PREDICTION OF COMPLEX TRAITS

From pedigrees and DNA to phenotypes

Daniel Gianola

Hans Fischer Senior Fellow TUM-IAS München

Sewall Wright Professor of Animal Breeding and Genetics

University of Wisconsin

Dairy Science

UW–MADISON ANIMAL SCIENCES

Biostatistics & Medical Informatics

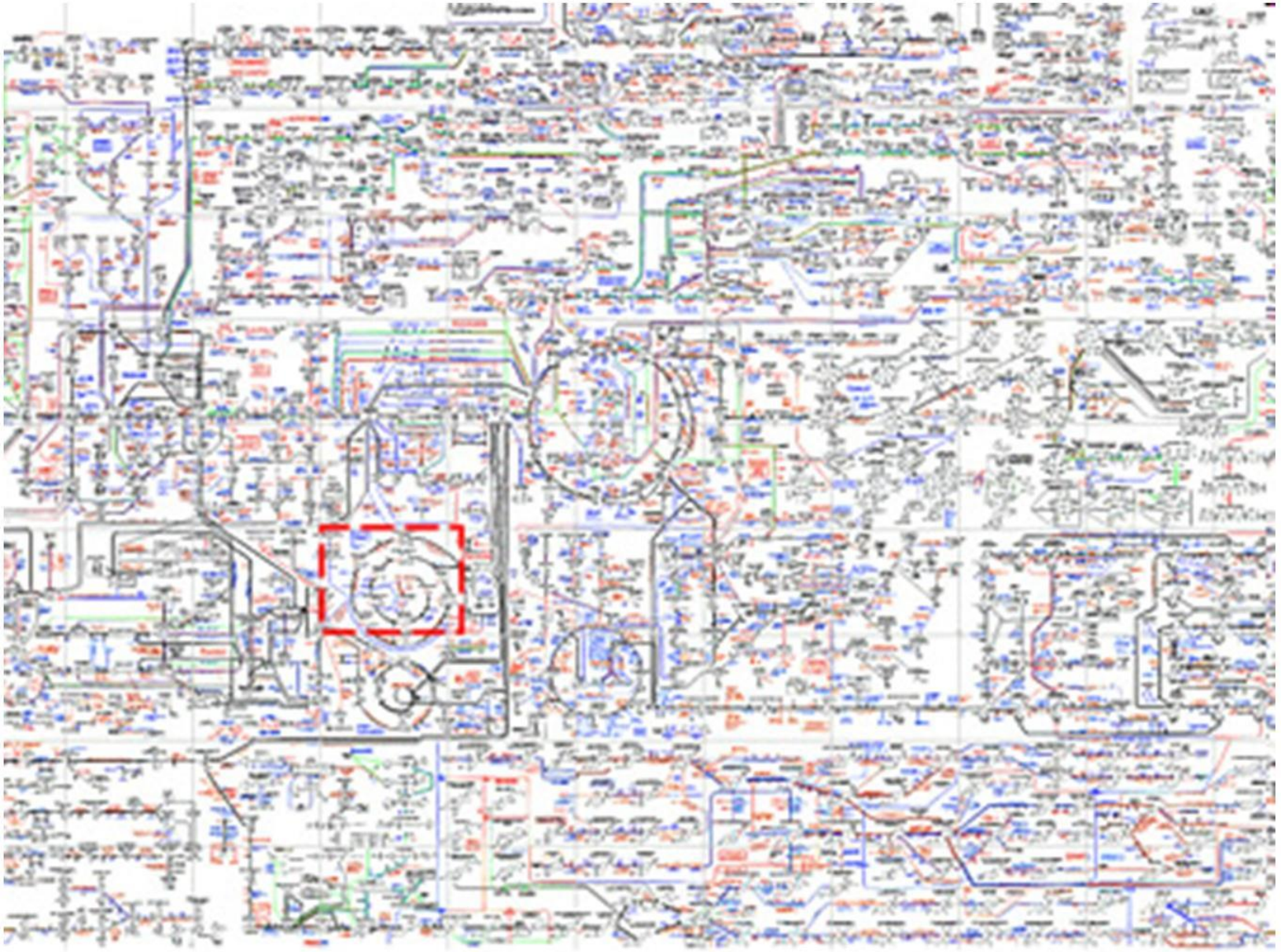
University of Wisconsin SCHOOL OF MEDICINE AND PUBLIC HEALTH

Proposition 1

It must be true that quantitative traits are “complex”, in any sense of the word.

Why?

