

Computational Systems Biology Deep Learning in the Life Sciences

6.802 6.874 20.390 20.490 HST.506

David Gifford
Lecture 8
April 4, 2017

Genotype to Phenotype



Massachusetts
Institute of
Technology

<http://mit6874.github.io>

Overall goal for today

- Identify genetic variants associated with traits as well as methods to measure the completeness of the variants we discover
- Identify genetic variants associated with traits in human where causal reasoning is more difficult

Today's lecture

- Fundamentals of heritability
 - Narrow sense heritability
 - Broad sense heritability
- Finding eQTLs in Yeast
 - Linear models
 - What do do when linear models break
- Finding eQTLs in human
 - Going beyond association

Part 1 - Model Organism Genetics

Today's Narrative Arc

1. Usually, you are more like your relatives than random people on the planet.
2. The heritability of a trait is the fraction of phenotypic variance that can be explained by genotype
3. Computational models that predict phenotype from genotype are key for understanding disease related genomic variants and the most effective therapy for a disease (pharmacogenomics)
4. We will computationally predict quantitative phenotypes by adding the contribution of individual loci (QTLs)
5. Typically our models can only predict a small fraction of phenotypic variance – the so called “missing heritability” problem

Today's Computational Approaches

1. Linear models of phenotype that use stepwise regression and forward feature selection
2. Test statistics for discovering significant QTLs
3. Measurement of narrow sense heritability (h^2), broad sense heritability (H^2), and environmental variance

OMIM - authoritative compendium of human genes and genetic phenotypes related to Mendelian Inheritance

The screenshot shows the OMIM website interface. At the top, there's a navigation bar with links for About, Statistics, Downloads, Contact Us, MIMmatch, Donate, Help, and a search icon. Below the navigation is a search bar with placeholder text "Search OMIM..." and a magnifying glass icon. To the right of the search bar is an "Options" dropdown menu.

OMIM Entry Statistics

Number of Entries in OMIM (Updated March 31st, 2017) :

| MIM Number Prefix | Autosomal | X Linked | Y Linked | Mitochondrial | Totals |
|---|-----------|----------|----------|---------------|--------|
| Gene description * | 14,744 | 717 | 49 | 35 | 15,545 |
| Gene and phenotype, combined + | 77 | 0 | 0 | 2 | 79 |
| Phenotype description, molecular basis known # | 4,620 | 319 | 4 | 31 | 4,974 |
| Phenotype description or locus, molecular basis unknown % | 1,479 | 124 | 5 | 0 | 1,608 |
| Other, mainly phenotypes with suspected mendelian basis | 1,676 | 111 | 2 | 0 | 1,789 |
| Totals | 22,596 | 1,271 | 60 | 68 | 23,995 |

NOTE: OMIM is intended for use primarily by physicians and other professionals concerned with genetic disorders, by genetics researchers, and by advanced students in science and medicine. While the OMIM database is open to the public, users seeking information about a personal medical or genetic condition are urged to consult with a qualified physician for diagnosis and for answers to personal questions.

OMIM® and Online Mendelian Inheritance in Man® are registered trademarks of the Johns Hopkins University.
Copyright © 1966-2017 Johns Hopkins University.

Statistics review

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)^2 = E[(X - \mu_x)^2]$$

$$\sigma_{xy}^2 = E[(X - \mu_x)(Y - \mu_y)]$$

Covariance
=0 when X and Y are independent

Genotype to Phenotype

- Genotype
 - Complete genome sequence (or an approximation)
 - Can be defined by markers at specific genomic sites that describe differences with a defined reference genome
- A phenotype is defined by one or more traits
 - Non-quantitative trait (dead/alive, etc.)
 - Quantitative Trait
 - Fitness (growth rate, lifespan, etc.)
 - Morphology (height, etc.)
 - Gene expression
- Quantitative Trait Loci – Genetic marker that is associated with a quantitative trait
 - eQTL – marker associated with gene expression

Common disease Common variant Model

- Assumes that common variants cause common phenotypes
- With this assumption we can only examine common variants; saves expense and complexity
- When we completely sequence our F1s we do not make this assumption

Why two heritabilities?

- Broad-sense
 - Describes the upper bound for phenotypic prediction by an optimal arbitrary model
 - Reveals complexity of molecular mechanism
- Narrow-sense
 - Describes the upper bound for phenotypic prediction by a linear model
 - Describes relative resemblance and utility of family disease history
 - Efficient genetic mapping studies

Key caveats

- Heritability is a property of population (segregating allele frequencies) and environment (noise component)
- “Heritability” in practice may refer to either broad- or narrow-sense (or an implicit assumption that they are the same)
- Estimation is difficult (matching environments and avoiding confounding)

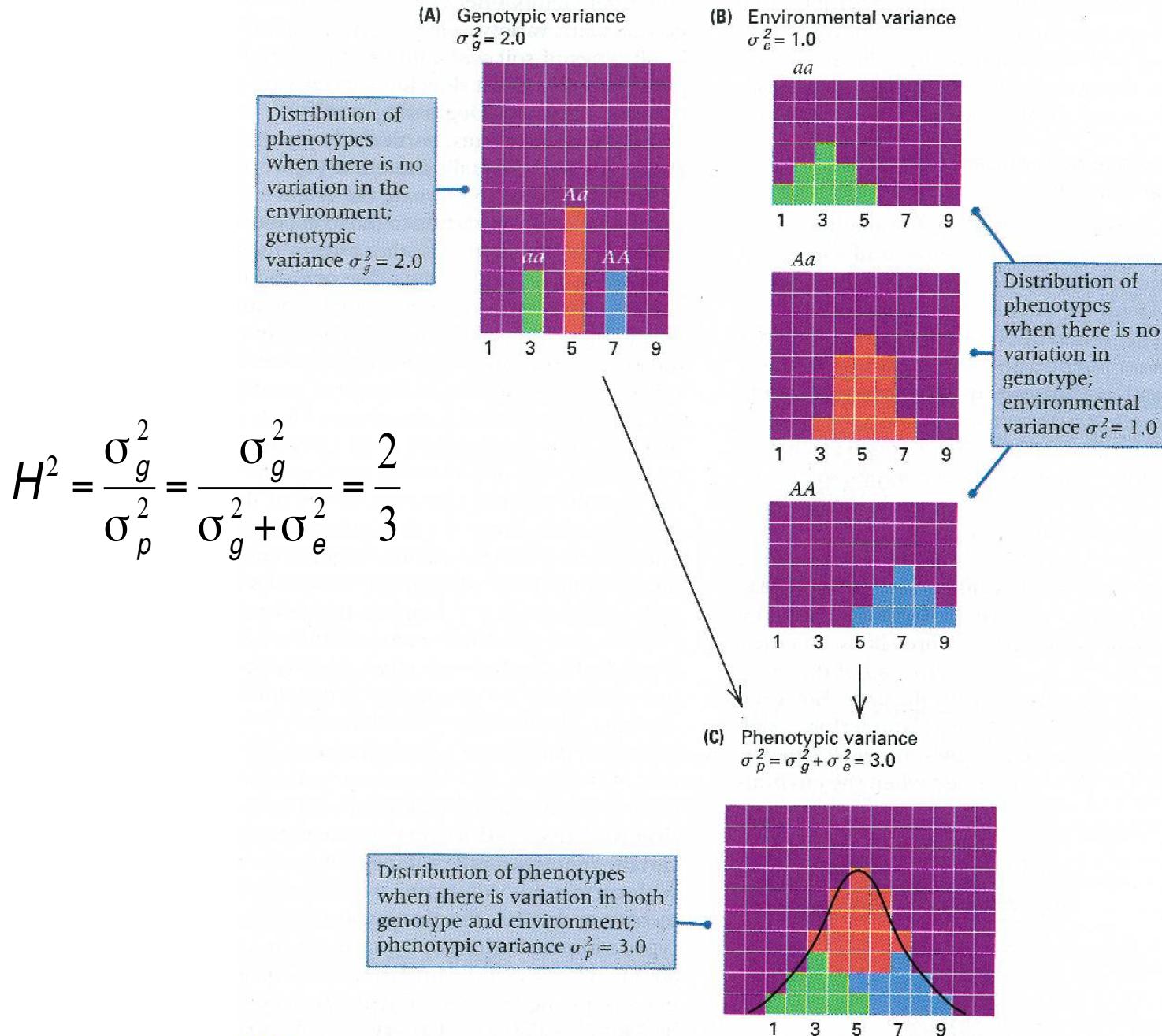
H^2 - Broad Sense heritability

- Fraction of phenotypic variance explained by genetic component

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} = \frac{\sigma_p^2 - \sigma_e^2}{\sigma_p^2}$$

- Can estimate σ_e^2 from identical twins or clones.

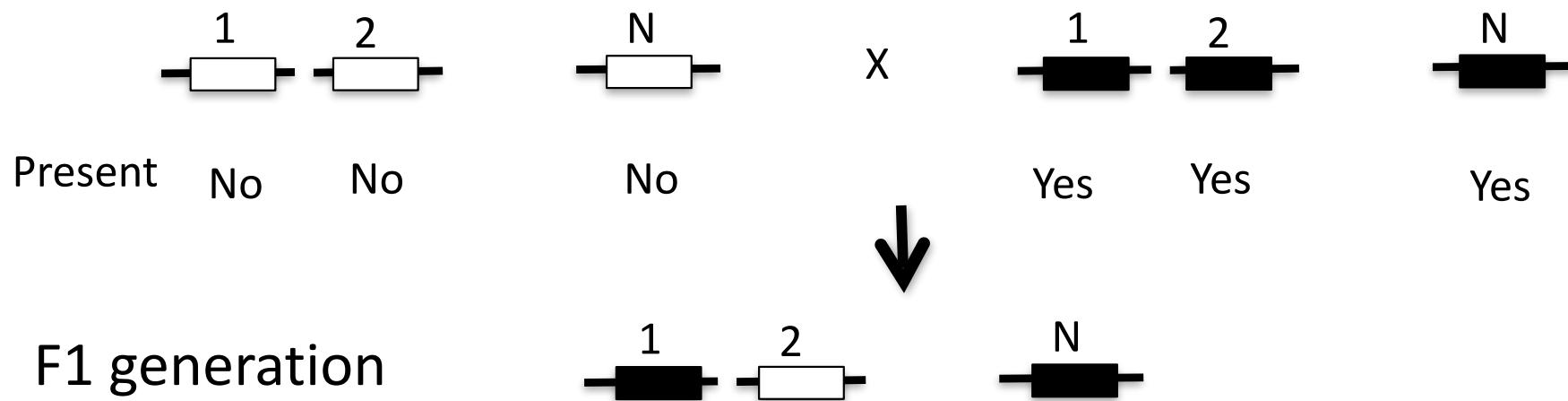
Broad heritability of a trait is fraction of phenotypic variance explained by genetic causes



h^2 - Narrow Sense heritability

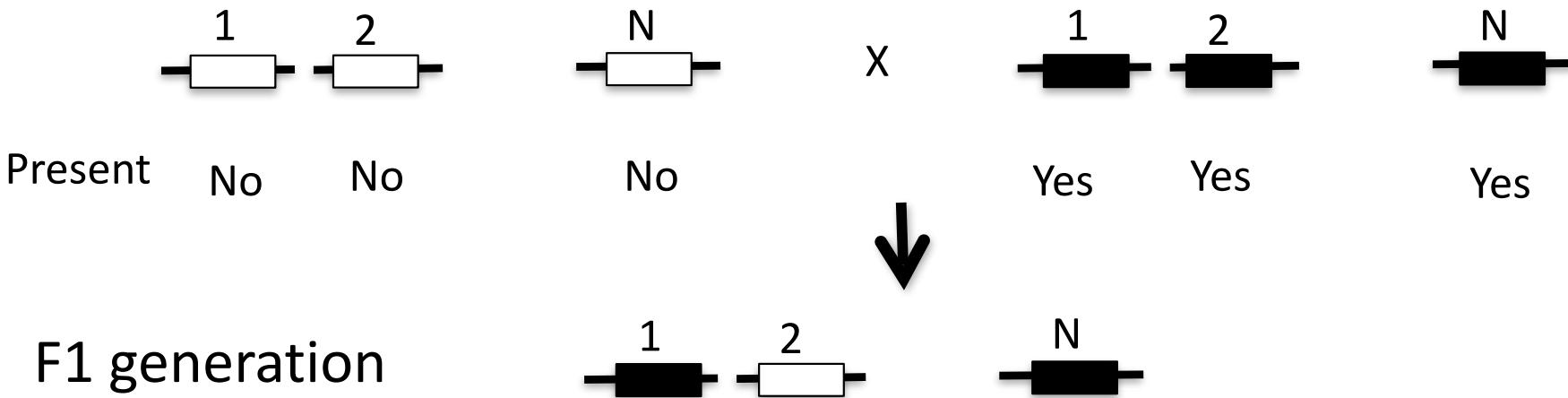
- Fraction of phenotypic variance explained by an additive model of markers
- $f_a(g_i)$ is additive model of genotypic components in g_i
- Difference between heritability explained by additive model and general model is one source of “missing heritability” in current studies

Binary haploid genetic model



Example Phenotype
Alive/Dead in a specific environment

Binary haploid genetic model



Example Phenotypes

Alive/Dead in a specific environment

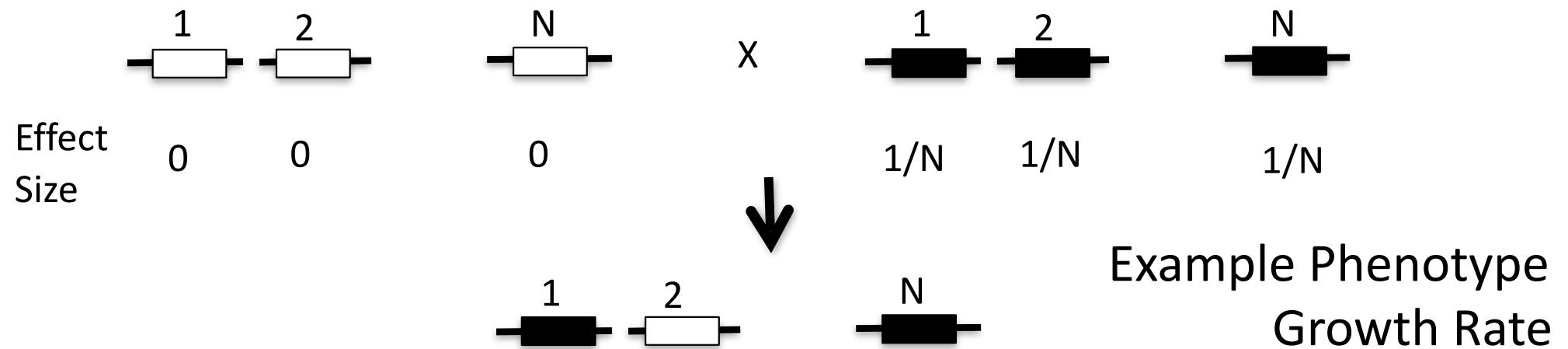
Resistant to a specific virus

Suppose we tested 128 F1s, 16 resistant.

