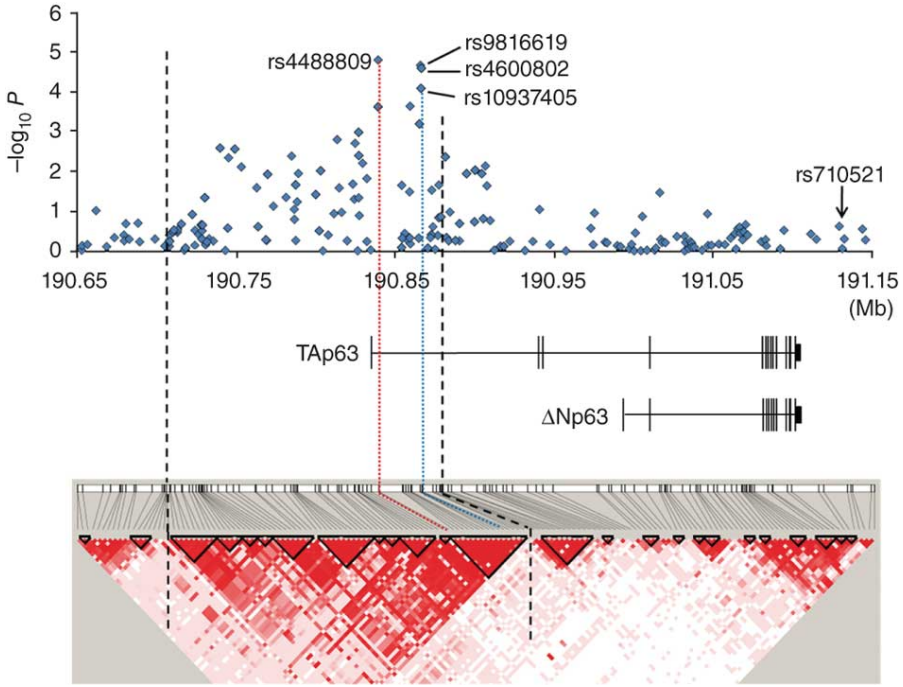
- PLINK
- snpStats



Step 4: Examine local region

Software:

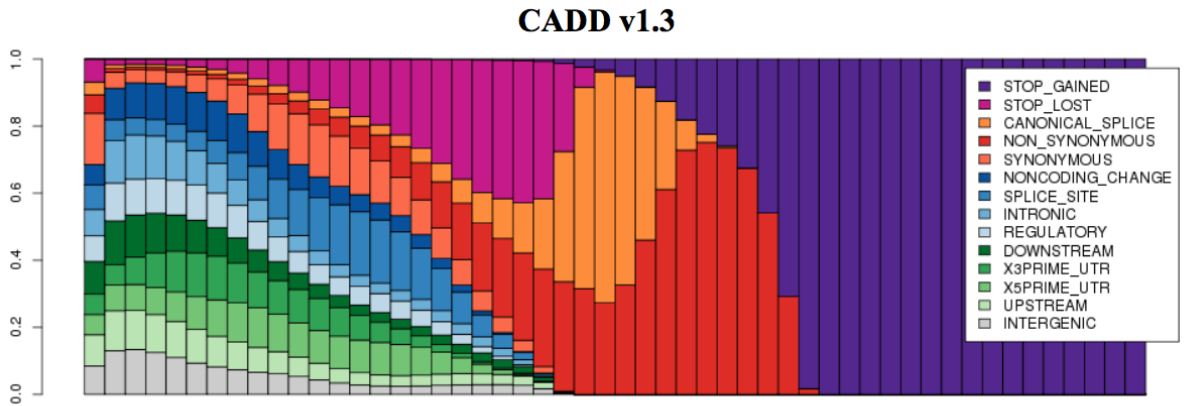
- PLINK
- Annotating Genomic Variants Workflow



http://www.nature.com/ng/journal/v42/n10/full/ng.667.html

Step 5: Annotation

http://cadd.gs.washington.edu/

Software:

- CADD
- variantAnnotation
- Annotating Genomic Variants Workflow