

The ‘omic coal face: classifying human genomic and epigenomic inheritance

Dr Miles Benton

IHBI Inspires: 23rd - 25th August 2017

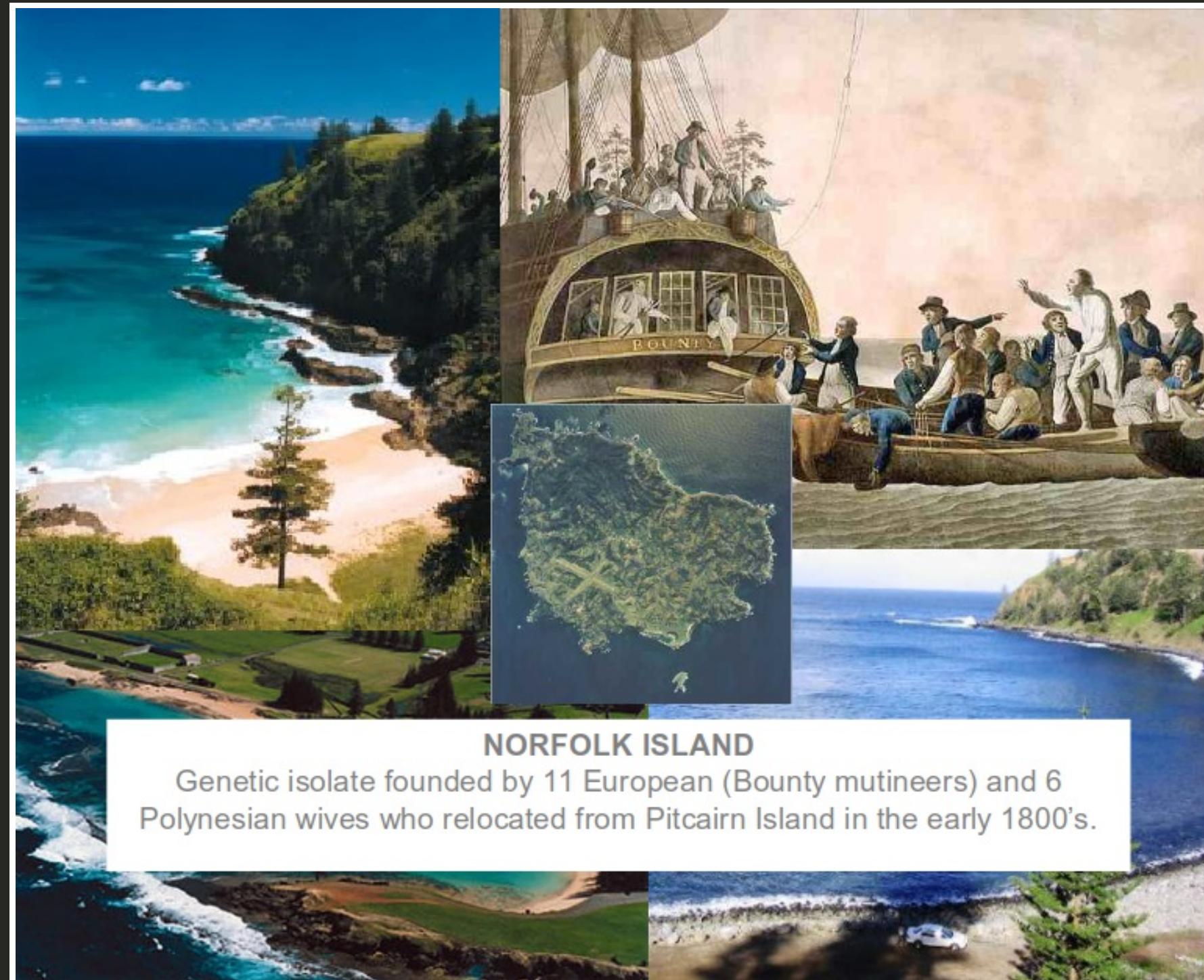
My presentation available online



sirselim.github.io/presentations

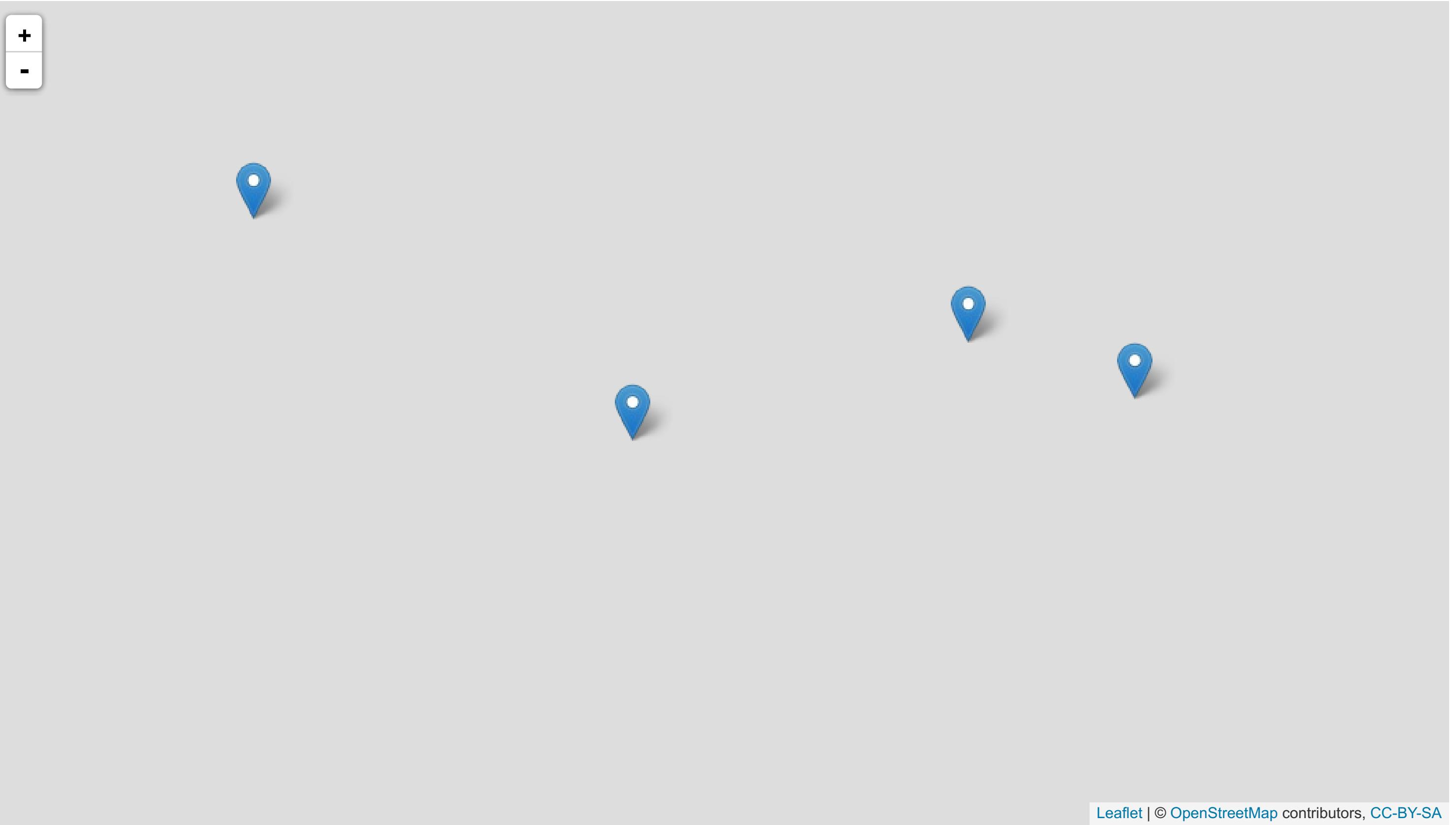
NORFOLK ISLAND

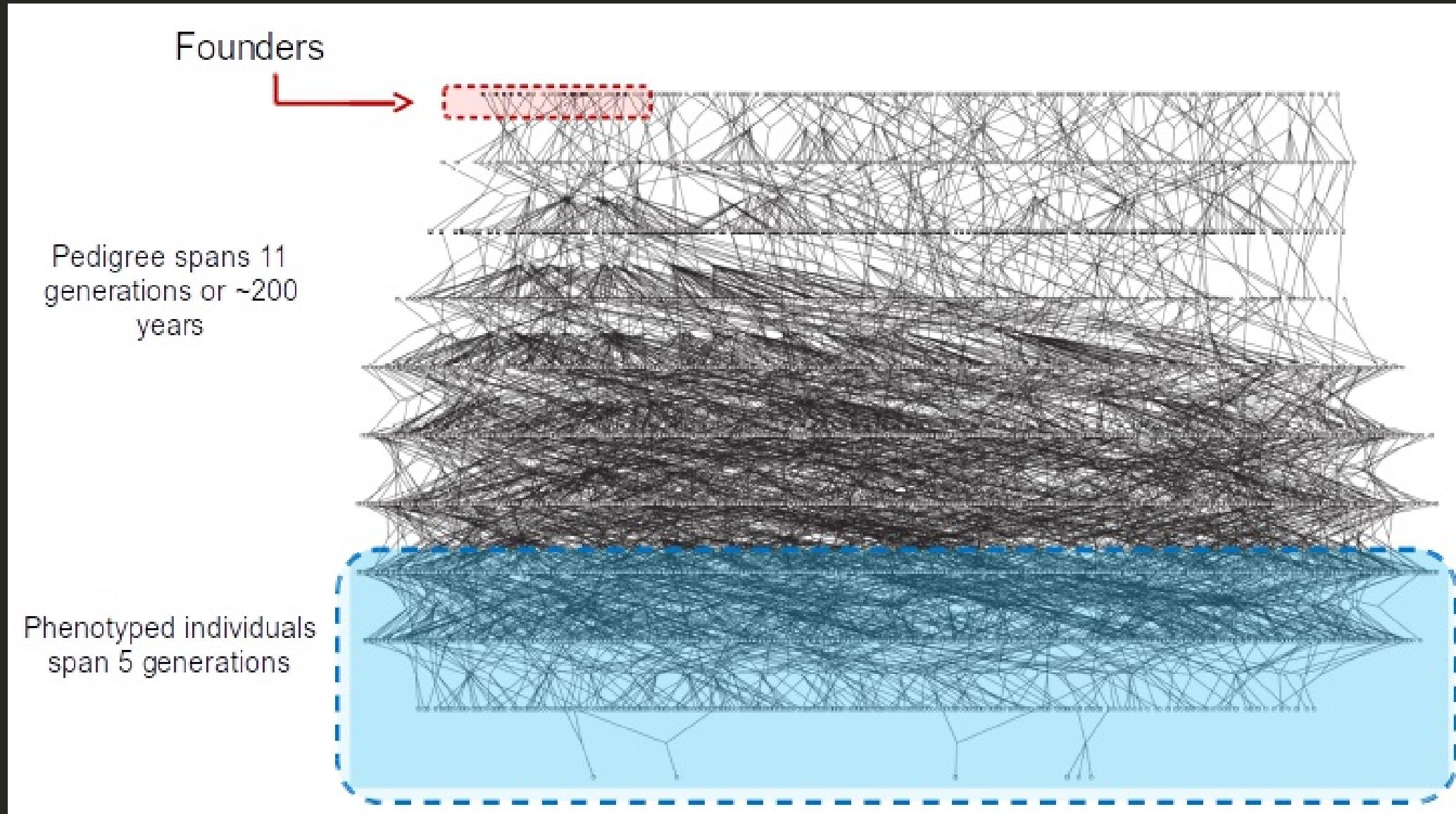
