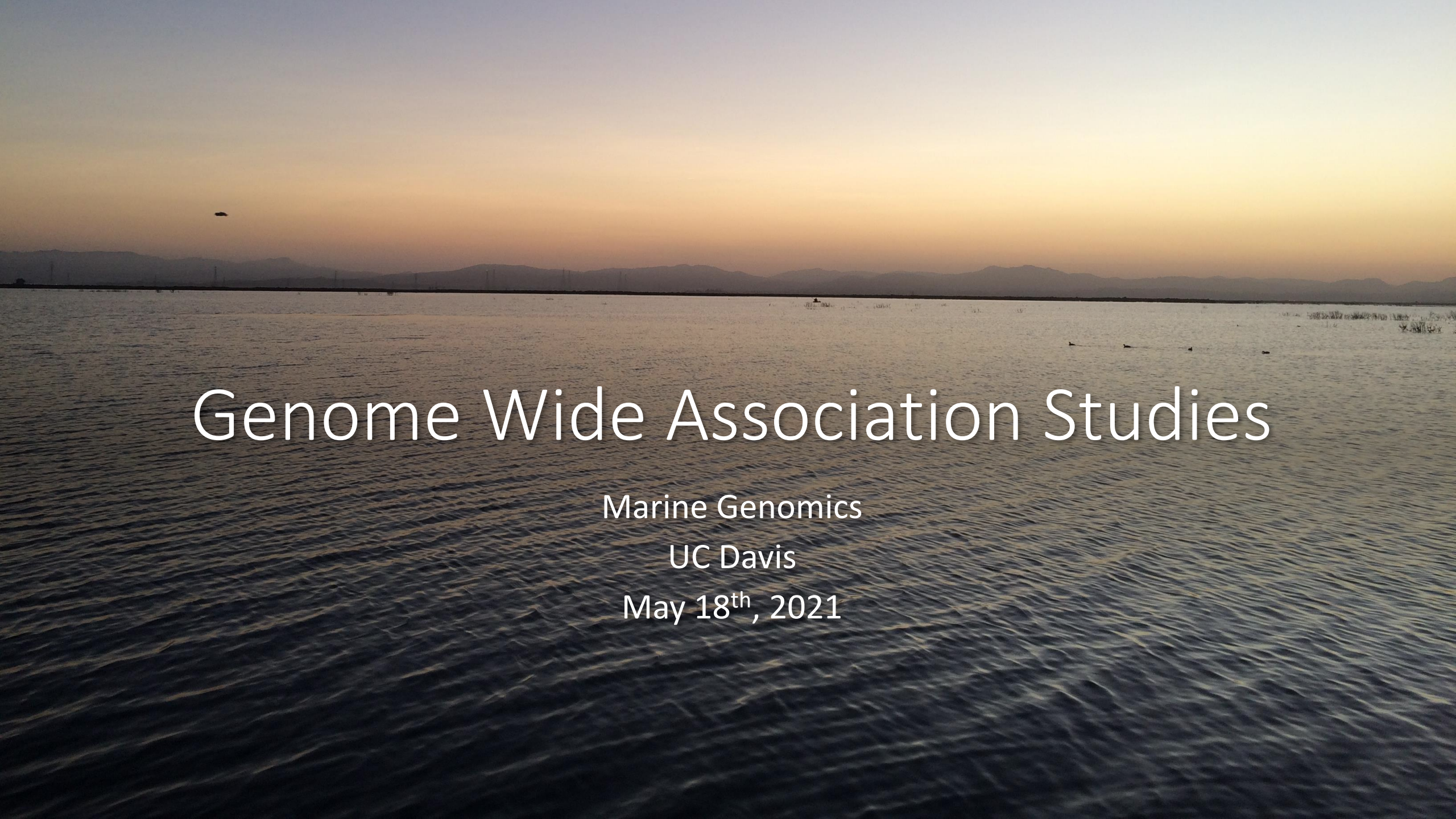


Genome Wide Association Studies

Marine Genomics

UC Davis

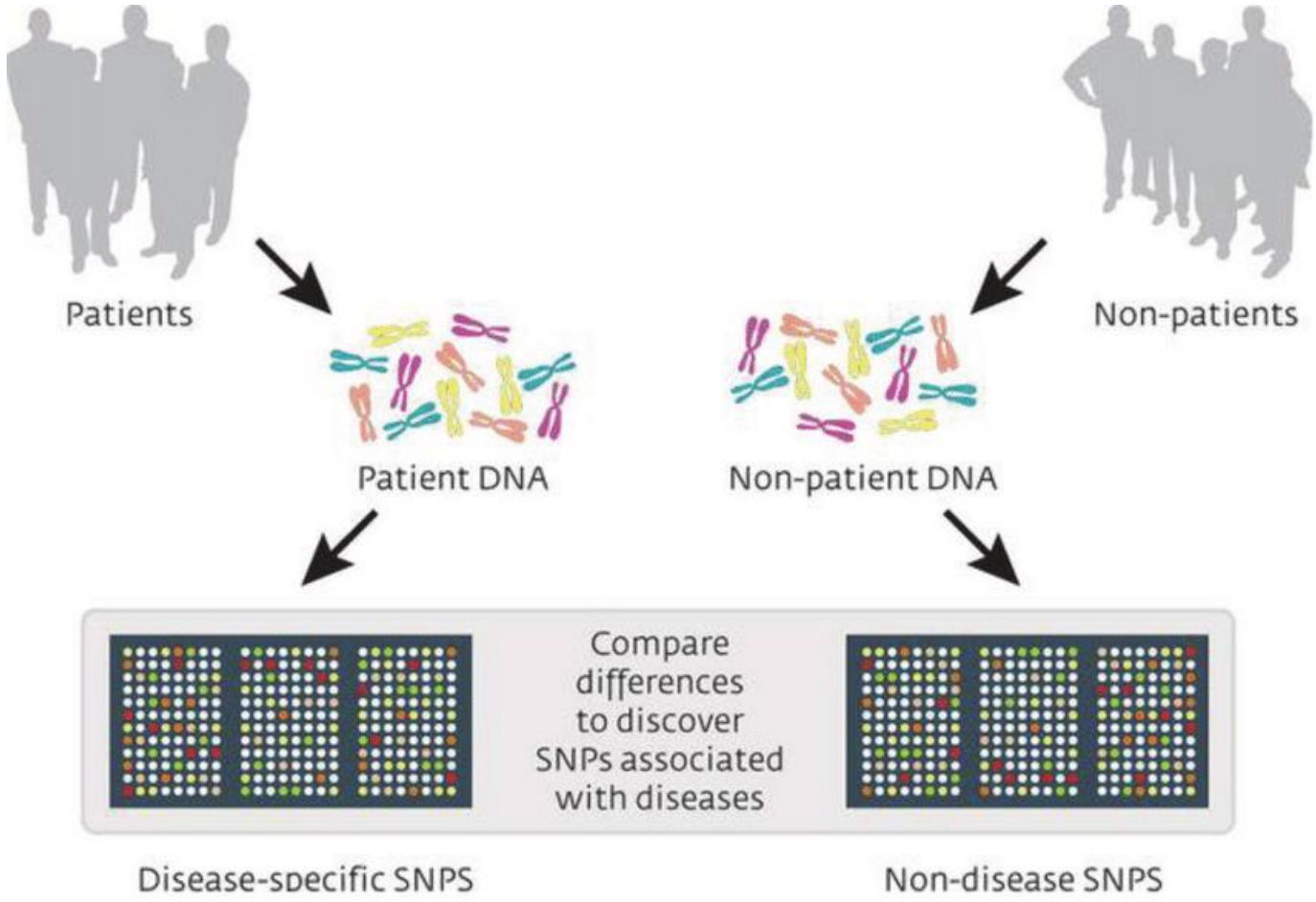
May 18th, 2021

What is a GWAS?

Genotype individuals for a large number of SNPs. Search for an association between those SNPs and a phenotype.

In humans used to find disease variants

https://www.ebi.ac.uk/gwas/diagram



Seems simple….