A “complex” trait involves many metabolic pathways: Roche’s Chart

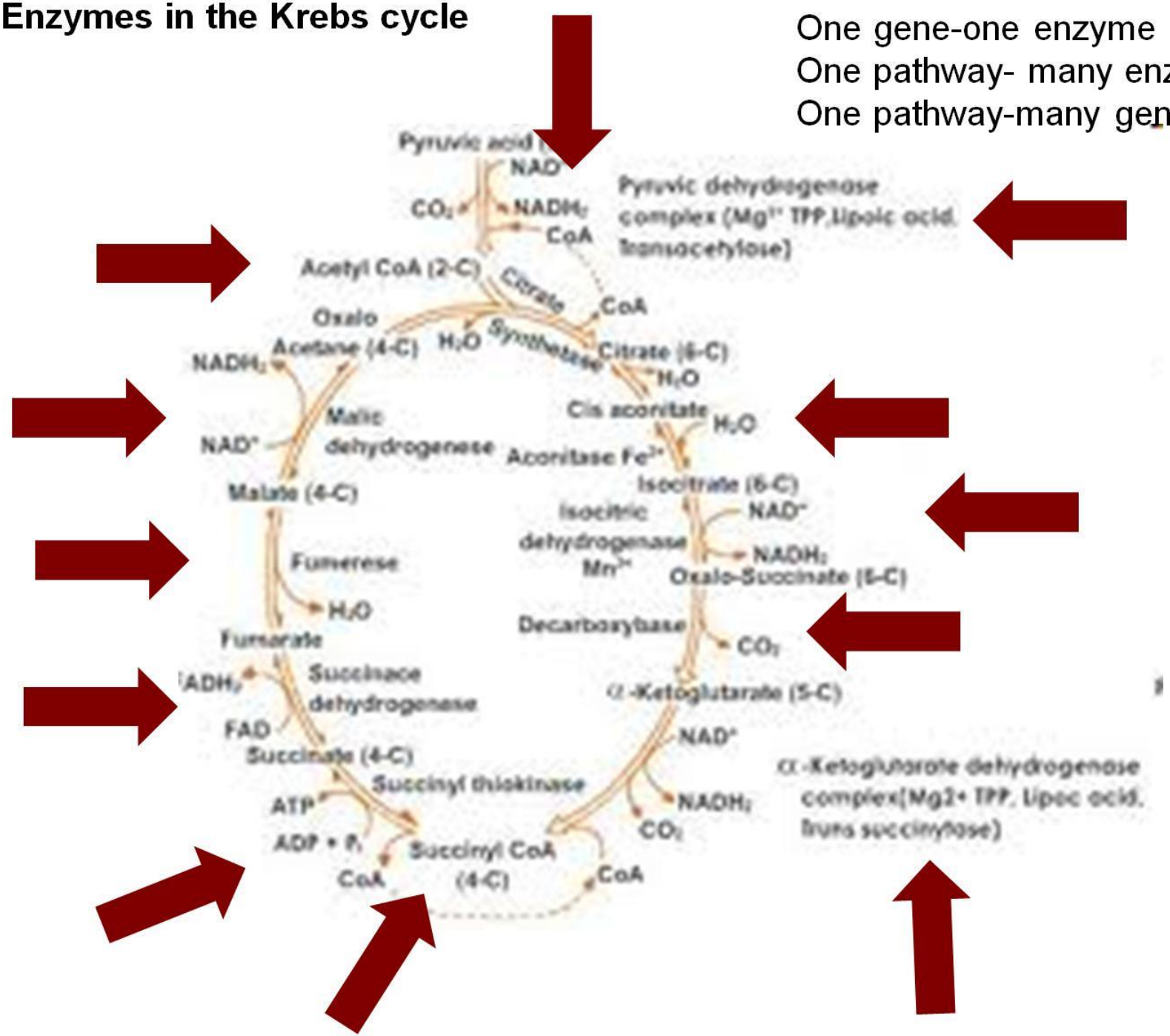


Proposition 2

It must be true that epistasis is pervasive

Enzymes in the Krebs cycle

One gene-one enzyme
One pathway- many enzymes
One pathway-many genes



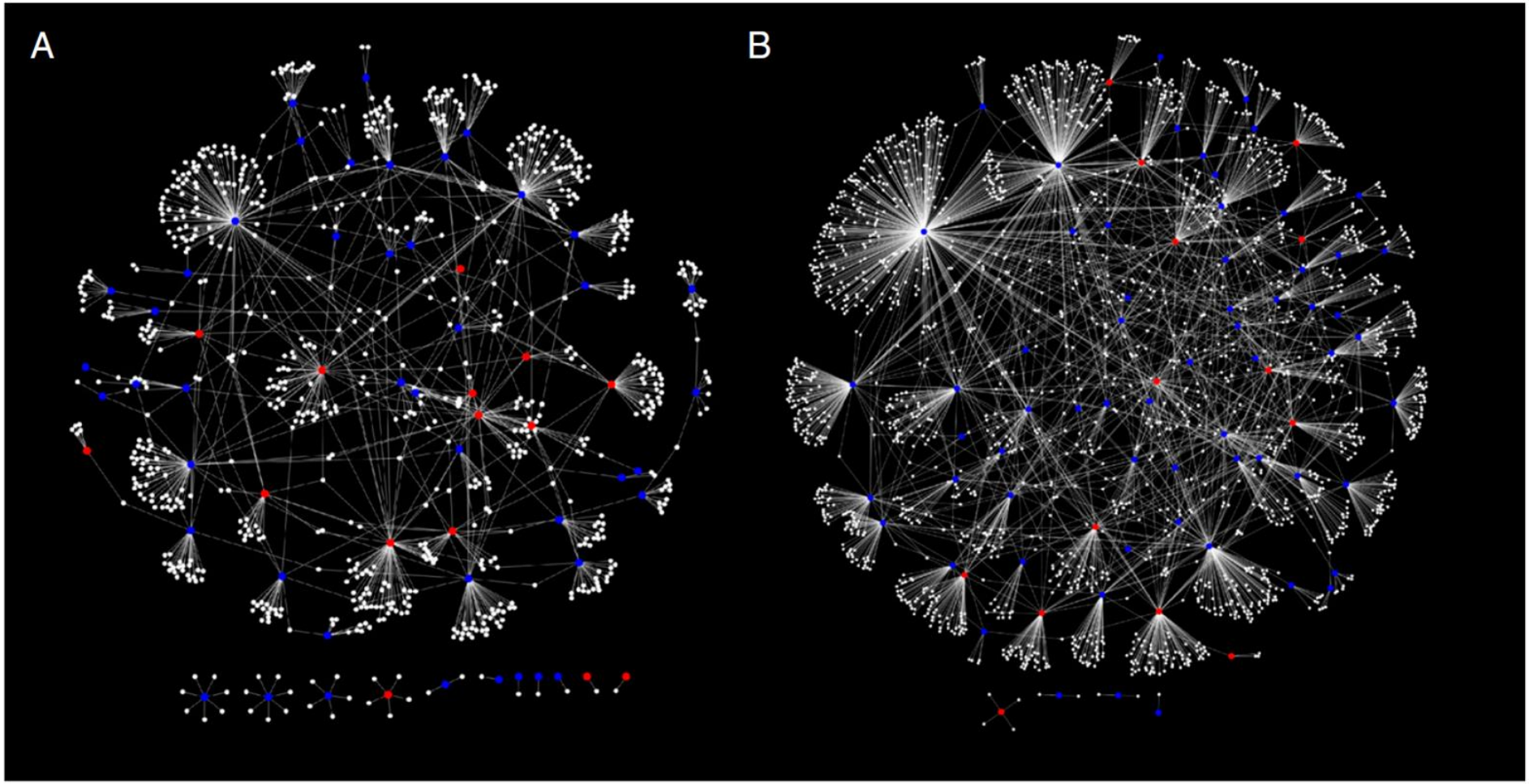


Fig. 5. Networks of epistatic interactions. Interaction networks are depicted for (*A*) starvation resistance and (*B*) chill coma recovery. Nodes depict genes, and edges significant interactions. Red nodes are genes containing significant SNPs from the Flyland analysis. Blue nodes are genes containing significant SNPs from DGRP analysis.

PNAS, 2012

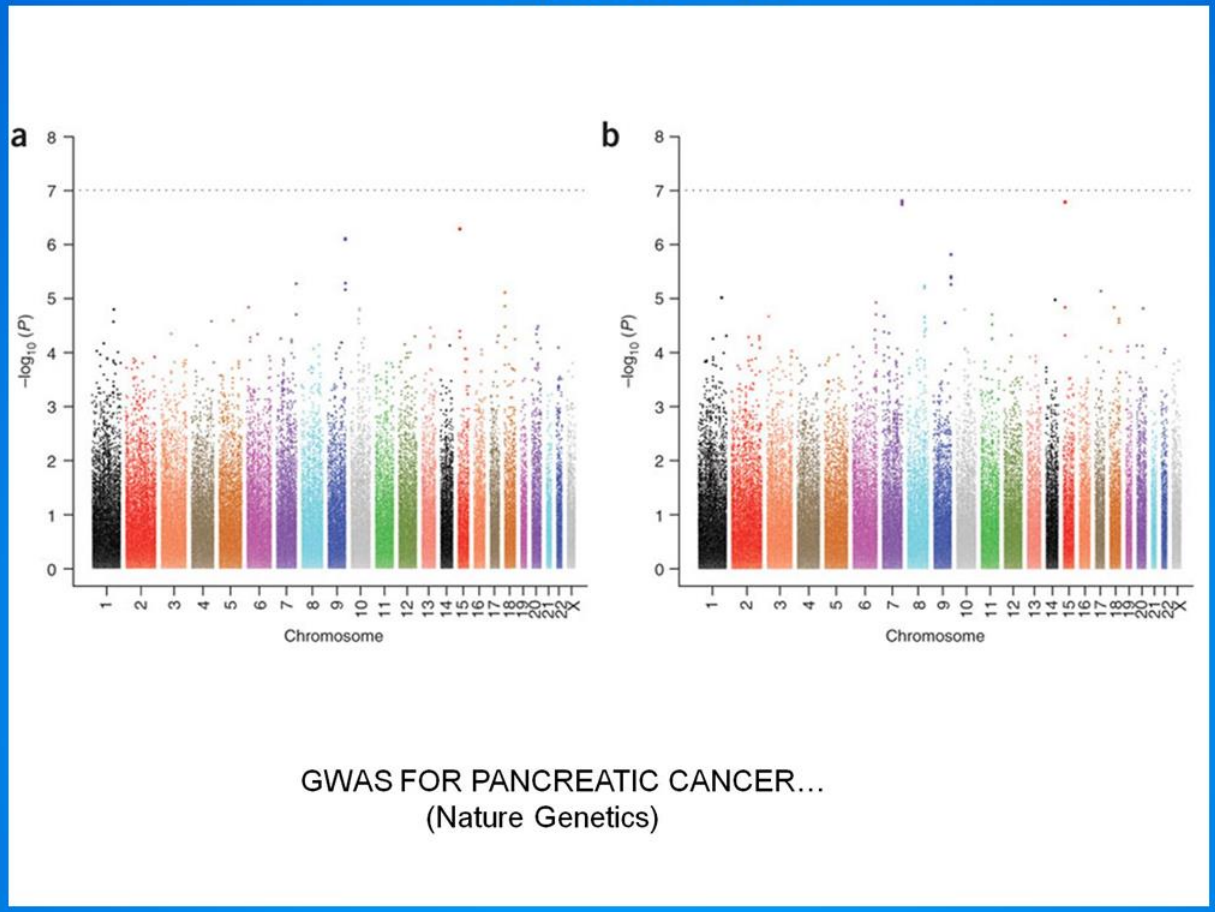
Epistasis dominates the genetic architecture of *Drosophila* quantitative traits

Wen Huang[a], Stephen Richards[b], Mary Anna Carbone[a], Dianhui Zhu[b], Robert R. H. Anholt[c], Julien F. Ayroles[a,1], Laura Duncan[a], Katherine W. Jordan[a], Faye Lawrence[a], Michael M. Magwire[a], Crystal B. Warner[b,2], Kerstin Blankenburg[b], Yi Han[b], Mehwish Javaid[b], Joy Jayaseelan[b], Shalini N. Jhangiani[b], Donna Muzny[b], Fiona Ongeri[b], Lora Perales[b], Yuan-Qing Wu[b,3], Yiqing Zhang[b], Xiaoyan Zou[b], Eric A. Stone[a], Richard A. Gibbs[b], and Trudy F. C. Mackay[a,4]

ABSTRACTION PARADIGM 1

GWAS: ***search for association between some marker or genomic region and a phenotype***

EXAMPLES



GWAS FOR PANCREATIC CANCER…
(Nature Genetics)

Genome-Wide Association Study to Identify Single Nucleotide Polymorphisms (SNPs) Associated With the Development of Erectile Dysfunction in African-American Men After Radiotherapy for Prostate Cancer. I*nternational Journal of Radiation Oncology, Biology, Physics 2010*

What about if there are epistatic QTLs and one fits *p* markers? (assume OLS is identified, with p<n)

Regression fitted

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & . & . & . & \mathbf{x}_p \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{bmatrix} + \mathbf{e},$$

"True" model

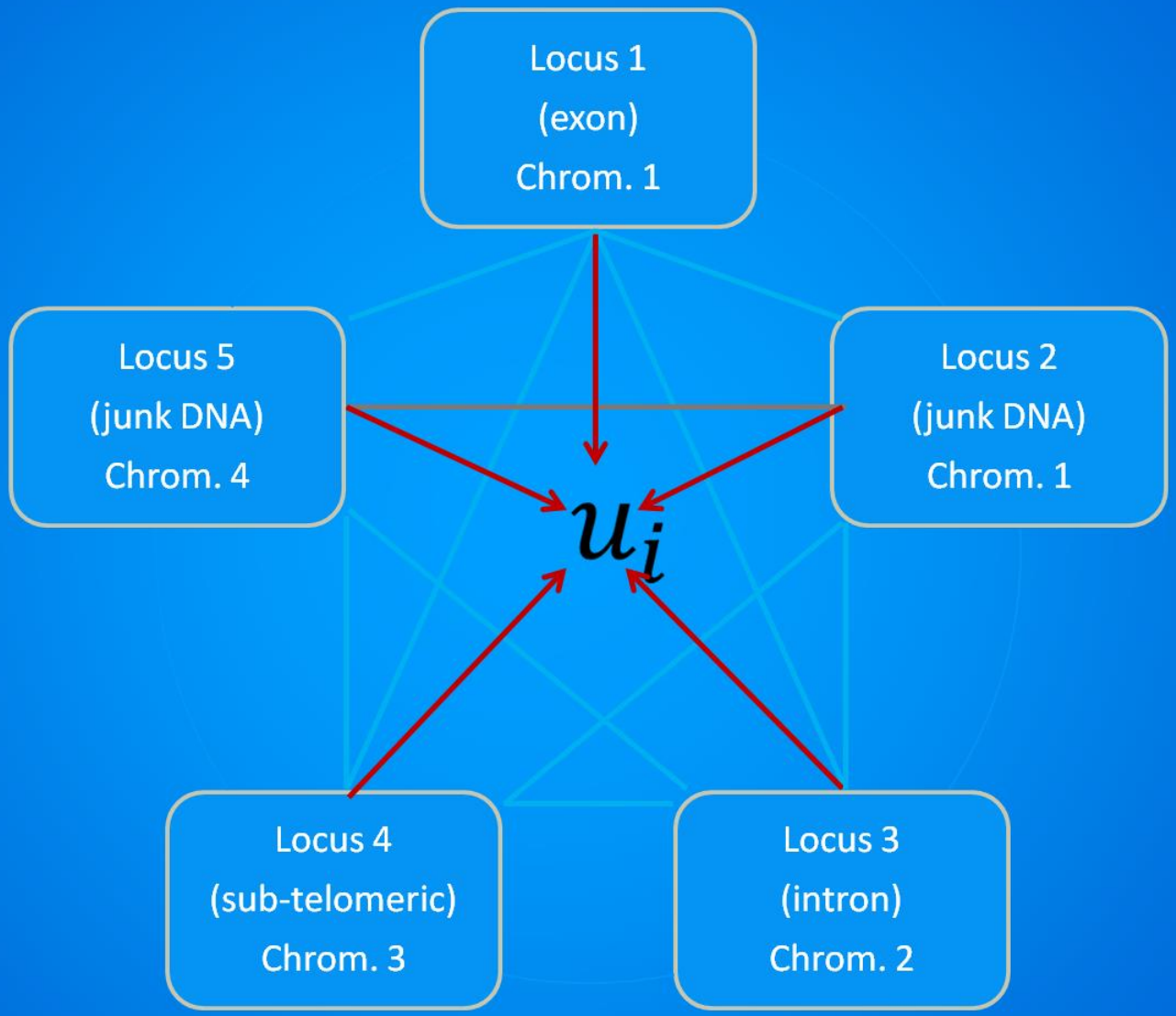
$$\mathbf{y} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_{12} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_{12} \end{bmatrix} + \mathbf{e} = \mathbf{Q}\boldsymbol{\alpha} + \mathbf{e},$$

$$E\left(\hat{\boldsymbol{\beta}}\right) = \begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 & \cdots & \mathbf{x}_1'\mathbf{x}_p \\ \cdot & \mathbf{x}_2'\mathbf{x}_2 & \cdots & \mathbf{x}_2'\mathbf{x}_p \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \text{symmetric} & \cdot & \cdots & \mathbf{x}_p'\mathbf{x}_p \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_1'\left(\mathbf{q}_1\alpha_1 + \mathbf{q}_2\alpha_2 + \mathbf{q}_{12}\alpha_{12}\right) \\ \mathbf{x}_2'\left(\mathbf{q}_1\alpha_1 + \mathbf{q}_2\alpha_2 + \mathbf{q}_{12}\alpha_{12}\right) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_p'\left(\mathbf{q}_1\alpha_1 + \mathbf{q}_2\alpha_2 + \mathbf{q}_{12}\alpha_{12}\right) \end{bmatrix}.$$

