

Steps in whole genome sequencing/GWAS

Jeff Leek

@jtleek

www.jtleek.com

Background on WGS/GWAS

DNA

ACTGACCTAGATCAGTGTAGCGATCGTATACGAGACCGATTTCATCGGCAT



transcription

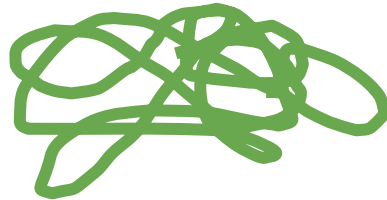
RNA

AUCAGUCGAUCACCGAU



translation

protein



ATGGGAATTCACGAATTCCTAGACCTGCCCCGGAAACCTACCGCCGCG



DNA molecule

ACCTGCCCCGGAAACCTACC

GCCGCG

ATGGGAATTCACGAATTCCTAG



Fragment DNA

The diagram illustrates a DNA fragment being inserted into a sequence. A dark gray arrow points from the 'Fragment DNA' box to the sequence 'ATGGGAATTCACGAATTCCTAG'. The sequence is written in black text on a white background. The fragment is represented by a dark gray rectangle with the text 'Fragment DNA' inside it.

ATGGGAATTCACGAATTCCTAG

Sequence DNA



ACACCTGCCCCGGGAAACC

ACACCTGCCCCGGGAAACC