Plant	Sequence	Phenotype
1	GAATTCCGCAATGCAGGTTAAGAGCTCTGTGAAAGAGGAAAACGAAAAAC	1
2	GAATTCCGCAATGCAGGTTAAGAGCTCTGTGAAAGAGGAAAACGAAAAAC	1
3	GAATTCCGCAATGCAGGTTAAGAGCTCTGTGAAAGAGGAAAACGAAAAAC	1
4	GAATTCCGCAATGCAGGTTAAGAGCTTTGTGAAAGAGGAAAACGAAAAAC	2
5	GAATTCCGCAATGCAGGTTAAGAGCTCTGTGAAAGAGGAAAACGAAAAAC	1
6	GAATTCCGCAATGCAGGTTAAGAGCTCTGTGAAAGAGGAAAACGAAAAAC	1
7	GAATTCCGCAATGCAGGTTAAGAGCTTTGTGAAAGAGGAAAACGAAAAAC	2
8	GAATTCCGCAATGCAGGTTAAGAGCTCTGTGAAAGAGGAAAACGAAAAAC	1

Figure modified from here: https://warwick.ac.uk/fac/sci/lifesci/research/vegin/geneticimprovement/geneticmarker/snp/

Problem with false positives

At a 95% significance rate we expect to find 5% of our data to be “significant”

Not all of the variants detected will be biologically meaningful.