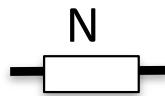
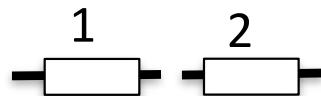
What is your estimate of N?

N is estimated by $\log_2 (\# \text{ F1s tested} / \# \text{ F1s with phenotype})$

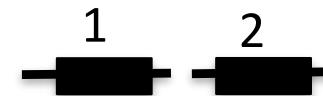
Quantitative haploid genetic model



Quantitative haploid genetic model



x



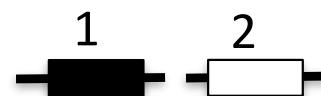
Effect
Size

0 0

0

$1/N$ $1/N$

$1/N$



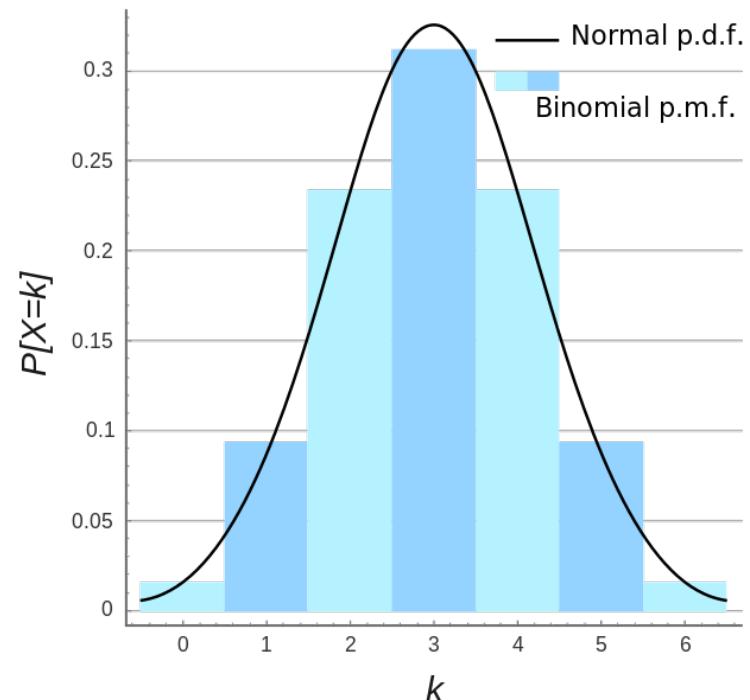
Example Phenotype
Growth Rate

$$y = x / N$$

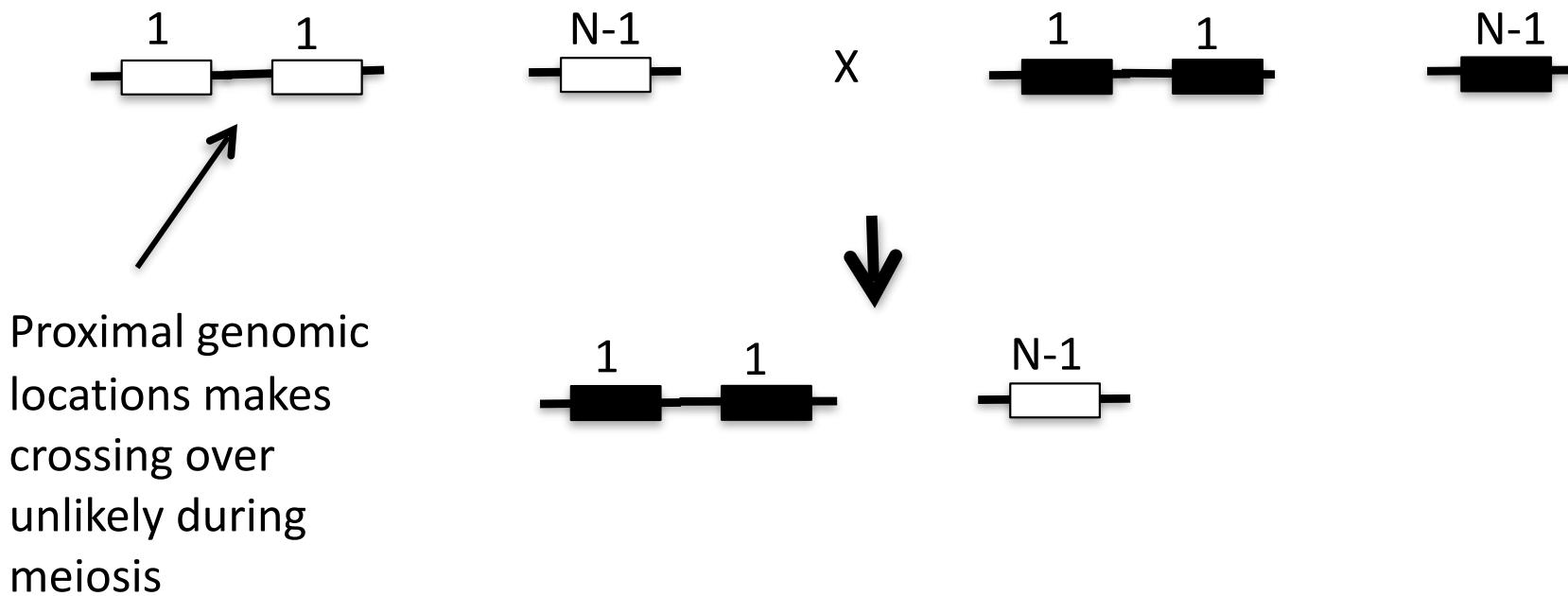
$$p(x, N) = \binom{N}{x} (1 - .5)^{N-x} .5^x$$

$$E[x] = N/2 \quad E[y] = 1/2$$

$$\sigma_x^2 = N/4 \quad \sigma_y^2 = 1/(4N)$$



Genetic linkage causes marker correlation



Phenotype is a function of genotype plus an environmental component

- i – individual in $[1 .. N]$
- g_i – genotype of individual i
- p_i – quantitative phenotype of individual i (single trait)
- e_i – environmental contribution to p_i

Phenotype is a function of genotype plus an environmental component

- i – individual in $[1 .. N]$
- g_i – genotype of individual i
- p_i – quantitative phenotype of individual i (single trait)
- e_i – environmental contribution to p_i

$$p_i = f(g_i) + e_i \quad \sigma_p^2 = \frac{1}{N} \sum_{i=1}^N (p_i - \mu_p)^2$$

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2 + 2\sigma_{ge}^2 \quad E[e] = 0 \quad E[e^2] = \sigma_e^2$$

g and e assumed or made independent yields

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

Additive model of phenotype

g_{ij} is marker j for individual i with values {0,1}

Quantitative trait loci (QTLs) are discovered for each trait

$$f_a(g_i) = \sum_{j \in QTL} \beta_j g_{ij} + \beta_0$$

$$E[f_a(g_i)] = \frac{f_a(p_1)}{2} + \frac{f_a(p_2)}{2}$$

Children tend to midpoint of parents for additive traits as they are expected to get an equal number of loci from each parent

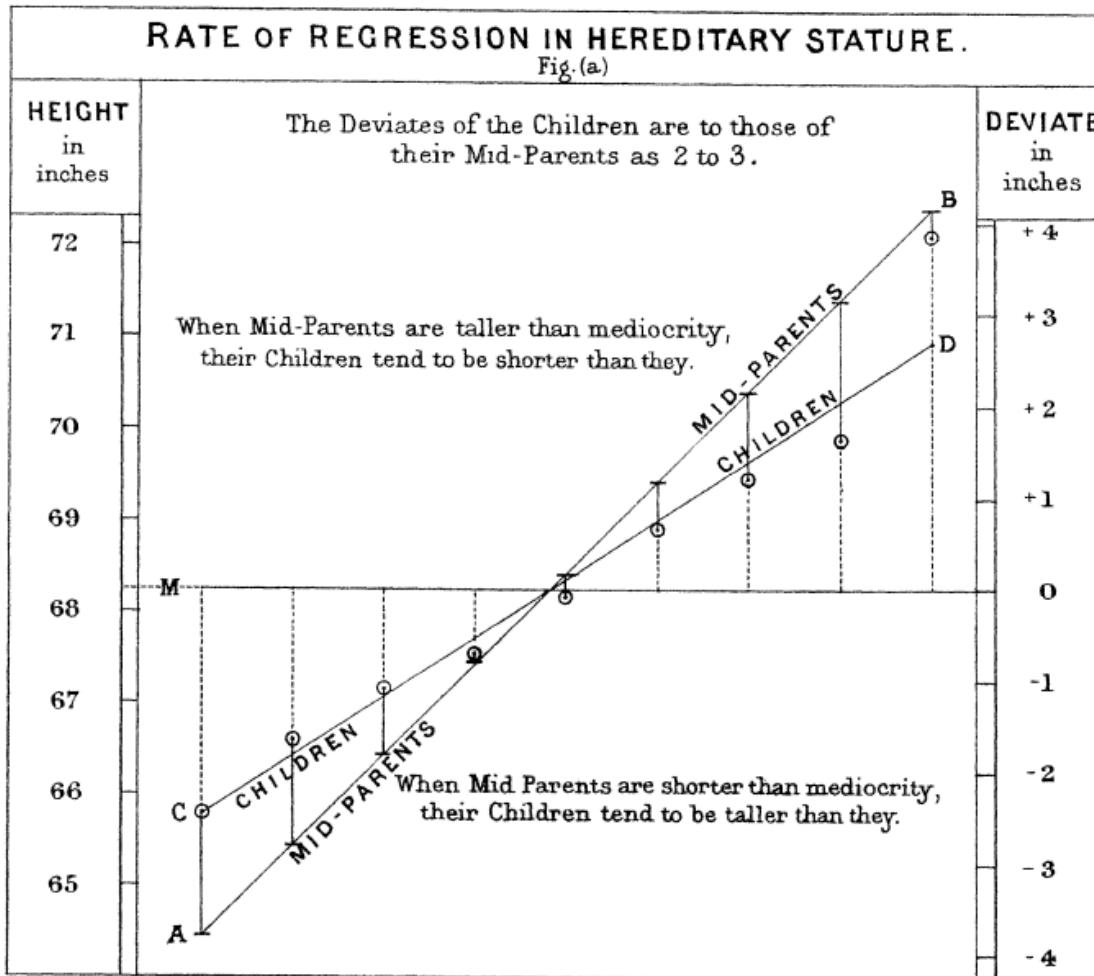
h^2 - Narrow Sense heritability

- Fraction of phenotypic variance explained by an additive model of markers
- $f_a(g_i)$ is additive model of genotypic components in g_i
- Difference between heritability explained by additive model and general model is one source of “missing heritability” in current studies

$$p_i = f_a(g_i) + e_i \quad \sigma_a^2 = \sigma_p^2 - \frac{1}{N} \sum_{i=1}^N (p_i - f_a(g_i))^2$$

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2}$$

Historical heritability example



Galton, "Regression towards mediocrity in hereditary stature" (1886)