...a little story time...



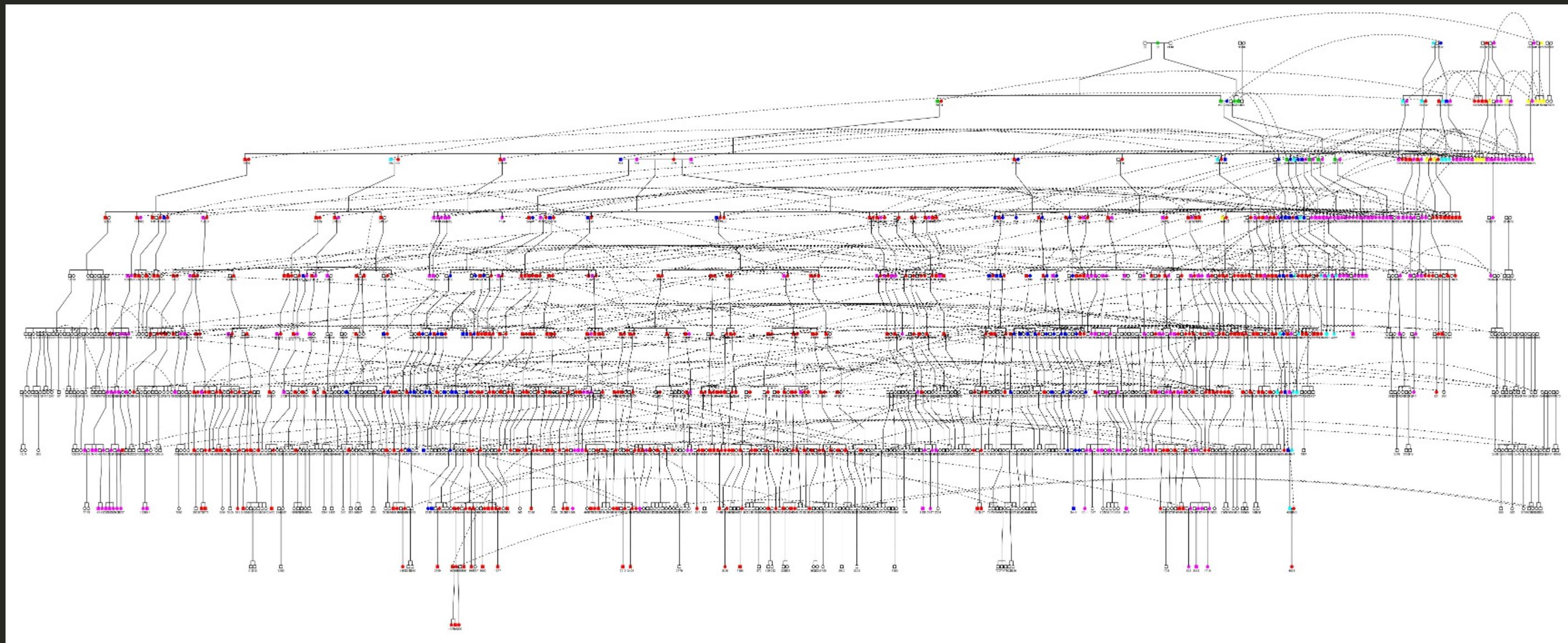
NORFOLK ISLAND

Genetic isolate founded by 11 European (Bounty mutineers) and 6 Polynesian wives who relocated from Pitcairn Island in the early 1800's.





Macgregor S et al.,: *Legacy of mutiny on the Bounty: founder effect and admixture on Norfolk Island*. Eur J Hum Genet. 2010; 18: 67–72.