- But some of them may be and can be verified via functionalization studies

We can limit the number of false positives by test corrections

- FDR (False Discovery Rate) one of the most common approaches

Problem with population structure

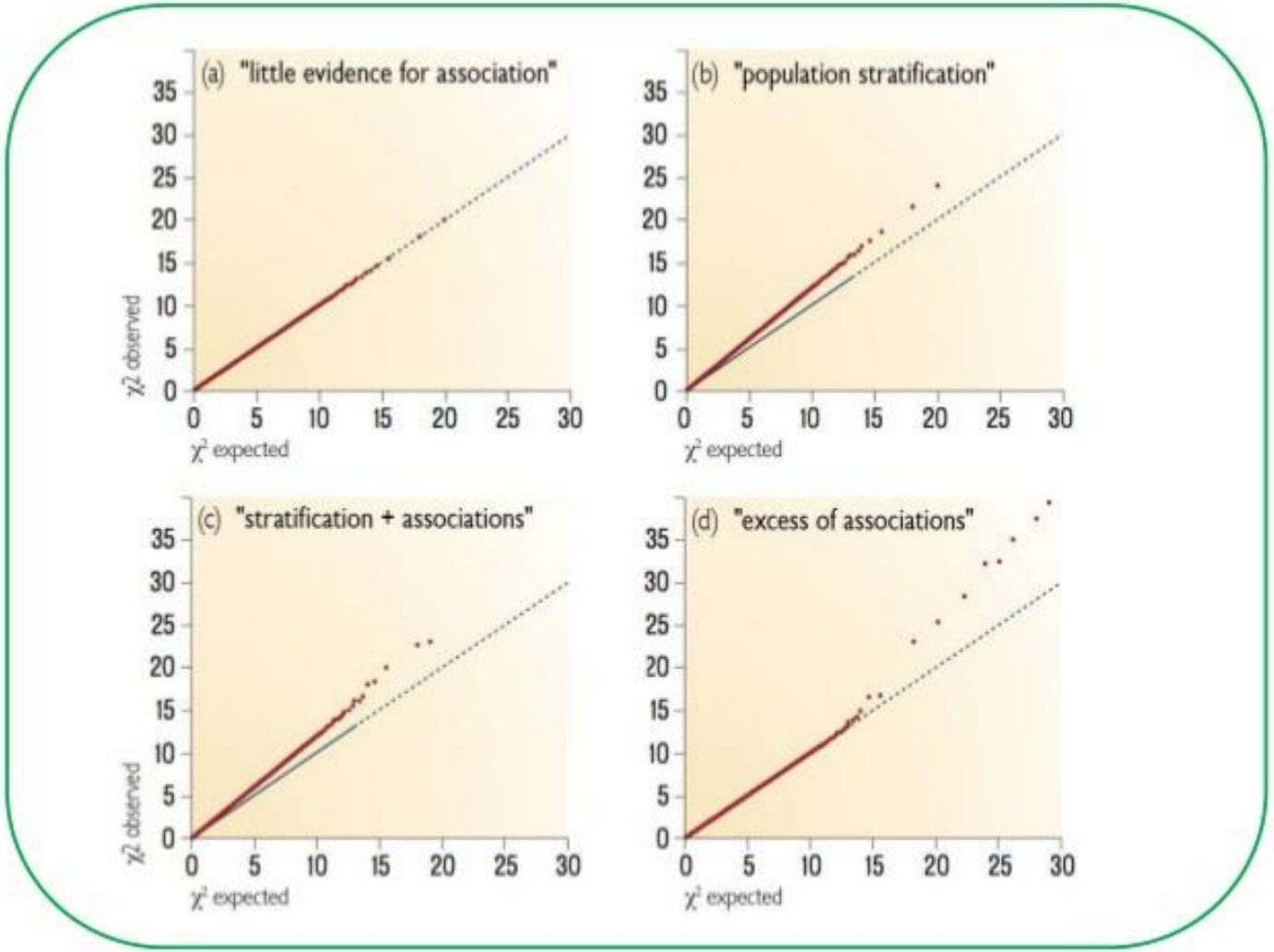


Figure credit: https://www.slideshare.net/LekkiFrazierWood/lecture-7-gwas-full

Ways to get around population structure:

sample one population
may miss informative SNPs

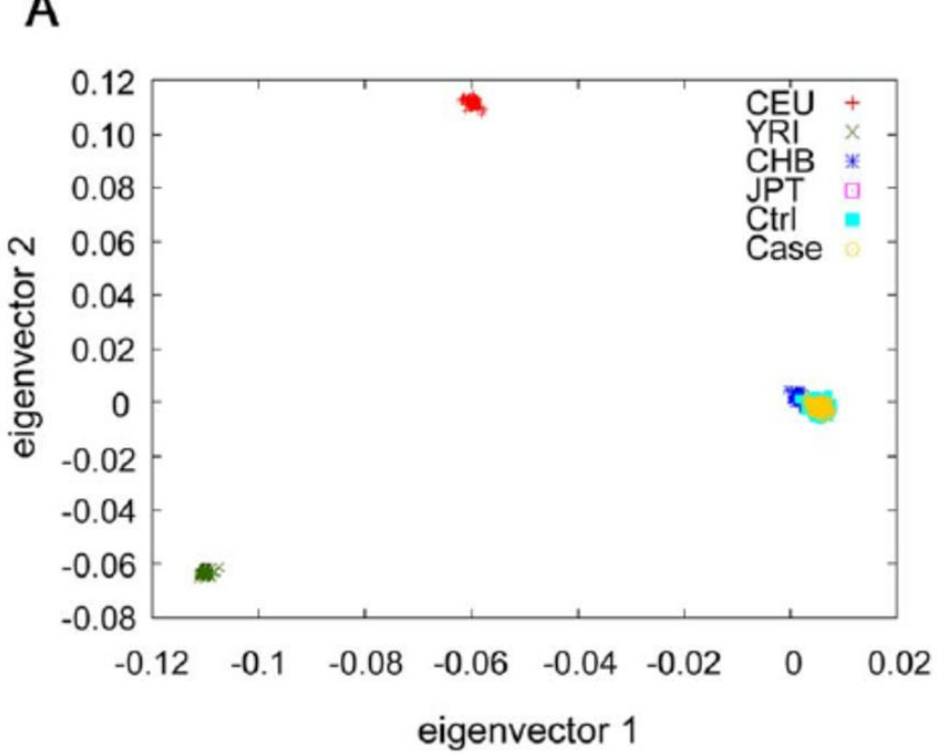
Critical for human data for disease variants
need to sample multiple populations

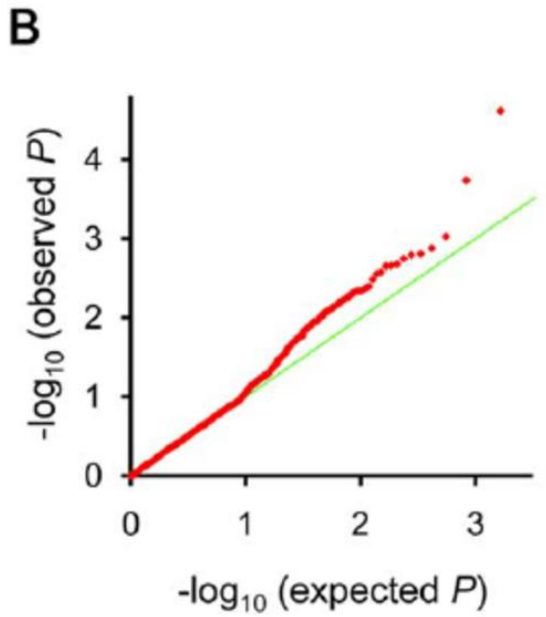
Must incorporate stratification into your model