BIAS AFFECTED BY

- **ALL** LD RELATIONSHIPS AMONG MARKERS
- **ALL** LD RELATIONSHIPS AMONG MARKERS AND **ALL** QTLs

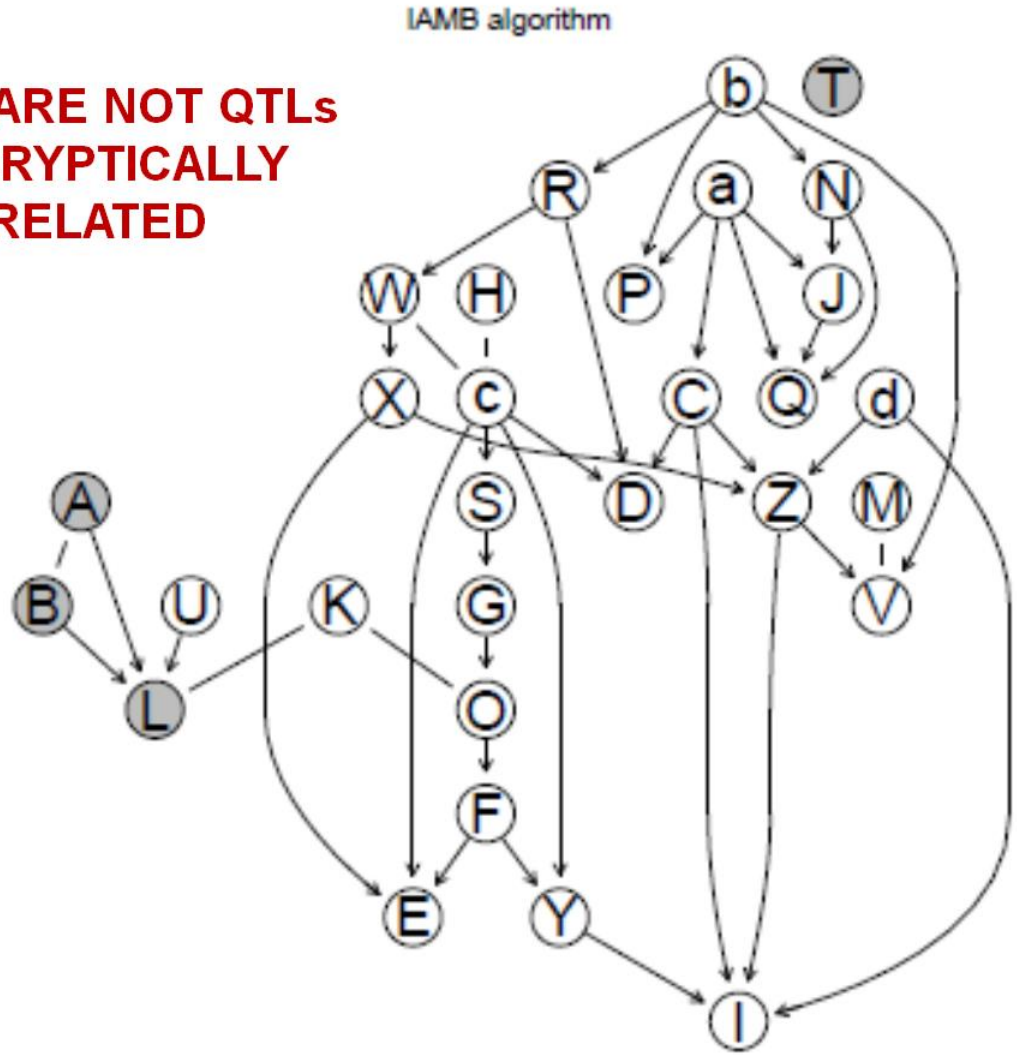
Figure 1. Five locus system in linkage disequilibrium. Arrows represent direct effects on additive genetic value (*u*) ; undirected lines and arcs represent correlations between genotypes stemming from linkage disequilibrium.



$$u = QTL_1 + QTL_2 + \cdots + QTL_5$$

How many QTLs? "Honey I shrunk epistasis!"

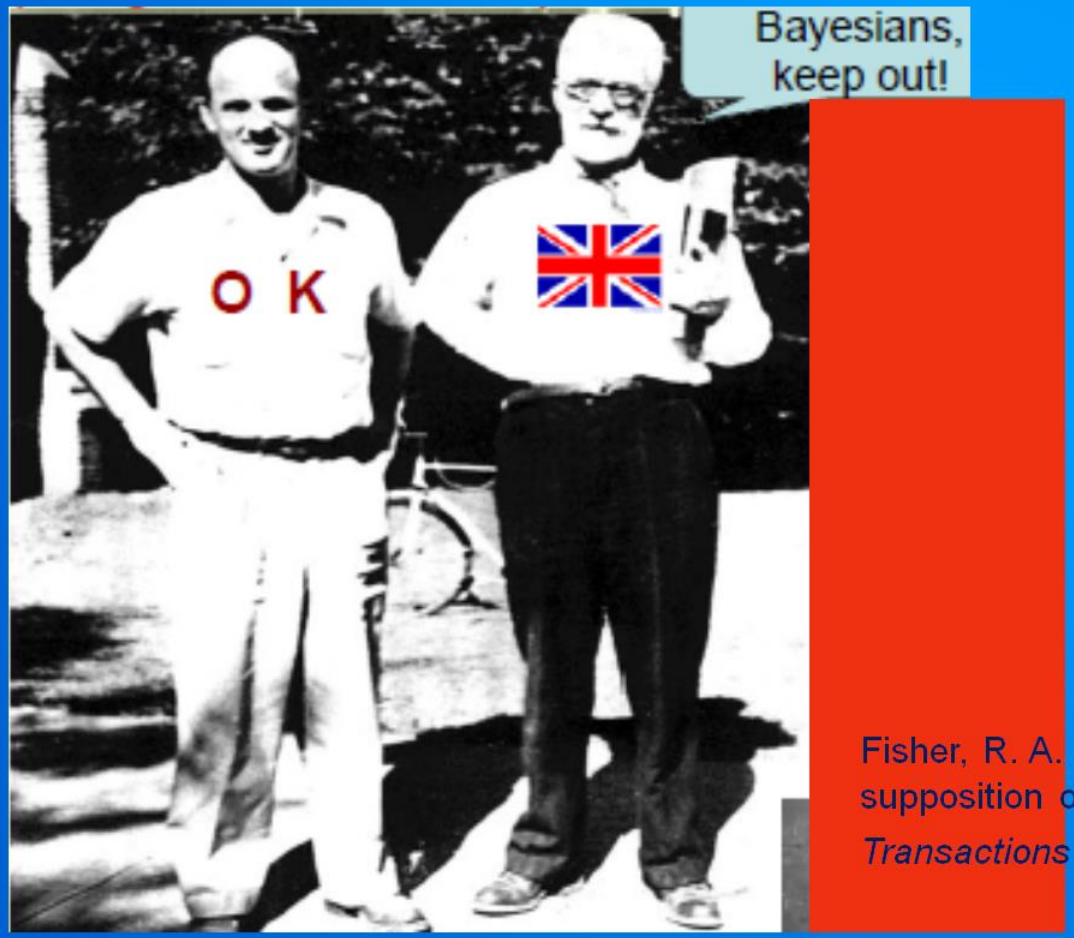
MARKERS ARE NOT QTLs AND ARE CRYPTICALLY INTER-RELATED

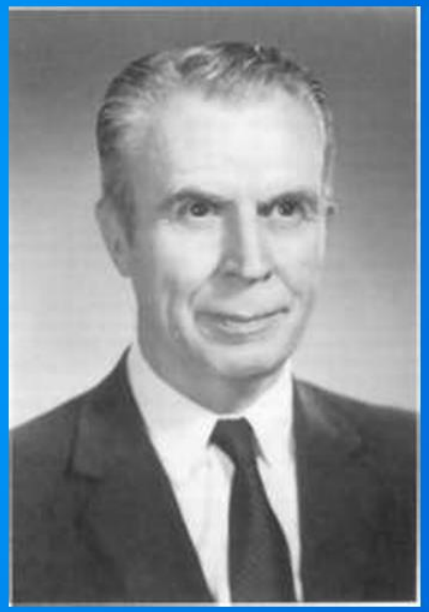


BAYESIAN NETWORK OF LINKAGE DISEQUILIBRIUM
(30 SNPs with effects on milk protein content, cows)
Morota el at. (2012)

ABSTRACTION PARADIGM 2

Fisher's infinitesimal model of additive effects
(extended vectorially by C. R. Henderson, animal breeder)





Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance.

Transactions of the Royal Society of Edinburgh, *52*, 399-433.