40% of current population haplogroup B4a1a[...]

Benton MC et al.: “Mutiny on the Bounty”: the genetic history of Norfolk Island reveals extreme gender-biased admixture. Investigative Genetics 2015, 6:11.

Founder Effect Variants

Other contributors: Rod Lea, David Eccles, Donia Macartney-Coxson, Heidi Sutherland, Larisa Haupt, Lyn Griffiths

Whole Genome Sequencing

N=108 core pedigree individuals sequenced

Platform – **Illumina HiSeq-X10 (Garvan)**

Bioinformatics:

- BOWTIE2 → SAMTOOLS → VCF annotation → dbSNP → dbNSFP → VEP → custom beds

Coverage >25X

Whole Genome Sequencing

N=108 core pedigree individuals sequenced

Platform – **Illumina HiSeq-X10 (Garvan)**

Bioinformatics:

- BOWTIE2 → SAMTOOLS → VCF annotation → dbSNP → dbNSFP → VEP → custom beds

Coverage >25X

Functional Founder Effect Variants

Functional = Predicted damaging in *in silico* tests:

- SIFT, POLYPHEN2, MUTATIONASTER, PROVEAN, MUTATION ASSESSOR, CADD

Founder effect = increased allele freq in NI compared to general population

- (MAF<0.01% in 1000G >5% in NI)

Variant = single nucleotide variant (SNV)

Show 10 entries

Search:

chr	position	id	aa_sub	variant	gene	MutTaster	CADD
chr1	2503934	.	p.Leu814Pro	missense	PLCH2	D	5.28
chr1	2503936	.	p.Ser992Pro	missense	PLCH2	D	3.896
chr1	3836493	rs770445221	.	intron	CEP104	D	23.5
chr1	8324522	.	p.Asn99His	missense	SLC45A1	D	25
chr1	8360571	.	p.His711Pro	downstream	RERE	D	4.689
chr1	9720654	.	p.Glu529Gly	missense	PIK3CD	D	17.03
chr1	9720657	.	p.Glu530Gly	missense	PIK3CD	D	23.9
chr1	9720660	.	p.Glu531Gly	missense	PIK3CD	D	23.4
chr1	9720663	.	.	downstream	PIK3CD	D	23.2
chr1	9720862	rs780269932	p.Arg572Trp	downstream	PIK3CD	D	26.7

Showing 1 to 10 of 788 entries

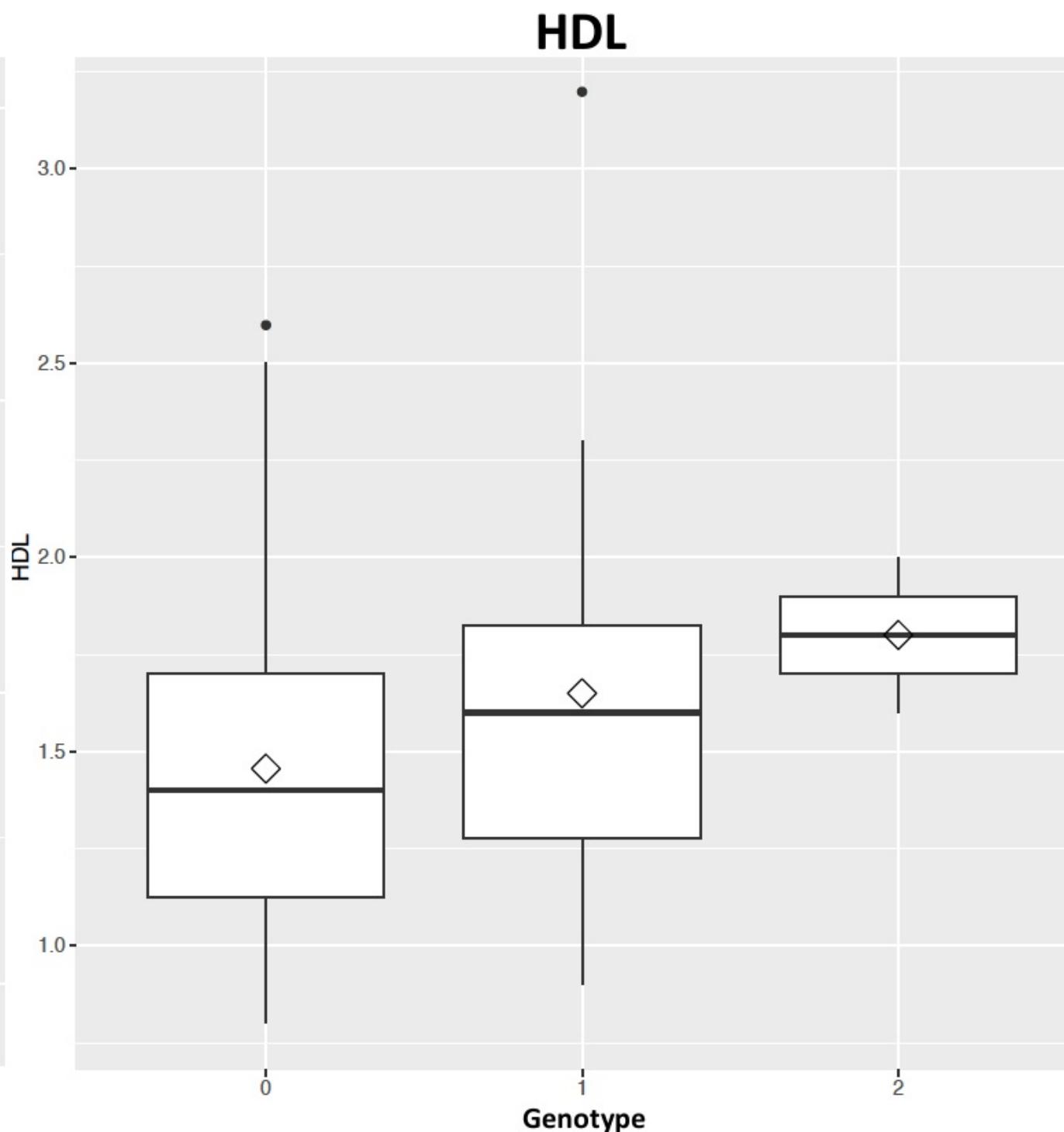
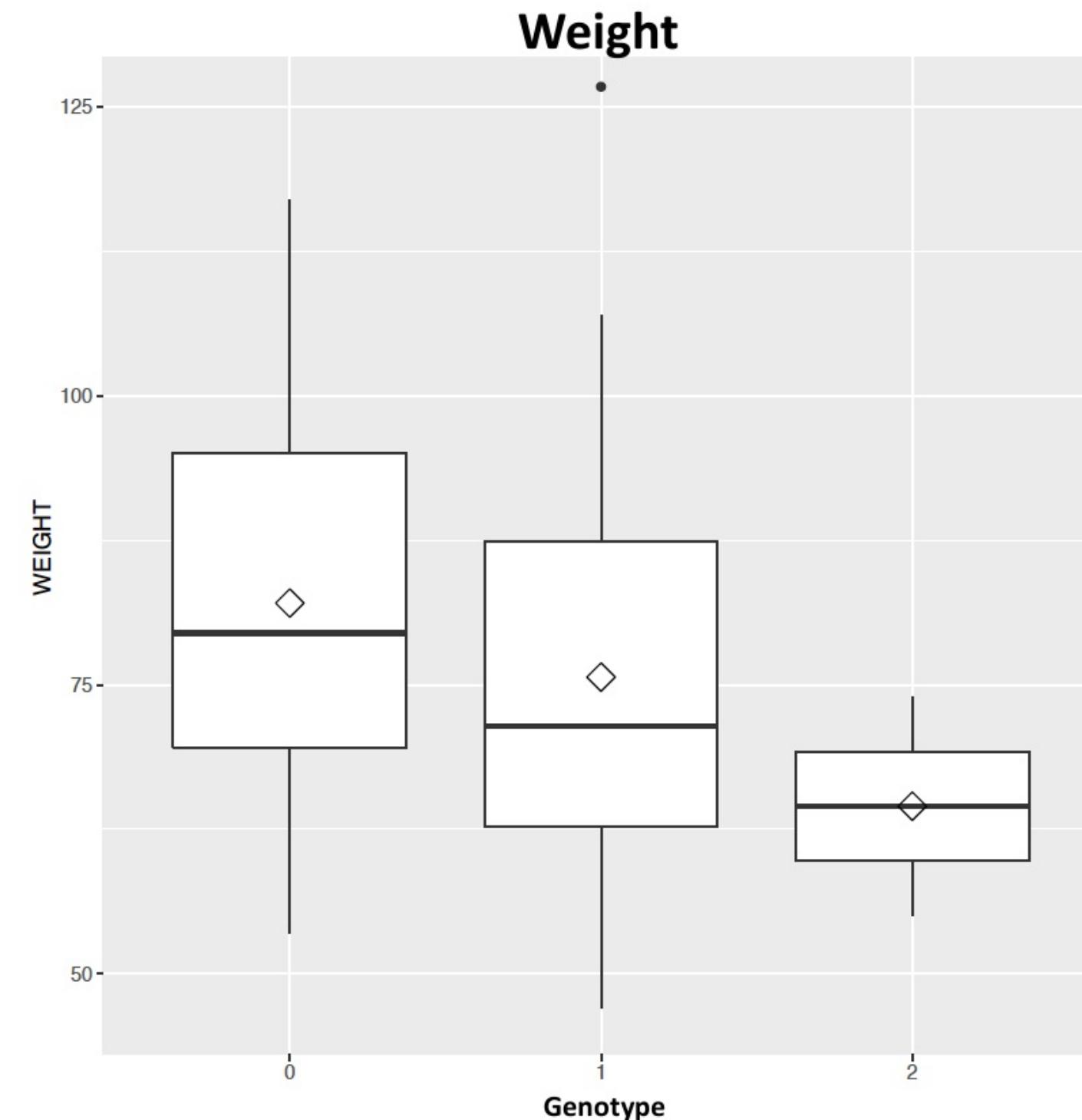
Previous