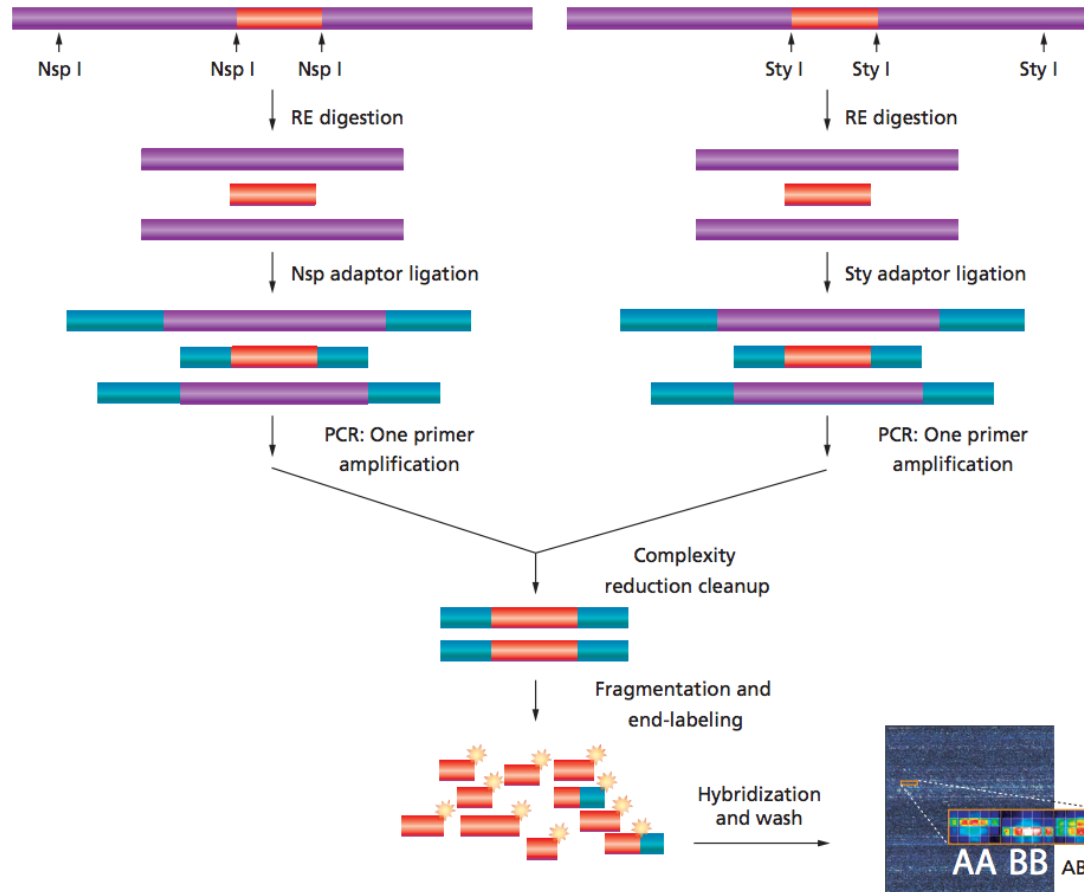
TCCTAGACCTGCCCCGG

AATTCCTAGACCTGCCCC

CGAATTCCTACACCTG

ATGGGAATTCACGAATTCCTAGACCTGCCCCGGGAAACCTACCGCCGCG

Genome



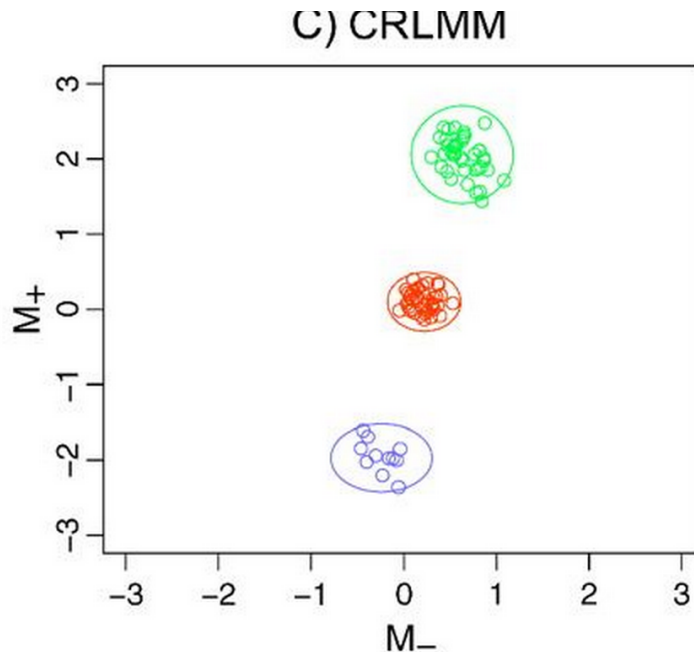
Steps

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](#)

	Ref	Variant Region		Variant Region		
Reads	TACCGAT	CATTGGATCA	CGATTCC...GCATTGC	AAAAAAA-	GACCGCA	
	TACCGAT	CATTGGATCA	CGATTCC...GCATTGC	-AAAAAA-	GACCGCA	
	ACCGAT	TATTGCATCG	CGATTCC...GCATTGC	-AAAAAA-	GACCGCA	
	ACCGAT	CATTGGATCA	CGATTCC...GCATTGC	AAAAAA-A	GACCGCA	
	ACCGAT	TATTGGATCG	CGATTCC...GCATTGC	-AAAAAAA	GACCGCA	
	CCGAT	C-TTGGATCA	CGATTCC...GCATTGC	AAAAAAA-	GACCGCA	
	CCGAT	CATGGGATCA	CGATTCC...GCATTGC	AAAAAAA	GACCGCA	
	
Observed Haplotypes	CATTGGATCA		x8	(A) ₇		x10
	TATTGGATCG		x9	(A) ₆		x7
	CTTGGATCA		x1	(A) ₅		x1
	CATGGGATCA		x1	(A) ₈		x1
		