Example trait heritabilities – h^2

Morphological Traits

Human height ~ .8

Cattle Yearling Weight ~ .35

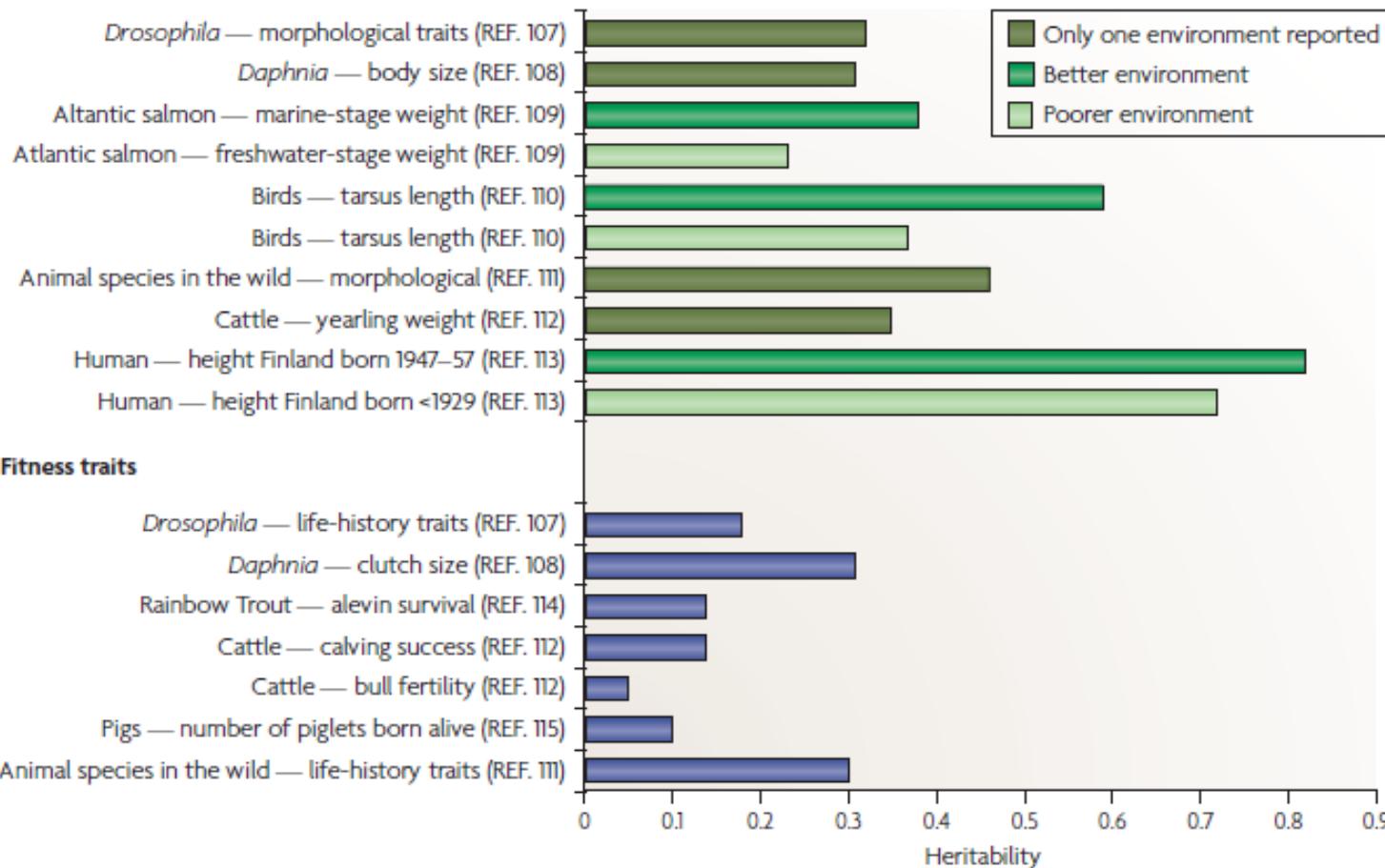
Fitness Traits

Drosophila life history ~ .2

Wild animal life history ~ .3

Example trait heritabilities

Morphological traits



Fitness traits

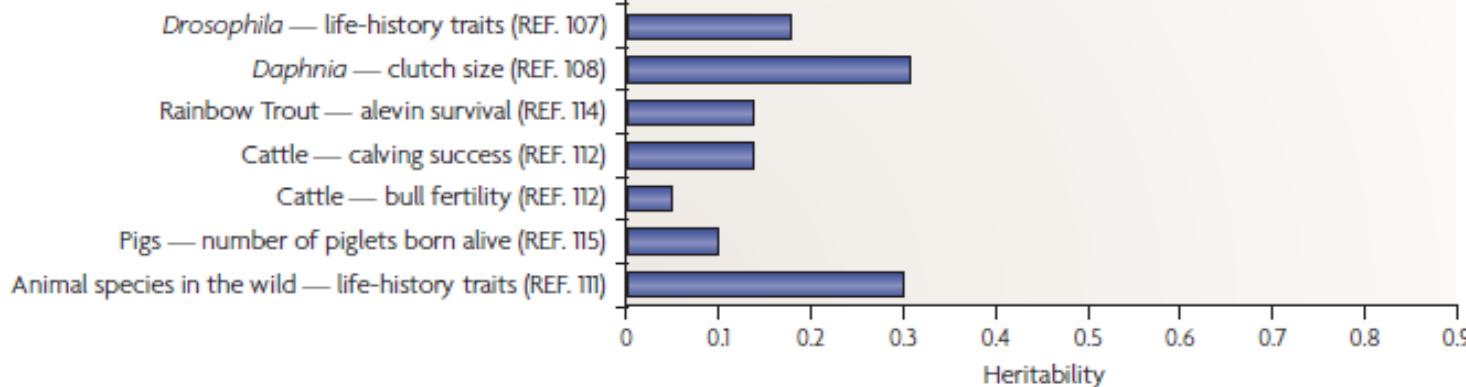


Figure 1 | Examples of estimates of heritabilities of morphological and fitness traits. Where possible, the estimates of heritability were taken from Reviews, and are the mean across a number of studies. The examples show that, on average, heritability estimates are larger for morphological traits than for fitness-related traits, and that heritability tends to be larger in better environments when compared with poorer environments.

h^2 from Visscher et al. 2008

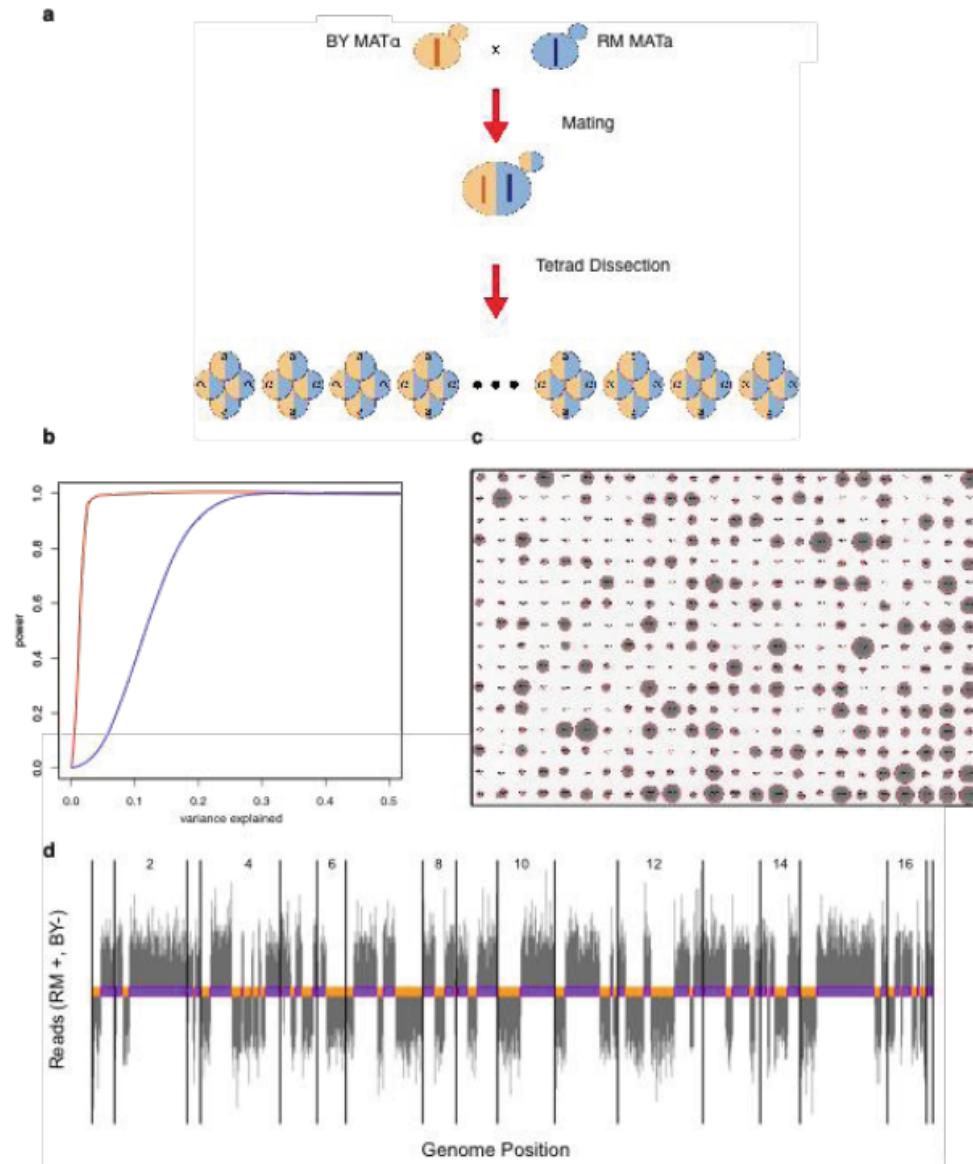
Can we predict phenotype in a haploid yeast system?

Finding the sources of missing heritability in a yeast cross

Joshua S. Bloom^{1,2}, Ian M. Ehrenreich^{1,3}, Wesley T. Loo^{1,2}, Thúy-Lan Võ Lite^{1,2} & Leonid Kruglyak^{1,4,5}

NATURE | VOL 494 | 14 FEBRUARY 2013

Study heritability of 46 traits in ~1000 segregants

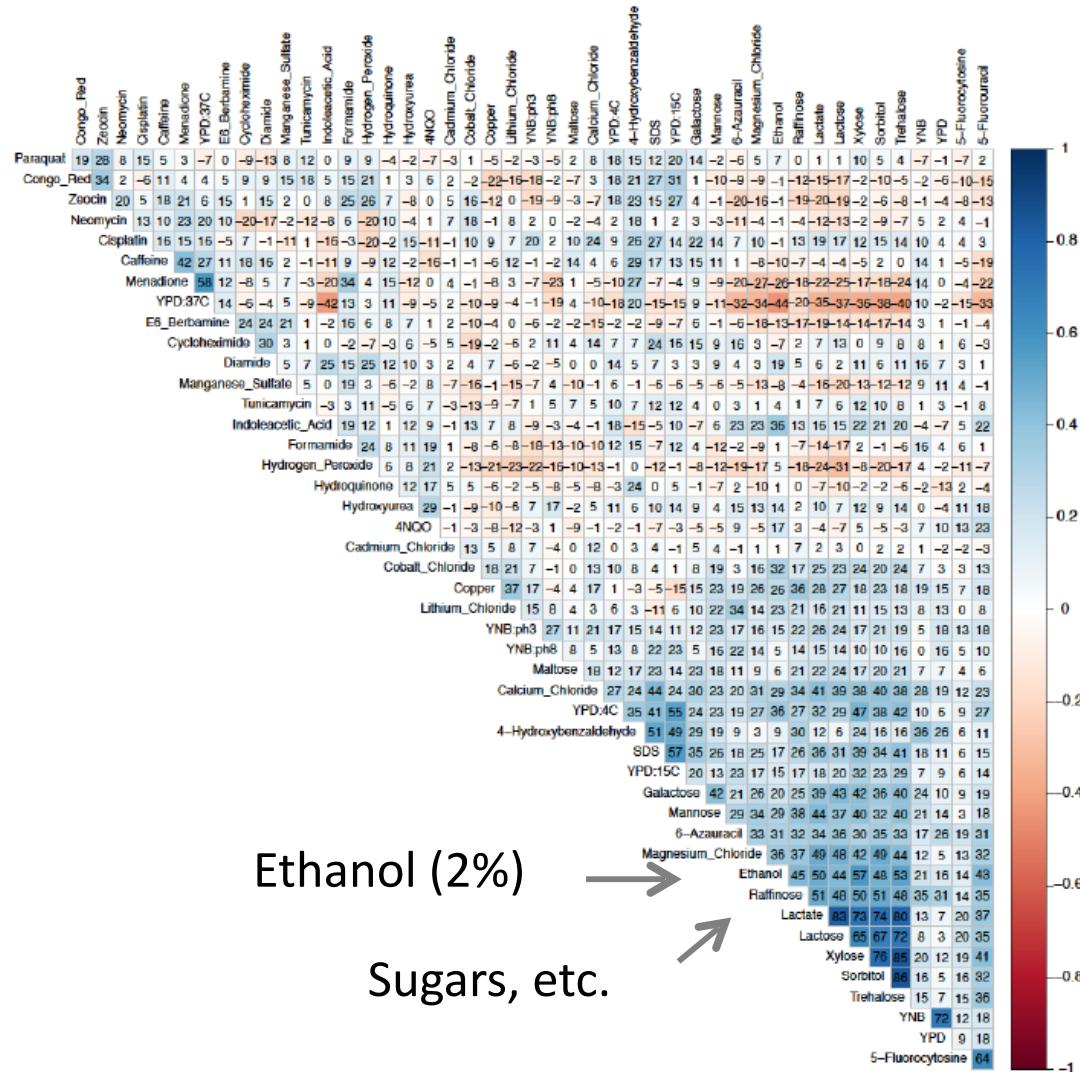


Key advance: large panel (~1000 segregants), many phenotypes (46)

BY and RM parents

Figure S1. The design of the segregant panel is shown in (A). (B) Curves illustrating statistical power are shown for mapping populations of 100 (blue) and 1000 (red) segregants at a genome-wide significance threshold. (C) An image of endpoint colony growth is shown for 384 segregants, with the outlines of colonies, as detected by our image processing software, indicated in red. (D) Counts of sequencing reads at SNP sites are plotted (Y-axis) against genome position (X-axis) for a representative segregant; the orange (BY) and purple (RM) bars indicate parental haplotype calls, and the vertical black bars delineate chromosomes.

Certain phenotypes are related



Bloom et al. 2013

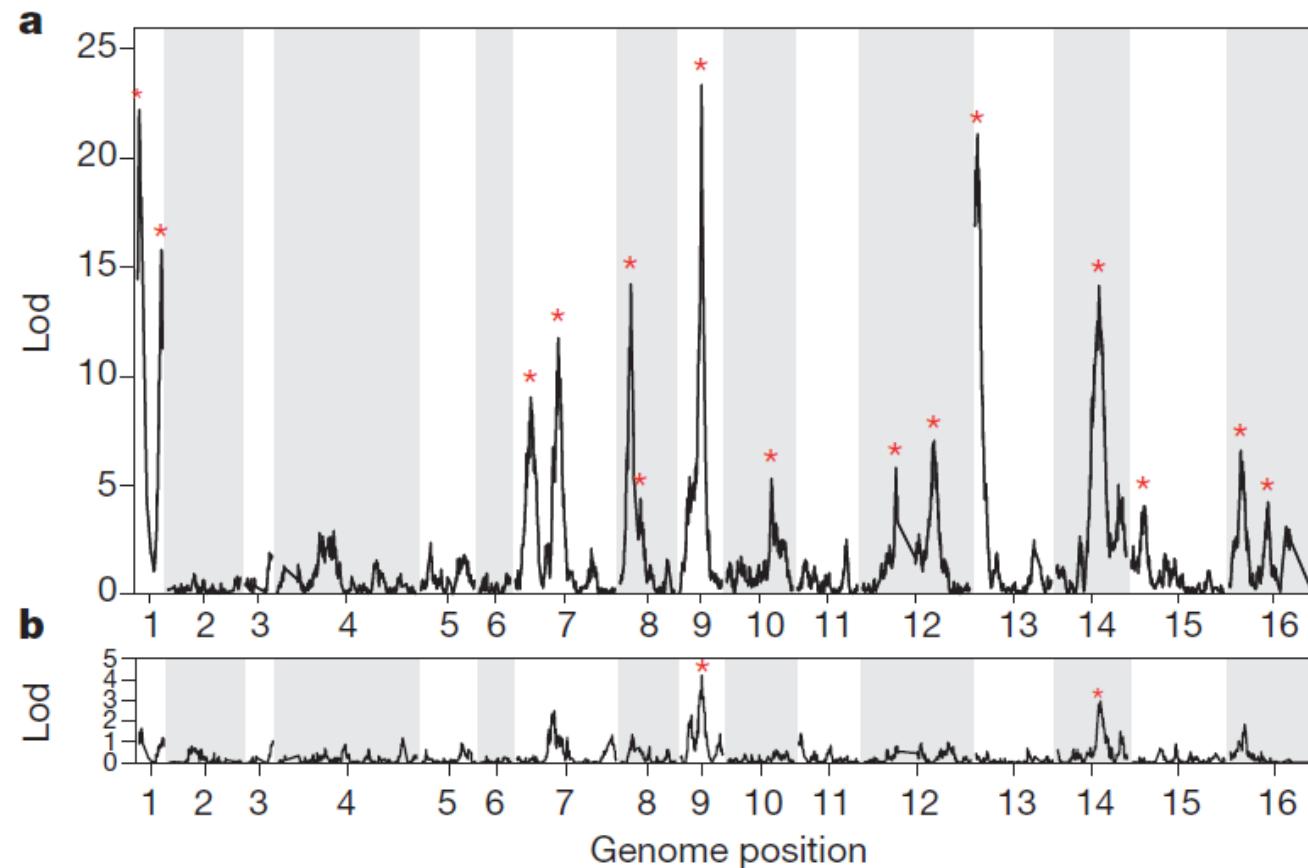
Figure S2. Spearman correlation coefficients for all pairs of traits are shown. Numbers in table cells indicate (100 * correlation coefficient).