1 2 3 4 5 ... 79 Next

ACOT4 (Acyl-CoA Thioesterase 4)

- **regulation of lipid metabolism**
- SNV = rs77408762 (T/A)
- Chr14, pos 73593812 (Hg38)
- Amino acid = p.TYR190ASN
- freq A = 0.0099 (EUR), 0 (AFR) and 0 (EAS)
- freq A in NI = 0.26

Rare allele most likely came from England via Bounty Mutineers and then increased in frequency in NI due to founder effect



de novo SNVs

3 variants meeting the 'damaging' criteria, >5% MAF in NI, all missense

- *JPH2* (Junctophilin 2)
 - chr20:44186654; c.52T>G; p.Trp18Gly; [48 hets, **MAF=0.22**]
 - mutations previously linked to inherited cardiomyopathy (Landstrom *et al*, 2007)
- *EPS15L1* (Epidermal growth factor receptor substrate 15-like 1)
 - Chr19:16425252; c.623T>C; p.Leu208Pro; [35 hets, **MAF=0.16**]
- *UGT2B4* (UDP glucuronosyltransferase 2 family, polypeptide B4)
 - chr4:69480794; c.1427T>A; p.Leu476His; [12 hets, **MAF=0.06**]

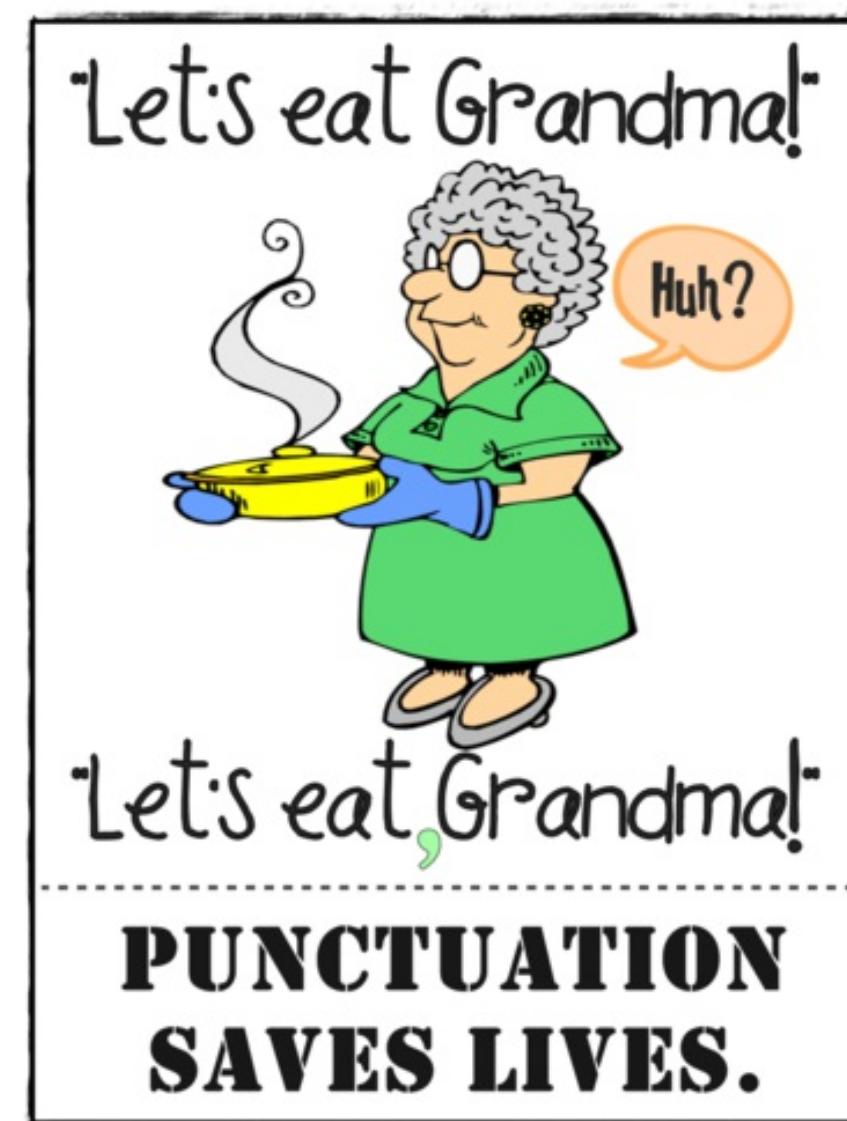
Summary

- Using WGS and founder effect in NI may reveal disease-related variants that are rare in general populations
- Focusing on “functional” SNVs has identified strong candidate variants
 - performing genotyping in the rest of the pedigree
- Some early evidence of association with T2D and CVD-related traits

Allele-Specific Methylation

Other contributors: Rod Lea, Donia Macartney-Coxson, Nicole White, Daniel Kennedy, Heidi Sutherland, Larisa Haupt, Kerrie Mengersen, Lyn Griffiths

Methylation - the punctuation of the genome



Allele-specific methylation (ASM):

- same cytosine is differentially methylated on the **two alleles** of a diploid organism

ASM is a major mechanism of **genomic imprinting** (aberrations can lead to disease)