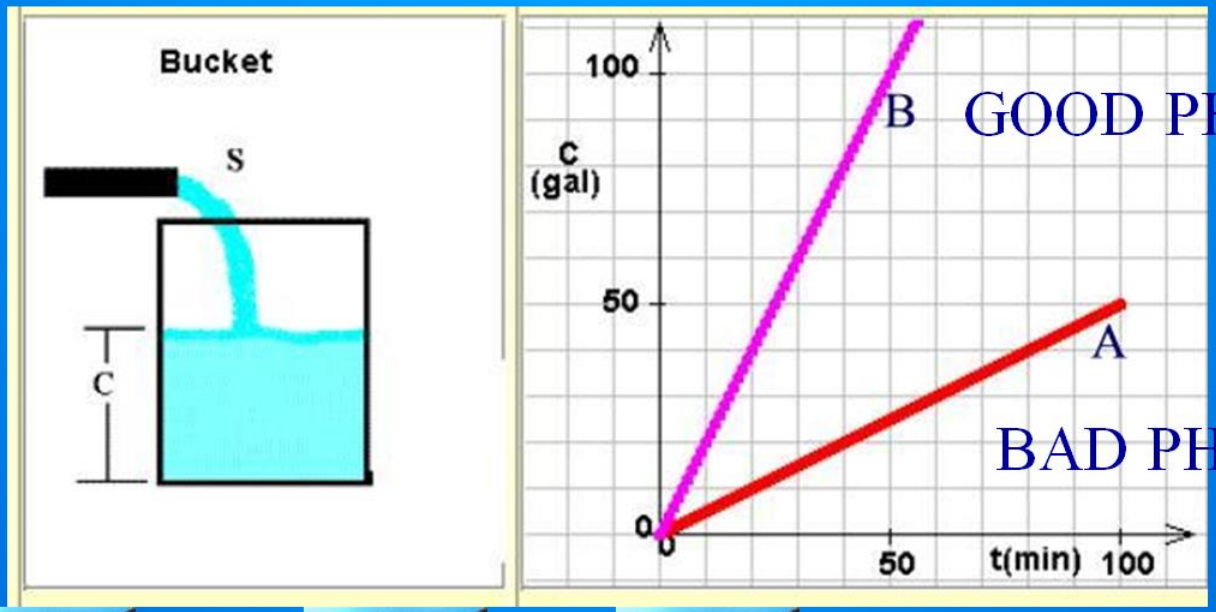
THE CENTRAL DOGMA OF QUANTITATIVE GENETICS:

the additive genetic model

$$u_i = W_{i1}a_1 + W_{i2}a_2 + \ldots + W_{iK}a_K$$

$$W_{ij}a_j = \begin{cases} -a_j & \text{if } W_{ij} = -1\,(aa);\ \Pr(W_{ij} = -1) = (1-p_j)^2 \\ 0 & \text{if } W_{ij} = 0\,(Aa);\ \Pr(W_{ij} = 0) = 2p_j(1-p_j) \\ a_j & \text{if } W_{ij} = 1\,(AA);\ \Pr(W_{ij} = 1) = p_j^2 \end{cases}$$

Genome



= 'additive genetic value'

EMULATE FISHER'S MODEL USING MOLECULAR MARKERS

A (slightly) less naïve form of approximating G is the whole-genome linear model:

$$G = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_px_p$$

Where the x's are either pedigree relationships, or marker genotype codes or whatever the latest fad in genomic data is

Bayes A
Bayes B
Bayes C (with or without π)
Bayesian Lasso
NON-BAYESIAN REGULARIZED: Lasso, Elastic Net

LEADS TO (EXTRAORDINARILY) SHRUNKEN ESTIMATES OF EFFECTS, BUT GOOD PREDICTIONS OF "TOTAL SIGNAL"

PARADIGM 2 IS NAIVE

-IT PRODUCES (CONDITIONALLY) BIASED AND INCONSISTENT ESTIMATES

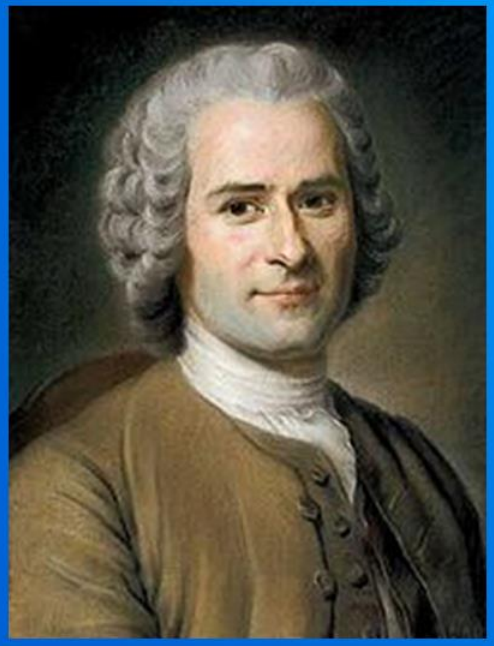
-ORDER PIZZA FOR 500 AND 1 MILLION EAT

-THERE IS AN IDENTIFICATION PROBLEM IN THE $n<<p$ CASE. NOT TRUE THAT DIFFERENT BAYESIAN MODELS (A, B, C,…, ETC.) ARE INFORMATIVE ABOUT "GENETIC ARCHITECTURE"

- AT BEST PRODUCES A LOCAL APPROXIMATION TO EPISTASIS

ROUSSEAU ON THE ADDITIVE GENETIC MODEL

"…denier ce que est, et d'expliquer ce qui n'est pas…"
Rousseau "Nouvelle Heloise"



Geneve 1712- Ermenonville 1778

Dealing with epistatic interactions and non-linearities

gene x gene

gene x gene x gene

gene x gene x gene x gene

.....

(Alice in Wonderland)

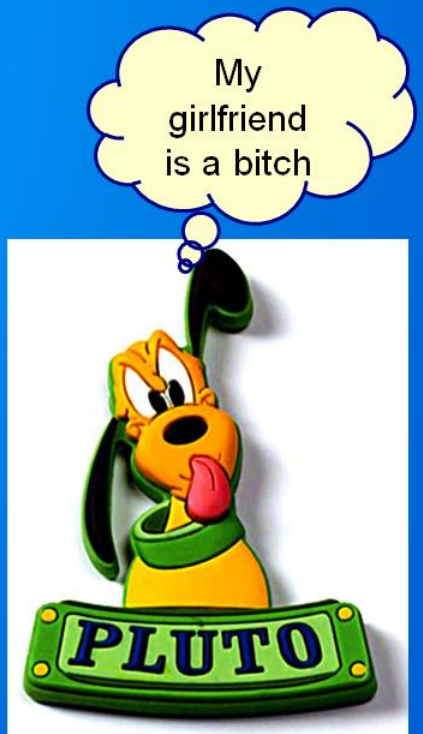
DO THESE ASSUMPTIONS HOLD?

RANDOM EFFECTS MODELS FOR ASSESSING EPISTASIS REST ON: Cockerham (1954) and Kempthorne (1954)

--Orthogonal partition of genetic variance into additive, dominance additive x additive, etc. **ONLY** if

- ❑No selection
- ❑No inbreeding
- ❑No assortative mating
- ❑No mutation
- ❑No migration
- ❑Linkage equilibrium+ no linkage

ALL ASSUMPTIONS VIOLATED!



EXAMPLE: 2 LOCUS MODELS

ADDITIVE

$$y_i^A = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$$

$$\frac{\partial y_i^A}{\partial x_{i1}} = \beta_1$$

ADDITIVE+ DOMINANCE

$$y_i^{A+D} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + e_i$$

$$\frac{\partial y_i}{\partial x_{i1}} = \beta_1 + 2\beta_{11} x_{i1}$$

ADDITIVE+ DOMINANCE EPISTASIS

$$y_i^{A+D+I} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \beta_{122} x_{i1} x_{i2}^2 + \beta_{112} x_{i1}^2 x_{i2} + \beta_{1122} x_{i1}^2 x_{i2}^2 + e_i$$

$$\frac{\partial y_i^{A+D+I}}{\partial x_{i1}} = \beta_1 + 2\beta_{11} x_{i1} + \beta_{12} x_{i2} + \beta_{122} x_{i2}^2 + 2\beta_{112} x_{i1} x_{i2} + 2\beta_{1122} x_{i1} x_{i2}^2$$

Additive x additive

Additive x dominance

Dominance x additive

Dominance x Dominance

EXAMPLE OF ADDITIVITY AS AN "EMERGENT PROPERTY" OF EPISTASIS

TRUE GENETIC SIGNAL

$$g = (x_1, x_2) = \beta_{12}x_1x_2 + \beta_{122}x_1x_2^2 + \beta_{112}x_1^2x_2 + \beta_{1122}x_1^2x_2^2$$

1) Use a first-order (linear) Taylor series approximations in the neighborhood of:
$x_1 = x_2 = -1$ (most genotypes *aa* and *bb)*

$$g = (x_1, x_2) \approx \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$$
$$\alpha_0 = 2\beta_{112} - \beta_{12} + 2\beta_{122} - 3\beta_{1122}$$
$$\alpha_1 = (\beta_{12} + 2\beta_{112} + \beta_{122} - 2\beta_{1122})$$
$$\alpha_2 = (-\beta_{12} + \beta_{112} + 2\beta_{122} - 2\beta_{1122})$$

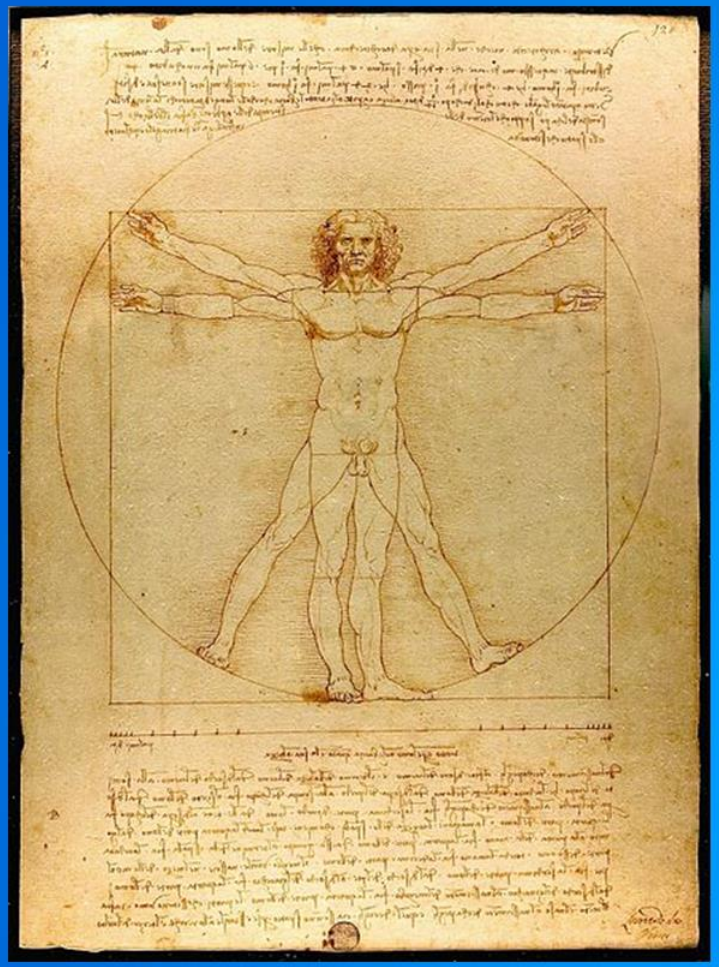
2) Now, in the neighborhood of $x_1 = x_2 = 0$ (most genotypes *Aa* and *Bb)*: $g = (x_1, x_2) \approx 0$

3) Now, in the neighborhood of $x_1 = x_2 = 1$ (most genotypes *AA* and *BB)*:

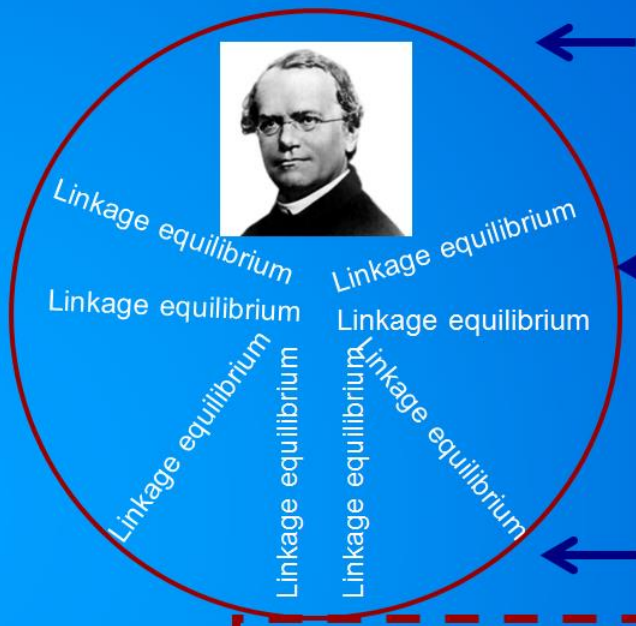
$$g = (x_1, x_2) \approx \alpha_0' + \alpha_1' x_1 + \alpha_2' x_2$$
$$\alpha_0' = -\beta_{12} - 2\beta_{112} - 2\beta_{122} - 3\beta_{1122}$$
$$\alpha_1' = \beta_{12} + 2\beta_{112} + \beta_{122} + 2\beta_{1122}$$
$$\alpha_2' = \beta_{12} + 2\beta_{112} + \beta_{122} + 2\beta_{1122}$$

MORAL: BEWARE! A STRAIGHT LINE WILL CONVEY MISLEADING MECHANISTIC INFO!