LOD scores to discover QTLs

$$LOD = \log_{10} \prod_{i=1}^N \frac{P(p_i | g_{ij}, \mu_0, \mu_1, \sigma)}{P(p_i | \mu, \sigma)}$$

- Use trait means conditioned on marker j in individual i vs. unconditioned mean for trait to test if marker j is a QTL
- Permute genotypes 1000 times and each time compute LOD scores to estimate null LOD distribution
- Determine null LOD score that describes FDR = 0.05
- Use this threshold on unpermuted LOD scores to find QTLs for each trait
- Fit linear model to discovered QTLs
- Repeat finding QTLs predicting residuals from existing model (3 times)

1005 segregants detect more QTLs than 100 segregants



Bloom et al. 2013

Figure 3 | QTL detection for a complex trait. LOD score is plotted against the genetic map. Red asterisks indicate statistically significant QTL. **a**, LOD score plot with 1,005 segregants for growth in E6 berbamine. **b**, LOD score plot with 100 segregants for growth in E6 berbamine. The 15 significant QTL in **a** explain 78% of the narrow-sense heritability, compared with 21% for the 2 significant QTL in **b**. Alternating shaded bands denote chromosome boundaries.

Phenotype prediction works well with identified QTLs

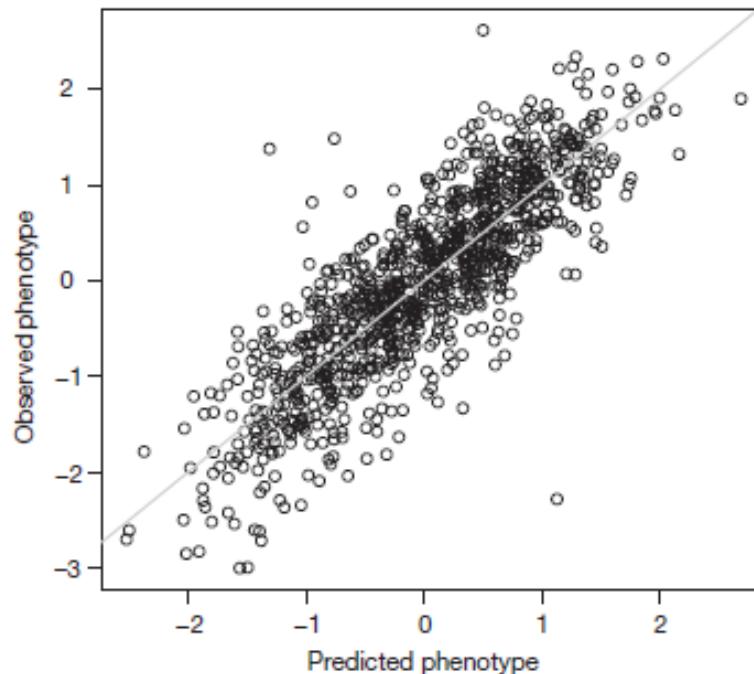


Figure 4 | Prediction of segregant trait values from QTL phenotypes. The observed phenotypic values for growth in lithium chloride are plotted against the predicted phenotypic values based on a cross-validated additive model of 22 QTL. The additive QTL model explains 88% of the narrow-sense heritability. The diagonal line represents $(\text{observed phenotype}) = (\text{predicted phenotype})$ and is shown as a visual guide.

Most identified QTLs have small effects

5-29 QTLs per trait
(median of 12),
reported at 5% FDR

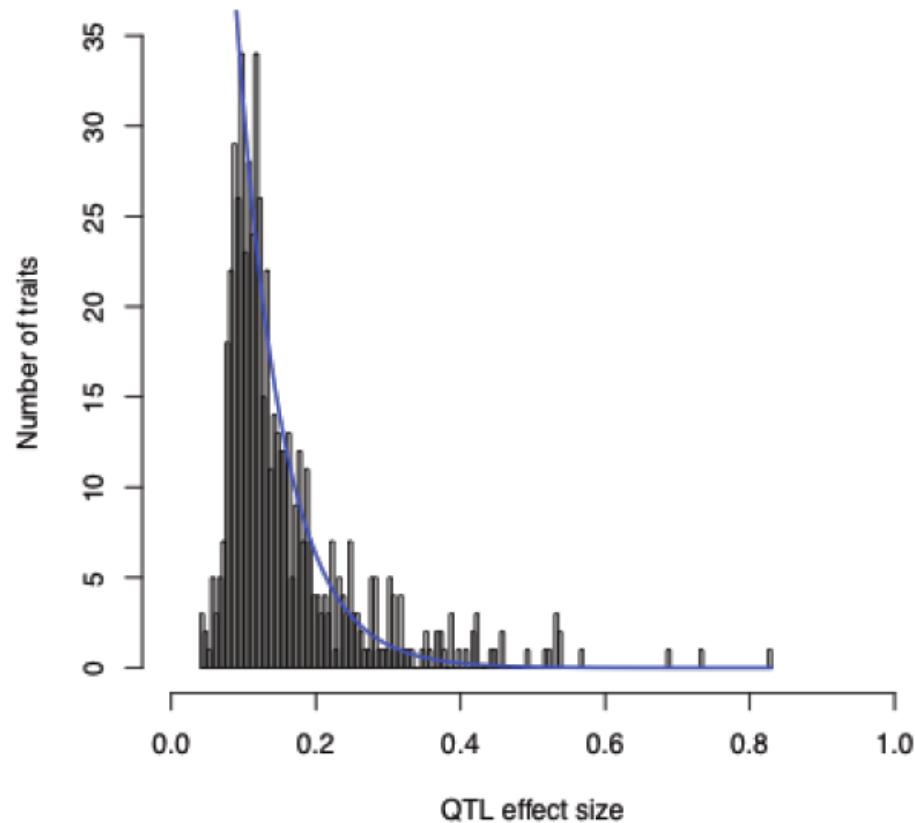
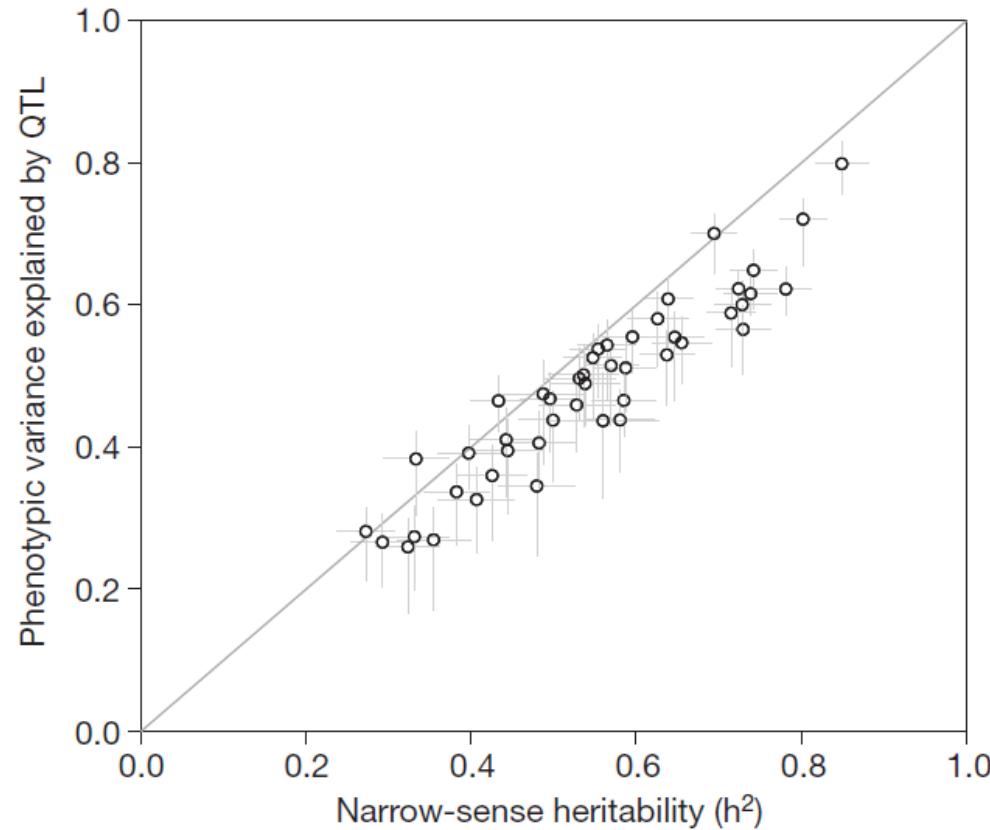


Figure S3. A histogram of QTL effect sizes across all traits is plotted, showing that most detected QTL have small effects. Effect size here is the absolute value of the standardized difference in allelic means for each QTL. The blue line indicates a fit of a truncated exponential distribution of effect sizes.

Identified QTLs explain most additive heritability

QTLs explain
72-100% of
narrow-sense
heritability

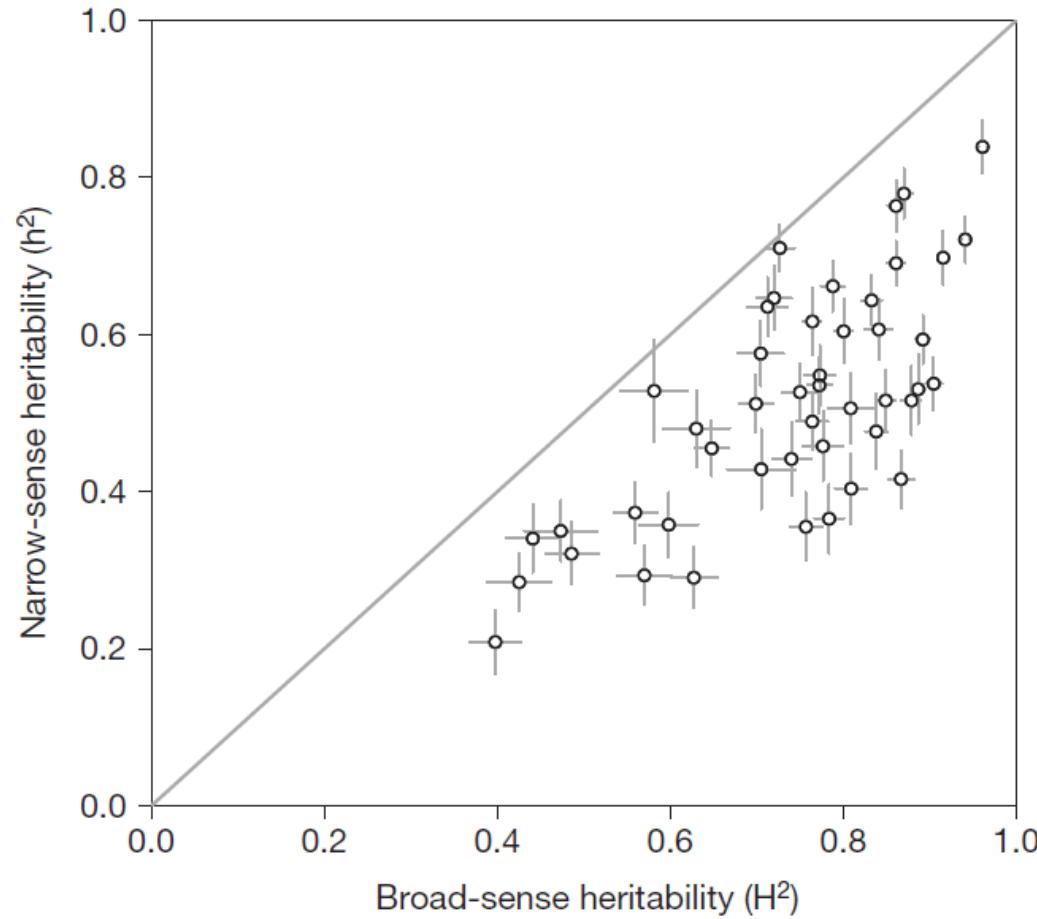


Bloom et al. 2013

Figure 2 | Most additive heritability is explained by detected QTL. a, The total variance explained by detected QTL for each trait is plotted against the narrow-sense heritability (h^2). Error bars show \pm s.e. The diagonal line represents $(\text{variance explained by detected QTL}) = h^2$ and is shown as a visual guide.

“Missing Heritability” exists with our linear model

Vertical gap
represents non-
additive
genetic
contributions



Bloom et al. 2013

Figure 1 | Heritability for 46 yeast traits. The narrow-sense heritability (h^2) for each trait is plotted against the broad-sense heritability (H^2). Error bars show \pm s.e. in heritability estimates. The diagonal line represents $h^2 = H^2$ and is shown as a visual guide.

What causes missing heritability?

Possible explanations (non-exclusive):

- Incorrect heritability estimates
- Non-chromosomal elements
- Rare variants
- Structural variants
- Many common variants of low effect
- Epistasis

What causes missing heritability?

- Consider
 - $f(ab) = 0$
 - $f(aB) = f(Ab) = 1$
 - $f(AB) = 0$
- A and B will not be detected as QTLs as individually they have no effect on phenotype
- Assuming no environmental noise $H^2=1$ and $h^2=0$.
- Non-additive interactions can result from gene-gene interactions (epistasis)
 - Can be more than pairwise!
 - Considering all combinations of markers is in general not tractable because of multi-hypothesis limits
- Broad sense heritability includes additive genetic factors, dominance effects, gene-gene interactions, gene-environment interactions, non genomic inheritance

Remaining sources of heritability

- Gap between narrow- and broad-sense heritability implies genetic interactions
- For most traits, gaps not explained by found pairwise interactions
- Exception: maltose (71% of gap explained by one pairwise interaction)

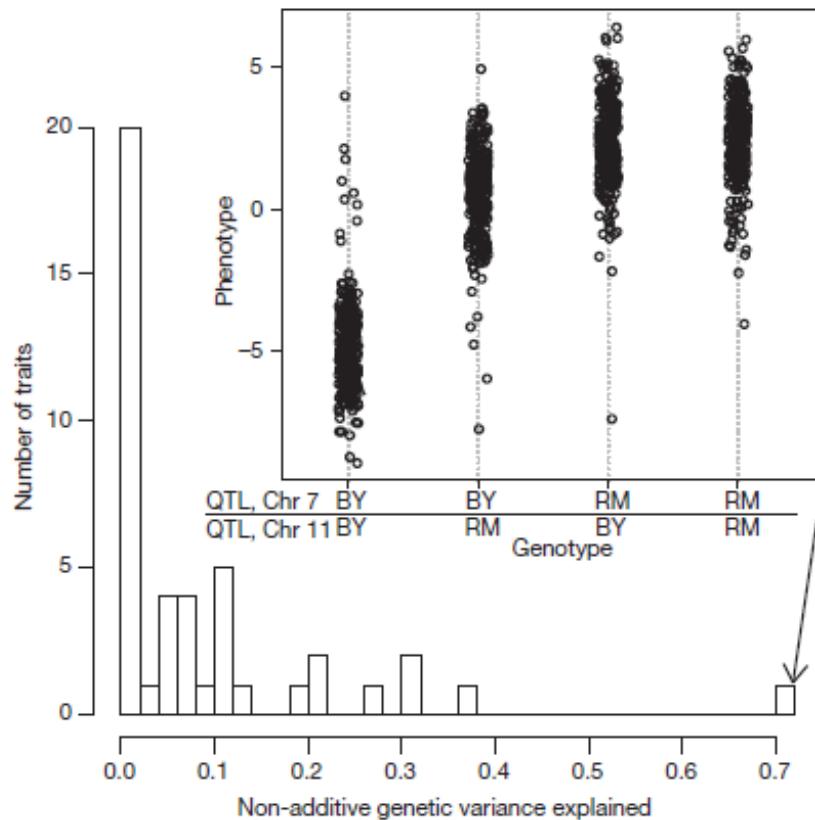


Figure 5 | Non-additive genetic variance explained by QTL–QTL interactions. A histogram of the fraction of non-additive genetic variance explained by detected QTL–QTL interactions per trait is plotted. The histogram is restricted to traits for which at least 10% of the total genetic variation is non-additive. Inset, phenotypes for growth in maltose are shown, grouped by two-locus genotypes at the two interacting QTL on chromosomes 7 and 11. This QTL–QTL interaction explained 71% of the difference between broad-sense and narrow-sense heritability.