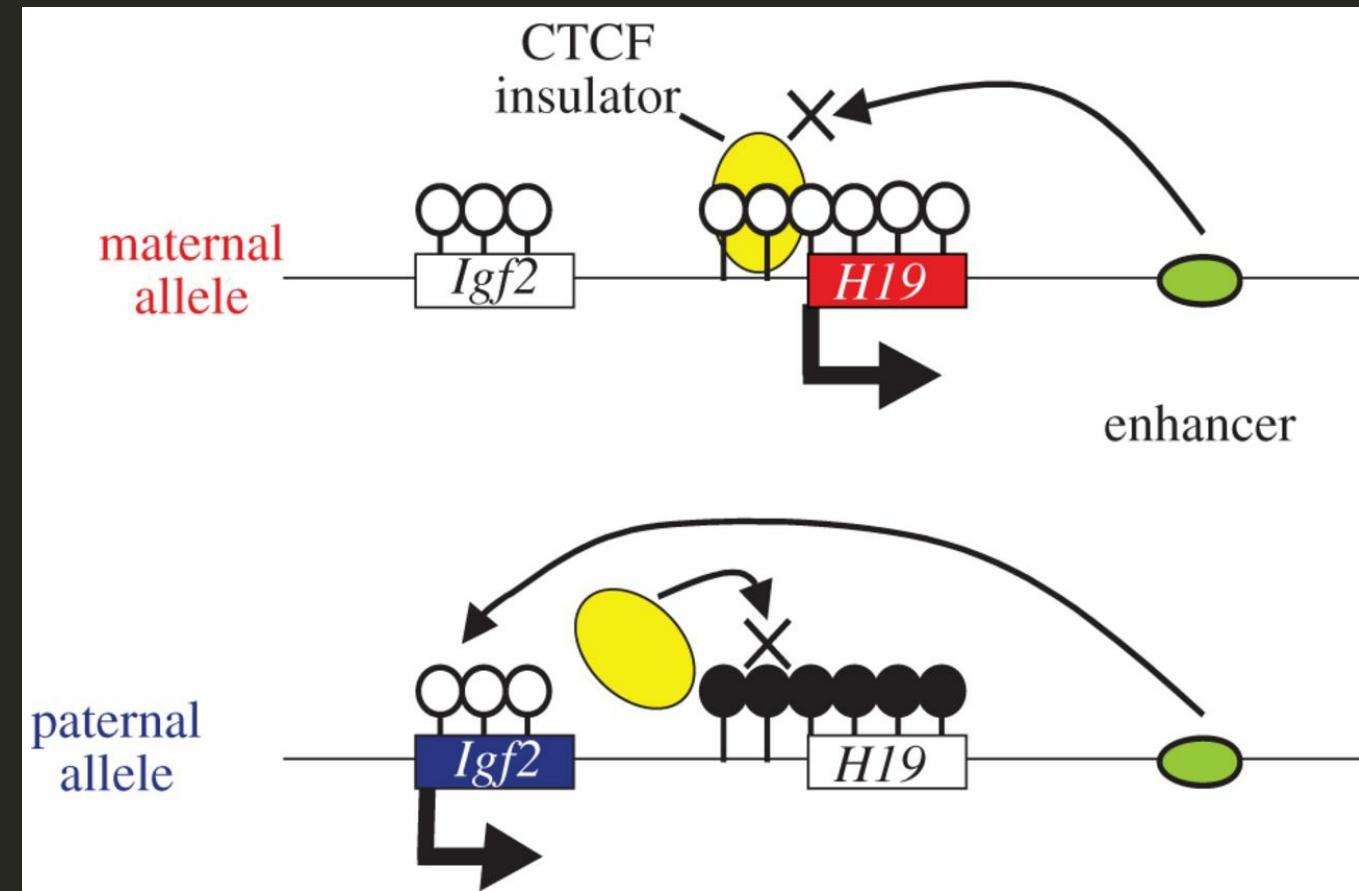


image rights: Renfree MB et al.: "The origin and evolution of genomic imprinting and viviparity in mammals. 2012

Identification of allele-specific methylation profiles across generations

measuring genome-wide allele-specific methylation (ASM)

- NGS bisulphite sequencing
- SeqCap Epi CpGiant (Illumina HiSeq)

collected data for **108** NI individuals

- comprising a close 3 generation pedigree

Identification of allele-specific methylation profiles across generations

measuring genome-wide allele-specific methylation (ASM)

- NGS bisulphite sequencing
- SeqCap Epi CpGiant (Illumina HiSeq)

collected data for **108** NI individuals

- comprising a close 3 generation pedigree

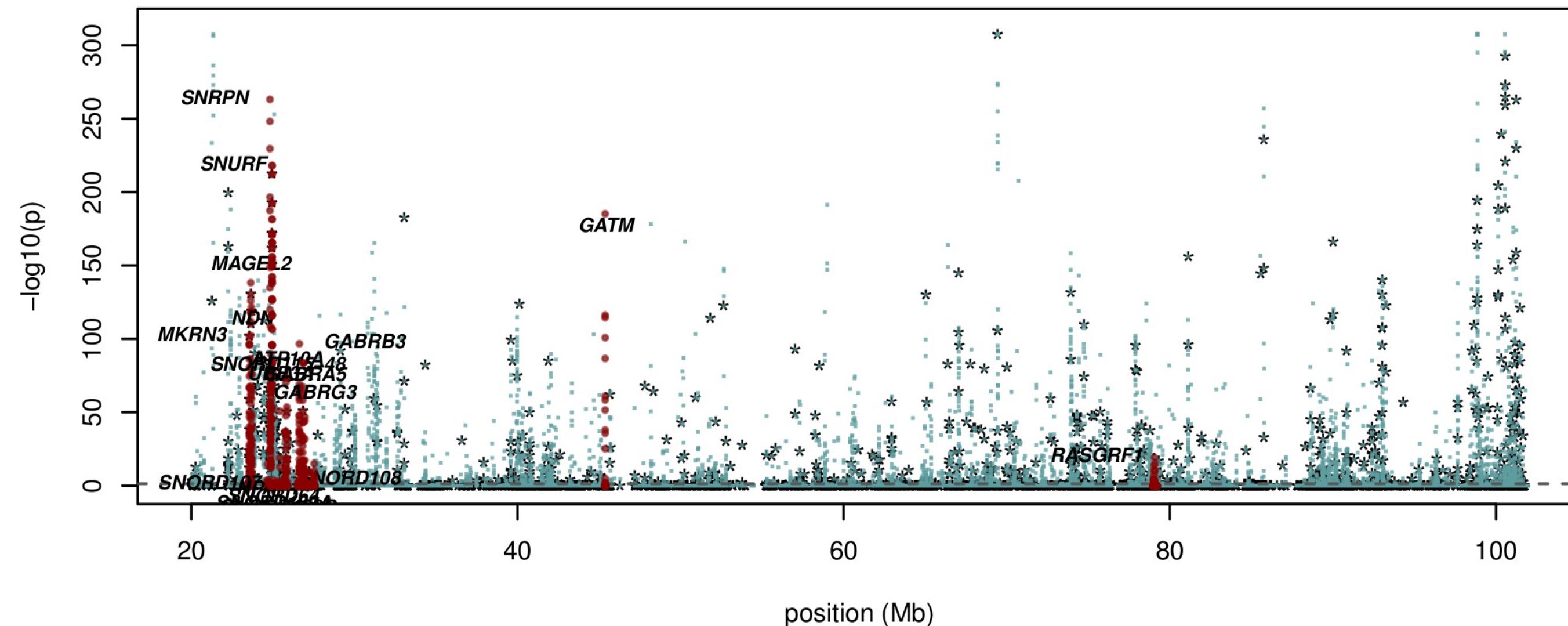
fully customised QC and analysis pipeline:^{*}