Leonardo da Vinci ‘s l’uommo Vitruviano



CRITICAL ASSUMPTIONS
Do they hold?



Wind here

HARDY-WEINBERG

L'***Uomo vitruviano*** è un disegno a matita e inchiostro su carta (34x24 cm) di Leonardo da Vinci, databile al 1490 circa e conservato nel Gabinetto dei Disegni e delle Stampe delle Gallerie dell'Accademia di Venezia. Celeberrima rappresentazione delle proporzioni ideali del corpo umano, dimostra come esso possa essere armoniosamente inscritto nelle due figure "perfette" del cerchio e del quadrato.

Leak here

A VIEW OF LINEAR MODELS (as employed in q. genetics)

Mathematically, can be viewed as a "local" approximation of a complex process

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2} + \frac{f^{(3)}(a)}{3!} + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots$$

Linear approximation

Quadratic approximation

n^{th} order approximation

FELDMAN and LEWONTIN (1975)
CHEVALET (1994)

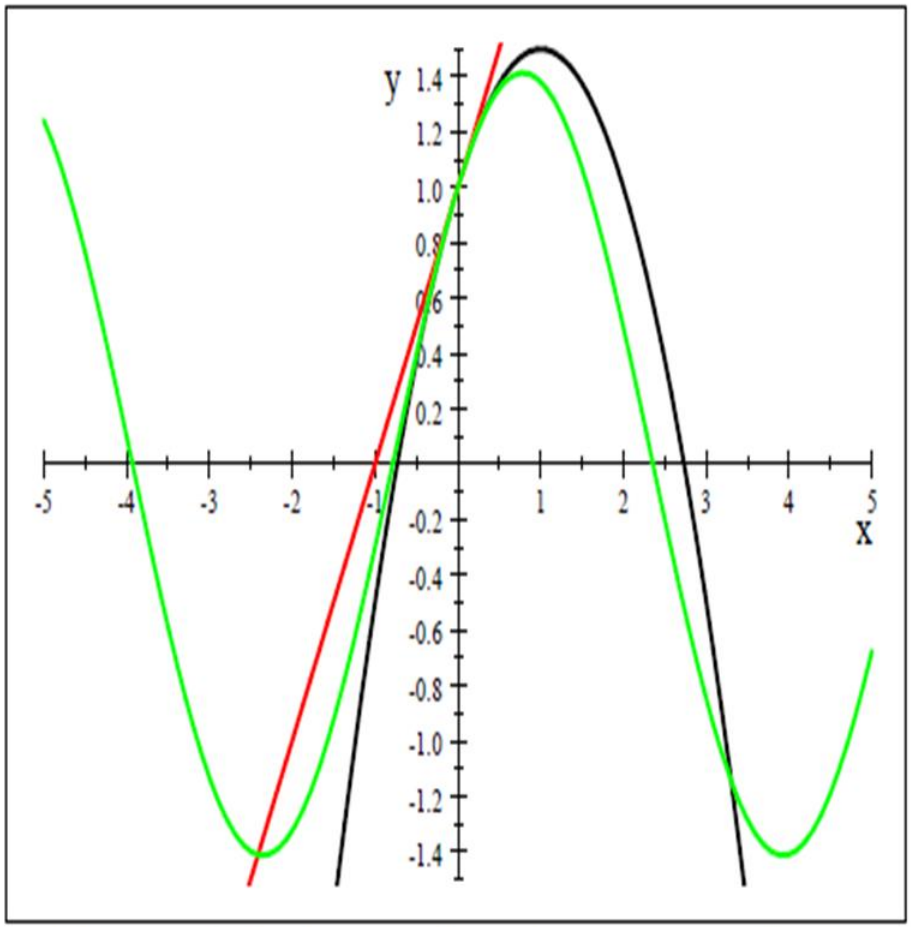


Figure 2A. Approximation at $x = 0$.of $\sin(x) + \cos(x)$ with linear and quadratic Taylor series.

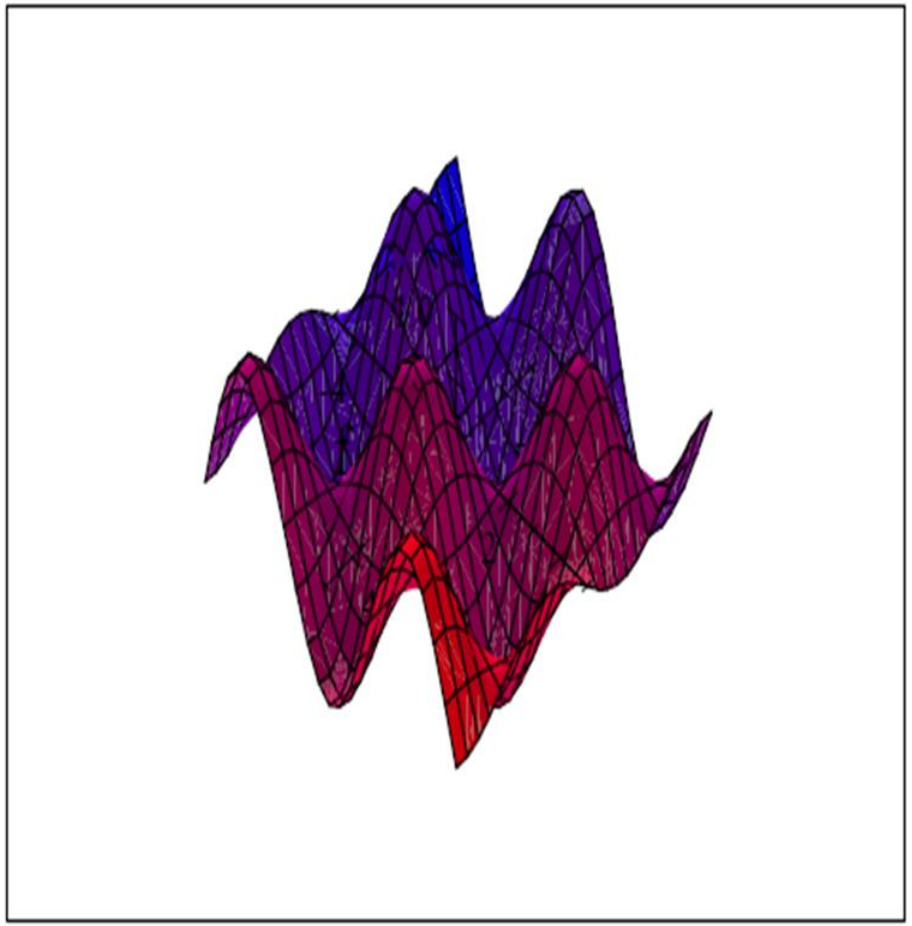


Figure 2B. Genetic signal resulting from risk variables x and y, with the signal being $f = \sin(x)\sin(y)$.

IS MY MODEL "RIGHT"?