Recent context

Problem: missing heritability for human diseases after hundreds of GWAS studies

Table 1 | Estimates of heritability and number of loci for several complex traits

| Disease | Number of loci | Proportion of heritability explained |
|---|----------------|--------------------------------------|
| Age-related macular degeneration ⁷² | 5 | 50% |
| Crohn's disease ²¹ | 32 | 20% |
| Systemic lupus erythematosus ⁷³ | 6 | 15% |
| Type 2 diabetes ⁷⁴ | 18 | 6% |
| HDL cholesterol ⁷⁵ | 7 | 5.2% |
| Height ¹⁵ | 40 | 5% |
| Early onset myocardial infarction ⁷⁶ | 9 | 2.8% |
| Fasting glucose ⁷⁷ | 4 | 1.5% |

* Residual is after adjustment for age, gender, diabetes.

“Found” h^2 from Manolio et al. 2009

Discovering what is missing

- Use other data to determine relevance of markers (SNPs in enhancers, non-sense mutations, etc.) to reduce marker search space
- When relevant marker space is simplified can consider non-linear interactions
- Consider non-chromosomal genetic elements
- Use complementary data to determine marker interactions (protein-protein interaction data, etc.)
- Your research goes here!

Part 2 - Human Genetics

Today's Narrative Arc

1. We can discover human variants that are associated with a phenotype by studying the genotypes of case and control populations
2. We can prioritize variants based upon their estimated importance
3. Follow up confirmation is important because correlation is not equivalent to causality

Today's Computational Approaches

1. Contingency tables for allelic association tests and genotypic association tests.
2. Methods of testing - Chi-Square tests, Fisher's exact test
3. Likelihood based tests of case/control posterior genotypes

Out of scope for today

1. Non-random genotyping failure
2. Methods to correct for population stratification
3. Structural variants (SVs) and copy number variations (CNVs)

Age-related macular degeneration

Cohort – 2172 unrelated European descent individuals at least 60 years old

2004: Little known about cause of AMD

934
controls



1238
cases



Contingency Tables – Fisher’s Exact Test

| Allele | Cases (with AMD) | Controls (without AMD) | Total Alleles |
|------------------|---------------------|---------------------------|------------------|
| C | a | b | a+b |
| T | c | d | c+d |
| Total Alleles | a+c | b+d | a+b+c+d |

$$P = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{a+b+c+d}{a+c}}$$

Sum all probabilities for observed and all more extreme values with same marginal totals to compute probability of null hypothesis

SNP rs1061170

1238 individuals with AMD and 934 controls

2172 individuals / 4333 alleles

| Allele | Cases (with AMD) | Controls (without AMD) | Total Alleles |
|------------------|---------------------|---------------------------|------------------|
| C | 1522 (a) | 670 (b) | 2192 |
| T | 954 (c) | 1198 (d) | 2152 |
| Total Alleles | 2476 | 1868 | 4344 |

$$p(a,b,c,d) = \frac{\begin{pmatrix} a+b \\ a \end{pmatrix} \begin{pmatrix} c+d \\ c \end{pmatrix}}{\begin{pmatrix} a+b+c+d \\ a+c \end{pmatrix}}$$
$$p-value = \sum_{i=0}^{670} p(1522+i, 670-i, 954-i, 1198+i)$$

Contingency Tables – χ^2 test

| Allele | Cases (with AMD) | Controls (without AMD) | Total Alleles |
|------------------|---------------------|---------------------------|------------------|
| C | a | b | a+b |
| T | c | d | c+d |
| Total Alleles | a+c | b+d | a+b+c+d |

$$E_1 = \frac{(a+b)(a+c)}{(a+b+c+d)}$$

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$$Df = (2 \text{ rows}-1) \times (2 \text{ columns}-1) = 1$$

SNP rs1061170

1238 individuals with AMD and 934 controls

2172 individuals / 4333 alleles

| Allele | Cases (with AMD) | Controls (without AMD) | Total Alleles |
|------------------|---------------------|---------------------------|------------------|
| C | 1522 (a) | 670 (b) | 2192 |
| T | 954 (c) | 1198 (d) | 2152 |
| Total Alleles | 2476 | 1868 | 4344 |

$$\chi^2 = \frac{(ad - bc)^2(a + b + c + d)}{(a + b)(c + d)(b + d)(a + c)}$$

$$\chi^2 = 279$$

$$Df = (2 \text{ rows}-1) \times (2 \text{ columns}-1) = 1$$

$$P\text{-value} = 1.2 \times 10^{-62}$$

Multiple hypothesis correction

n – number of hypotheses being tested

α – target false positive probability

Bonferroni correction – reject null (true positive) when

$$p_i \leq \frac{\alpha}{n}$$

Fixing a false discovery rate (FDR)

Benjamani Hochberg Multiple Hypothesis Correction (FDR method)

Assume p_i ordered small to large $p_1 p_2 p_3 p_4 \dots$

Find largest k such that

$$p_k \leq \frac{k\alpha}{n}$$

Reject null (true positive) for i less than or equal to k

Expected False Discovery Rate (FDR) is α

Does the affected or control group exhibit Population Stratification?

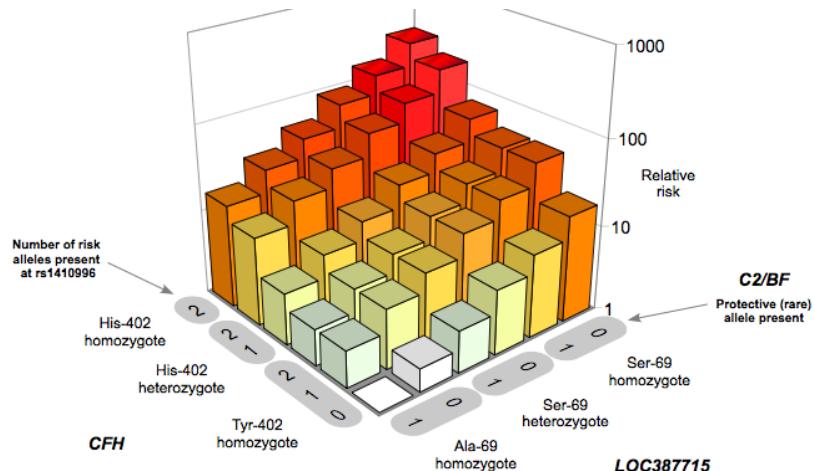
- Population stratification is when subpopulations exhibit allelic variation because of ancestry
- Can cause false positives in an association study if there are SNP differences in the case and control population structures
- Control for this artifact by testing control SNPs for general elevation in χ^2 distribution between cases and controls

Age-related macular degeneration



2004: Little known about cause of AMD

2006: Three genes (5 common variants)
Together explain >50% of risk



Relative risk plotted as a function of the genetic load of the five variants that influence risk of AMD. Two variants are in the CFH gene on chromosome 1: Y402H and rs1410996. Another common variant (A69S) is in hypothetical gene LOC387715 on chromosome 10. Two relatively rare variants are observed in the C2 and BF genes on chromosome 6. We find no evidence for interaction between any of these variants, suggesting an independent mode of action.

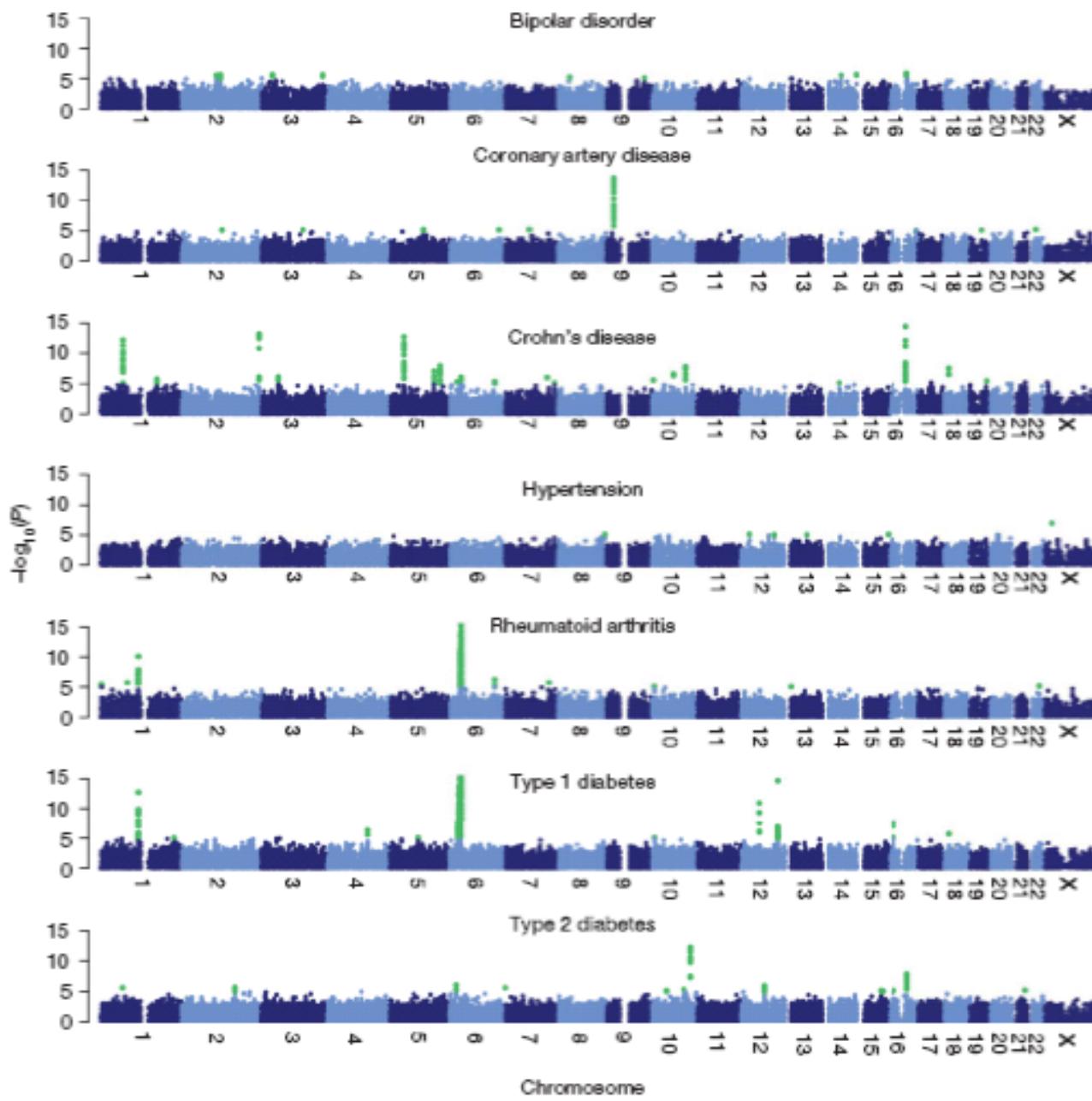


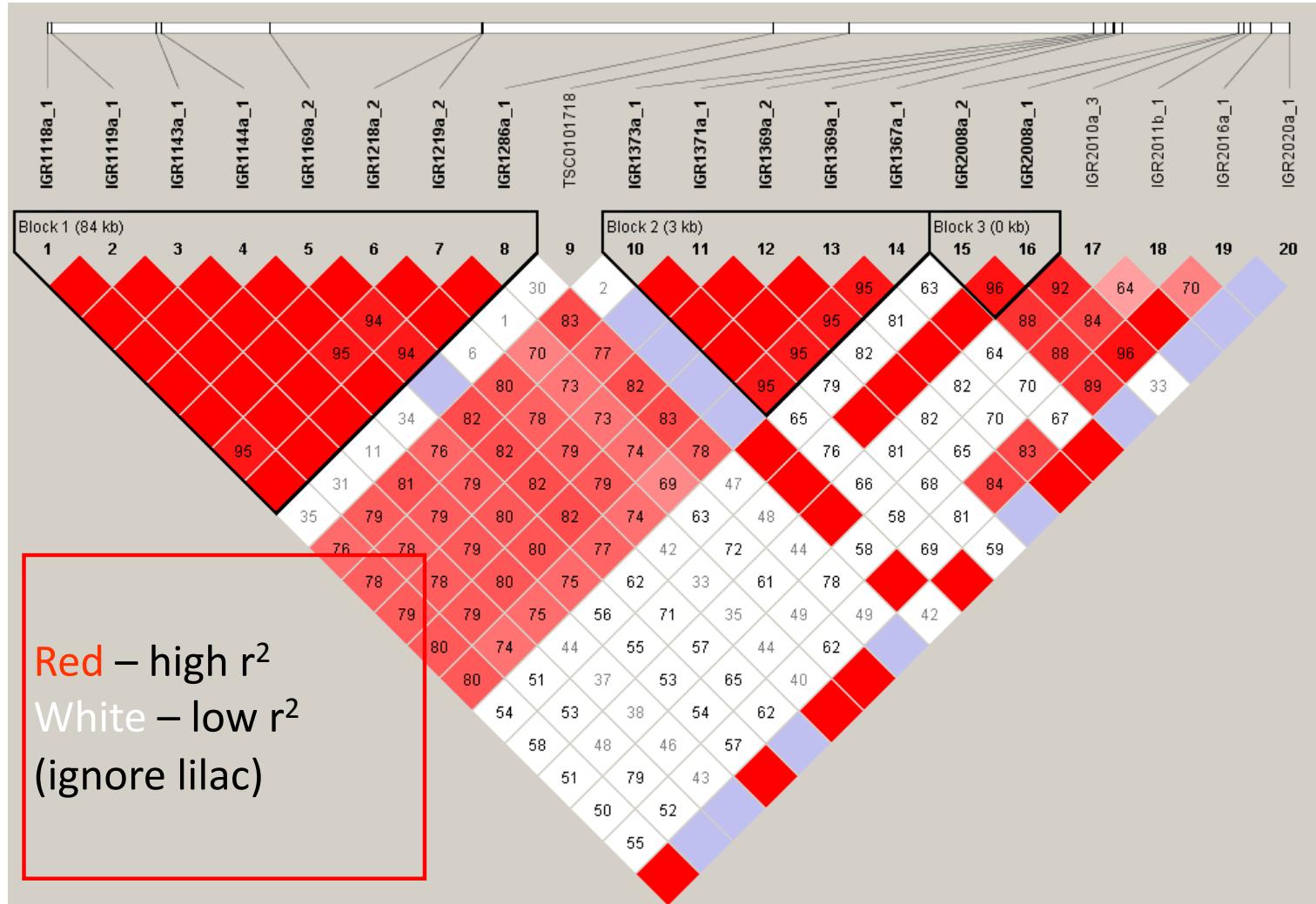
Figure 4 | Genome-wide scan for seven diseases. For each of seven diseases $-\log_{10}$ of the trend test P value for quality-control-positive SNPs, excluding those in each disease that were excluded for having poor clustering after visual inspection, are plotted against position on each chromosome.