- fastqc, trimgalore
- bismark, sambamba, picard tools
- methpipe (ASM estimation)
- MethylDackel (originally PileOMeth), R and methylkit
- Shiny webserver visualisation

parallel processing enabled for local and remote machines

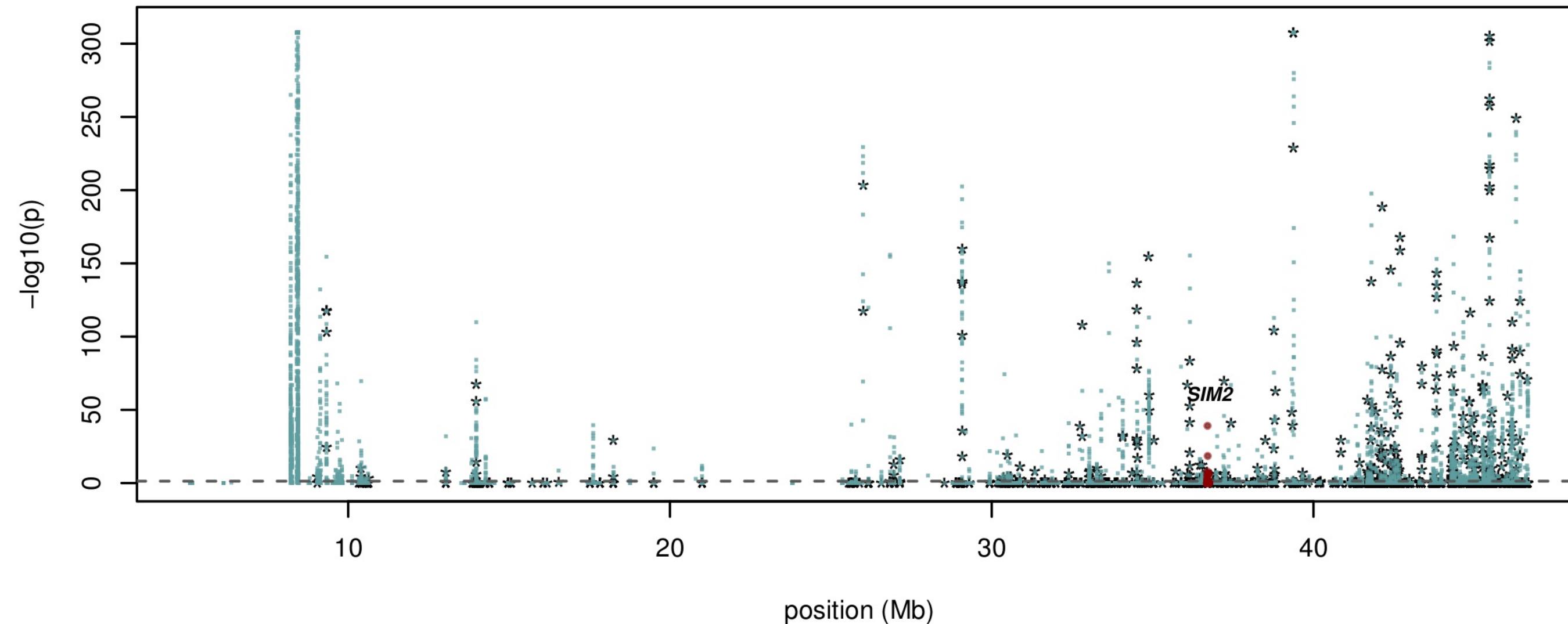
**once wrangled into shape scripts will be accessible via GitHub*

Chromosome 15



15q11-q13: Angelman syndrome (maternally active allele) | Prader-Willi syndrome (paternally active allele)

Chromosome 21



WRB is found to be maternally imprinted

Allele-specific Methylated Regions (AMRs)

Overlap with known imprinting genes

- there are **91** known imprinted genes with strong evidence
- we have identified **12761** total AMR (≥ 2 CpG sites)
- we recoup 72/91 **79%**) of the above known genes

Allele-specific Methylated Regions (AMRs)

Overlap with known imprinting genes

- there are **91** known imprinted genes with strong evidence
- we have identified **12761** total AMR (≥ 2 CpG sites)
- we recoup **72/91 79%** of the above known genes

Small AMRs (≥ 2 CpG sites and ≤ 200 bp length)

- there are **4731 small AMR** which collectively span 6987 genomic features
- **35.8%** promoter related

Allele-specific Methylated Regions (AMRs)

Overlap with known imprinting genes

- there are **91** known imprinted genes with strong evidence
- we have identified **12761** total AMR (≥ 2 CpG sites)
- we recoup **72/91 79%** of the above known genes

Small AMRs (≥ 2 CpG sites and ≤ 200 bp length)

- there are **4731 small AMR** which collectively span 6987 genomic features
- **35.8%** promoter related

Large AMRs (≥ 200 bp length)

- **11 large AMR** (spanning some really interesting regions, i.e. *PCDH*, *HOX*)

Allele-specific Methylated Regions (AMRs)

Overlap with known imprinting genes

- there are **91** known imprinted genes with strong evidence
- we have identified **12761** total AMR (≥ 2 CpG sites)
- we recoup **72/91 79%** of the above known genes

Small AMRs (≥ 2 CpG sites and ≤ 200 bp length)

- there are **4731 small AMR** which collectively span 6987 genomic features
- **35.8%** promoter related

Large AMRs (≥ 200 bp length)

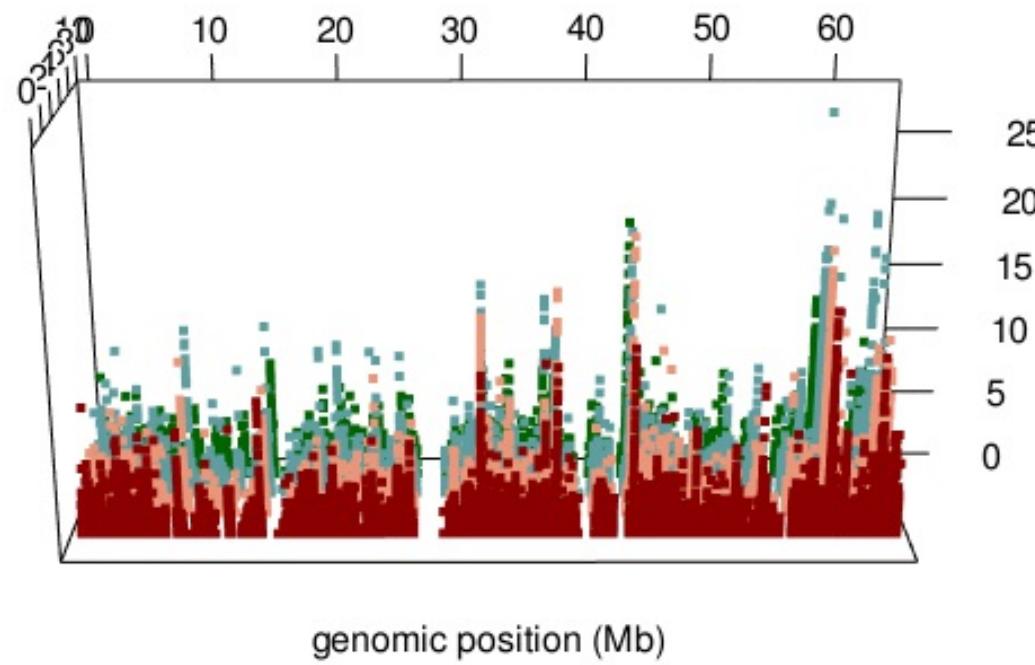
- **11 large AMR** (spanning some really interesting regions, i.e. *PCDH*, *HOX*)