Another example of population stratification

Common Variants in a Novel Gene, FONG on Chromosome 2q33.1 Confer Risk of Osteoporosis in Japanese

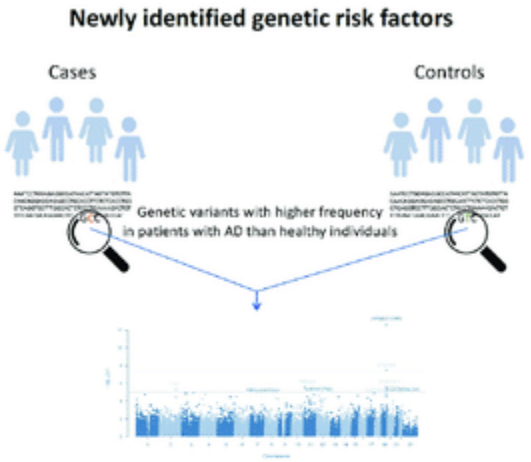
Kou et al. 2011 PloS ONE

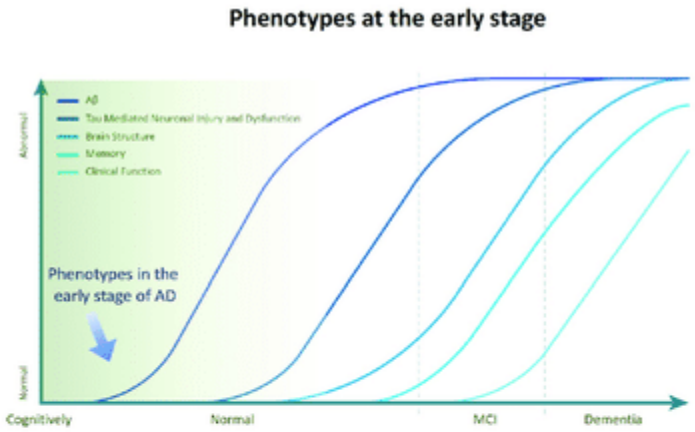


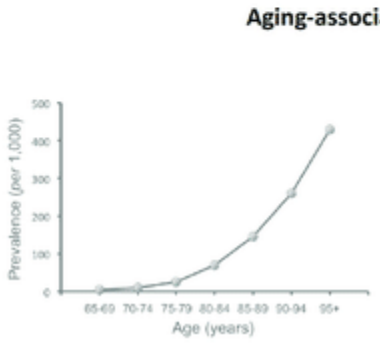


Commonly used terms

- Risk factor
- Effect size







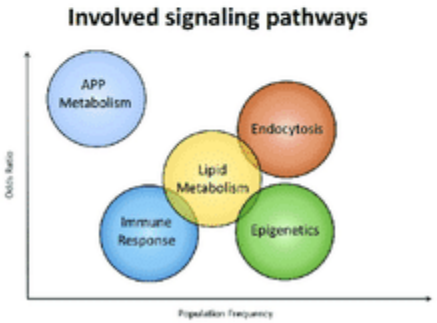


Figure credit: Jung et al 2018 Genomics: New Light on Alzheimer's Disease Research

Article | Open Access | Published: 04 April 2018

Evolution of Sex Determination Loci in Atlantic Salmon

James Kijas, Sean McWilliam, Marina Naval Sanchez, Peter Kube, Harry King, Bradley Evans, Torfinn Nome, Sigbjørn Lien & Klara Verbyla