Chromosomes are shown in alternating colours for clarity, with P values $<1 \times 10^{-5}$ highlighted in green. All panels are truncated at $-\log_{10}(P\text{ value}) = 15$, although some markers (for example, in the MHC in T1D and RA) exceed this significance threshold.

Haplotype blocks

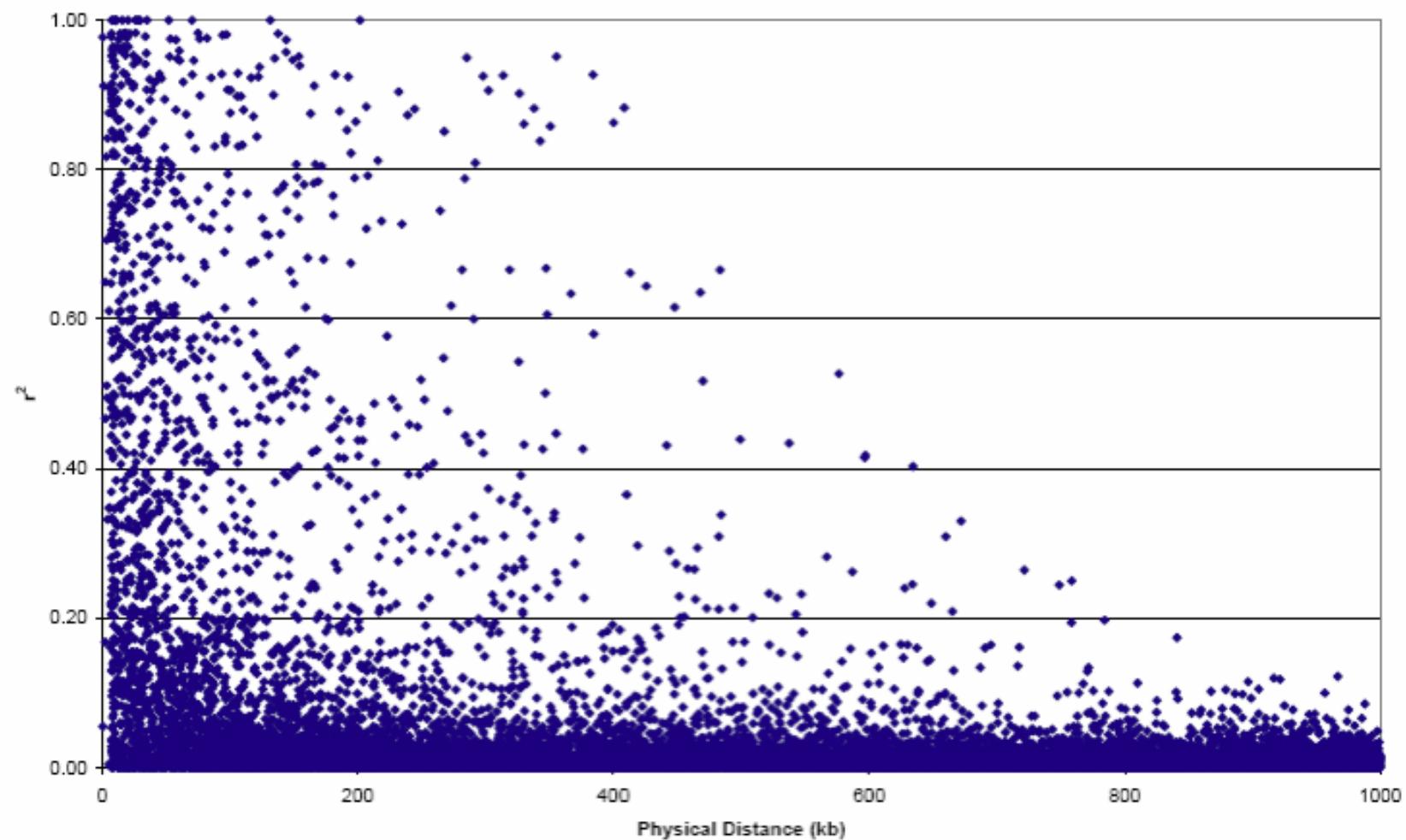
1. A *haplotype block* is a contiguous region of the genome that shows low historical recombination (high linkage disequilibrium - LD)
2. The HapMap project used many individuals to create a draft map of haplotype blocks
3. Phasing variants tells us their chromosome (parent) of origin
4. Phasing is important to understand functional consequence of variants
5. HapMap can help us phase variants by observing their occurrence in haplotype blocks

LD organizes the genome into haplotype blocks

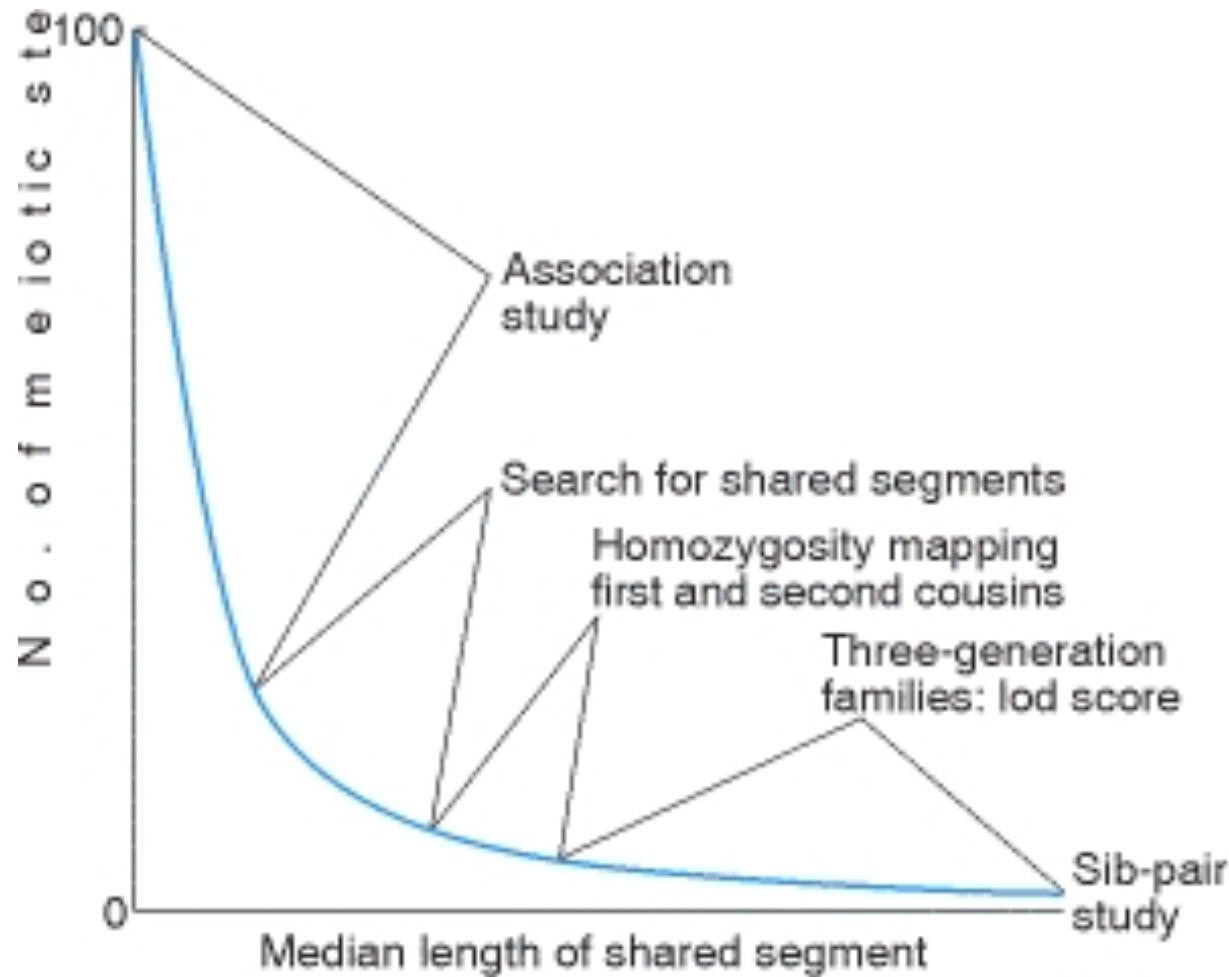


Human genome 5q31 region (associated with Inflammatory Bowel Disease)

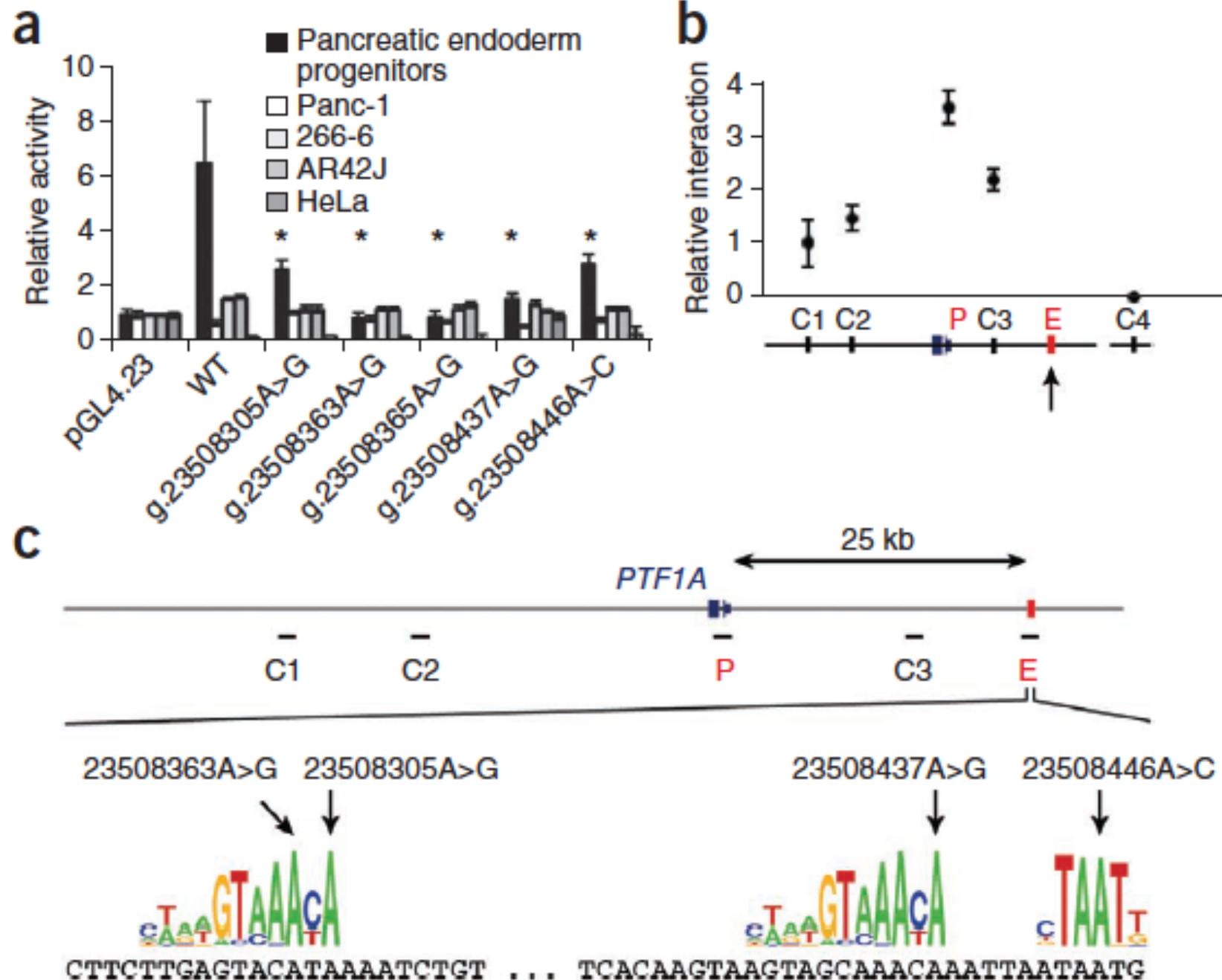
r^2 from human chromosome 22



The length of shared segments vs time



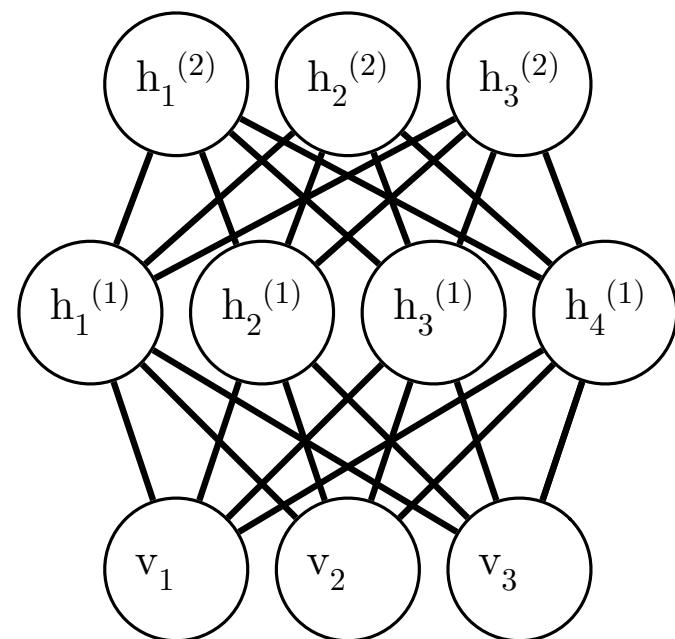
Confirmation of variant function



Part 3 - Approaches to Deep Learning

What drives success in ML?