Genetic variation across AMR

- **2,388** AMR have no common SNPs (little genetic variation)

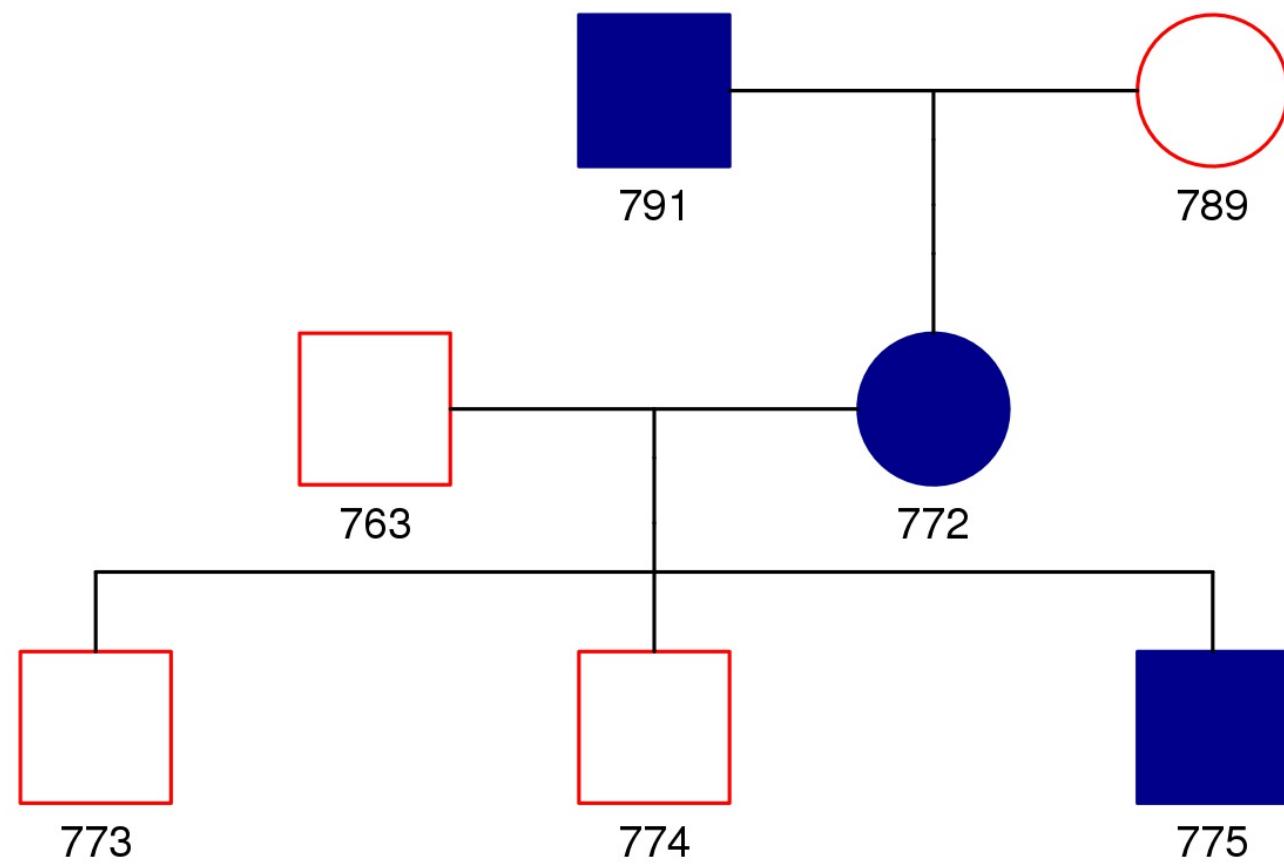
annotation	observed	expected	fold	pvalue
ChIP_seq_region	15496	3766.844	4.113	0.001
five_prime_utr	12347	3371.693	3.6612	0.001
start_codon	6851	1890.07	3.6233	0.001
miRNA	406	128.725	3.1374	0.001
TF_binding_site	34330	14187.176	2.4197	0.001
three_prime_utr	10341	4325.146	2.3906	0.001
exon	43713	19129.451	2.285	0.001
gwascatalog	1326	749.187	1.7689	0.001
gene	10570	6342.178	1.6665	0.001
transcript	10578	6349.741	1.6658	0.001
lincRNA	1409	887.126	1.5876	0.001
piRNA	37108	30761.966	1.2063	0.001

Inheritance of Allele-Specific Methylation?

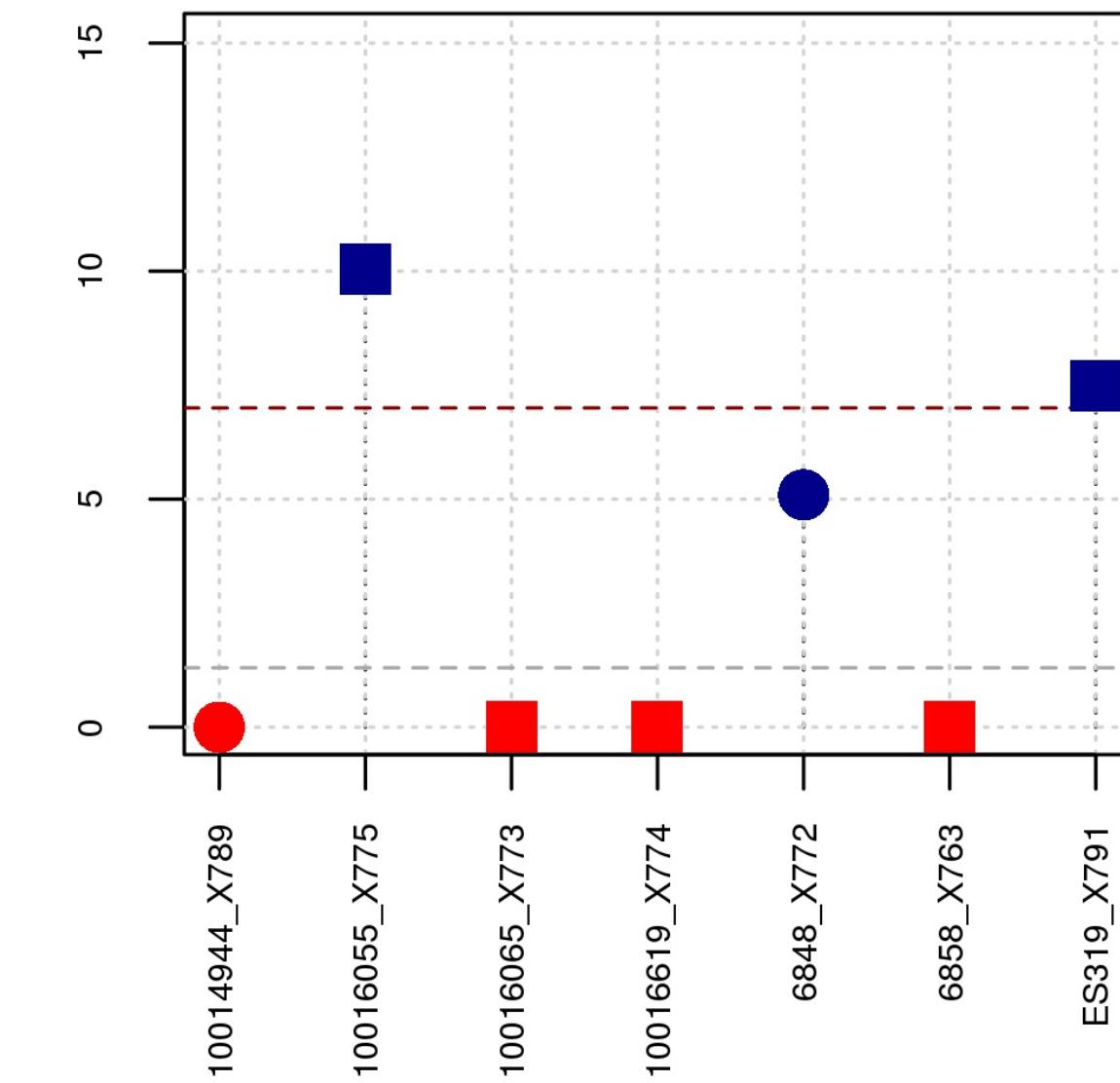


ASM plot of chromosome 20 for a nuclear family (father, mother, son, daughter)