TAKING MODEL UNCERTAINTY INTO ACCOUNT BY MODEL AVERAGING

$$p(\theta|y) = \sum p(\theta|y,M)p(M|y)$$

FEW MODELS

$$= \int p(\theta|y,M)p(M|y)dM$$

MANY MODELS

THE PUNCH LINE: VARIANCE OF PREDICTION ERRORS TAKING MODEL UNCERTAINTY INTO ACCOUNT

$$Var(\theta|y) = E_M[Var(\theta|y,M)] + Var[E_M[\theta|y,M]]$$

Average "prediction Error" variance

Variance among predictions from different models

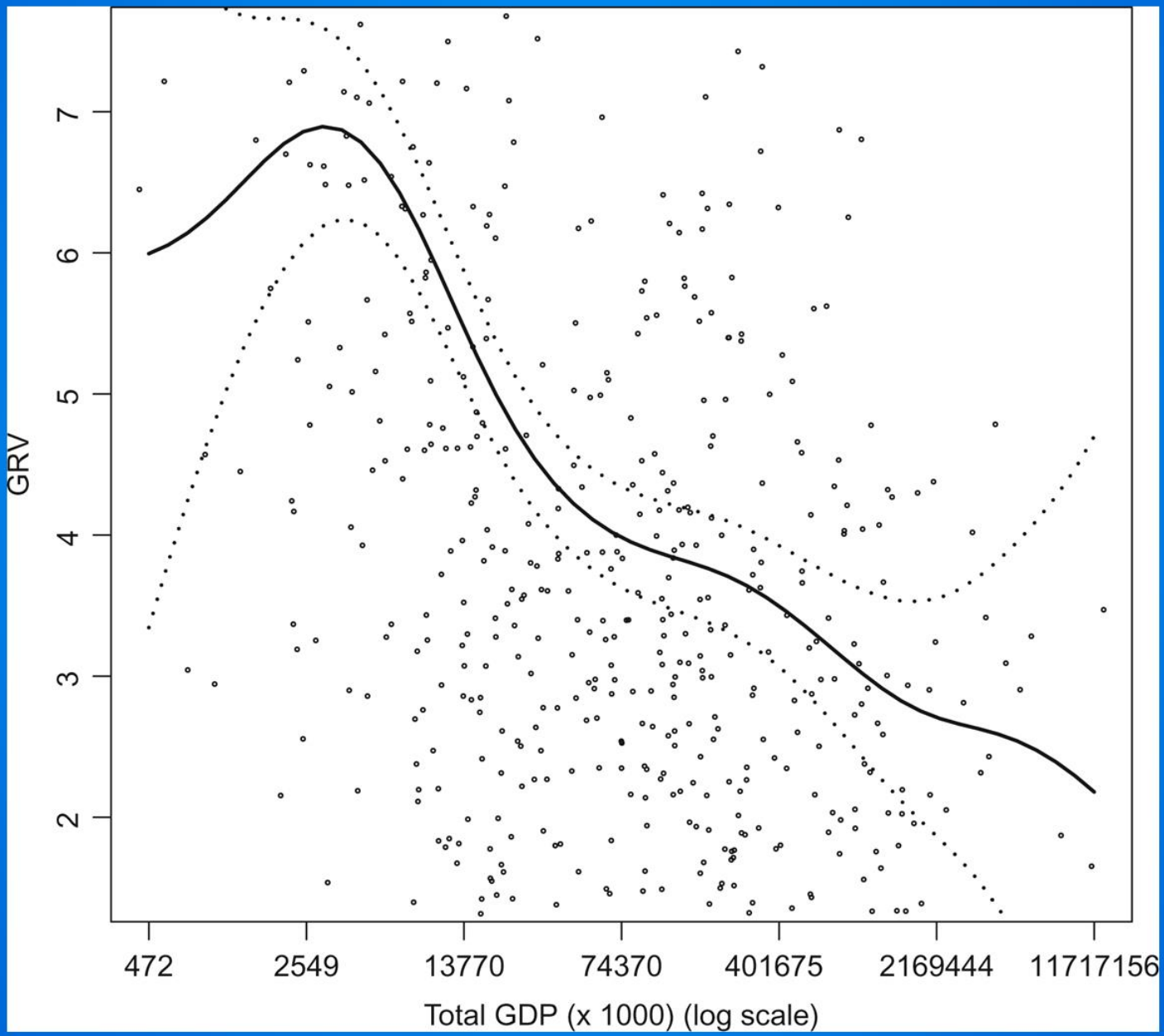
PARADIGM 3
(machine learning)

Distinctive aspects of non-parametric fitting

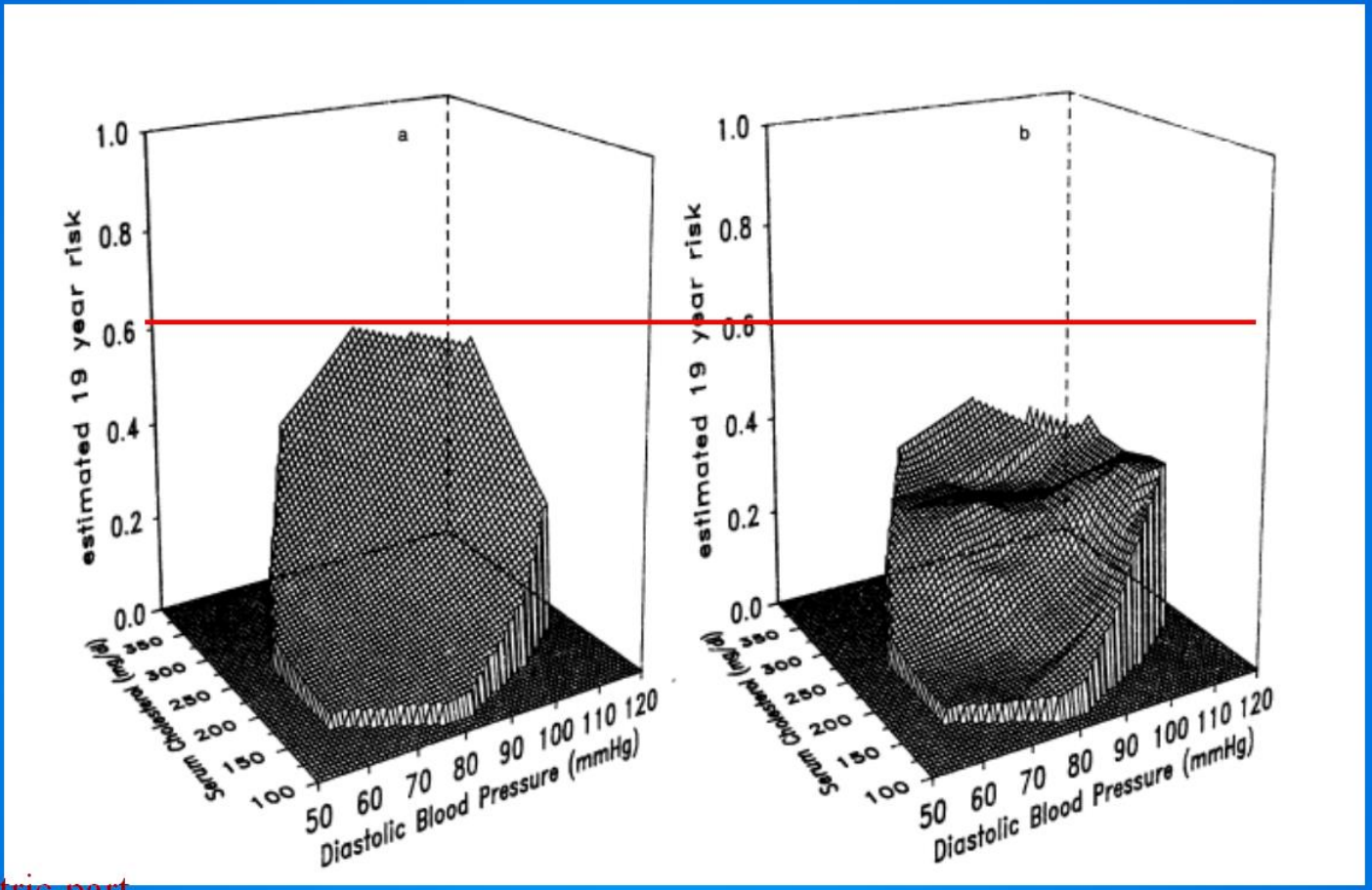
- **I**nvestigate patterns free of strictures imposed by parametric models
- **R**egression coefficients appear but (typically) do not have an obvious interpretation
- **O**ften: very good predictive performance in cross-validation
- **T**uning methods and algorithms (maximization, MCMC) similar to those of parametric methods
- **O**ften produce surprising results


Pattern recognition
Neural networks
Data mining
Universal approximators
Machine learning
Cross-validation designs
Kernel methods
Random forest algorithms
Sampling methods
Support vector machines
Ensemble Methods: bagging
Ensemble Methods: boosting
Non-parametric prediction
Bayesian networks

Economic growth volatility (many theories, none credible. Sounds like economics…)



Logistic regression with thin-plate splines



parametric part

$$f(\mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \sum_{j=1}^{N} \alpha_j \left[\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 \right] \log\left[\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 \right]$$

Risk of heart attack after 19 years as a function of cholesterol level and blood pressure. Left: logistic regression model. Right: thin plate spline fit. Wahba (2007)

CROSS-VALIDATION

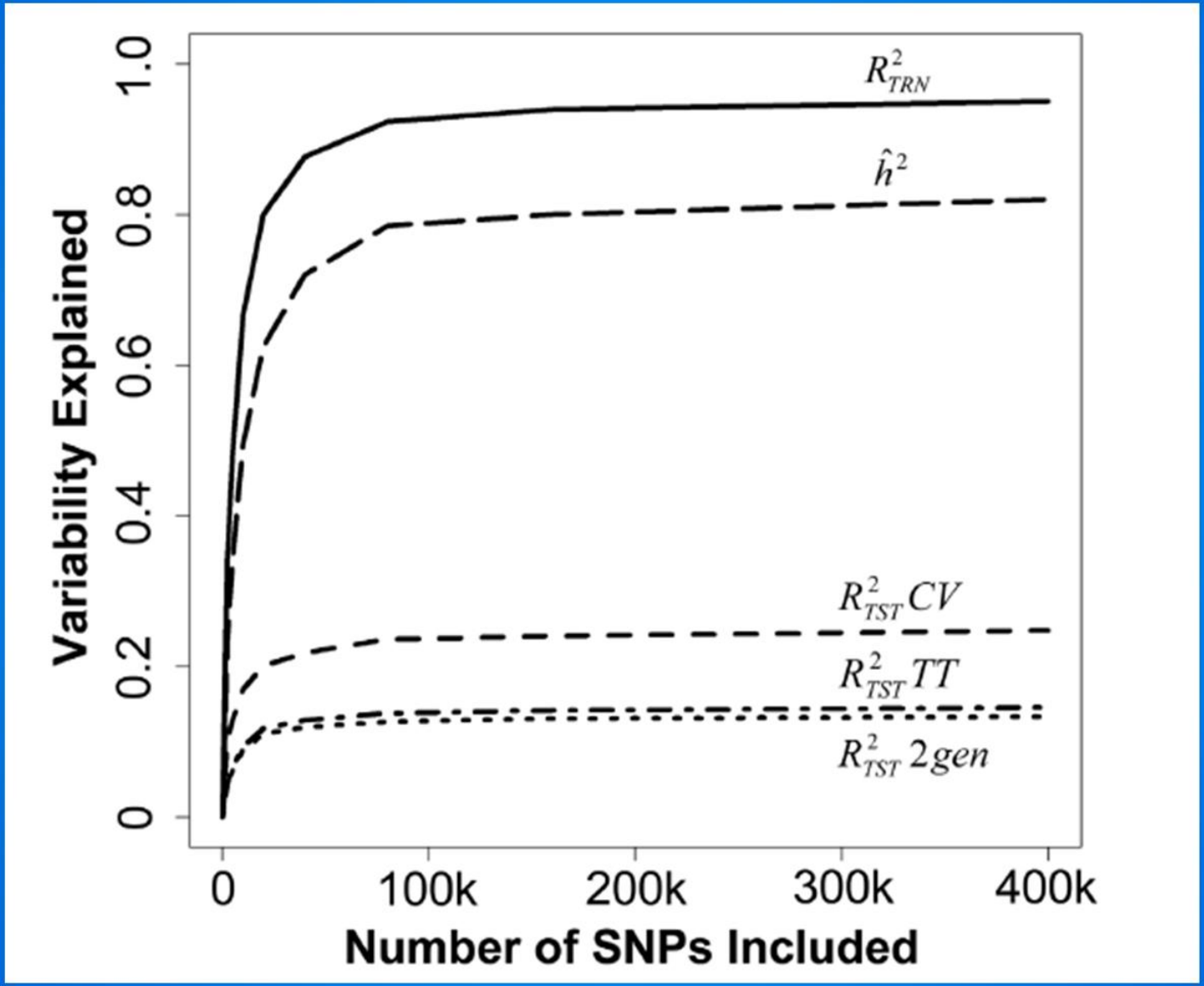
(take model uncertainty into account:
seldom done in animal breeding in the BLUP era. Often absent in GWAS and medical studies)

➔A. Prediction and goodness of fit are different ball games: a model that fits well to training data may have atrocious predictive ability

➔B. Any cross-validation scheme (e.g., k-folds) has a cross-validation distribution

THIS IS THE DISTRIBUTION THAT MATTERS AND NOT A MODEL DERIVED QUANTITY, THAT IGNORES UNCERTAINTY ABOUT THE MODEL!!!!!!