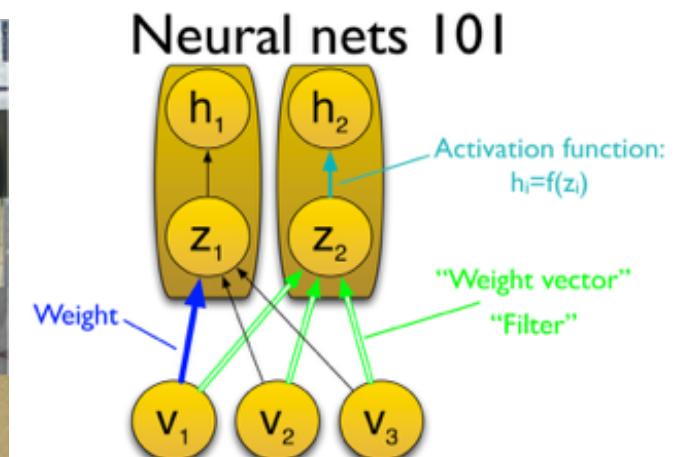
Arcane knowledge
of dozens of
obscure algorithms?



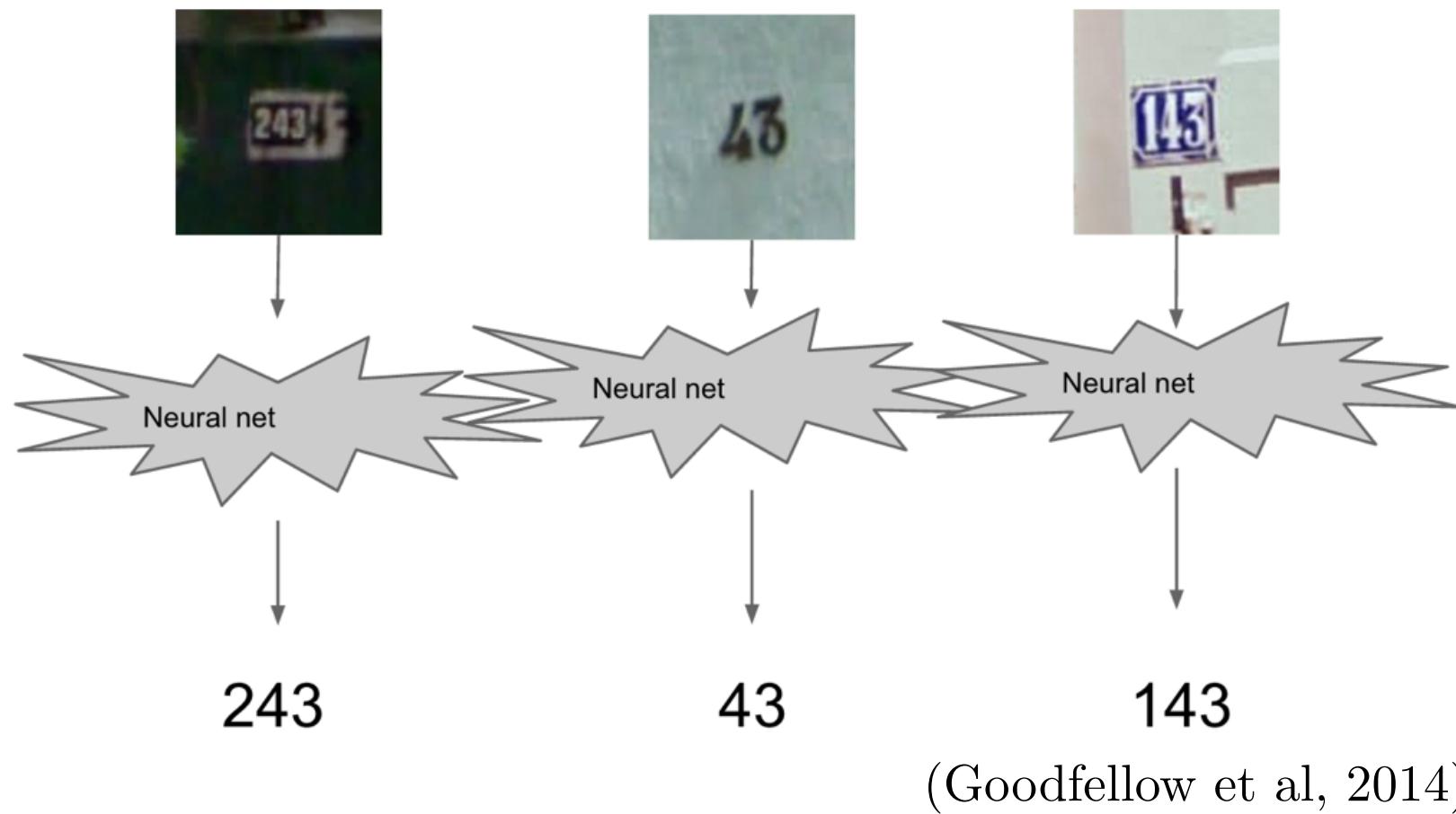
Mountains
of data?



Knowing how
to apply 3-4
standard techniques?



Example: Street View Address Number Transcription



Three Step Process

- Use needs to define metric-based goals
- Build an end-to-end system
- Data-driven refinement

Identify Needs

- High accuracy or low accuracy?
- Surgery robot: high accuracy
- Celebrity look-a-like app: low accuracy

Choose Metrics

- Accuracy? (% of examples correct)
- Coverage? (% of examples processed)
- Precision? (% of detections that are right)
- Recall? (% of objects detected)
- Amount of error? (For regression problems)

End-to-end System

- Get up and running ASAP
- Build the simplest viable system first
- What baseline to start with though?
 - Copy state-of-the-art from related publication

Deep or Not?

- Lots of noise, little structure -> not deep
- Little noise, complex structure -> deep
- Good shallow baseline:
 - *Use what you know*
 - Logistic regression, SVM, boosted tree are all good

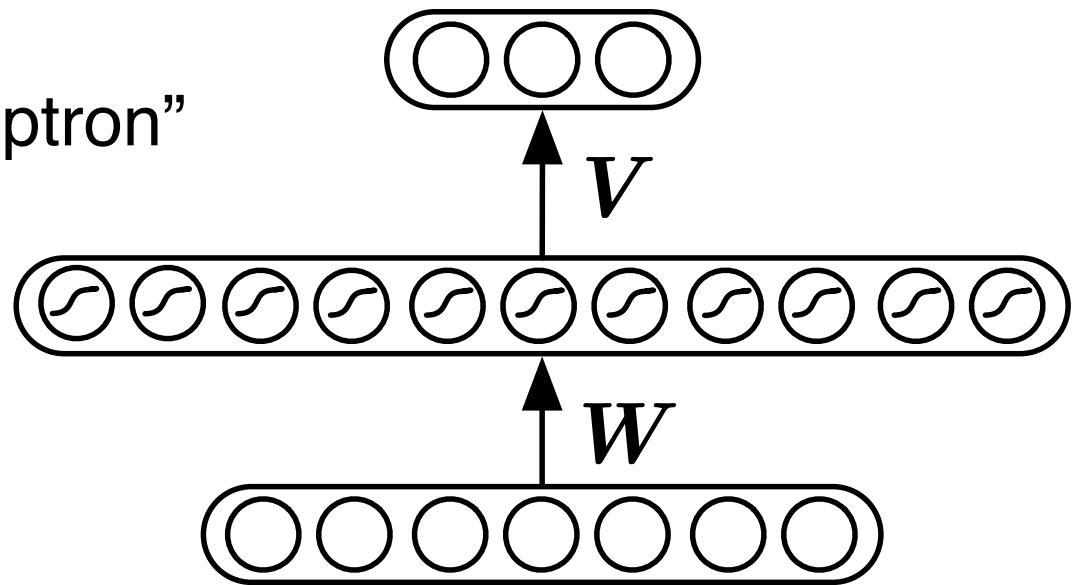
<http://scikit-learn.org/>

Choosing Architecture Family

- No structure -> fully connected
- Spatial structure -> convolutional
- Sequential structure -> recurrent

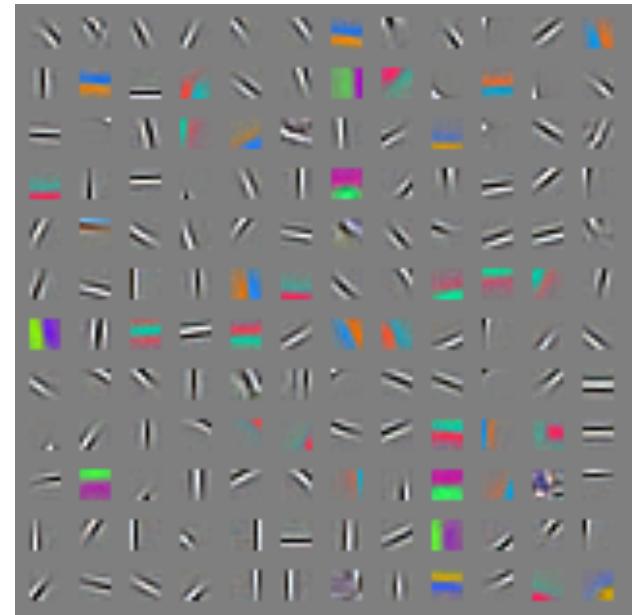
Fully Connected Baseline

- 2-3 hidden layer feed-forward neural network
 - AKA “multilayer perceptron”
- Rectified linear units
- Batch normalization
- Adam
- Maybe dropout



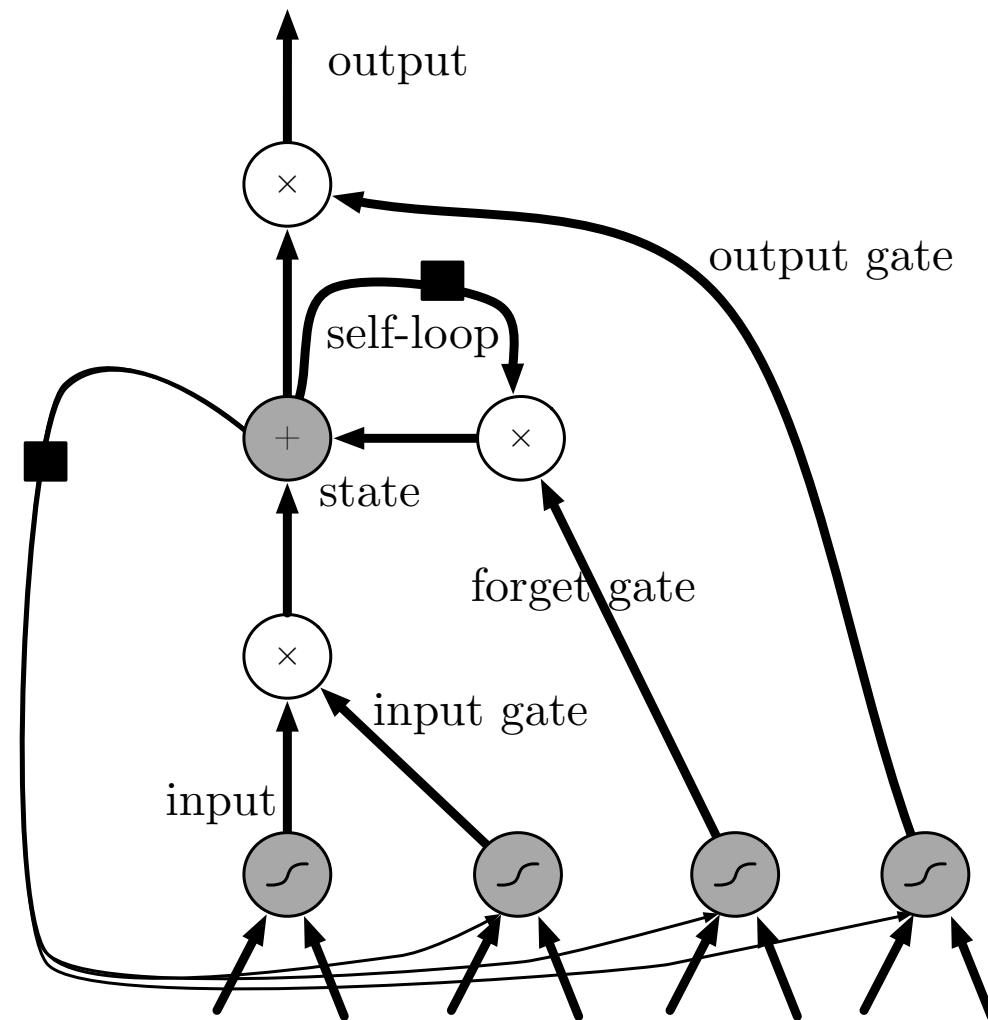
Convolutional Network Baseline

- Download a pretrained network
- Or copy-paste an architecture from a related task
 - Or:
 - Deep residual network
 - Batch normalization
 - Adam



Recurrent Network Baseline

- LSTM
- SGD
- Gradient clipping
- High forget gate bias



Data-driven Adaptation

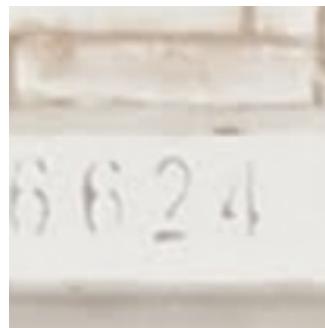
- Choose what to do based on data
- Don't believe hype
- Measure train and test error
 - “Overfitting” versus “underfitting”

High Train Error

- Inspect data for defects
- Inspect software for bugs
 - Don't roll your own unless you know what you're doing
- Tune learning rate (and other optimization settings)
- Make model bigger

Checking Data for Defects

- Can a human process it?