<https://github.com/ekg/freebayes>

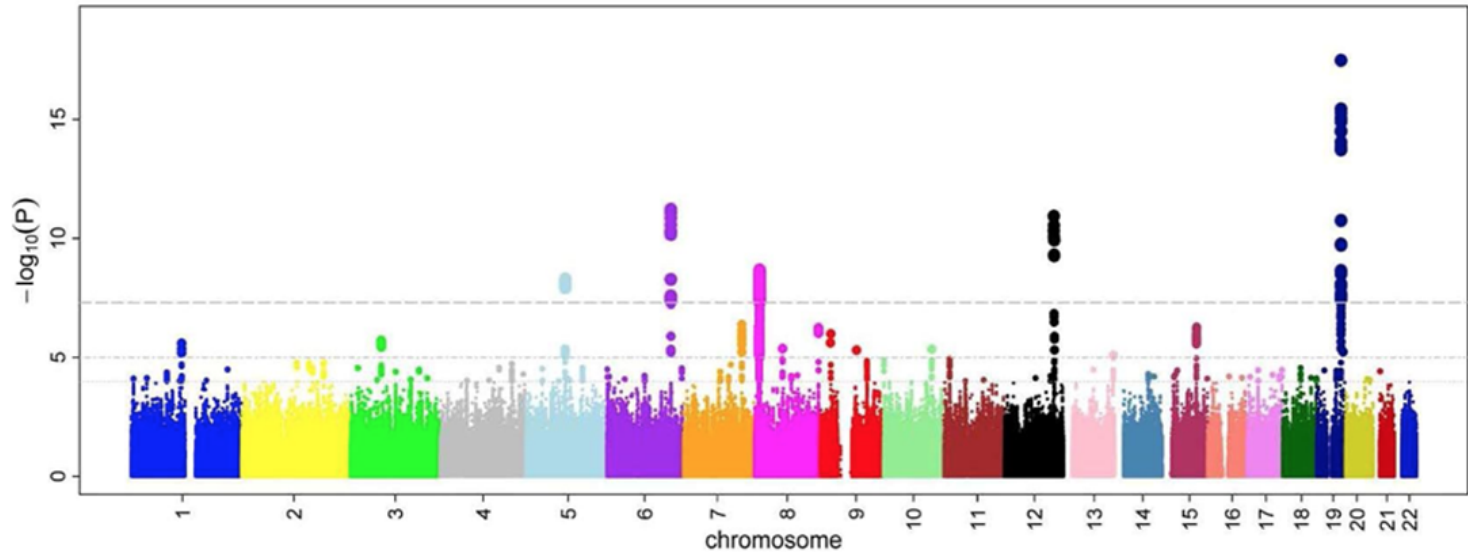
Software:

-

Step 3: Statistical tests

Software:

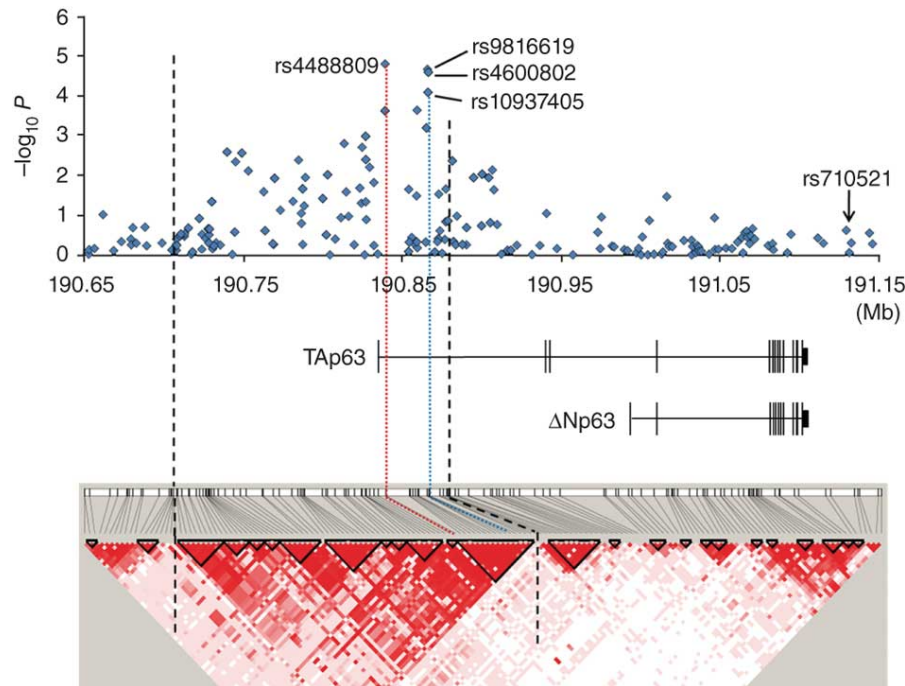
- [PLINK](#)
- [snpStats](#)



Step 4: Examine local region

Software:

- [PLINK](#)
- [Annotating Genomic Variants Workflow](#)



Step 5: Annotation

Software:

- [CADD](#)
- [variantAnnotation](#)
- [Annotating Genomic Variants Workflow](#)