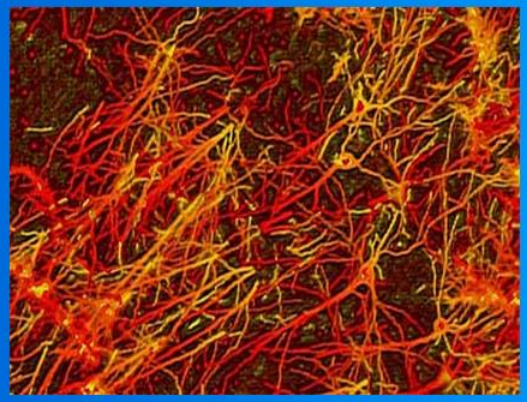
GOODNESS OF FIT (TRAINING= TRN) vs. **PREDICTIVE ABILITY** (TESTING= TST)

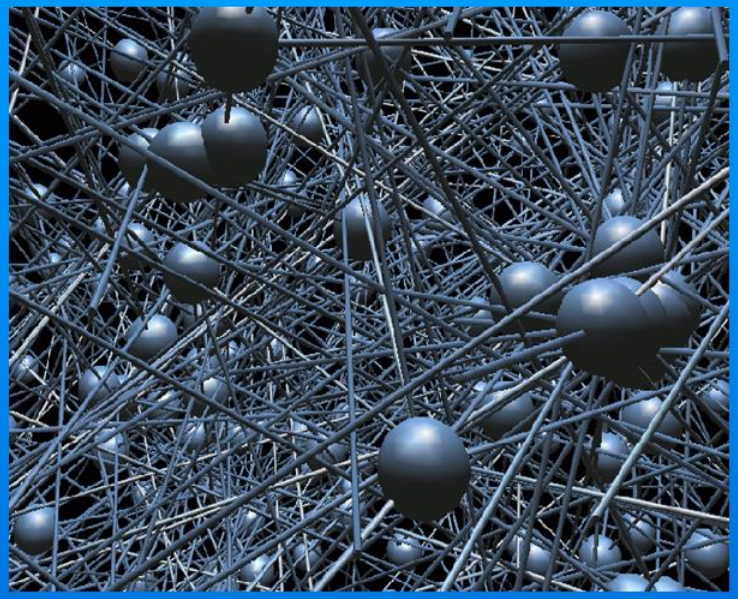


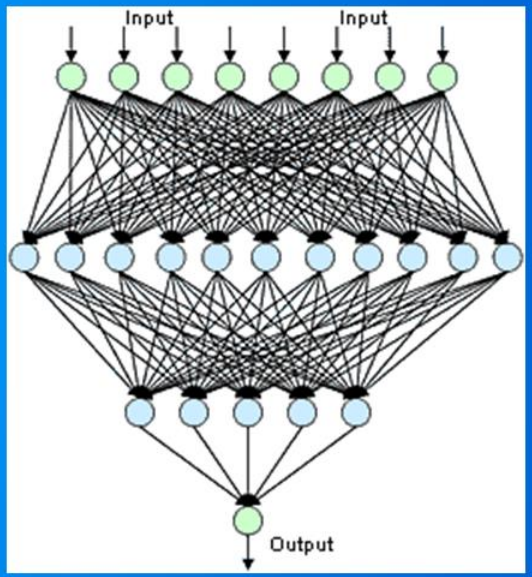
HUMAN STATURE: MAKOWSKY et al. , Plos Genetics 2011

PARADIGM3

More universal prediction machines:
A. Regularized Neural Networks







B. Reproducing Kernel Hilbert spaces mixed model regression

Function of molecular information **x (vector of SNP variables)**

$$SS[g(\mathbf{x}),\lambda] = \sum_{i=1}^{n}[y_i - \mathbf{w}_i'\boldsymbol{\beta} - \mathbf{z}_i'\mathbf{u} - g(x_i)]^2 + \lambda\|g(\mathbf{x})\|_H^2$$

Smoothing parameter (λ)

"Penalized sum of squares"

Some norm under Hilbert space (*H*) of functions

Variational problem: find *g(**x**)* over entire space of functions minimizing SS(.)

SOME CASE STUDIES WITH PARADIGM 3

WHEAT DATA SET: 599 lines (480 training-119 testing, 50 random repeats)
1279 binary markers

ANN architectures	Linear	1 neuron	2 neurons	3 neurons	4 neurons
Criterion					
Effective number of parameters	299±5.5	260±6.1	253±5.9	238±5.5	220±2.8
Correlations in testing set	BENCHMARKS: BAYESIAN LASSO 0.50 4 SVM MODELS 0.50-0.58				
	0.48±0.03	0.54±0.03	056±0.02	0.57±0.02	0.59±0.02
Mean squared error in testing set	0.99±0.04	0.77±0.03	0.74±0.03	0.71±0.02	0.72±0.02

PEREZ et al. (2012, G3): wheat

Table 2 Average correlation (SE in parentheses) between observed and predicted values for grain yield (GY) and days to heading (DTH) in 12 environments for seven models

Trait	Environment	BL	BRR	Bayes A	Bayes B	RKHS	RBFNN	BRNN
DTH	1	0.59 (0.11)	0.59 (0.11)	0.59 (0.11)	0.56 (0.11)	0.66 (0.09)	0.66 (0.10)	0.64 (0.11)
	2	0.58 (0.14)	0.57 (0.14)	0.61 (0.12)	0.57 (0.13)	0.63 (0.13)	0.61 (0.13)	0.62 (0.13)
	3	0.60 (0.13)	0.60 (0.12)	0.62 (0.11)	0.60 (0.12)	0.68 (0.10)	0.69 (0.10)	0.67 (0.11)
	4	0.02 (0.18)	0.07 (0.17)	0.06 (0.17)	0.06 (0.17)	0.12 (0.18)	0.16 (0.18)	0.02 (0.19)
	5	0.65 (0.09)	0.64 (0.10)	0.66 (0.09)	0.66 (0.09)	0.69 (0.08)	0.68 (0.08)	0.68 (0.08)
	8	0.36 (0.15)	0.37 (0.15)	0.36 (0.15)	0.35 (0.14)	0.46 (0.13)	0.46 (0.14)	0.39 (0.15)
	9	0.59 (0.12)	0.59 (0.11)	0.53 (0.12)	0.52 (0.11)	0.62 (0.11)	0.63 (0.11)	0.61 (0.12)
	10	0.54 (0.14)	0.52 (0.14)	0.56 (0.13)	0.54 (0.14)	0.61 (0.13)	0.62 (0.12)	0.57 (0.13)
	11	0.52 (0.15)	0.52 (0.16)	0.53 (0.13)	0.51 (0.13)	0.58 (0.14)	0.59 (0.13)	0.55 (0.14)
	12	0.45 (0.19)	0.42 (0.18)	0.45 (0.18)	0.45 (0.18)	0.47 (0.18)	0.39 (0.19)	0.35 (0.19)
	Average	0.59 (0.12)	0.58 (0.12)	0.60 (0.12)	0.57 (0.12)	0.65 (0.10)	0.48 (0.14)	0.48 (0.14)
GY	1	0.48 (0.13)	0.43 (0.14)	0.48 (0.13)	0.46 (0.13)	0.51 (0.12)	0.51 (0.12)	0.50 (0.13)
	2	0.48 (0.14)	0.41 (0.17)	0.48 (0.14)	0.48 (0.14)	0.50 (0.14)	0.43 (0.16)	0.43 (0.16)
	3	0.20 (0.21)	0.29 (0.22)	0.20 (0.22)	0.18 (0.22)	0.37 (0.20)	0.42 (0.21)	0.32 (0.24)
	4	0.45 (0.15)	0.46 (0.13)	0.43 (0.15)	0.42 (0.15)	0.53 (0.12)	0.55 (0.11)	0.49 (0.14)
	5	0.59 (0.14)	0.56 (0.16)	0.75 (0.11)	0.74 (0.12)	0.64 (0.13)	0.66 (0.13)	0.63 (0.13)
	6	0.70 (0.10)	0.67 (0.11)	0.73 (0.08)	0.71 (0.08)	0.73 (0.08)	0.71 (0.08)	0.69 (0.10)
	7	0.46 (0.14)	0.50 (0.14)	0.42 (0.14)	0.40 (0.15)	0.53 (0.13)	0.54 (0.14)	0.50 (0.14)
	Average	0.62 (0.10)	0.57 (0.14)	0.69 (0.10)	0.70 (0.09)	0.67 (0.09)	0.56 (0.12)	0.65 (0.10)

Fitted models were Bayesian LASSO (BL), RR-BLUP (BRR), Bayes A, Bayes B, reproducing kernel Hilbert spaces regression (RKHS), radial basis function neural networks (RBFNN) and Bayesian regularized neural networks (BRNN) across 50 random partitions of the data with 90% in the training set and 10% in the validation set. The models with highest correlations are underlined.

over 35 wheat and maize trials (Crossa et al. 2011)

M-BL	M-RKHS	M-RBFNN
14%	34%	52%
5	12	18

Any concerns about the predictive ability of non-parametric methods, relative to those that "*help to understand genetic architecture*"?

Comparison among methods in plants (Heslot et al., 2012)

Table 2. Accuracy for each trait and model, average non-cross-validated correlation for each model, and average MSE for each model.