Scientific Reports **8**, Article number: 5664 (2018) | Cite this article

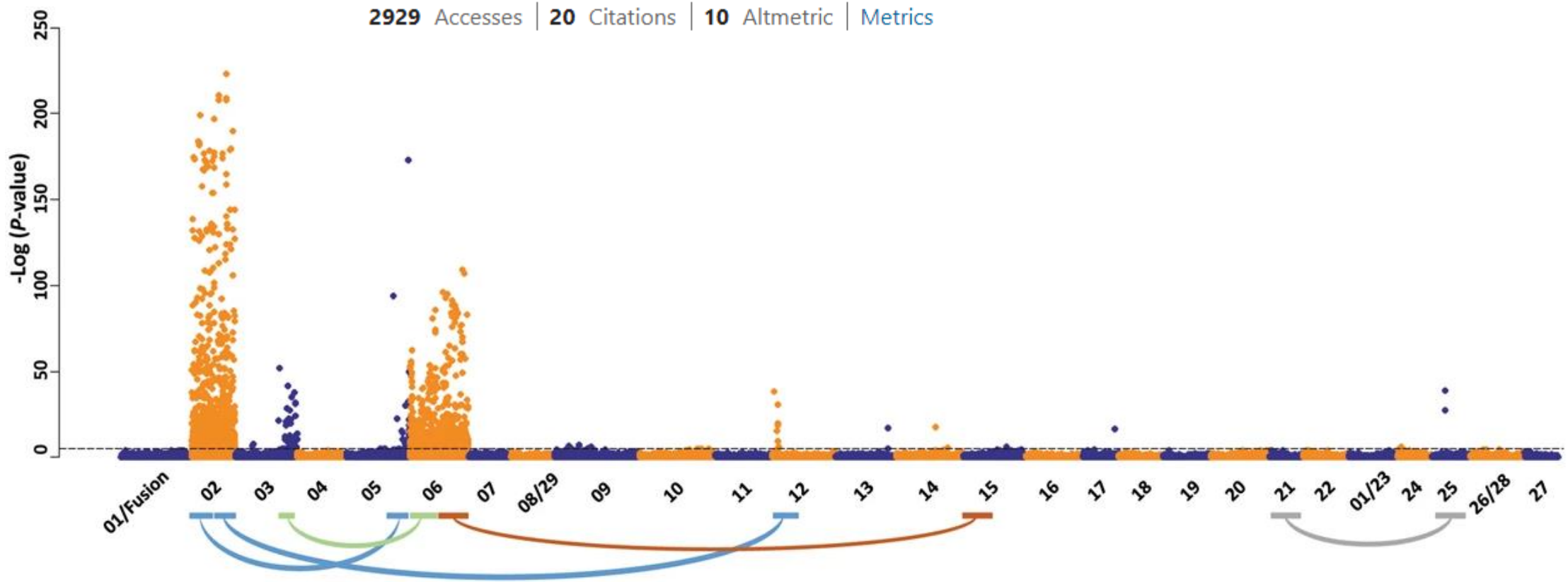


Table 1 Salmon used for whole genome sequencing. Phenotypic assignment of sex (PSEX) is given along with the sex lineage (SL) assigned by either simple sequence repeat (SSR) segregation data[7] or using whole genome sequence data in this study (SL_WGS). Missing data is indicated as 'nd'. The depth of coverage and number of SNP per animals is given following application of quality filtering.

From: Evolution of Sex Determination Loci in Atlantic Salmon

Animal	PSEX	SL_SSR	SL-WGS	Coverage	SNP
1_2012	M	Ssa06	Ssa03/06	44.0	6,927,595
2_2005	M	nd	Ssa03/06	48.5	7,010,428
3_2005	F	Ssa06	nd	42.0	6,901,270
5_2006	F	nd	nd	43.6	6,920,473
6_2007	M	Ssa03	Ssa03/06	42.8	6,926,011
7_2007	M	Ssa02	Ssa02	47.1	6,955,804
8_2007	F	Ssa03	nd	32.9	6,674,337
9_2009	F	Ssa06	nd	37.1	6,857,336
10_2013	M	Ssa06	Ssa03/06	28.9	6,567,065
11_2006	M	Ssa06	Ssa03/06	36.6	6,774,991
12_2007	M	Ssa06	Ssa03/06	36.2	6,743,704
13_2009	F	Ssa02	nd	33.6	6,762,881
14_2010	F	Ssa02	nd	38.0	6,796,035
15_2011	M	Ssa02	Ssa02	39.6	6,869,394
16_2012	F	Ssa06	nd	44.1	6,969,672
17_2013	M	nd	Ssa03/06	48.2	7,012,754
18_2013	M	Ssa03	Ssa03/06	44.7	6,959,577
19_2015	nd	nd	Ssa02	45.0	6,978,524
20_2015	nd	nd	nd	47.9	6,997,752