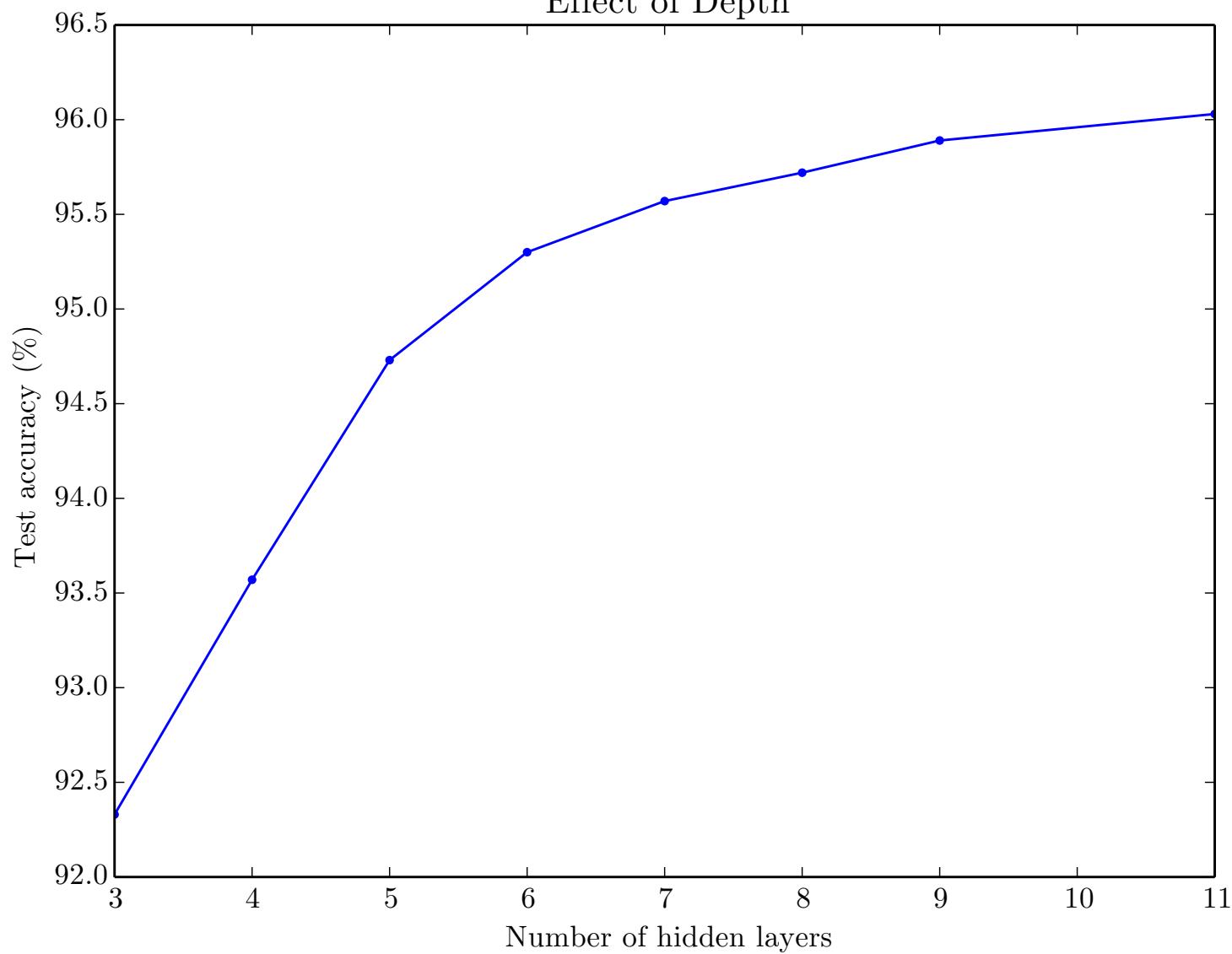


26624



Increasing Depth

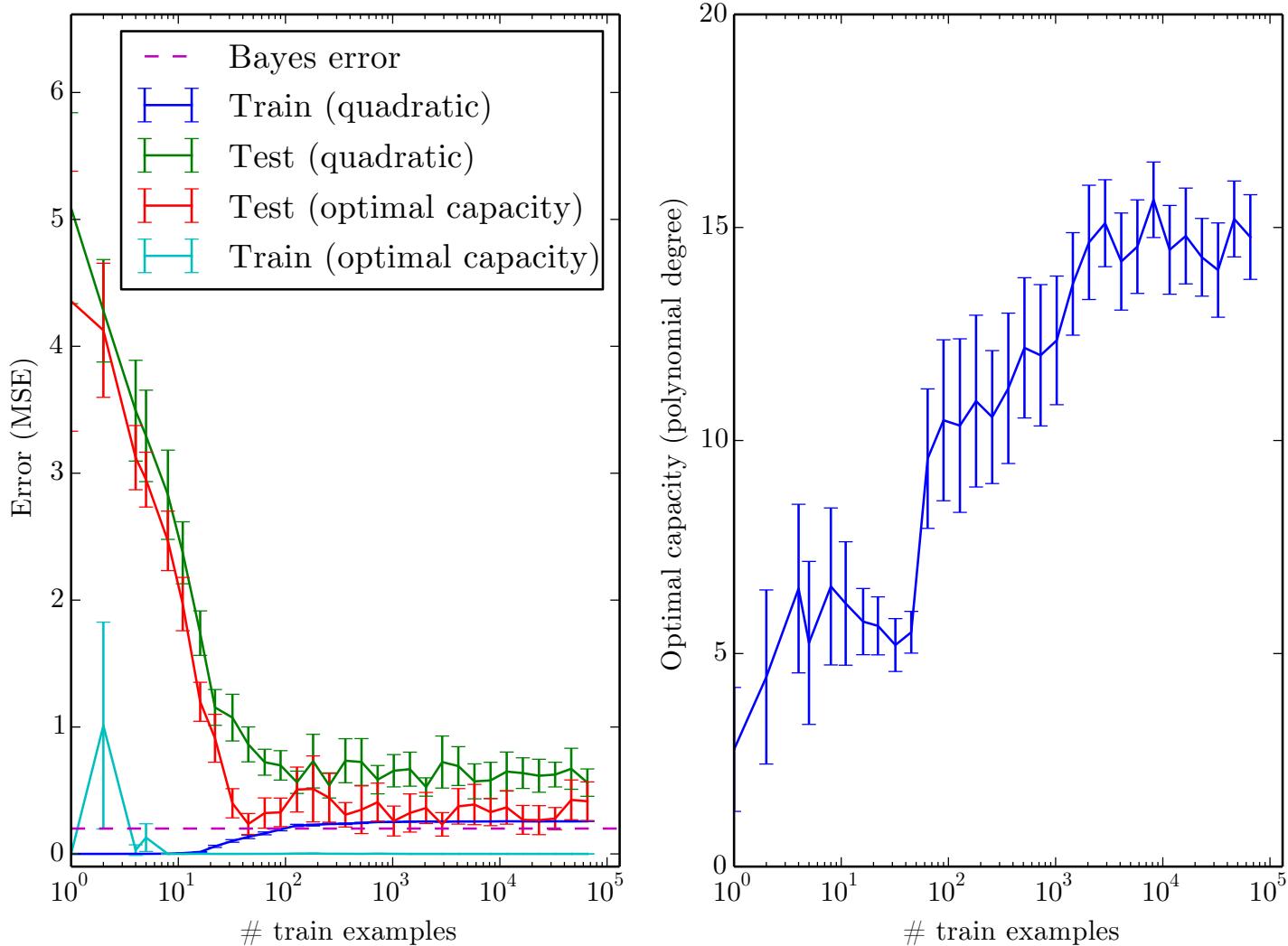
Effect of Depth



High Test Error

- Add dataset augmentation
- Add dropout
- Collect more data

Increasing Training Set Size



Tuning the Learning Rate

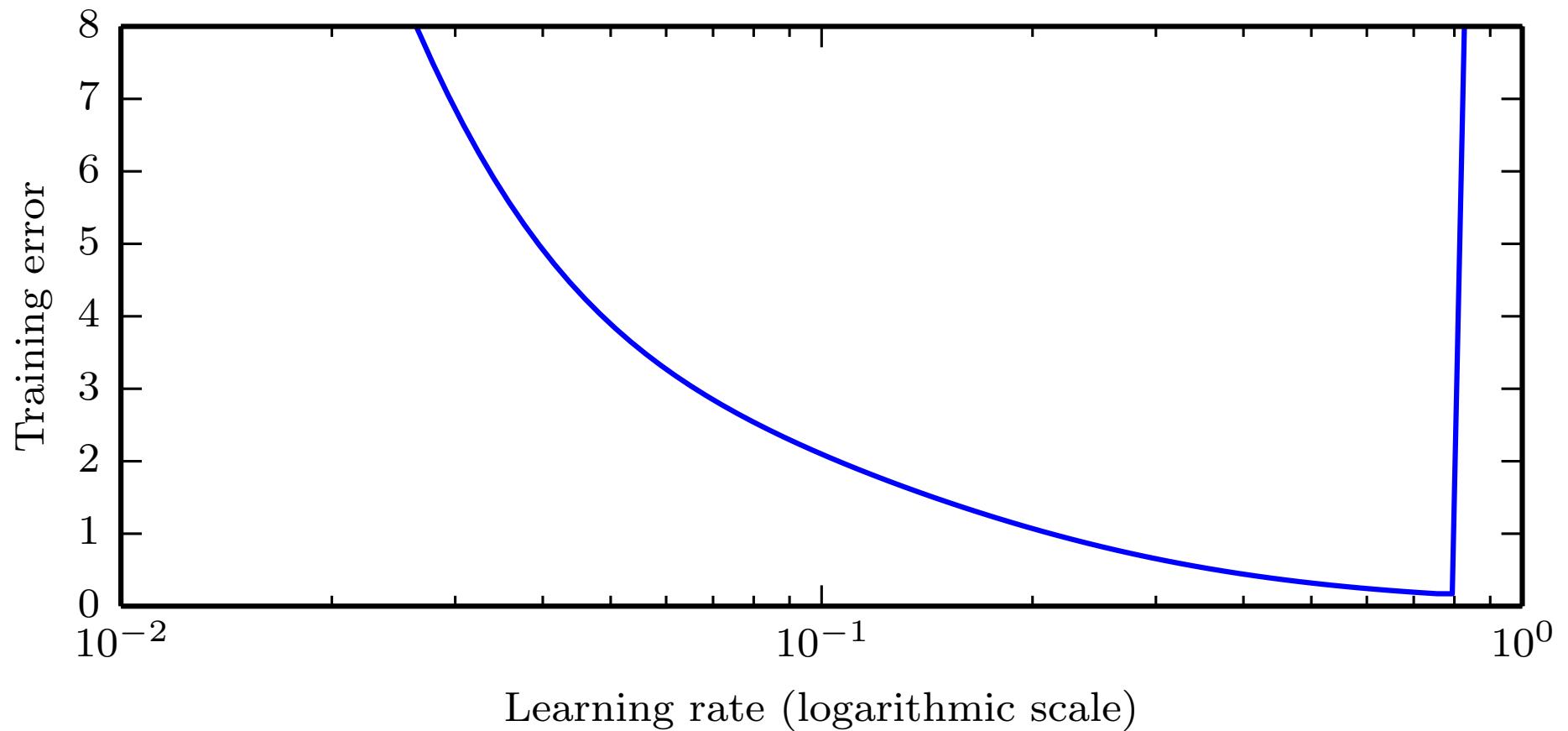


Figure 11.1

(Goodfellow 2016)

Reasoning about Hyperparameters

| Hyperparameter | Increases capacity when... | Reason | Caveats |
|------------------------|----------------------------|---|--|
| Number of hidden units | increased | Increasing the number of hidden units increases the representational capacity of the model. | Increasing the number of hidden units increases both the time and memory cost of essentially every operation on the model. |

Table 11.1

(Goodfellow 2016)

Hyperparameter Search

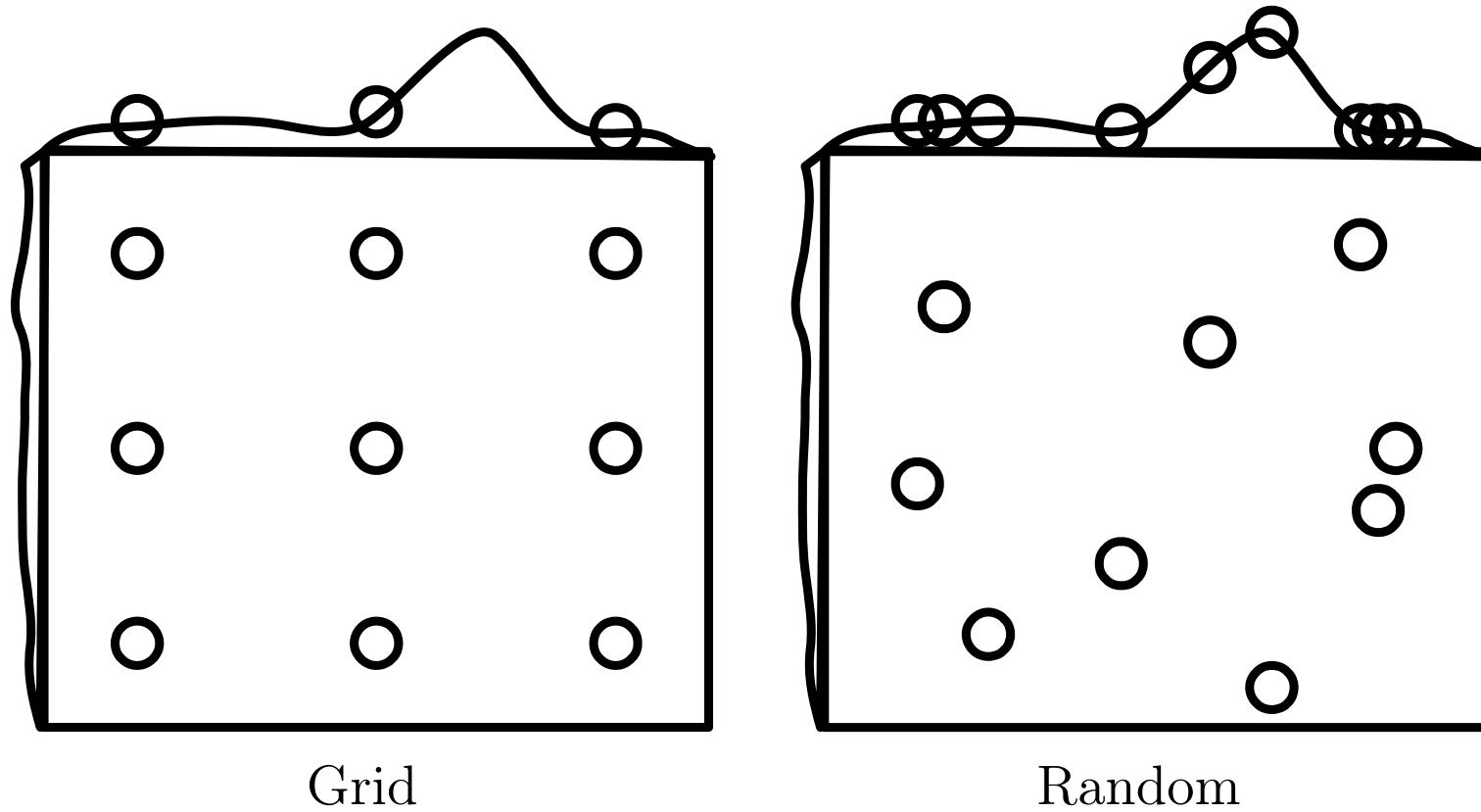


Figure 11.2

FIN - Thank You