Dataset†	Trait‡	RR-BLUP§	BL	Elastic net	wBSR	BayesCπ	E-Bayes	RKHS	SVM	RF	NNET
Barley 1	Yield	0.53	0.55	0.52	0.53	0.53	0.53	0.6	0.43	0.56	0.51
Barley CAP	Betaglucan	0.57	0.57	0.57	0.57	0.57	0.57	0.6	0.35	0.55	0.54
Bay × Sha (Bay-0 × Shahdara)	FLOSD	0.82	0.82	0.83	0.83	0.82	0.82	0.83	0.8	0.85	0.82
	DM10	0.63	0.63	0.63	0.64	0.63	0.63	0.64	0.56	0.57	0.56
	DM3	0.4	0.39	0.40	0.4	0.39	0.4	0.41	0.33	0.38	0.35
Panel maize	Moisture	0.75	0.75	0.75	0.76	0.75	0.73	0.79	0.45	0.73	0.73
	Yield	0.63	0.63	0.61	0.63	0.63	0.59	0.64	0.32	0.6	0.59
Diallel maize	Moisture	0.74	0.74	0.72	0.73	0.74	0.73	0.75	0.56	0.61	0.72
	Yield	0.52	0.52	0.49	0.51	0.52	0.51	0.5	0.29	0.49	0.48
Wheat CIMMYT	YLD1	0.51	0.5	0.46	0.48	0.51	0.49	0.59	0.36	0.52	0.54
	YLD2	0.5	0.49	0.45	0.5	0.5	0.46	0.52	0.36	0.43	0.51
	YLD4	0.38	0.37	0.35	0.36	0.38	0.36	0.43	0.32	0.38	0.43
	YLD5	0.44	0.47	0.42	0.47	0.44	0.39	0.52	0.27	0.46	0.44
Wheat Cornell	Yield	0.36	0.35	0.37	0.37	0.34	0.26	0.28	0.22	0.36	0.36
	Height	0.45	0.44	0.41	0.44	0.44	0.41	0.55	0.37	0.46	0.45
Wheat diallel	Height	0.64	0.66	0.68	0.67	0.66	0.67	0.73	0.51	0.62	0.67
	TKW	0.6	0.57	0.59	0.6	0.59	0.59	0.68	0.41	0.54	0.65
	Yield	0.53	0.52	0.51	0.52	0.53	0.51	0.58	0.39	0.52	0.57
Average accuracy (cross-validated)		0.56	0.56	0.54	0.56	0.55	0.54	0.59	0.41	0.54	0.55
Average non-cross-validated correlation		0.77	0.79	0.75	0.77	0.77	0.93	0.99	0.89	0.76	0.85
Average MSE		0.67	0.67	0.69	0.68	0.68	0.76	0.64	1.36	0.72	10.54

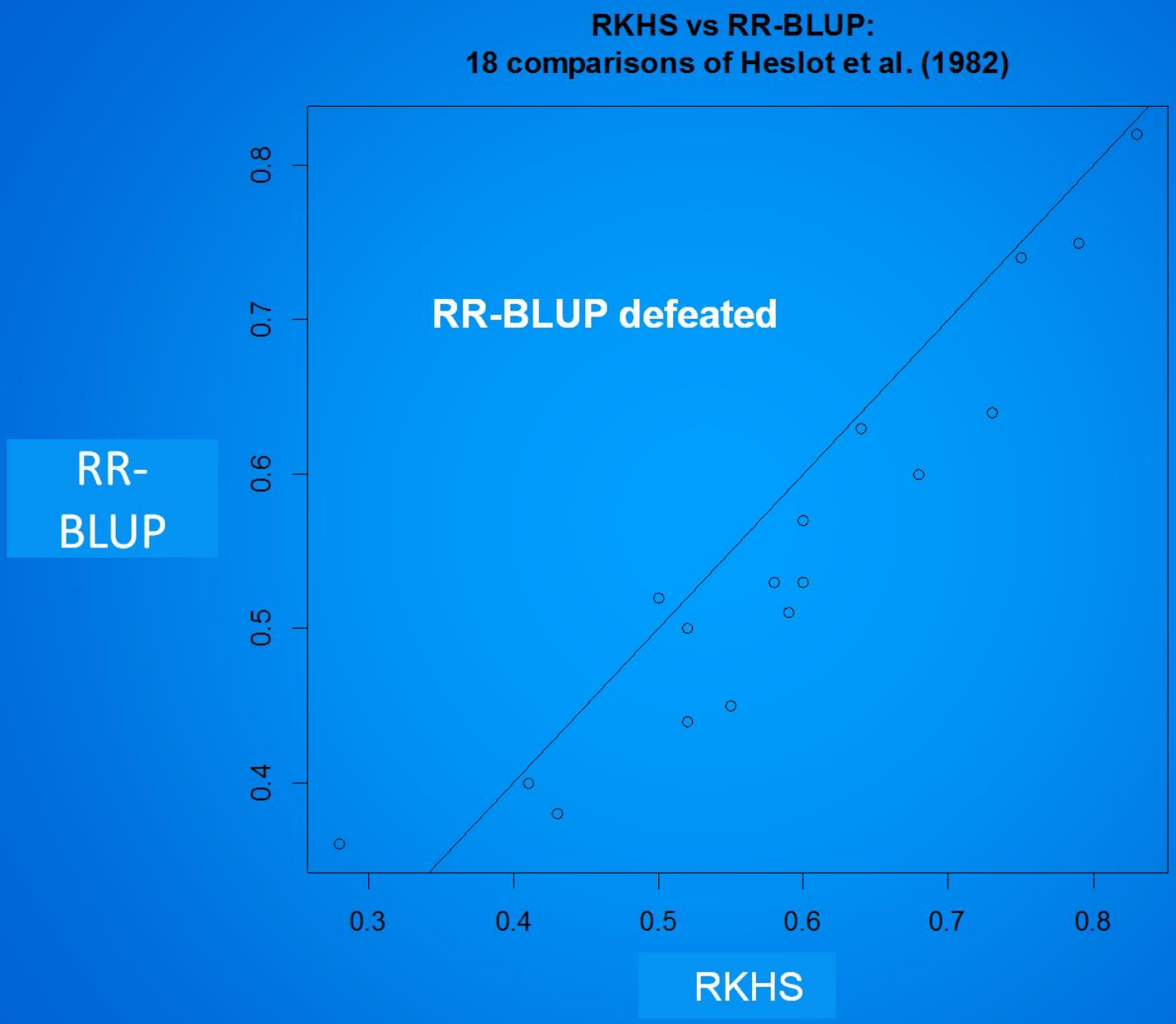
†Barley 1, Limagrain Europe, Riom, France; Barley CAP (Barley Coordinated Agricultural Project, 2011); Bay Sha (Loudet et al. 2002); Panel maize, Limagrain Europe; Diallel maize, Limagrain Europe; Wheat CIMMYT (Crossa et al., 2010); Wheat Cornell (Heffner et al., 2011); Wheat diallel, Limagrain Europe.

‡Betaglucan, betaglucan content; FLOSD, flowering time in short days; DM10, dry matter in nonlimiting N conditions; DM3, dry matter in limiting N conditions; YLD1 to YLD5 refers to the yield traits reported in Crossa et al. (2010); TKW, thousand kernel weight.

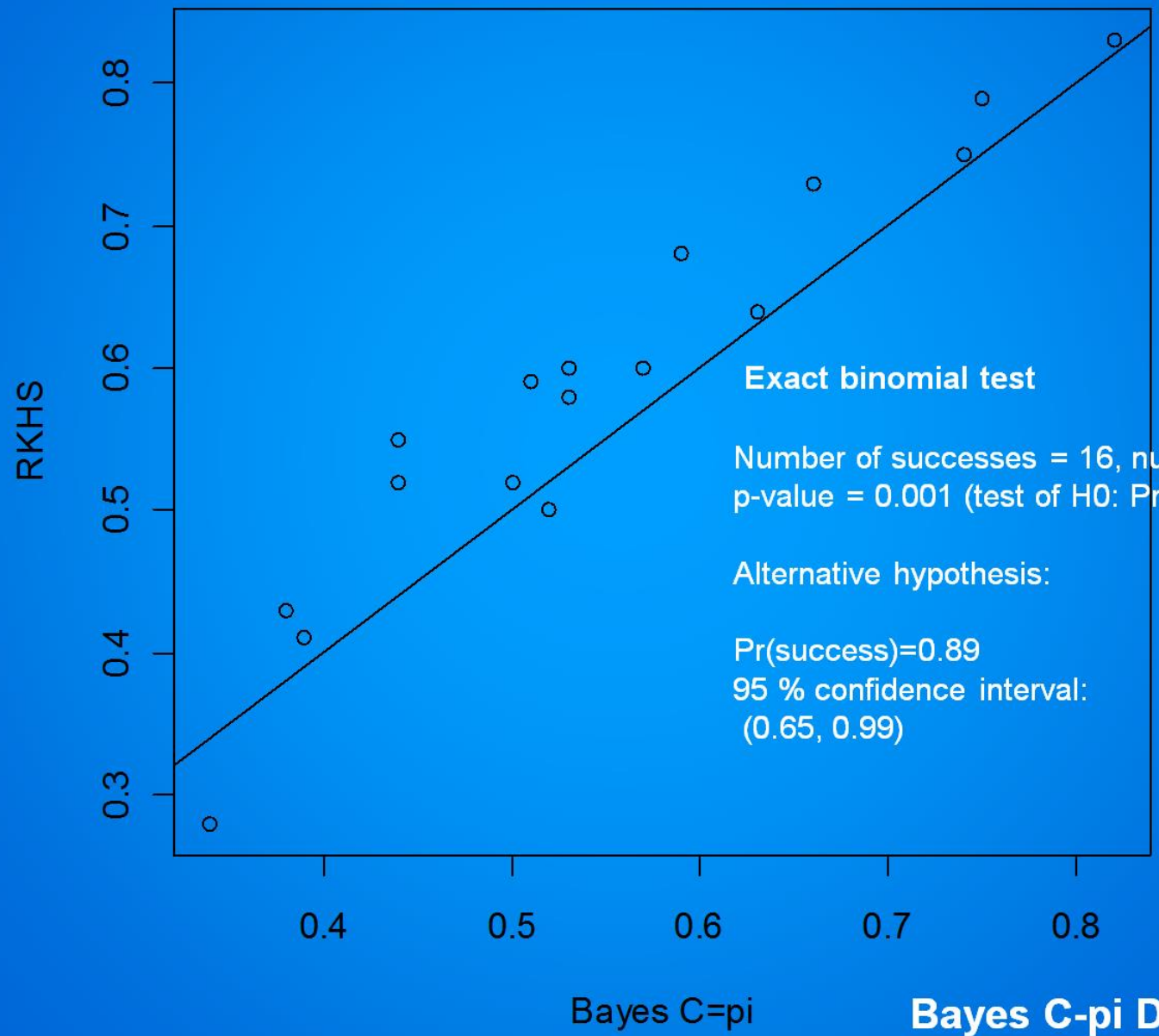
TENTATIVE CONCLUSION: choice of method does not make a difference, in practice

Really?

Single malt rankings:
DAILUAINE 14 YEARS CONNOISSEUR'S CHOICE (90)
GLEN GARIOCH FOUNDER'S RESERVE (85)
BOWMORE 12 YEAR OLD (80)
OBAN 14 YEARS (73)…


RKHS vs RR-BLUP:
18 comparisons of Heslot et al. (1982)
RR-BLUP defeated
RR-BLUP
RKHS
0.3
0.4
0.5
0.6
0.7
0.8
0.4
0.5
0.6
0.7
0.8

RKHS vs Bayes C-pi:
18 comparisons of Heslot et al. (2012)



Bayes C-pi DEFEATED

TAKE HOME THOUGHTS CONNECTED WITH PARADIGM 3…

"Would you refuse your dinner because you do not understand the digestive system?"

quote by British mathematician in "*The emperor of the maladies: a biography of cancer*",2010, by *Siddhartha Mujkherjee*

THERE IS NO UNIVERSALLY 'BEST' MACHINE

Change the species, trait and environment and the ranking in predictive ability of methods will probably vary