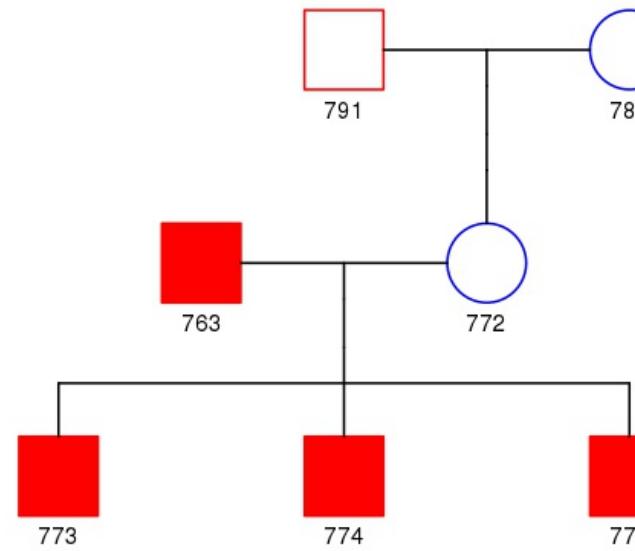
Pedigree



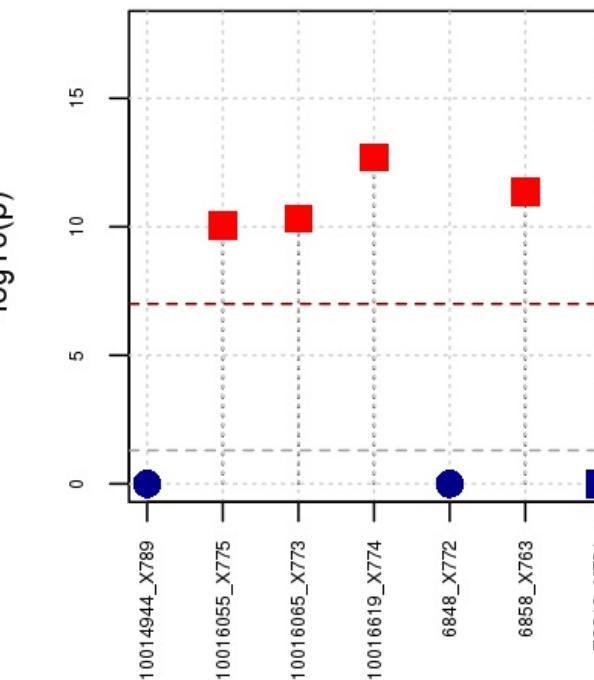
chr10:45224346



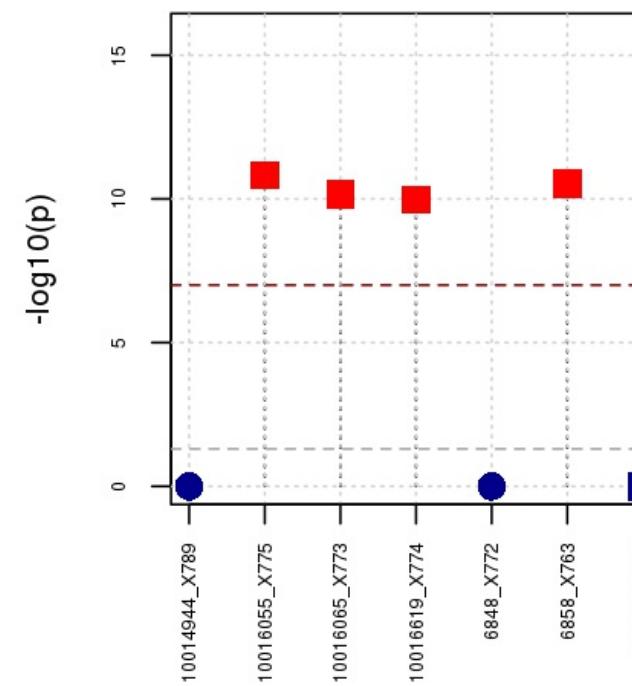
Pedigree



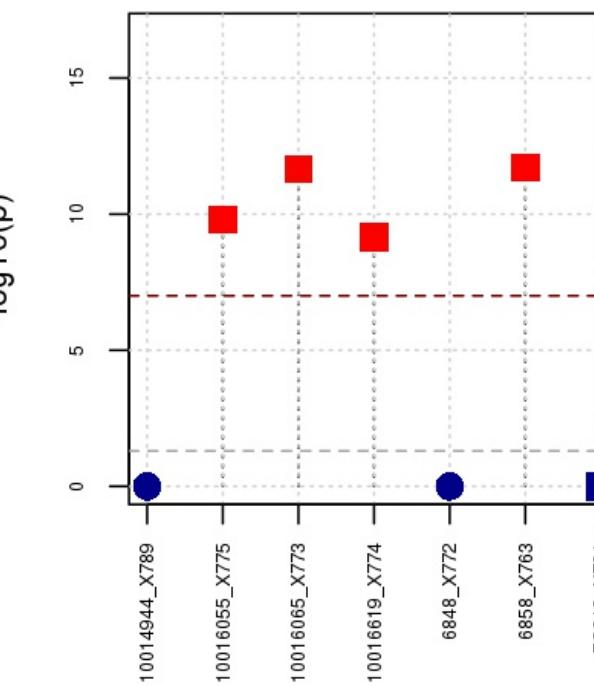
chr17:5771317



chr17:5771320



chr17:5771338



Summary

- implemented a custom bioinformatic pipeline
- generated a genome-wide 'map' of AMRs
 - allows exploration of the genomic landscape and features
- identified **ASM hotspots** overlapping non-coding RNA and other important functional loci
 - **might account for the number of GWAS hits outside of genes?**

Acknowledgments

QUT

Lyn Griffiths

Rod Lea

Larisa Haupt

Heidi Sutherland

Michelle Hanna

...the rest of the GRC IHBI lab group

Nicole White

Daniel Kennedy

Kerrie Mengersen

WGS sequencing: Garvan Institute of Medical Research /
Kinghorn Centre for Clinical Genomics

Funding: NHMRC

STDOI @ UTRGV</span?

John Blangero

Joanne Curran

Harald Goring

Greg Gibson (Georgia Institute of Technology, USA)

Melanie Carless (Texas Biomedical Research Institute)

Claire Bellis (Genome Institute of Singapore)

NZ collaborators

Donia Macartney-Coxson (ESR)

David Eccles (Gringene Bioinformatics)

Geoff Chambers (VUW)

The people of Norfolk Island who volunteered for this study.



Thanks, any questions?